

UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO  
DEPARTAMENTO DE COMUNICAÇÕES

## **Proposta de métodos de separação cega de fontes para misturas convolutivas e não-lineares**

### **Autor**

Ricardo Suyama

### **Orientador**

Prof. Dr. João Marcos Travassos Romano (FEEC/UNICAMP)

Banca Examinadora:

Prof. Dr. João Marcos Travassos Romano (FEEC/UNICAMP)

Profa. Dra. Mariane Rembold Petraglia (COPPE/UFRJ)

Prof. Dr. Charles Casimiro Cavalcante (GTEL/UFC)

Prof. Dr. Amauri Lopes (FEEC/UNICAMP)

Prof. Dr. Reginaldo Palazzo Jr. (FEEC/UNICAMP)

Prof. Dr. Romis Ribeiro de Faissol Attux (FEEC/UNICAMP)

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

Campinas, 24 de agosto de 2007

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Su97p      Suyama, Ricardo  
Proposta de métodos de separação cega de fontes para  
misturas convolutivas e não-lineares. / Ricardo Suyama. --  
Campinas, SP: [s.n.], 2007.

Orientador: João Marcos Travassos Romano  
Tese (doutorado) - Universidade Estadual de Campinas,  
Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais. 2. Processamento de sinais  
adaptativo. 3. Teoria da informação. I. Romano, João  
Marcos Travassos. II. Universidade Estadual de Campinas.  
Faculdade de Engenharia Elétrica e de Computação. III.  
Título.

Título em Inglês: Proposal of blind source separation methods for convolutive and  
nonlinear mixtures

Palavras-chave em Inglês: Blind source separation, Independent component analysis,  
Digital signal processing, Convolutive mixtures, Nonlinear  
mixtures, Information theory

Área de concentração: Telecomunicações e Telemática

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Mariane Rembold Petraglia, Charles Casimiro Cavalcante,  
Amauri Lopes, Reginaldo Palazzo Júnior, Romis Ribeiro de  
Faissol Attux

Data da defesa: 24/08/2007

Programa de Pós-Graduação: Engenharia Elétrica

## COMISSÃO JULGADORA - TESE DE DOUTORADO

**Candidato:** Ricardo Suyama

**Data da Defesa:** 24 de agosto de 2007

**Título da Tese:** "Proposta de Métodos de Separação Cega de Fontes para Misturas Convolutivas e Não-Lineares"


Prof. Dr. João Marcos Travassos Romano (Presidente):



Profa. Dra. Mariane Rembold Petraglia:



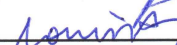
Prof. Dr. Charles Casimiro Cavalcante:



Prof. Dr. Amauri Lopes:



Prof. Dr. Romis Ribeiro de Faissol Attux:



Prof. Dr. Reginaldo Palazzo Júnior:



# Resumo

O problema de separação cega de fontes (BSS - *Blind Source Separation*) vem despertando o interesse de um número crescente de pesquisadores. Esse destaque é devido, em grande parte, à formulação abrangente do problema, que torna possível o uso das técnicas desenvolvidas no contexto de BSS nas mais diversas áreas de aplicação.

O presente trabalho tem como objetivo propor novos métodos de solução do problema de separação cega de fontes, nos casos de mistura convolutiva e mistura não-linear. Para o primeiro caso propomos um método baseado em predição não-linear, cujo intuito é eliminar o caráter convolutivo da mistura e, dessa forma, separar os sinais utilizando ferramentas bem estabelecidas no contexto de misturas lineares sem memória.

No contexto de misturas não-lineares, propomos uma nova metodologia para separação de sinais em um modelo específico de mistura denominado modelo com não-linearidade posterior (PNL - *Post Nonlinear*). Com o intuito de minimizar problemas de convergência para mínimos locais no processo de adaptação do sistema separador, o método proposto emprega um algoritmo evolutivo como ferramenta de otimização, e utiliza um estimador de entropia baseado em estatísticas de ordem para avaliar a função custo. A eficácia de ambos os métodos é verificada através de simulações em diferentes cenários.



# Abstract

The problem of blind source separation (BSS) has attracted the attention of a growing number of researchers, mostly due to its potential applications in a significant number of different areas.

The objective of the present work is to propose new methods to solve the problem of BSS in the cases of convolutive mixtures and nonlinear mixtures. For the first case, we propose a new method based on nonlinear prediction filters. The nonlinear structure is employed to eliminate the convolutive character of the mixture, hence converting the problem into an instantaneous mixture, to which several well established tools may be used to recover the sources.

In the context of nonlinear mixtures, we present a new methodology for signal separation in the so-called post-nonlinear mixing models (PNL). In order to avoid convergence to local minima, the proposed method uses an evolutionary algorithm to perform the optimization of the separating system. In addition to that, we employ an entropy estimator based on order-statistics to evaluate the cost function. The effectiveness of both methods is assessed through simulations in different scenarios.

# Agradecimentos

Desejo manifestar meus mais sinceros agradecimentos a todas as pessoas que de uma forma ou de outra ajudaram nesta jornada. Em especial, agradeço:

Ao professor João Marcos Travassos Romano, pela orientação dedicada, atenção e apoio dispensados desde a época de iniciação científica.

Aos meus pais, Cecília e Jorge, e à minha irmã, Bruna, por todo o carinho e compreensão.

Aos meus familiares, pelo incentivo e apoio.

Aos colegas do Laboratório de Processamento de Sinais para Comunicações - DSP-Com, pelas discussões sempre proveitosas, e também pela amizade e companheirismo em todos os momentos.

Aos professores da banca de defesa, pela leitura cuidadosa e importantes considerações sobre a tese.

À Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, pelo apoio financeiro.

# Índice

<b>Resumo</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Agradecimentos</b>	<b>ix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos e organização da tese . . . . .	4
<b>2 Separação Cega de Fontes: Princípios</b>	<b>7</b>
2.1 Definição do problema . . . . .	7
2.2 BSS através da Análise por Componentes Independentes . . . . .	9
2.2.1 Correlação x Independência . . . . .	11
2.3 Outras abordagens para Separação Cega de Fontes . . . . .	14
2.3.1 Separação Cega de Sinais utilizando apenas estatísticas de segunda ordem . . . . .	14
2.3.2 Análise por componentes independentes não-negativas . . . . .	16
2.3.3 Análise de componentes esparsos . . . . .	16
2.3.4 Abordagem Bayesiana . . . . .	19
2.4 Sumário . . . . .	19
<b>3 Análise por Componentes Independentes: Critérios e Estratégias</b>	<b>21</b>
3.1 Critérios para ICA . . . . .	21
3.1.1 Informação mútua . . . . .	22
3.1.2 Estatísticas de ordem superior . . . . .	24
3.1.3 Correlação não-linear . . . . .	25
3.1.4 Maximização da não-gaussianidade . . . . .	26
3.1.5 Princípio Infomax e a abordagem de Máxima Verossimilhança	29
3.2 Comentários gerais sobre as diferentes abordagens . . . . .	31
3.3 Algoritmos para Separação Cega de Fontes . . . . .	33
3.3.1 Algoritmo de Héroult, Jutten e Ans . . . . .	33
3.3.2 Algoritmo Infomax e Infomax estendido . . . . .	34

3.3.3	PCA não-linear . . . . .	35
3.3.4	Método do Gradiente Relativo / Natural . . . . .	37
3.3.5	Método do ponto fixo - Algoritmo FastICA . . . . .	40
3.3.6	Otimização utilizando métodos de Lie . . . . .	41
3.4	Sumário . . . . .	43
<b>4</b>	<b>Misturas Lineares Convolutivas</b>	<b>45</b>
4.1	Mistura Linear Convolutiva . . . . .	46
4.1.1	Inversibilidade de estruturas MIMO . . . . .	47
4.1.2	Separação dos sinais através da independência . . . . .	49
4.1.3	Separação dos sinais no domínio da frequência . . . . .	51
4.2	Uma proposta para separação de sinais baseada em predição . . . . .	54
4.3	Solução através de filtros fuzzy . . . . .	57
4.3.1	Filtros Fuzzy . . . . .	58
4.3.2	Estrutura fuzzy aplicada ao problema de predição . . . . .	61
4.3.3	Preditor de Mínimo Erro Quadrático Médio e o Preditor Fuzzy . . . . .	63
4.3.4	Algoritmo de Clusterização - K-means + Iterated Local Search . . . . .	67
4.3.5	Adaptação dos pesos de saída da estrutura fuzzy . . . . .	71
4.4	Resumo da proposta . . . . .	72
4.5	Resultados de Simulações . . . . .	72
4.6	Discussão . . . . .	80
4.7	Sumário . . . . .	81
<b>5</b>	<b>Misturas Não-Lineares</b>	<b>83</b>
5.1	Misturas não-lineares . . . . .	84
5.1.1	Técnicas para o caso geral . . . . .	84
5.1.2	ICA não-linear . . . . .	85
5.2	Mistura <i>Post-Nonlinear</i> . . . . .	88
5.2.1	Separabilidade de misturas PNL . . . . .	89
5.2.2	Minimização da informação mútua . . . . .	89
5.2.3	Abordagem de gaussianização . . . . .	91
5.2.4	Outras abordagens . . . . .	91
5.3	Misturas <i>Post Nonlinear</i> . . . . .	92
5.3.1	Estimação de entropia baseada em estatísticas de ordem . . . . .	93
5.3.2	Algoritmo de otimização . . . . .	95
5.3.3	Primeiros Resultados . . . . .	99
5.4	Uma solução modificada . . . . .	100
5.4.1	Avaliação da solução modificada . . . . .	102
5.4.2	Estrutura linear por partes . . . . .	103
5.5	Sumário . . . . .	105
<b>6</b>	<b>Conclusões e Perspectivas</b>	<b>109</b>

Trabalhos publicados durante o doutorado	111
--	-----

Referências Bibliográficas	115
----------------------------	-----

# 1

## Introdução

Imagine-se em uma festa conversando com um amigo. O sinal que chega aos seus ouvidos é uma composição (mistura) de vários sinais, que podem ser vozes de outras pessoas na festa chegando diretamente aos seus ouvidos, ou reflexões das mesmas nas paredes do ambiente, ruído de fundo, música ambiente, além da voz de seu interlocutor. Mesmo em meio a esta cacofonia de conversas e ruído conseguimos distinguir o sinal de interesse (voz do amigo) dos outros sinais interferentes. Seria possível desenvolver um sistema artificial que realizasse a mesma tarefa<sup>1</sup>?

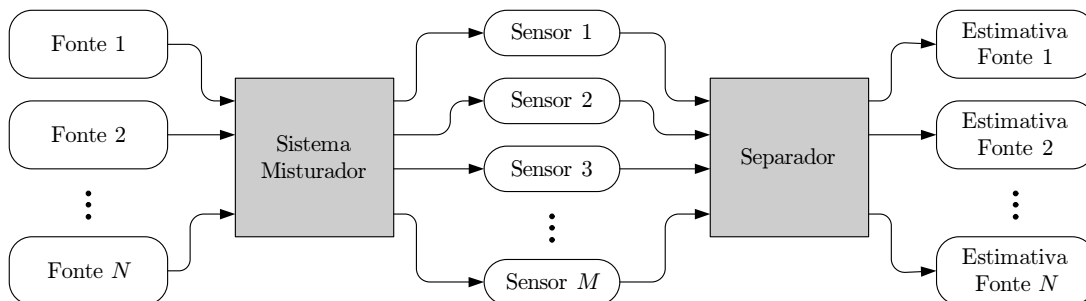
Embora o processo pelo qual lidamos com o problema possa ser complexo demais para ser inteiramente mimetizado, podemos dizer que, de uma forma geral, nosso cérebro é capaz de separar os sinais de voz sem conhecer a forma como foram misturados. A busca por técnicas capazes de realizar tarefa semelhante está diretamente relacionada à área de pesquisa conhecida como *Separação Cega de Fontes*, ou *BSS* (*Blind Source Separation*).

De maneira geral, o problema de BSS consiste em recuperar um conjunto de sinais, denominados *fontes*, a partir de observações que são misturas desconhecidas destas fontes (figura 1.1). O termo “cego” é usado para explicitar o fato de que as fontes

---

<sup>1</sup>O exemplo descrito é conhecido na literatura como *Cocktail Party Problem* [82]

não são observadas diretamente, e que o sistema de mistura é desconhecido.



**Figura 1.1:** Problema de separação de fontes

A falta de um conhecimento *a priori* sobre a mistura é de certa forma compensada pela hipótese de que os sinais das fontes são estatisticamente independentes entre si. Esta consideração, embora aparentemente restritiva, é uma hipótese plausível em muitos cenários práticos, e levou ao desenvolvimento da *Análise por Componentes Independentes*, ou ICA (*Independent Component Analysis*), uma das técnicas mais empregadas no contexto de separação cega de fontes.

O grande interesse por parte dos pesquisadores neste problema se deve ao fato deste mesmo cenário ser encontrado em diferentes contextos, dando margem ao uso das técnicas de BSS nas mais diversas aplicações, como por exemplo:

- **Tratamento de Sinais de Voz/Áudio:** As técnicas de BSS podem servir como ferramenta para o processamento de sinais de áudio gravados em ambientes ruidosos, como o descrito pelo *Cocktail Party problem*, no qual deseja-se extrair ou realçar algum sinal particular captado pelos sensores [99, 90, 21]. Podemos considerar que esta aplicação se insere em um campo de pesquisa muito mais amplo, denominado *Computational Auditory Scene Analysis*, cujo objetivo é justamente estudar e mimetizar diferentes aspectos de nossa habilidade de percepção auditiva [10, 50].

Outras aplicações incluem a transcrição musical automática [116, 92], onde deve-se separar os sons provenientes de cada instrumento musical a fim de se identificar o número e tipos de instrumentos presentes no trecho processado, reconhecimento e realce de fala, entre outros.

- **Processamento de Sinais Biomédicos:** No campo de engenharia biomédica existe uma variedade de problemas nos quais se aplicam as técnicas de separação cega de fontes. Nesta área, muitas vezes, os sinais de interesse não

podem ser obtidos de forma direta. A forma usual de se obter a medição é utilizar um conjunto de sensores o mais próximo possível das fontes, porém, neste caso, o sinal captado será necessariamente uma mistura dos sinais de interesse, configurando-se como um típico problema de separação de fontes.

Um exemplo clássico de aplicação é o de eletrocardiograma fetal [43, 31, 86], em que se pretende obter, da forma mais precisa possível, um eletrocardiograma do feto a partir de medidas feitas através de eletrodos na pele da mãe. Os sinais captados, portanto, são misturas do sinal proveniente do feto e da mãe, fazendo-se necessário o uso de técnicas de BSS para separá-los. Outros exemplos de aplicação envolvem a análise de eletroencefalogramas (EEG) [73, 141], análise de imagens obtidas por ressonância magnética funcional (fMRI) [47, 22] etc.

- **Sistemas de Comunicações:** Em sistemas de comunicações sem fio, em especial nos sistemas de telefonia celular, podemos ter, em uma dada frequência de operação, vários usuários transmitindo sinais para um único receptor, a estação rádio-base, de forma que o sinal resultante é uma combinação dos sinais emitidos ponderados pelos efeitos de propagação dos canais relativos a cada um dos usuários [30]. Assim, a fim de recuperar as informações transmitidas deve-se separar os sinais de cada um dos usuários, tarefa que pode ser realizada através das técnicas de BSS.

Outra possível aplicação é na equalização de canais de comunicação, onde busca-se combater o fenômeno de interferência intersimbólica. Este fenômeno, que decorre do espalhamento temporal dos pulsos transmitidos devido à limitação na largura de banda do canal [102], ocasiona uma distorção no sinal recebido de forma que a amostra do sinal no instante  $k$  representa uma mistura dos pulsos correspondentes a diferentes instantes de tempo. Assim, equalizar o canal pode ser visto como um processo de separação de sinais [136, 139, 34].

- **Análise de Dados Geofísicos:** Algumas aplicações interessantes para as ferramentas de BSS incluem o estudo de dados climáticos com o objetivo de identificar e extrair variáveis relevantes para o estudo e modelamento do clima [79]. As técnicas de BSS também podem ser utilizadas no contexto de análise de dados sísmicos para exploração da superfície terrestre [134, 54] etc.

Considera-se que a área de separação cega de fontes teve início em um trabalho publicado no início dos anos 80 [64]. Nele, os pesquisadores franceses Héroult, Jutten e Ans, motivados por um problema de separação de sinais de ativação musculares, apresentam um algoritmo inspirado em redes neurais artificiais para separação das



fontes sem o conhecimento do sistema misturador.

Embora sem uma justificativa teórica sólida, verifica-se que o algoritmo proposto buscava ajustar os pesos do sistema separador de forma a minimizar a correlação não-linear entre as estimativas, indicando que a simples decorrelação dos sinais, que utiliza apenas estatísticas de segunda ordem, não é suficiente para a solução do problema.

A formalização teórica para a abordagem apresentada em [64] foi feita posteriormente no trabalho de Pierre Comon [36], no qual discute-se a técnica de ICA, e sua aplicabilidade ao problema de separação de misturas lineares sem memória. Utilizando o teorema de Darmois-Skitovich, Comon mostrou que adaptar o sistema separador de forma a recuperar a independência entre os sinais levaria à separação das fontes.

Dessa forma, os critérios propostos para separação de fontes considerando a hipótese de independência entre as fontes deveriam expressar, direta ou indiretamente, a noção de independência entre os sinais à saída do sistema separador. O algoritmo de Héault, Jutten e Ans, conforme comentado anteriormente, expressa esta idéia através do uso de estatísticas de ordem superior, buscando minimizar a correlação não-linear entre os sinais. Outros critérios e algoritmos desenvolvidos para realizar a ICA serão descritos no capítulo 3.

Paralelamente aos trabalhos na linha de separação de misturas lineares sem memória, vemos o desenvolvimento de técnicas para modelos mais elaborados de mistura, para os quais ainda é possível empregar, com algumas restrições, a técnica de ICA para recuperação das fontes. Particularmente, em [138], os autores abordam o problema de misturas convolutivas e mostram que a abordagem de ICA ainda é válida para a separação dos sinais. Da mesma forma, em [126], mostra-se que a ICA também pode ser aplicada a um modelo específico de mistura não-linear, denominado mistura com não-linearidade posterior, ou *PNL (Post-Nonlinear)*.

## 1.1 Objetivos e organização da tese

A presente tese tem como objetivo apresentar novas abordagens voltadas para dois casos que se revestem de dificuldades particulares, no contexto de BSS. O primeiro se refere às misturas lineares com memória, ou convolutivas, onde o fenômeno de mistura e distorção dos sinais ocorrem conjuntamente. O segundo caso tratado inclui não-linearidades no processo de mistura, o que torna mais complexa a tarefa de recuperação dos sinais.

Em ambos os casos teremos por base a teoria de ICA, de forma que trataremos deste tema com certo detalhe nos dois capítulos subseqüentes. Sendo assim, o presente documento de tese está organizado da seguinte maneira:

No capítulo 2, apresentamos os fundamentos relacionados ao problema de separação de fontes, discutindo a técnica de análise por componentes independentes. Veremos também algumas outras abordagens utilizadas em BSS, que são aplicáveis quando hipóteses adicionais sobre as fontes são consideradas.

No capítulo 3, abordamos com maior detalhe a técnica de ICA, discutindo os principais critérios e algoritmos desenvolvidos para determinação das componentes independentes.

No capítulo 4, tratamos especificamente do problema de separação de misturas convolutivas. Muitos dos conceitos apresentados no capítulo 3 se aplicam diretamente a este caso. Comentamos algumas técnicas existentes, e apresentamos uma nova abordagem para separação dos sinais empregando filtros de erro de predição não-lineares. O uso de estruturas não-lineares tem como objetivo eliminar o caráter convolutivo da mistura e, dessa forma, realizar a separação efetiva das fontes por meio de técnicas consagradas desenvolvidas para o caso de mistura sem memória. Em particular, empregamos filtros de erro de predição baseados em sistemas fuzzy, que apresentam uma estrutura especialmente adequada para cenários nos quais as fontes apresentam distribuição discreta.

Já no capítulo 5 abordamos o problema de misturas não-lineares. Para o caso geral de mistura não-lineares, veremos que a metodologia da ICA não resulta na separação das fontes. Porém, para alguns tipos particulares de misturas, denominadas misturas com não-linearidade posterior, veremos que o método ainda garante a recuperação dos sinais. Uma dificuldade adicional do problema reside na necessidade de uma estrutura de separação não-linear, o que torna o processo de ajuste de parâmetros mais complicado do que no caso linear. Neste contexto, propomos um método para tentar contornar tal dificuldade, utilizando uma ferramenta de otimização evolutiva. O algoritmo utilizado, baseado em redes imunológicas artificiais, apresenta um bom desempenho em problemas de busca multimodal. Aliado ao método de otimização, empregamos um algoritmo de estimação de entropia baseado em estatísticas de ordem, a fim de avaliar a função custo a ser otimizada.

Finalmente, no capítulo 6, tecemos algumas conclusões sobre os problemas abordados e algumas perspectivas de continuação do trabalho.

# 2

## Separação Cega de Fontes: Princípios

Neste capítulo, abordaremos os princípios relacionados ao problema de separação cega de fontes. Veremos que é possível, a partir das hipóteses de que as fontes são independentes entre si e que possuem distribuição diferente de uma variável gaussiana, estabelecer um método para separação dos sinais originais. Esta abordagem, conhecida como *análise por componentes independentes*, ou ICA (*Independent Component Analysis*), é uma das principais ferramentas para solução do problema de BSS. Discutiremos também algumas outras abordagens, que se aplicam ao problema de separação caso hipóteses adicionais sobre as fontes e/ou sistema de mistura forem consideradas.

### 2.1 Definição do problema

Consideremos um conjunto de  $N$  fontes, cujas amostras compõem um vetor  $\mathbf{s}(k)$ , e um conjunto de  $M$  sinais captados pelos sensores, as observações, cujas amostras formam um vetor  $\mathbf{x}(k)$ . De maneira geral, podemos formular o vetor de observações  $\mathbf{x}(k)$  como sendo

$$\mathbf{x}(k) = \mathbf{F}(\mathbf{s}(k), \dots, \mathbf{s}(k-L), \mathbf{n}(k), k) \quad (2.1)$$

ou seja, ele é expresso por uma função  $\mathbf{F}$ , linear ou não (e possivelmente variante no tempo), que depende das fontes  $\mathbf{s}(k)$  e dos ruídos envolvidos  $\mathbf{n}(k)$ .

Evidentemente, se o mapeamento  $\mathbf{F}$  de mistura das fontes é conhecido, podemos estimar as fontes através do mapeamento inverso a  $\mathbf{F}$  (supondo que o processo é inversível). Ainda, se tivermos acesso, por um certo período de tempo, a amostras dos sinais das fontes, é possível identificar de maneira supervisionada o processo de mistura, e assim obter o mapeamento inverso. No entanto, o objetivo das técnicas de separação cega de fonte é estimar os sinais das fontes sem o conhecimento do processo de mistura, utilizando somente o vetor de observações.

O problema colocado dessa forma geral tem imenso apelo prático, já que em um grande número de aplicações podemos encontrar modelos que são casos particulares da formulação expressa por (2.1). Porém, infelizmente, a solução para o problema considerando um modelo tão abrangente ainda é desconhecida e, normalmente, adotam-se modelos simplificados que expressam casos particulares de (2.1). Em geral, podemos classificar os modelos adotados de acordo com algumas características específicas:

**Sistemas Lineares ou Não-Lineares** Um sistema misturador é dito linear se o mapeamento  $\mathbf{F}(\cdot)$  atende ao princípio da superposição, ou seja, nas situações em que:

$$\mathbf{F}(\alpha_1 \mathbf{s}_1(k) + \alpha_2 \mathbf{s}_2(k)) = \alpha_1 \mathbf{F}(\mathbf{s}_1(k)) + \alpha_2 \mathbf{F}(\mathbf{s}_2(k)), \quad (2.2)$$

para quaisquer constantes  $\alpha_1$  e  $\alpha_2$ , e vetores de sinais  $\mathbf{s}_1(k)$  e  $\mathbf{s}_2(k)$ . Caso contrário, o sistema misturador é dito não-linear. Veremos com mais detalhes no capítulo 5 o problema de separação de misturas não-lineares.

**Sistemas Instantâneos e com Memória** Caso os sinais observados representem uma mistura de amostras atuais e passadas das fontes, ou seja,  $L > 0$ , o sistema misturador é dito convolutivo, ou com memória. Nas situações em que  $L = 0$ , o caso mais simples, o sistema é dito instantâneo.

**Com Relação ao Número de Fontes e de Sensores** Na situação na qual o número de sensores é maior que o número de fontes ( $M > N$ ), temos o chamado caso de mistura sobre-determinada. Analogamente, quando o número de sensores é inferior ao número de fontes ( $M < N$ ), dizemos que se trata de uma mistura sub-determinada.

Neste primeiro momento consideraremos o modelo mais simples, no qual o processo

de mistura das fontes é representado por um sistema linear sem memória e recebe o nome de *mistura linear instantânea*. O modelo é completamente caracterizado por uma matriz de mistura  $\mathbf{A}_{M \times N}$ , de maneira que o vetor de observações  $\mathbf{x}$  é expresso por

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \quad (2.3)$$

ou seja, os sinais observados são resultado de uma combinação linear dos sinais das fontes.

A fim de expor os princípios relacionados ao problema de separação de fontes, consideraremos que as observações não contêm ruído, e que o número de sensores é igual ao de fontes ( $\mathbf{A}_{N \times N}$ ). Além disso, a menos que seja mencionado explicitamente, consideraremos que os sinais das fontes são processos estocásticos estacionários de média nula [109]. Dessa forma, sempre que possível, omitiremos o índice temporal.

## 2.2 BSS através da Análise por Componentes Independentes

Sendo o modelo um mapeamento linear, e considerando que a matriz de mistura é inversível, pode-se recuperar os sinais das fontes através de uma matriz  $\mathbf{W}$  tal que

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{W}\mathbf{x}(k) \\ &= \mathbf{W}\mathbf{A}\mathbf{s}(k) \quad , \\ &= \mathbf{s}(k) \end{aligned} \quad (2.4)$$

ou seja,  $\mathbf{W} = \mathbf{A}^{-1}$ . Porém, tendo em mente o objetivo das técnicas de BSS, como podemos obter os coeficientes de  $\mathbf{W}$  sem utilizar nenhum conhecimento explícito dos sinais das fontes e/ou da matriz de mistura  $\mathbf{A}$ ?

A resposta para esta pergunta reside no conjunto de hipóteses sobre os sinais das fontes. Uma vez que não se tem acesso diretamente aos sinais, o que possibilitaria o uso de técnicas supervisionadas para obtenção do sistema separador, deve-se ao menos considerar que alguma informação estatística sobre as fontes esteja disponível.

A hipótese central sobre as fontes que levou ao desenvolvimento de diferentes algoritmos é a de que as fontes são *estatisticamente independentes entre si*, uma hipótese até certo ponto restritiva, mas plausível em muitas aplicações práticas. Matematicamente, um vetor de variáveis aleatórias  $\mathbf{s}$  possui elementos estatisticamente indepen-

dentes entre si se

$$p_{\mathbf{s}}(\mathbf{s}) = p_{s_1}(s_1)p_{s_2}(s_2) \dots p_{s_N}(s_N), \quad (2.5)$$

onde  $s_1, s_2, \dots, s_N$  são as componentes do vetor  $\mathbf{s}$ , e  $p_{s_i}(\cdot)$  denota a função densidade de probabilidade, ou pdf (*Probability Density Function*) da componente  $s_i$ .

Considerando tal hipótese, Comon [36] mostrou em seu trabalho que sob certas condições é possível separar as fontes adaptando a matriz de separação de forma a restituir a independência entre os sinais estimados. Para tanto, utilizou o teorema de Darmois-Skitovich [77], que pode ser enunciado da seguinte maneira:

**Teorema 2.2.1** (Darmois-Skitovich). *Sejam  $s_1, s_2, \dots, s_N$  um conjunto de variáveis aleatórias de média nula e independentes entre si. Definindo-se duas variáveis  $x_1$  e  $x_2$  por*

$$\begin{aligned} x_1 &= a_1 s_1 + a_2 s_2 + \dots + a_N s_N \\ x_2 &= b_1 s_1 + b_2 s_2 + \dots + b_N s_N \end{aligned} \quad (2.6)$$

*Se  $x_1$  e  $x_2$  forem independentes entre si, então todos os  $s_i$ 's, para os quais  $a_i b_i \neq 0$ , serão variáveis com distribuição gaussiana.*

Através deste teorema nota-se que não seria possível obter *variáveis independentes* entre si a partir de misturas lineares de fontes *não-gaussianas*. Conseqüentemente, se considerarmos os coeficientes  $a_i$  e  $b_i$  como sendo os parâmetros da composição mistura/sistema separador  $\mathbf{WA}$ , veremos que as estimativas somente serão independentes, assumindo que nenhuma (ou apenas uma) delas é gaussiana, se efetivamente não houver mistura das fontes.

Portanto, considerando-se a restrição quanto a fontes com distribuição gaussiana, é possível efetuar a separação dos sinais utilizando a ICA, que pode ser definida formalmente como:

**Definição 2.2.1** (ICA). *A ICA de um vetor aleatório  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_M]^T$  consiste na determinação de uma matriz  $\mathbf{W}$  de tal maneira que os elementos do vetor aleatório  $\mathbf{y} = \mathbf{W}\mathbf{x}$  sejam tão estatisticamente independentes quanto possível, no sentido de otimizar uma função custo que expresse, direta ou indiretamente, uma certa noção de independência entre os sinais<sup>a</sup>.*

---

<sup>a</sup>No contexto de separação de fontes, a função custo também recebe o nome de *função contraste* [36].

Portanto, vemos que a recuperação das fontes se apóia em dois conceitos importantes: o de *independência estatística* e o de *não-gaussianidade*. Pode-se dizer que os métodos de ICA exploram, de uma forma ou de outra, estas duas características para compor critérios válidos para a determinação da estrutura de separação. Veremos com maior detalhe no capítulo 3 alguns dos principais critérios e algoritmos desenvolvidos utilizando a abordagem da ICA.

É importante ressaltar algumas ambigüidades na determinação da matriz  $\mathbf{W}$ . Note que a independência estatística entre os elementos de um vetor  $\mathbf{y}$  não será alterada caso os seus componentes estiverem dispostos em uma ordem arbitrária, nem se a amplitude de cada elemento for modificada. Em outras palavras, se  $\mathbf{y}$  possuir componentes mutuamente independentes, o vetor

$$\mathbf{y} = \mathbf{\Lambda} \mathbf{P} \mathbf{s}, \quad (2.7)$$

onde  $\mathbf{\Lambda}$  é uma matriz diagonal, e  $\mathbf{P}$  é uma matriz de permutação, também apresentará componentes independentes entre si. Dessa forma, a equação (2.7) mostra uma indeterminação inerente à ICA. Embora tais ambigüidades possam parecer uma grande limitação da técnica de separação, em grande parte das aplicações a informação das fontes está contida na forma de onda, e não propriamente na amplitude ou na ordem na qual os sinais estão dispostos nas saídas do sistema [33]. As condições sob as quais as fontes podem ser recuperadas são resumidas através do seguinte teorema [36]:

**Teorema 2.2.2** (Separabilidade). *O sistema apresentado em (2.3) é separável por meio da ICA, ou seja, é possível determinar uma matriz  $\mathbf{W}$  tal que o vetor  $\mathbf{y}$  corresponda às fontes, a menos das ambigüidades de permutação e escala dos sinais se, e somente se, a matriz de mistura  $\mathbf{A}$  possui posto completo e a mistura contém, no máximo, uma fonte gaussiana.*

### 2.2.1 Correlação x Independência

A descorrelação representa um conceito menos restritivo do que o da independência estatística. Por definição, um vetor de variáveis aleatórias  $\mathbf{s}$  possui elementos descorrelacionados se sua matriz de covariância é uma matriz diagonal [109], i.e.,

$$E \left\{ [\mathbf{s} - \mathbf{m}_s] [\mathbf{s} - \mathbf{m}_s]^T \right\} = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_N), \quad (2.8)$$

onde  $\mathbf{m}_s = E\{\mathbf{s}\}$ .

Se o vetor  $\mathbf{s}$  apresenta componentes independentes, seus elementos serão descorrelacionados. Todavia, a recíproca só será verdadeira para variáveis com distribuição gaussiana, as quais são completamente descritas por seus momentos de primeira e segunda ordem [109]. Portanto, a descorrelação dos sinais à saída do sistema separador não implica em separação dos sinais, já que os mesmos são, *ex hypothesi*, não-gaussianos [24]. Entretanto, tornar os sinais descorrelacionados, ou branqueá-los<sup>1</sup>, pode servir como passo intermediário à separação completa das fontes [82].

Para melhor compreender esta afirmação, considere a matriz de autocorrelação<sup>2</sup> das observações, dada por

$$\begin{aligned}\mathbf{R}_{\mathbf{x}} &= E \left\{ \mathbf{x}(k) \mathbf{x}(k)^T \right\} \\ &= \mathbf{A} \mathbf{R}_{\mathbf{s}} \mathbf{A}^T, \\ &= \mathbf{A} \mathbf{A}^T\end{aligned}\tag{2.9}$$

onde assumimos que a matriz de correlação das fontes é  $\mathbf{R}_{\mathbf{s}} = \mathbf{I}$ . Seja  $\mathbf{A} = \mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{V}^T$  a decomposição em valores singulares, ou SVD (*Singular Value Decomposition*), da matriz de mistura [57]. Podemos verificar que  $\mathbf{R}_{\mathbf{x}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ . Dessa forma, é possível definir uma matriz de transformação

$$\mathbf{T} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^T,\tag{2.10}$$

de forma que o vetor de dados transformados  $\bar{\mathbf{x}}(k) = \mathbf{T} \mathbf{x}(k)$  apresenta matriz de autocorrelação

$$\begin{aligned}\mathbf{R}_{\bar{\mathbf{x}}} &= E \left\{ \bar{\mathbf{x}}(k) \bar{\mathbf{x}}(k)^T \right\} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^T \mathbf{R}_{\mathbf{x}} \mathbf{U} \mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^T \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \mathbf{U} \mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{I}\end{aligned}\tag{2.11}$$

ou seja, os dados transformados são descorrelacionados.

Cabe ressaltar que a solução não é única, uma vez que toda transformação  $\mathbf{T} = \mathbf{Q} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{V}^T$ , para qualquer matriz ortogonal  $\mathbf{Q}$ , também tornará os dados descorrelacionados entre si. Além disso, como a descorrelação é uma condição necessária para a independência entre os sinais, entre as matrizes “branqueadoras” dos sinais existe

---

<sup>1</sup>Os dados descorrelacionados também são denotados como *esferizados*, ou *padronizados* [30].

<sup>2</sup>A matriz de autocorrelação se iguala à matriz de covariância para vetores com média nula.



aquela que não somente tornará as componentes descorrelacionadas, mas independentes entre si. Portanto, após o branqueamento dos dados observados utilizando  $\mathbf{T}$ , o espaço de busca da matriz de separação fica restrito ao grupo de matrizes ortogonais  $\mathbf{Q}$ . Em outras palavras, o processo de branqueamento reduz o problema a um no qual a matriz de mistura é ortogonal:

$$\begin{aligned}\bar{\mathbf{x}}(k) &= \mathbf{T}\mathbf{x}(k) \\ &= \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T\mathbf{A}\mathbf{s}(k) \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^T\mathbf{s}(k) \\ &= \mathbf{V}\mathbf{s}(k)\end{aligned}\quad (2.12)$$

Isto mostra, de outra forma, a razão pela qual restringe-se o número de fontes gaussianas presentes na mistura. Considere, por exemplo, um caso de mistura entre duas fontes com distribuição gaussiana de média nula. Nesta condição, tornar os sinais descorrelacionados é suficiente para garantir a independência, já que a distribuição gaussiana pode ser completamente caracterizada com estatísticas de segunda ordem [109]. Como a solução que branqueia a mistura não é única, não será possível determinar a matriz ortogonal de separação dos sinais.

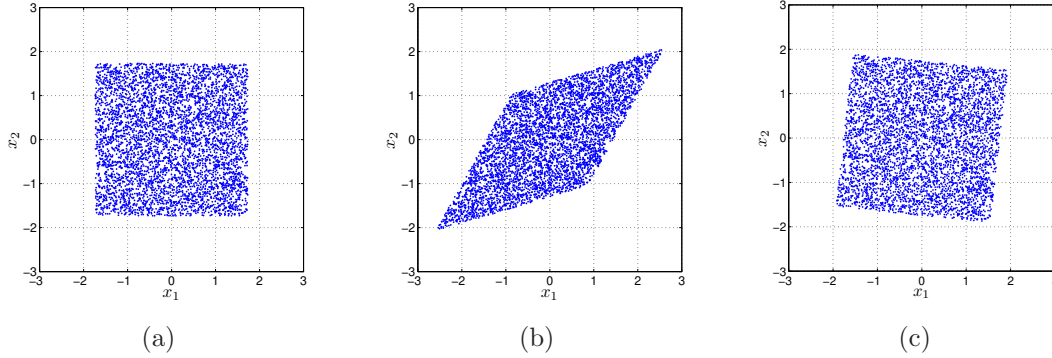
**Exemplo 2.2.1** (Exemplo com duas fontes). *Considere o caso em que temos duas fontes independentes, cada uma com distribuição uniforme no intervalo  $[-\sqrt{3}, \sqrt{3}]$ . A distribuição conjunta das fontes é dada por:*

$$p(x, y) = \begin{cases} \frac{1}{12}, & -\sqrt{3} \leq x, y \leq \sqrt{3} \\ 0, & \text{outros casos} \end{cases},$$

*e é mostrada graficamente na figura 2.1(a). A matriz de mistura utilizada é dada por:*

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.3 & 0.9 \end{bmatrix},$$

*e a distribuição dos dados misturados pode ser observada na figura 2.1 (b). Note que o efeito da mistura é tanto distorcer os eixos da figura representando a distribuição conjunta das fontes como ocasionar uma rotação na figura. O efeito do branqueamento pode ser visto como um passo intermediário à identificação completa da matriz de mistura. Note na figura 2.1(c) que após o branqueamento a forma da distribuição conjunta é restituída, restando apenas determinar a matriz ortogonal responsável pela rotação da pdf conjunta das fontes.*



**Figura 2.1:** Exemplo de mistura de duas fontes independentes com distribuição uniforme. (a) distribuição das fontes; (b) distribuição dos dados observados (após a mistura); (c) distribuição após o branqueamento.

## 2.3 Outras abordagens para Separação Cega de Fontes

Conforme estabelecido anteriormente, a solução do problema de separação cega de fontes por meio da ICA se baseia nas hipóteses de independência e não-gaussianidade das fontes. No entanto, dependendo da aplicação, pode-se ter acesso a mais informações *a priori* sobre a natureza das fontes, e até mesmo sobre o processo de mistura dos sinais.

Nesta seção, apresentaremos algumas abordagens para separação cega das fontes que consideram hipóteses adicionais sobre as fontes e/ou processo de mistura. Nestas condições, é possível obter métodos específicos para recuperação de fontes que, embora tenham aplicabilidade restrita, podem apresentar vantagens em relação à ICA.

### 2.3.1 Separação Cega de Sinais utilizando apenas estatísticas de segunda ordem

Uma classe de algoritmos simples e efetivos para separação de fontes pode ser obtida se as densidades espectrais de potência das fontes forem distintos entre si, i.e.,

$$E \{s_i(k) s_i(k-l)\} \neq E \{s_j(k) s_j(k-l)\}, \quad (2.13)$$

para  $i \neq j$ , e algum  $l \neq 0$ .

Diferentemente da abordagem utilizada na ICA, os algoritmos desta classe exploram a informação temporal dos sinais observados para efetuar a recuperação das fontes e, com isso, são capazes de realizar a separação dos sinais utilizando apenas estatísticas de segunda ordem.

Se considerarmos que os sinais observados foram “esferizados”, podemos dizer que a matriz de autocorrelação dos sinais observados para um atraso  $l$  será dada por

$$\mathbf{R}_{\bar{\mathbf{x}}}(l) = E \{ \bar{\mathbf{x}}(k) \bar{\mathbf{x}}^T(k-l) \} = \mathbf{Q} \mathbf{R}_{\mathbf{s}}(l) \mathbf{Q}^T, \quad (2.14)$$

onde  $\mathbf{R}_{\mathbf{s}}(l)$  denota a matriz de autocorrelação das fontes para um atraso  $l$ . Como, por hipótese, os sinais são independentes entre si (ou pelo menos descorrelacionados),  $\mathbf{R}_{\mathbf{s}}(l)$  será uma matriz diagonal, indicando que (2.14) representa a decomposição em auto-valores e auto-vetores de  $\mathbf{R}_{\bar{\mathbf{x}}}(l)$ . Dessa forma, realizar a decomposição de  $\mathbf{R}_{\bar{\mathbf{x}}}(l)$  resulta na obtenção da matriz de rotação a ser utilizada na recuperação das fontes. O método descrito corresponde ao algoritmo AMUSE [129] e depende da escolha correta do atraso  $l$  para o qual a matriz  $\mathbf{R}_{\mathbf{s}}(l)$  apresente valores distintos.

Uma extensão deste método é obtida se, ao invés de considerarmos apenas um único atraso  $l$ , incorporarmos a informação presente nas matrizes de covariância para diversos atrasos. Uma vez que a escolha correta deste atraso é crítica para o problema de separação, ao levarmos em consideração diferentes matrizes de covariância tornamos o método mais robusto a erros na estimativa do valor correto de  $l$ . Este é o princípio presente em algoritmos como SOBI [17] e TDSEP [143].

Os métodos mencionados provêem formas simples e eficientes de recuperação das fontes. No entanto, o ponto central destas técnicas reside na hipótese de que todas as fontes possuem espectros diferentes entre si. Ainda que os espectros sejam diferentes, porém similares entre si, a separação das fontes, embora teoricamente viável, não ocorre na prática [66, 34]. Por outro lado, nota-se que não é necessário que os sinais sejam independentes entre si, apenas descorrelacionados. E, mais importante, não é feita nenhuma restrição quanto a fontes gaussianas.

Outro cenário no qual é possível obter métodos para separação de sinais baseados apenas em estatísticas de segunda ordem ocorre quando os sinais das fontes são não-estacionários, como no caso de sinais de voz. Nesta situação mostra-se que a descorrelação entre os sinais estimados é uma condição suficiente para a separação das fontes [97, 111], não sendo necessário fazer uso de estatísticas de ordem superior para a determinação do sistema separador.

### 2.3.2 Análise por componentes independentes não-negativas

Outra classe de algoritmos para BSS pode ser obtida se as fontes não assumirem valores negativos (ou, equivalentemente, positivos), uma condição natural em aplicações como, por exemplo, análise de imagens [34].

A hipótese de não-negatividade ou de positividade de sinais está presente em métodos de análise de dados como a fatoração não-negativa de matrizes [85], ou NMF (*Non-negative Matrix Factorization*) e a fatoração positiva de matrizes [106], ou PMF (*Positive Matrix Factorization*). No entanto, nem sempre é possível garantir a unicidade da fatoração, e dessa forma a aplicação direta de tais métodos no problema de BSS requer que restrições adicionais sobre os sinais originais sejam consideradas.

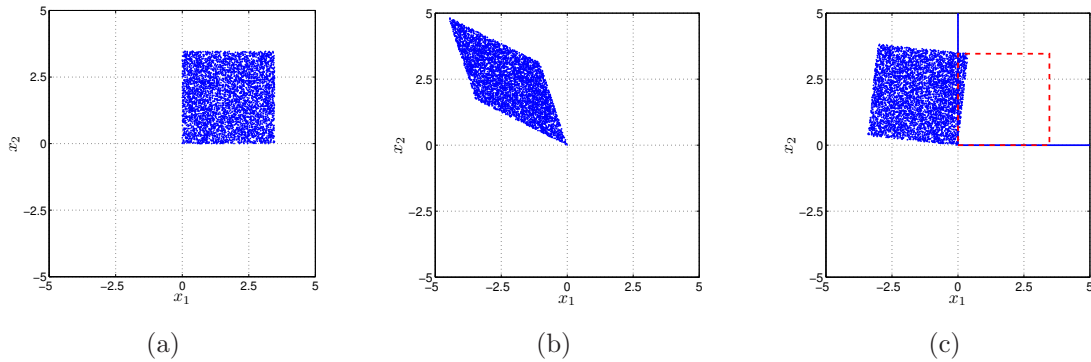
Neste sentido, a análise por componentes independentes não-negativas trabalha conjuntamente com as hipóteses de independência e não-negatividade das fontes, e estas considerações permitem derivar algoritmos simples e efetivos para separação dos sinais. O resultado central desta abordagem é resumido pelo seguinte teorema [112]:

**Teorema 2.3.1** (Separabilidade de mistura de fontes não-negativas). *Seja  $\mathbf{s}$  um vetor de componentes independentes entre si, sendo que cada um dos componentes possui variância unitária e assuma somente valores não-negativos, com  $P(s_i < \delta) > 0$  para qualquer  $\delta > 0$ . Seja também  $\mathbf{y} = \mathbf{U}\mathbf{s}$ , onde  $\mathbf{U}$  é uma matriz quadrada ortonormal ( $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ ). Então  $\mathbf{U}$  será uma matriz de permutação se, e somente se, todos os componentes de  $\mathbf{y}$  forem não-negativos. Portanto, nesta condição, o vetor  $\mathbf{y}$  conterá os sinais das fontes, a menos de uma ambigüidade de permutação de seus elementos.*

Dessa forma, é possível separar as fontes construindo funções custo que expressem a condição de não-negatividade dos elementos de  $\mathbf{y}$ . A figura 2.2 mostra um exemplo de duas fontes com distribuição uniforme no intervalo  $[0, 1]$ . Após a etapa de branqueamento dos dados, recuperar as fontes seria equivalente a determinar a rotação dos dados que torna as estimativas novamente não-negativas (c). Alguns exemplos de algoritmos derivados a partir desta abordagem podem ser encontrados em [113, 105].

### 2.3.3 Análise de componentes esparsos

A análise por componentes esparsos, ou SCA (*Sparse Component Analysis*), assim como as outras abordagens descritas nas seções anteriores, se apóia em características



**Figura 2.2:** Exemplo de mistura de duas fontes que não assumem valores negativos. (a) Distribuição das fontes; (b) Distribuição das observações; (c) Distribuição após o branqueamento.

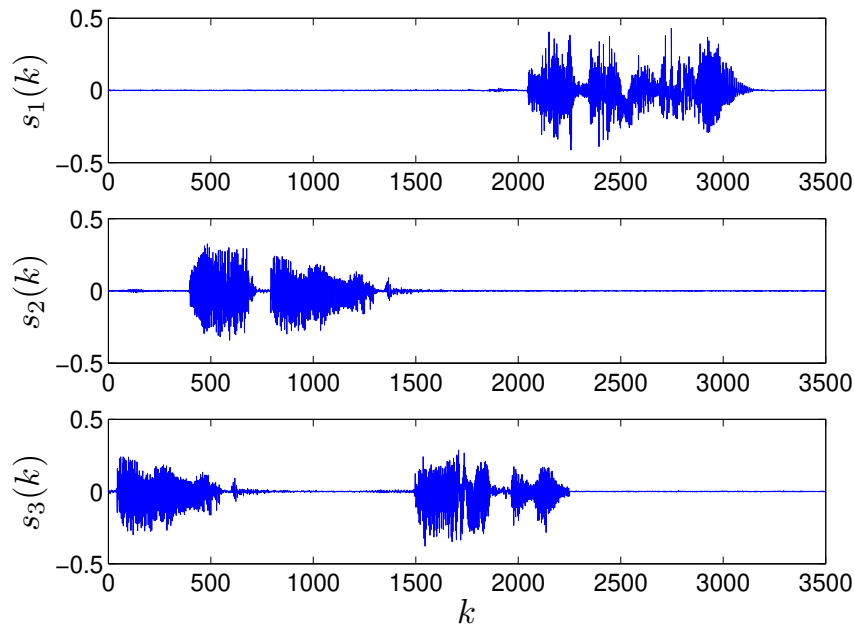
adicionais das fontes para realizar a separação dos sinais. Conforme o próprio nome indica, a abordagem baseia-se na hipótese de esparsidade das fontes, ou seja, na maior parte do tempo as fontes assumem valores próximos a zero [20], como na figura 2.3. O grande interesse nesta ferramenta por parte dos pesquisadores justifica-se principalmente por ela prover meios de separação de sinais no caso em que o número de sensores é inferior ao número de fontes [55, 88].

Considere o exemplo apresentado na figura 2.4, onde são mostradas as distribuições conjuntas das observações em um caso com 3 fontes e 2 sensores. Na figura 2.4(a) as fontes possuem distribuição uniforme, e na figura 2.4(b) as fontes são esparsas.

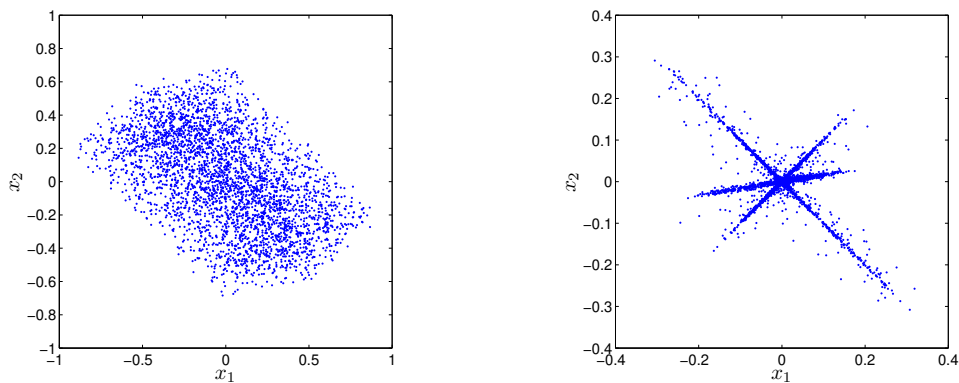
É possível notar que no caso de fonte esparsas a distribuição conjunta dos dados tende a se acumular ao longo de algumas direções específicas, que em particular representam as colunas da matriz de mistura. Devido à esparsidade das fontes, é bem provável que apenas uma das fontes seja diferente de zero em um dado instante de tempo, e como consequência os dados observados formarão aglomerados ao longo da direção alinhada à coluna da matriz de mistura associada a esta fonte. Dessa forma, identificar tais direções na distribuição conjunta dos dados corresponderá a identificar a matriz de mistura, processo que pode ser efetuado por técnicas de clusterização [62]. Outro ponto a ser observado é que ainda seria possível identificar, sem conhecimento *a priori*, o número de sinais que compõem a mistura [8, 9].

É possível empregar a SCA para realizar a separação de fontes que não são temporalmente esparsas, mas que apresentem esta característica em outros domínios, como por exemplo no domínio freqüencial ou tempo-freqüencial [55, 119, 104]. Para isso

aplica-se uma transformação linear ao conjunto de dados de maneira que, no domínio transformado, seja possível empregar a SCA e, posteriormente, obter as estimativas das fontes no domínio temporal através da transformada inversa.



**Figura 2.3:** Exemplo de fontes esparsas



(a) Distribuição das misturas de fontes não-esparsas.

(b) Distribuição das misturas de fontes esparsas.

**Figura 2.4:** Saídas de um sistema misturador sub-determinado (3 fontes e 2 sensores).

### 2.3.4 Abordagem Bayesiana

Os métodos descritos até o momento dependem da validade de hipóteses específicas, restringindo a aplicabilidade das técnicas desenvolvidas. A abordagem Bayesiana [100], por outro lado, segue um caminho ligeiramente diferente, e permite que informações *a priori* sejam incorporadas naturalmente ao problema de separação de fontes.

De maneira geral, a abordagem Bayesiana consiste em obter estimativas das fontes e da matriz de mistura a partir da maximização da função de probabilidade *a posteriori*, dada por  $p(\mathbf{A}, \mathbf{s}|\mathbf{x})$ . Utilizando a regra de Bayes [109] pode-se observar que

$$p(\mathbf{A}, \mathbf{s}|\mathbf{x}) \propto P(\mathbf{x}|\mathbf{A}, \mathbf{s})P(\mathbf{A})P(\mathbf{s}), \quad (2.15)$$

onde  $p(\mathbf{s})$  e  $p(\mathbf{A})$  denotam, respectivamente, as funções densidade de probabilidades das fontes e da matriz de mistura, e  $p(\mathbf{x}|\mathbf{A}, \mathbf{s})$  a função de verossimilhança [81, 80] dos dados.

Nota-se a partir de (2.15) que a abordagem é extremamente abrangente. Através dos termos  $p(\mathbf{A})$  e  $p(\mathbf{s})$  é possível incluir no problema informações adicionais sobre o processo de mistura e sobre as fontes, possibilitando o uso da abordagem em uma ampla gama de situações. Outro ponto a ser ressaltado é que, embora a formulação apresentada em (2.15) considere o modelo de mistura linear, a abordagem Bayesiana permite o uso de modelos mais elaborados, tornando possível, por exemplo, tratar o caso com ruído e de misturas com mais fontes do que sensores [133, 140].

Em contrapartida, a maximização de (2.15) representa um grande desafio. A função a ser otimizada é, em geral, não-convexa, e conta com um grande número de parâmetros a serem ajustados.

## 2.4 Sumário

Neste capítulo, apresentamos os fundamentos da abordagem conhecida como análise por componentes independentes, que explora a hipótese de independência entre as fontes para realizar a separação. O método consiste em determinar a estrutura de separação buscando reestabelecer a independência entre os sinais em suas saídas. A abordagem é efetiva na solução do problema de BSS desde que haja, no máximo, uma fonte com distribuição gaussiana.

Apresentamos também alguns métodos alternativos, que consideram outras hipóteses

no problema de separação. Isto possibilita o desenvolvimento, em alguns casos, de ferramentas específicas, capazes de tratar casos não abordados pela ICA e, em outros, como é o caso da abordagem Bayesiana, expandir o horizonte de aplicações, tendo em contrapartida uma maior complexidade computacional.



# 3

## Análise por Componentes Independentes: Critérios e Estratégias

A essência da técnica de análise por componentes independentes, conforme discutido na seção 2.2, consiste em determinar a estrutura de separação de forma a obter estimativas independentes entre si, o que garante a recuperação das fontes caso não haja mais do que uma fonte gaussiana.

Neste capítulo, descreveremos as principais estratégias desenvolvidas para ICA, visando a separação de fontes no caso de misturas lineares instantâneas, uma vez que tais técnicas estão na base dos problemas a serem tratados nos capítulos subsequentes. Apresentaremos os critérios e algoritmos de otimização da estrutura de separação, fazendo algumas considerações sobre as relações existentes entre as diferentes propostas.

### 3.1 Critérios para ICA

Os critérios desenvolvidos para ICA exploram direta ou indiretamente as hipóteses de independência e não-gaussianidade das fontes. Nesta seção apresentaremos as principais abordagens desenvolvidas nesse contexto.

### 3.1.1 Informação mútua

A definição de um critério sólido que reflita de forma direta a condição de independência entre os sinais estimados pode ser obtida através de conceitos provenientes da *Teoria da Informação*, em particular dos conceitos de *Entropia* e de *Informação Mútua*.

**Definição 3.1.1** (Entropia Diferencial). *Seja uma variável aleatória  $\alpha$  com função densidade de probabilidade  $p_\alpha(\alpha)$ . Defina-se a entropia diferencial, ou simplesmente entropia,  $H(\alpha)$  por*

$$H(\alpha) = -E[\log p_\alpha(\alpha)] = - \int_{-\infty}^{\infty} p_\alpha(\tau) \log p_\alpha(\tau) d\tau. \quad (3.1)$$

A entropia<sup>1</sup> de uma variável aleatória fornece uma idéia da incerteza sobre o valor que esta variável assume [38]. Pode-se mostrar que a distribuição gaussiana apresenta a maior entropia dentre as distribuições com mesma variância. Outro resultado importante é que, se restringirmos a comparação a pdf's definidas apenas em um intervalo finito, a distribuição uniforme neste intervalo apresentará a máxima entropia.

De maneira similar, define-se a *entropia condicional*:

**Definição 3.1.2** (Entropia Condicional). *Seja a pdf condicional de  $\alpha$  dado  $\beta$ , representada por  $p_{\alpha|\beta}(\alpha | \beta)$ . A entropia condicional é dada por:*

$$H(\alpha | \beta) = -E[\log p_{\alpha|\beta}(\alpha | \beta)] = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\alpha,\beta}(\tau, v) \log p_{\alpha|\beta}(\tau | v) d\tau dv. \quad (3.2)$$

A entropia condicional fornece uma idéia sobre a incerteza restante em  $\alpha$  após observarmos o valor de  $\beta$ . Assim, é possível definir a *Informação Mútua*  $I(\alpha, \beta)$ , que reflete a quantidade de informação que  $\beta$  contém sobre  $\alpha$ , e vice-versa:

---

<sup>1</sup>A quantidade  $H(\alpha)$  também pode ser definida para variáveis discretas. Neste caso  $H(\alpha) = -\sum_i p_i \log p_i$ , e apresenta propriedades semelhantes à entropia diferencial [38].

**Definição 3.1.3** (Informação Mútua). *A informação mútua entre duas variáveis aleatórias  $\alpha$  e  $\beta$  é definida como*

$$I(\alpha, \beta) = H(\alpha) - H(\alpha | \beta) = H(\beta) - H(\beta | \alpha). \quad (3.3)$$

Note que, utilizando (3.1) e (3.2), podemos mostrar que a informação mútua também pode ser expressar por

$$\begin{aligned} I(\alpha, \beta) &= \int p_{\alpha, \beta}(\alpha, \beta) \log \frac{p_{\alpha, \beta}(\alpha, \beta)}{p_{\alpha}(\alpha) p_{\beta}(\beta)} d\alpha d\beta, \\ &= D(p_{\alpha, \beta}(\alpha, \beta) \| p_{\alpha}(\alpha) p_{\beta}(\beta)) \end{aligned} \quad (3.4)$$

onde  $D(p \| q)$  denota a *divergência de Kulback-Leibler* entre as distribuições  $p$  e  $q$ . Embora nos seja útil encará-la como uma medida de distância entre as distribuições, cabe mencionar que a divergência não é matematicamente uma distância, uma vez que não é simétrica e não satisfaz a desigualdade do triângulo.

Neste contexto, a informação mútua pode ser interpretada como uma medida de proximidade entre a distribuição conjunta de  $\alpha$  e  $\beta$ , e o produto de suas distribuições marginais. Isto, aliado ao fato de que a divergência de Kullback-Leibler entre duas distribuições é sempre maior ou igual a zero, e nula somente se as distribuições forem iguais, demonstra que a informação mútua entre duas variáveis aleatórias *se iguala a zero se, e somente se,  $\alpha$  e  $\beta$  forem independentes*. Isto proporciona uma certa noção do grau de dependência estatística entre as variáveis, de forma que a minimização da informação mútua configura-se como um critério para a adaptação da estrutura de separação.

Utilizando esta mesma idéia de comparação entre a distribuição conjunta e o produto das distribuições marginais, a extensão do conceito de informação mútua entre duas variáveis para um vetor com  $N$  componentes é feita de forma direta:

$$I(\alpha_1, \dots, \alpha_N) = D(p_{\alpha_1, \dots, \alpha_N}(\alpha_1, \dots, \alpha_N) \| p_{\alpha_1}(\alpha_1) \cdots p_{\alpha_N}(\alpha_N)), \quad (3.5)$$

que ainda pode ser expressa por

$$I(\alpha_1, \alpha_2, \dots, \alpha_n) = \sum_{i=1}^N H(\alpha_i) - H(\alpha_1, \alpha_2, \dots, \alpha_n), \quad (3.6)$$

indicando que minimizar a informação mútua equivale a tornar a entropia do vetor

$[\alpha_1, \alpha_2, \dots, \alpha_n]$  a mais próxima possível da soma das entropias marginais [82].

Dessa forma, ao calcular a informação mútua entre as saídas do sistema separador, podemos obter a seguinte expressão:

$$\begin{aligned} I(\mathbf{y}) &= \sum_{i=1}^N H(y_i) - H(\mathbf{y}) \\ &= \sum_{i=1}^N H(y_i) - H(\mathbf{x}) - \log |\det \mathbf{W}| \end{aligned} \quad (3.7)$$

Note que dentre os termos presentes em (3.7), a entropia do vetor  $\mathbf{x}$  não depende dos parâmetros da matriz  $\mathbf{W}$ . Assim, aplicar o critério de minimização da informação mútua no problema de BSS resulta no seguinte problema de otimização:

$$\min_{\mathbf{W}} \sum_i H(y_i) - \log |\det \mathbf{W}|. \quad (3.8)$$

### 3.1.2 Estatísticas de ordem superior

Uma vez que o uso de estatísticas de segunda ordem não possibilita, em geral, a formulação de um critério completo de BSS, alguns pesquisadores consideraram o uso de estatísticas de ordem superior a 2 como forma de medir a dependência estatística dos sinais, empregando os *cumulantes conjuntos* das estimativas [23, 36].

**Definição 3.1.4** (Cumulante Conjunto). *Dado o conjunto de variáveis aleatórias  $x_1, x_2, \dots, x_p$ , o cumulante conjunto de ordem  $r$  destas variáveis pode ser definido como [103]*

$$cum(x_1^{k_1}, x_2^{k_2}, \dots, x_p^{k_p}) \triangleq (-j)^r \frac{\partial^r \Psi(\omega_1, \dots, \omega_p)}{\partial \omega_1^{k_1} \dots \partial \omega_p^{k_p}} \bigg|_{\omega_1 = \dots = \omega_p = 0}, \quad (3.9)$$

onde  $\Psi(\omega_1, \dots, \omega_p)$  representa o logaritmo da função característica conjunta, e  $k_1 + k_2 + \dots + k_p = r$ .

Assim como a correlação, os cumulantes conjuntos serão nulos para qualquer ordem  $r$  se os dados forem independentes entre si [109].

Embora a independência esteja, a rigor, ligada a todos os cumulantes, uma abordagem

presente na literatura consiste em considerar apenas alguns cumulantes [23, 36, 84]. Em [23], por exemplo, considera-se o *tensor de cumulantes*<sup>2</sup> de quarta ordem, no qual o elemento  $ijkl$  denota o valor  $cum(x_i, x_j, x_k, x_l)$ . Assim, minimizar a dependência entre as estimativas consiste em minimizar os elementos  $ijkl$  do tensor, para  $i \neq j \neq k \neq l$ , i.e., os elementos fora da diagonal principal, em consonância com a idéia de diagonalização utilizada no processo de branqueamento dos dados. Entre os algoritmos baseados nessa idéia destaca-se o JADE (*Joint Approximate Diagonalization of EigenMatrices*) [23, 27], na qual um método engenhoso, empregando o algoritmo de diagonalização de Jacobi, é derivado para a minimização dos cumulantes conjuntos das estimativas.

### 3.1.3 Correlação não-linear

Outra forma de se incorporar a informação presente em estatísticas de ordem superior é através do uso de funções não-lineares na medida de correlação entre os sinais analisados. Desta maneira, a regra de adaptação para os coeficientes da estrutura de separação visa minimizar a *correlação não-linear* entre as saídas  $y_i$  e  $y_j$ , que é definida como:

$$E \{ f(y_i) g(y_j) \}, \quad (3.10)$$

onde  $f(\cdot)$  e  $g(\cdot)$  são funções não-lineares ímpares.

A razão para o sucesso desta abordagem torna-se mais clara ao considerarmos as expansões em série de Taylor das funções  $f$  e  $g$ . De fato, podemos reescrever (3.10) como

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} f_k g_l E \{ y_i^k y_j^l \} = 0 \quad \forall i, j, \quad (3.11)$$

onde  $f_k$  e  $g_l$  correspondem, respectivamente, ao  $k$ -ésimo e ao  $l$ -ésimo coeficiente da expansão de  $f(\cdot)$  e  $g(\cdot)$ , respectivamente. Vemos com isso que, implicitamente, o uso de funções não-lineares introduz estatísticas de ordem superior ao critério de adaptação. Em particular, note que (3.11) pode ser nula caso todos momentos cruzados entre  $y_i$  e  $y_j$  forem iguais a zero, impondo uma condição mais restritiva do que simplesmente a decorrelação entre os sinais. Obviamente não há garantias que um certo par de não-linearidades leve à condição de independência das fontes.

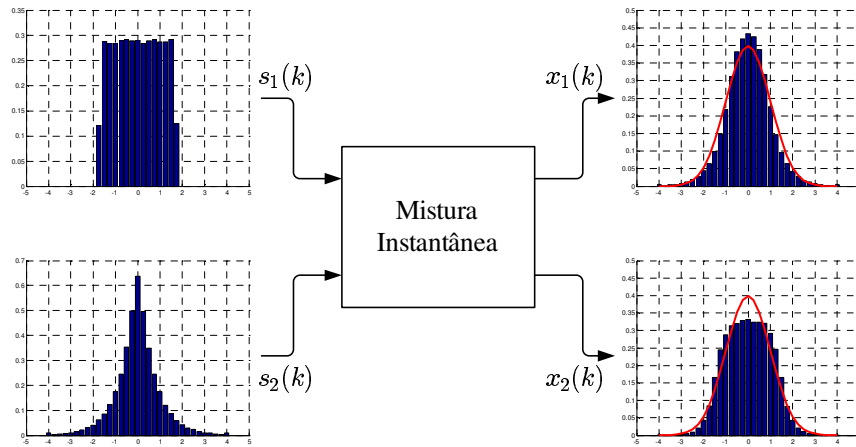
---

<sup>2</sup>O tensor é uma estrutura que pode ser compreendida aqui como uma extensão do conceito de matriz, onde cada elemento é indexado através de não dois, mas um número maior de índices.

### 3.1.4 Maximização da não-gaussianidade

Outra estratégia efetiva na separação de fontes consiste em adaptar a matriz de separação de forma a maximizar a não-gaussianidade dos sinais estimados. Embora não seja evidente, o princípio leva à estimação de sinais independentes entre si, mostrando-se um método apropriado para a ICA.

A abordagem pode ser compreendida através do teorema central do limite [109], um resultado clássico da teoria de probabilidade. Segundo o teorema, a função densidade de probabilidade da soma de variáveis aleatórias independentes, sob certas condições, tende à de uma variável gaussiana [37]. Portanto, de certa forma, a pdf da soma de duas variáveis aleatórias independentes está mais “próxima” de uma distribuição gaussiana do que qualquer uma das variáveis originais [66] (vide figura 3.1).



**Figura 3.1:** Exemplo de mistura de duas fontes independentes, uma com distribuição uniforme, e outra com distribuição laplaciana. Nota-se que as observações possuem perfil mais próximo ao de uma variável aleatória gaussiana do que as variáveis originais.

Tomemos o modelo dado por (2.3), onde as observações são o resultado de uma mistura linear instantânea. Digamos que estejamos interessados em recuperar apenas uma das fontes e, para isso, utilizaremos um vetor de separação  $\mathbf{w}_1$ :

$$y_1 = \mathbf{w}_1^T \mathbf{x} = \mathbf{q}^T \mathbf{s} = \sum_{i=1}^N q_i s_i \quad , \quad (3.12)$$

onde  $\mathbf{q} = \mathbf{w}_1^T \mathbf{A}$  representa o efeito conjunto da mistura e do sistema separador. Note que  $y_1$ , uma estimativa da fonte original, é uma combinação linear do vetor de fontes

s. Assim, para que se tenha uma estimativa do sinal original, basta que  $\mathbf{q}$  seja um vetor com apenas um elemento não nulo.

Como a estimativa  $y_1 = \mathbf{q}_1^T \mathbf{s}$  é uma soma ponderada dos sinais das fontes, podemos dizer que a sua distribuição será mais próxima à de uma variável gaussiana do que aquela de um dos sinais  $s_i$ . Portanto, pensando de forma inversa,  $y_1$  será o menos gaussiana possível quando sua distribuição for igual à de qualquer uma das fontes  $s_i$ , ou seja, quando  $\mathbf{q}$  possuir apenas um elemento não nulo. Dessa forma, obter o vetor  $\mathbf{w}_1$  que *maximize a não-gaussianidade de  $\mathbf{w}_1^T \mathbf{x}$*  configura-se como uma possível estratégia para recuperação da fonte.

A forma clássica de se averiguar se uma variável possui distribuição gaussiana é através de sua *curtose* [109], definida, de acordo com a notação em (3.9), como:

$$K(x) = \text{cum}(x, x, x, x) = E\{x^4\} - 3(E\{x^2\})^2. \quad (3.13)$$

Para uma ampla gama de distribuições a curtose apresenta valores diferentes de zero, sendo a distribuição gaussiana uma das poucas exceções. De fato, é usual classificar as distribuições com relação aos valores de suas curtoses: caso  $K(x) > 0$ , diz-se que  $x$  possui distribuição *super-gaussiana*; caso  $K(x) < 0$  diz-se que  $x$  possui distribuição *sub-gaussiana*. Portanto, um critério para se maximizar a não-gaussianidade do sinal estimado seria

$$\max_{\mathbf{w}} |K(y_i)|, \quad (3.14)$$

tendo como principal atrativo sua simplicidade. Sua estimação, a partir de um conjunto de dados, pode ser feita através de uma

Uma outra opção de medida de não-gaussianidade, mais robusta no que diz respeito a *outliers*, é a *Negentropia*. Lembrando que *uma variável aleatória com distribuição gaussiana possui a maior entropia dentre as demais distribuições com mesma variância* [109], define-se a negentropia:

**Definição 3.1.5** (Negentropia). *A negentropia de uma variável aleatória (ou equivalentemente, um vetor de v.a.'s) corresponde a*

$$J_{\text{Negentropia}}(y) = H(y_{\text{gauss}}) - H(y), \quad (3.15)$$

*onde  $y_{\text{gauss}}$  representa uma variável aleatória com distribuição gaussiana e mesma média e variância de  $y$ .*

Uma vez que a entropia de uma variável aleatória é máxima para a distribuição gaussiana, a negentropia sempre assumirá valores maiores ou iguais a zero, ocorrendo a igualdade somente quando  $y$  possuir distribuição normal. Dessa forma a negentropia fornece uma idéia da distância entre a distribuição de  $y$  e a variável gaussiana. O conceito pode ser estendido diretamente para um vetor  $\mathbf{y}$  de variáveis aleatórias, sendo  $J_{Negentropia}(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$ , onde  $\mathbf{y}_{gauss}$  representa um vetor aleatório com distribuição gaussiana e matriz de covariância igual à de  $\mathbf{y}$ .

### Extração cega de fontes

Diferentemente das estratégias anteriores, a abordagem de maximização da não-gaussianidade, por se basear em uma medida que depende somente de um dos sinais estimados, pode ser utilizada para estimação individual das componentes independentes. As técnicas que se baseiam nesta idéia são comumente associadas ao problema de *Extração Cega de Fontes* [89, 39], muito semelhante ao problema de BSS.

No problema de extração de fontes o objetivo não é obter todos os sinais das fontes, e sim conseguir extrair da mistura apenas os sinais de interesse, mas ainda considerando que estes são independentes dos demais sinais presentes na mistura. Se o número de sinais extraídos for igual ao número total de fontes presentes veremos que os dois problemas são equivalentes.

O procedimento para extração de mais de uma fonte pode ser efetuado utilizando-se pelo menos duas estratégias distintas, porém ambas explorando a idéia de que os vetores  $\mathbf{w}_i$ , obtidos a partir de dados branqueados, serão necessariamente ortogonais:

1. **Estimação serial:** Nesta abordagem as componentes são estimadas seqüencialmente. Inicia-se estimando, sem restrição, a primeira componente independente, e com isso obtém-se um vetor  $\mathbf{w}_1$  (que deve ter norma unitária). A partir da segunda componente, a adaptação dos vetores  $\mathbf{w}_i$  deve ser feita sempre realizando um passo intermediário de ortonormalização do vetor sendo ajustado em relação aos demais já obtidos, o que pode ser feito com o método de Gram-Schmidt [57]. Esta abordagem também é conhecida como *separação de fontes por deflação* (do inglês *Deflation approach*).
2. **Estimação paralela:** Neste caso um certo número de fontes será estimado ao mesmo tempo, adaptando-se paralelamente os vetores  $\mathbf{w}_i$ . No entanto, a cada passo executado, é necessário garantir que as soluções encontradas são ortonormais entre si, o que também pode ser efetuado com o método de Gram-Schmidt.



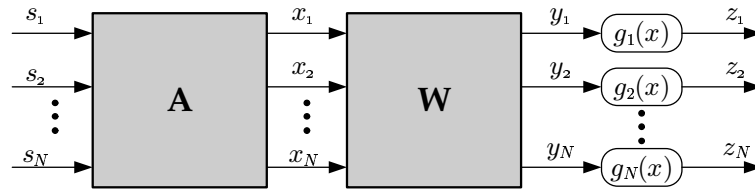
No entanto, é possível utilizar uma técnica de ortogonalização simétrica [66], dada por

$$\tilde{\mathbf{W}} \leftarrow \left( \tilde{\mathbf{W}} \tilde{\mathbf{W}}^T \right)^{1/2} \tilde{\mathbf{W}}, \quad (3.16)$$

onde  $\tilde{\mathbf{W}}$  representa uma matriz composta pelos vetores  $\mathbf{w}_i$ .

### 3.1.5 Princípio Infomax e a abordagem de Máxima Verossimilhança

Uma abordagem interessante para a obtenção do sistema de separação das fontes está relacionada a um sólido paradigma de treinamento não-supervisionado de redes neurais artificiais, conhecido como princípio *Infomax* (*Information Maximization*).



**Figura 3.2:** Estrutura de uma rede neural artificial.

Considere a estrutura apresentada na figura 3.2, que representa uma rede neural artificial. O bloco  $\mathbf{g}(\cdot)$  aplica uma não-linearidade a cada saída  $y_i$ , ou seja:

$$\mathbf{g}(\mathbf{y}) = \begin{bmatrix} g_1(y_1) & g_2(y_2) & \cdots & g_N(y_N) \end{bmatrix}^T, \quad (3.17)$$

onde  $g_i(\cdot)$  são funções não-lineares monotonicamente crescentes, com  $g_i(-\infty) = 0$  e  $g_i(\infty) = 1$ .

A abordagem infomax no contexto de separação de fontes foi apresentada por Bell e Sejnowski [16], embora os resultados principais tenham sido estudados de forma independente por outros pesquisadores [101]. Segundo o princípio, os coeficientes da rede neural devem ser ajustados de forma a *maximizar a transferência de informação entre suas entradas e saídas*, o que, em outras palavras, corresponde a adaptar a matriz  $\mathbf{W}$  a fim de se *maximizar a informação mútua* entre  $\mathbf{x}$  e  $\mathbf{z}$ .

Na situação em que o ruído é desprezível, mostra-se que a abordagem de maximização da informação mútua equivale à *maximização da entropia da saída*, estratégia conhecida na literatura como *MaxEnt* [82]. Isto ocorre porque na ausência de ruído o

mapeamento entre  $\mathbf{x}$  e  $\mathbf{z}$  torna-se determinístico, e nesta condição a entropia condicional  $H(\mathbf{z} | \mathbf{x})$  não dependerá dos parâmetros da rede [16].

Note que, reescrevendo (3.7), obtemos

$$H(\mathbf{z}) = \sum_{i=1}^N H(z_i) - I(\mathbf{z}), \quad (3.18)$$

revelando que a maximização da entropia conjunta está ligada tanto à maximização das entropias marginais como à minimização da informação mútua entre os elementos de  $\mathbf{z}$ . Com isso, a maximização da transferência de informação tenderá, em geral, a reduzir a dependência estatística entre as saídas da rede neural e, conseqüentemente, dos elementos do vetor  $\mathbf{y}$ .

A importância das não-linearidades neste problema se revela de duas formas. Primeiramente, ela limita o valor das variáveis  $z_i$ , e assim restringe o valor de suas entropias, que a princípio poderiam crescer indefinidamente. Em segundo lugar, verifica-se que a obtenção de variáveis independentes à saída da rede está intimamente ligada à escolha adequada das funções não-lineares, e relaciona-se aos tipos de fontes presentes na mistura.

Considerando a estrutura descrita na figura 3.2, pode-se verificar que

$$H(\mathbf{z}) = H(\mathbf{x}) + E \left\{ \sum_{i=1}^N \log(g'_i(\mathbf{w}_i \mathbf{x})) \right\} + \log(|\det(\mathbf{W})|), \quad (3.19)$$

onde  $g'_i$  representa a derivada da função  $g_i$  e  $\mathbf{w}_i$  denota a  $i$ -ésima linha de  $\mathbf{W}$ . Como a entropia conjunta dos dados  $\mathbf{x}$  não depende da matriz de separação, não influirá no processo de adaptação e, portanto, a aplicação do princípio Infomax no problema de BSS resulta no problema de maximização da seguinte função custo:

$$J_{Infomax}(\mathbf{W}) \triangleq E \left\{ \sum_{i=1}^N \log(g'_i(\mathbf{w}_i \mathbf{x})) \right\} + \log(|\det(\mathbf{W})|). \quad (3.20)$$

### Máxima verossimilhança

Cabe comentar que, embora desenvolvida a partir de um princípio diferente, a abordagem Infomax possui estreita relação com o método de estimação por máxima verossimilhança [24, 25]. Em [25], Cardoso mostra que, no contexto de separação de

fontes, a função de verossimilhança logarítmica assume a seguinte forma

$$J_{Verossimilhança}(\mathbf{W}) \propto E \left\{ \sum_{n=1}^N \log (p_{s_n}(\mathbf{w}_n \mathbf{x}(j))) \right\} + \log (|\det(\mathbf{W})|). \quad (3.21)$$

Comparando (3.20) e (3.21) nota-se claramente que a abordagem de máxima verossimilhança e o princípio infomax possuem a mesma função custo, a menos das funções não-lineares utilizadas. De fato, a igualdade ocorre quando  $g_i$  for igual à função de distribuição cumulativa, ou cdf (*Cumulative Distribution Function*) da fonte  $i$ , e nesta condição a distribuição de  $g_i(y_i)$  será uniforme no intervalo  $[0, 1]$  quando  $y_i$  for igual à fonte  $s_i$ , ou a alguma outra com mesma pdf [82].

Outra visão interessante do problema de separação através da estimação por máxima verossimilhança, e conseqüentemente pelo princípio infomax, pode ser obtida ao reescrever a função de verossimilhança em termos da divergência de Kullback-Leibler [82]:

$$J_{Verossimilhança}(\mathbf{W}) = -D(p_{\mathbf{y}}(\mathbf{y}) \| p_{\mathbf{s}}(\mathbf{s})). \quad (3.22)$$

A partir de (3.22) podemos interpretar a abordagem de estimação por máxima verossimilhança das fontes como um critério de “casamento” de pdf’s: deve-se determinar a matriz  $\mathbf{W}$  de forma a tornar a distribuição das estimativas tão “próximas” quanto possível da distribuição das fontes reais. No caso da abordagem infomax, as distribuições reais das fontes são substituídas pelas derivadas das funções não-lineares  $g_i$ . Felizmente, mesmo que não haja um casamento perfeito entre as funções  $g_i$  e as pdf’s das fontes, é possível separar os sinais. Voltaremos a comentar sobre isto na seção 3.3.2.

É interessante notar que esta interpretação de “casamento” de pdf’s evidencia uma estreita relação entre os critérios desenvolvidos para o problema de BSS e o teorema de Benveniste-Goursat-Ruget [18], um dos pilares da teoria de equalização cega de canais de comunicação. Esta e outras relações existentes são discutidas em maior detalhe em [13, 14].

## 3.2 Comentários gerais sobre as diferentes abordagens

Os critérios apresentados, apesar de diferentes em sua concepção, possuem estreita ligação. A primeira relação evidente entre critérios já foi previamente discutida na

seção 3.1.5, onde mostrou-se que o princípio infomax, aplicado ao problema de separação de fontes, e a abordagem de máxima verossimilhança resultam em funções objetivo semelhantes, sendo equivalentes quando as funções não-lineares  $g_i$  na camada de saída da rede neural artificial forem tais que representem as cdf's das fontes.

Outra relação de extrema relevância é a equivalência entre a abordagem de minimização da informação mútua e a maximização da não-gaussianidade. Observando o problema ao qual se resume o critério de informação mútua entre as saídas do sistema separador, dado em (3.8), e assumindo que os dados foram devidamente branqueados, verifica-se que a função objetivo do critério reduz-se a

$$J(\mathbf{y}) = \sum_{i=1}^N H(y_i), \quad (3.23)$$

pois a matriz de separação será necessariamente ortogonal, o que anula o último termo referente ao logaritmo de seu determinante. Isto indica que a adaptação da matriz de separação será efetuada de forma a reduzir a entropia das estimativas. Por outro lado, nesta condição pode-se notar que a negentropia (3.15) e a informação mútua diferem apenas pelo termo  $H(\mathbf{y}_{gauss})$ , que permanece constante pois a matriz de covariância das estimativas não é alterada. Assim podemos escrever

$$I(\mathbf{y}) = \text{Constante} - \sum_{i=1}^N J_{Negentropia}(y_i), \quad (3.24)$$

o que estabelece uma relação fundamental entre a informação mútua e a negentropia. Dessa forma mostra-se que a abordagem de *minimização de informação mútua entre as saídas do sistema separador é equivalente à maximização da soma das negentropia das estimativas, i.e., a não-gaussianidade das mesmas*. Isto justifica, de uma maneira mais rigorosa, a idéia de maximização da não-gaussianidade, apresentada na seção 3.1.4 utilizando a idéia intuitiva fornecida pelo teorema central do limite.

Um outro ponto importante diz respeito à implementação prática de alguns dos critérios apresentados. Por exemplo, a negentropia pode, teoricamente, ser calculada utilizando-se a definição de entropia (3.1). Para isso é necessário estimar as pdf's dos sinais, o que poderia ser realizado utilizando-se métodos de kernel, como o estimador de Parzen [30], reconhecidos por sua notável capacidade de aproximação.

Porém, o processo pode ser simplificado se considerarmos aproximações da pdf que empregam expansões em séries ortonormais, como as séries de Gram-Charlier e Edgeworth [66] [30]. Neste métodos, a pdf é representada através de séries polinomiais,

com coeficientes que se relacionam com os momentos e cumulantes das variáveis. A partir destas aproximações obtém-se meios simples para estimação da negentropia, sem a necessidade de se efetuar a integração de funções, conforme a definição (3.1).

Tomando novamente a negentropia como exemplo, a partir da expansão de Gram-Charlier da pdf de uma variável  $y_i$ , pode-se obter a seguinte aproximação [66]:

$$J_{Negentropia}(y_i) \approx \frac{1}{12}E\{y_i^3\}^2 + \frac{1}{48}K(y_i)^2 \quad (3.25)$$

Note que se a variável  $y_i$  for simétrica, o primeiro termo da aproximação será nulo, e a maximização de (3.25) será, em essência, equivalente à maximização da curtose de  $y_i$ , revelando-se equivalente a uma das abordagens descritas na seção 3.1.4.

Outras aproximações, derivadas ou não a partir das expansões em séries, podem ser encontradas na literatura. Por exemplo, em [65] é proposta uma aproximação não-polinomial para a negentropia de uma variável aleatória, dada por:

$$J(y) = \alpha(E\{G(y)\} - E\{G(\nu)\})^2, \quad (3.26)$$

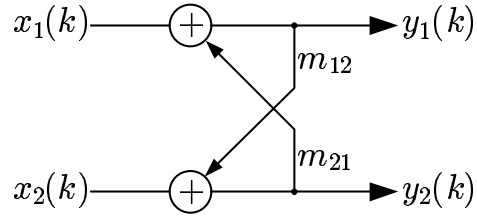
onde  $G(\cdot)$  é uma função não-linear não-quadrática,  $\alpha$  é uma constante e  $\nu$  é uma variável aleatória gaussiana de média e zero e variância unitária. Como consequência da utilização de uma gaussiana normalizada neste caso, é necessário restringir, durante a etapa de adaptação, a potência de cada uma das estimativas das fontes, assumindo por exemplo que  $E\{y_i\} = E\{\mathbf{w}_i^T \mathbf{x}\} = 1$ . A acuidade da aproximação, no entanto, depende da escolha adequada da função não-linear  $G(\cdot)$ .

### 3.3 Algoritmos para Separação Cega de Fontes

Uma vez definidos os principais critérios a serem seguidos, se fazem necessários métodos práticos para implementá-los. Nesta seção veremos alguns dos algoritmos propostos para a ICA baseados nas idéias expostas anteriormente.

#### 3.3.1 Algoritmo de Héroult, Jutten e Ans

O trabalho pioneiro de Héroult e Jutten [63] tinha como inspiração modelos de redes neurais artificiais para recuperação dos sinais, de acordo com a figura 3.3.



**Figura 3.3:** Estrutura de separação utilizada por Héault e Jutten.

A estrutura apresentada na figura é essencialmente uma rede com realimentação, de forma que suas saídas podem ser descritas através da seguinte equação:

$$\mathbf{y} = \mathbf{x} - \mathbf{M}\mathbf{y}, \quad (3.27)$$

onde  $\mathbf{M}$  representa a matriz com os pesos  $m_{ij}$ , sendo  $m_{ij} = 0$  para  $i = j$ . Note que a estrutura utilizada pode ser descrita na forma  $\mathbf{y} = (\mathbf{I} + \mathbf{M})^{-1} \mathbf{x}$ , seguindo a idéia expressa em (2.4).

Embora sem grandes formalismos teóricos, os autores propuseram um algoritmo de adaptação para a matriz  $\mathbf{M}$ , resumido através da regra

$$m_{ij} \leftarrow m_{ij} - \mu E\{f(y_i)g(y_j)\}, \quad (3.28)$$

que, conforme comentado na seção 3.1.3, busca reduzir a correlação não-linear entre as saídas da estrutura de separação. O algoritmo, embora capaz de separar os sinais em alguns casos, apresenta alguns problemas de convergência e em geral é eficiente para um número de fontes bastante limitado. Algumas análises sobre a convergência deste algoritmo, assim como melhorias que lhe foram acrescentadas, podem ser encontradas em [124, 45, 35].

### 3.3.2 Algoritmo Infomax e Infomax estendido

Em [16], Bell e Sejnowski apresentam a abordagem de separação de fontes baseada no princípio infomax e propõem um algoritmo, obtido através do cálculo do gradiente de (3.20), dado por

$$\frac{\partial H(\mathbf{z})}{\partial \mathbf{W}} = E\{\mathbf{f}(\mathbf{W}\mathbf{x})\mathbf{x}^T\} + (\mathbf{W}^T)^{-1}, \quad (3.29)$$

onde  $\mathbf{f}(\cdot) = [f_1(\cdot) \dots f_N(\cdot)]$  é um vetor de funções de modo que  $f_i(x) = d \log(g'_i(x))/dx$ . Diante disso, obtém-se a seguinte regra de atualização de  $\mathbf{W}$

$$\mathbf{W} \leftarrow \mathbf{W} + \mu \{E\{\mathbf{f}(\mathbf{W}\mathbf{x})\mathbf{x}^T\} + (\mathbf{W}^T)^{-1}\}, \quad (3.30)$$

onde  $\mu$  corresponde ao passo de adaptação. A versão estocástica (adaptativa) do algoritmo pode ser obtida simplesmente desconsiderando o operador de esperança presente nesta expressão. O algoritmo é normalmente referenciado como algoritmo Infomax, ou ainda algoritmo Bell-Sejnowski (BS).

Na sua forma original, o algoritmo Infomax considera uma não-linearidade fixa, de maneira que os cenários nos quais o algoritmo é capaz de separar os sinais fica restrito a alguns tipos particulares de fontes. A influência da escolha destas não-linearidades é assunto de alguns artigos encontrados na literatura. No entanto, argumenta-se que, em geral, é possível obter bons resultados selecionando-se apenas algumas funções.

Esta idéia foi o ponto de partida para uma das modificações do algoritmo Infomax, apresentado em [87]. No trabalho introduziu-se um parâmetro adicional para determinar qual o tipo de não-linearidade deve ser utilizada. Empregando apenas duas não-linearidades:  $g(y) = y - \tanh(y)$ , para fontes sub-gaussianas, e  $g(y) = y + \tanh(y)$ , para fontes super-gaussianas, é possível obter o algoritmo denominado *Infomax Estendido*, capaz de trabalhar em cenários com fontes de diferentes classes. Com isso, a atualização da matriz  $\mathbf{W}$ , considerando um ajuste baseado no gradiente natural (o qual será comentado na seção 3.3.4), é reescrita como

$$\mathbf{W} \leftarrow \mathbf{W} + \mu (\mathbf{I} - E\{\mathbf{M} \tanh(\mathbf{y})\mathbf{y}^T - \mathbf{y}\mathbf{y}\}), \quad (3.31)$$

onde  $\mathbf{M} = \text{diag}(m_1, \dots, m_n)$  é uma matriz diagonal tal que, idealmente,  $m_i = 1$  para fontes super-gaussianas e  $m_i = -1$  para fontes sub-gaussianas. A adaptação dos parâmetros  $m_i$  é feita através da seguinte equação:

$$m_i = \text{sign}(E\{\text{sech}^2(y_i)\}E\{y_i^2\} - E\{\tanh(y_i)y_i\}), \quad (3.32)$$

onde  $\text{sign}(\cdot)$  representa a função sinal.

### 3.3.3 PCA não-linear

A PCA (*Principal Component Analysis*) é uma técnica consagrada na análise de dados, sendo utilizada em aplicações como compressão de dados e extração de ca-

racterísticas. Diferentemente da ICA, a PCA utiliza apenas estatísticas de segunda ordem que, conforme já vimos, não é suficiente para realizar a separação dos sinais.

Uma das possíveis definições da PCA emerge sob a ótica do problema de compressão de dados, na qual o objetivo é obter representações de um vetor aleatório  $\mathbf{x} = [x_1, \dots, x_M]^T$  em um subespaço de dimensão inferior a  $M$ , de tal forma que o erro entre esta nova representação e o vetor original seja mínima. Neste contexto, podemos definir a PCA como ferramenta que busca o conjunto de vetores ortonormais que formam a base deste subespaço.

Matematicamente temos que o objetivo da PCA é determinar um conjunto de vetores base  $\mathbf{w}_i$  tal que a função

$$J_{PCA}(\mathbf{W}) = E\left\{\left\|\mathbf{x} - \sum_{i=1}^N (\mathbf{w}_i^T \mathbf{x}) \mathbf{w}_i\right\|^2\right\}, \quad (3.33)$$

com  $\mathbf{w}_i^T \mathbf{w}_j = \delta_{ij}$ , seja minimizada. Diante desta formulação, nota-se que, quanto maior for o número de componentes principais considerados na compressão, menor será o erro de compressão.

Assumindo que o vetor  $\mathbf{x}$  possui média nula, é possível mostrar que uma solução deste problema é dada pelos autovetores associados à matriz de correlação  $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^T\}$ , considerando que  $\mathbf{x}$  é um vetor de média nula. Além disso, os dados projetados nas direções dos vetores bases, i.e., as *componentes principais*, serão necessariamente *descorrelacionados entre si*. Isto, aliado ao fato de existirem métodos adaptativos para determinação das componentes principais [46], tornam a PCA uma ferramenta atraente para realizar o branqueamento das observações, conforme comentado na seção 2.2.1.

A limitação quanto ao uso de estatísticas de ordem 2, que impede seu uso para a separação de fontes, pode ser suplantada introduzindo-se não-linearidades, a exemplo do que vimos na abordagem de descorrelação não-linear, na função custo (3.33), o que dá origem à técnica de *PCA não-linear*, ou NPCA, (*Nonlinear Principal Component Analysis*) [66]. Reformulando (3.33), obtemos a função custo da técnica de NPCA:

$$J_{NPCA}(\mathbf{W}) = E\left\{\left\|\mathbf{x} - \sum_{i=1}^N (g_i(\mathbf{w}_i^T \mathbf{x})) \mathbf{w}_i\right\|^2\right\}, \quad (3.34)$$

onde  $g_i(\cdot)$  é uma função não-linear. Neste caso, a não-linearidade é introduzida na projeção das bases  $\mathbf{w}_i$  no conjunto de misturas  $\mathbf{x}$ . Todavia, é importante ressaltar



que há outras variantes não-lineares da PCA [66]. A equação (3.34) ainda pode ser reescrita da seguinte forma matricial:

$$J_{NPCA}(\mathbf{W}) = E\{\|\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2\}, \quad (3.35)$$

onde  $\mathbf{g}(\cdot) = [g_1(\cdot) \dots g_N(\cdot)]$ .

Se considerarmos que os dados foram pré-processados, i.e. branqueados, a matriz  $\mathbf{W}$  a ser determinada será ortogonal. Incluindo esta restrição em (3.35), e após algum desenvolvimento matemático, obtém-se que

$$J_{NPCA}(\mathbf{W}) = \sum_{i=1}^N E\{[y_i - g_i(y_i)]^2\}. \quad (3.36)$$

Esta formulação da função custo da NPCA apresenta grande similaridade com os critérios de Bussgang [102, 60], amplamente utilizados no problema de equalização cega de canais de comunicação.

A função custo (3.35) pode ser, via de regra, minimizada utilizando-se quaisquer método de otimização, como por exemplo o método do gradiente descendente [66]. Outra possibilidade, que apresenta algumas vantagens em relação ao método anterior, consiste em utilizar o consagrado algoritmo de mínimos quadrados recursivo, ou RLS (*Recursive Least Squares*), para adaptação dos parâmetros [108]. Um resumo do algoritmo NPCA utilizando tal abordagem é mostrado no quadro 1.

Uma característica inerente a este tipo de abordagem é a possibilidade de se recuperar as fontes de forma individual, de maneira semelhante à maximização da não-gaussianidade.

### 3.3.4 Método do Gradiente Relativo / Natural

Algumas das contribuições significativas ao estudo de BSS não estão necessariamente relacionadas a novos critérios, mas sim à maneira inovadora como o processo de otimização do sistema separador é efetuado. Um exemplo disso é o método do *gradiente relativo*, que deu origem a um algoritmo denominado EASI (*Equivariant Adaptive Source Separation*) [26].

O algoritmo clássico para se determinar o mínimo (ou máximo) de uma dada função custo, utilizado no algoritmo Infomax, é o *método do Gradiente*. A idéia é utilizar a informação de direção fornecida pelo gradiente da função e, iterativamente, buscar

1. Inicializar  $\mathbf{W}_0$  e  $\mathbf{P}_0$  com matrizes unitárias;
2. enquanto não convergir, executar as seguintes etapas de adaptação:

$$\begin{aligned}
\mathbf{h}_k &= \mathbf{P}_{k-1} \bar{\mathbf{x}}(k) \\
\mathbf{m}_k &= \mathbf{h}_k / (\lambda + \bar{\mathbf{x}}^T(k) \mathbf{h}_k) \\
\mathbf{P}_k &= \lambda^{-1} \Upsilon[\mathbf{P}_{k-1} - \mathbf{m}_k \mathbf{h}_k^T] \\
\mathbf{e}_k &= \bar{\mathbf{x}}(k) - \mathbf{W}_{k-1}^T \bar{\mathbf{x}}(k) \\
\mathbf{W}_k &= \mathbf{W}_{k-1} + \mathbf{m}_k \mathbf{e}_k^T,
\end{aligned} \tag{3.37}$$

onde  $\Upsilon[\mathbf{J}]$  representa um operador que gera uma nova matriz simétrica cuja parte triangular superior é a mesma da matriz  $\mathbf{J}$ ,  $\bar{\mathbf{x}}(k)$  denota o vetor de dados branqueados, e  $\lambda$  o fator de esquecimento do algoritmo RLS.

**Quadro 1:** Versão do algoritmo de PCA não-linear utilizando o RLS.

o ponto de mínimo, ou máximo, da função. A forma geral deste tipo de algoritmo, considerando que estamos adaptando uma certa matriz  $\mathbf{W}$ , é dada por:

$$\mathbf{W} \leftarrow \mathbf{W} \pm \mu \left. \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \right|_{\mathbf{w}}, \tag{3.38}$$

onde o sinal depende se estamos buscando o mínimo ou o máximo da função custo.

Uma variação deste algoritmo é apresentada em [26], onde os pesquisadores definem o *gradiente relativo*. Em seu trabalho, Cardoso e Laheld utilizam o que chamaram de regra de *adaptação serial*, que consiste em realizar a atualização da matriz de separação de acordo com

$$\mathbf{W} \leftarrow (\mathbf{I} - \lambda \Psi(\mathbf{y})) \mathbf{W}, \tag{3.39}$$

onde  $\Psi(\cdot)$  mapeia um vetor em uma matriz e  $\lambda$  corresponde ao passo de adaptação. Comparando esta formulação com a regra derivada pelo método do gradiente, vemos que o incremento a cada iteração não é mais feito de forma aditiva, mas sim através da multiplicação à esquerda da matriz de separação. Note que esta idéia é consistente com a estrutura de grupo do conjunto das matrizes inversíveis, em particular das matrizes ortogonais.

A forma imposta pela adaptação serial sugere que o algoritmo iterativo derivado deste modelo deve utilizar uma modificação do conceito de gradiente. No caso padrão, o

gradiente em um dado ponto  $\mathbf{W}$  pode ser visto como o termo de primeira ordem da expansão da série de Taylor de  $J(\mathbf{W} + \mathbf{D})$ :

$$J(\mathbf{W} + \mathbf{D}) \approx J(\mathbf{W}) + \text{tr} \left( \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}^T \mathbf{D} \right), \quad (3.40)$$

onde  $\text{tr}(\cdot)$  denota o traço e  $\mathbf{D}$  representa um incremento. Por sua vez, o *gradiente relativo*, é definido de forma equivalente a partir da expansão da função  $J(\mathbf{W} + \mathbf{D}\mathbf{W})$

$$\begin{aligned} J(\mathbf{W} + \mathbf{D}\mathbf{W}) &\approx J(\mathbf{W}) + \text{tr} \left( \mathbf{W} \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}^T \mathbf{D} \right) \\ &\approx J(\mathbf{W}) + \text{tr} \left( \frac{\partial_R J(\mathbf{W})}{\partial \mathbf{W}}^T \mathbf{D} \right), \end{aligned} \quad (3.41)$$

ou seja, o gradiente relativo  $\frac{\partial_R \phi(\mathbf{W})}{\partial \mathbf{W}} = \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T$ .

Utilizando uma função custo dada por  $J(\mathbf{W}) = E\{\mathbf{f}(\mathbf{y})\}$ , é possível mostrar que a regra de adaptação empregando o gradiente relativo é dada por

$$\mathbf{W} \leftarrow \mathbf{W} - \lambda E\{\mathbf{f}'(\mathbf{y})\mathbf{y}^T\} \mathbf{W}, \quad (3.42)$$

onde  $\mathbf{f}'(y) = [f'(y_1) \dots f'(y_N)]$ . Além disso, é possível incorporar a esta regra de adaptação a restrição de ortogonalidade da matriz  $\mathbf{W}$ , resultando na seguinte regra de adaptação

$$\mathbf{W} \leftarrow \mathbf{W} - \lambda E\{\mathbf{y}\mathbf{y}^T - \mathbf{I} + \mathbf{f}'(\mathbf{y})\mathbf{y}^T - \mathbf{y}\mathbf{f}'(\mathbf{y})^T\} \mathbf{W}, \quad (3.43)$$

que descreve a regra de adaptação presente no algoritmo EASI.

Interessante observar que, de maneira independente, em [5] o pesquisador Amari desenvolve o conceito de *gradiente natural*, cuja expressão leva à mesma regra de adaptação desenvolvida por Cardoso e Laheld. O enfoque seguido por Amari foi o de se definir a noção de gradiente levando-se em conta a estrutura subjacente ao espaço dos parâmetros, no caso o de matrizes não-singulares.

Neste espaço, a direção de maior variação da função custo será dada por:

$$\frac{\partial_{\text{Natural}} J(\mathbf{W})}{\partial \mathbf{W}} = \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}, \quad (3.44)$$

uma expressão bastante similar ao gradiente relativo. No entanto, é necessário res-

saltar que, ao considerar a regra de adaptação (3.38), substituindo o gradiente pelo gradiente natural, obtemos essencialmente a mesma regra obtida a partir do modelo de atualização serial.

### 3.3.5 Método do ponto fixo - Algoritmo FastICA

Um outro exemplo de contribuição cuja inovação está preponderantemente ligada ao método de otimização é o algoritmo FastICA [67, 65], muito popular na área de separação de fontes.

Suponha uma função  $f(x)$  definida para  $x \in \mathbb{R}$ . Um número  $p$  será denominado ponto fixo de  $f(x)$  se  $p = f(p)$ . Considere então um algoritmo cuja iteração é dada por  $x_{k+1} = f(x_k)$ , para um certo valor inicial  $x_0$ . Sob certas condições demonstra-se que o algoritmo converge para o ponto fixo ( $x_{k \rightarrow \infty} \rightarrow p$ ) [63]. Veja que é possível reformular o algoritmo para encontrar a raiz de  $f(x)$ . Basta considerar uma nova função,  $g(x) = f(x) + x$ , e utilizar a mesma iteração descrita acima.

Inspirando-se nessa idéia, Hyvärinen e Oja [67] propuseram o algoritmo FastICA. No contexto de maximização, ou minimização, de uma função custo o algoritmo do ponto fixo, assim como o algoritmo do gradiente, deve fornecer os pontos em que o gradiente da função custo se anula. Utilizando a iteração do ponto fixo, e considerando a função  $g(x)$  como sendo o gradiente da função custo, Hyvärinen e Oja obtiveram um algoritmo de otimização extremamente eficaz.

Posteriormente [65], uma derivação mais formal do algoritmo foi apresentada. Utilizando a aproximação da negentropia dada pela equação (3.26), e assumindo que os dados foram branqueados, a maximização da negentropia pode ser reformulada como um problema de otimização com restrição sobre a norma do vetor de separação  $\mathbf{w}$ . Empregando o método de Newton para solucionar o problema, chega-se à iteração dada por

$$\mathbf{w} \leftarrow \mathbf{w} - [E \{ \mathbf{x} G'(\mathbf{w}^T \mathbf{x}) \} - \lambda \mathbf{w}] [E \{ G''(\mathbf{w}^T \mathbf{x}) \} - \lambda]^{-1}, \quad (3.45)$$

onde  $G'$  e  $G''$  denotam as derivadas de primeira e segunda ordem da função  $G$ , e  $\lambda = E \{ \mathbf{w}^T \mathbf{x} G'(\mathbf{w}^T \mathbf{x}) \}$ , o multiplicador de Lagrange. Adicionalmente, pode-se realizar um passo intermediário de normalização do vetor  $\mathbf{w}$  para melhorar a estabilidade do algoritmo. Finalmente, multiplicando ambos os lados de (3.45) por  $\lambda - E \{ \mathbf{w}^T \mathbf{x} G'(\mathbf{w}^T \mathbf{x}) \}$ , obtém-se a forma final da iteração utilizada pelo algoritmo

FastICA, que representa uma iteração típica de um algoritmo do ponto fixo:

$$\mathbf{w} \leftarrow E \{ \mathbf{x} G'(\mathbf{w}^T \mathbf{x}) \} - E \{ G''(\mathbf{w}^T \mathbf{x}) \} \mathbf{w}. \quad (3.46)$$

A versão do algoritmo para maximização da negentropia, considerando-se a recuperação simultânea de todas as fontes presentes na mistura, pode ser vista no quadro 2.

1. Pré-processamento dos dados: centralizar os dados a fim de tornar a média nula, e branqueá-los. Os dados resultantes formam um vetor  $\bar{\mathbf{x}}$ ;
2. Inicializar vetores  $\mathbf{w}_i$  aleatoriamente, mantendo a norma unitária;
3. Para todo  $i = 1, 2, \dots, N$ ,

$$\mathbf{w}_i \leftarrow E \{ \bar{\mathbf{x}} \cdot G'(\mathbf{w}_i^T \cdot \bar{\mathbf{x}}) \} - \mathbf{w}_i \cdot E \{ G''(\mathbf{w}_i^T \cdot \bar{\mathbf{x}}) \}$$

4. Organizar os vetores  $\mathbf{w}_i$  em uma matriz  $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N)^T$ , e realizar o processo de ortogonalização simétrica:

$$\mathbf{W} \leftarrow (\mathbf{W} \cdot \mathbf{W}^T)^{-\frac{1}{2}} \cdot \mathbf{W}$$

5. Se o algoritmo não convergiu, retornar para o passo 3.

**Quadro 2:** Versão do Algoritmo FastICA utilizando a negentropia como critério de otimização.

### 3.3.6 Otimização utilizando métodos de Lie

Conforme vimos anteriormente, realizar o branqueamento dos dados reduz o problema de BSS à busca de uma matriz de separação ortogonal. Muitos dos algoritmos existentes, como o FastICA, realizam a adaptação da matriz de separação em dois passos distintos: No primeiro, a atualização da matriz de separação é feita a partir da estimativa anterior, sem levar em conta a restrição de ortogonalidade; No segundo, impõe-se a restrição de ortogonalidade através de algum método, como a ortogonalização de Gram-Schmidt.

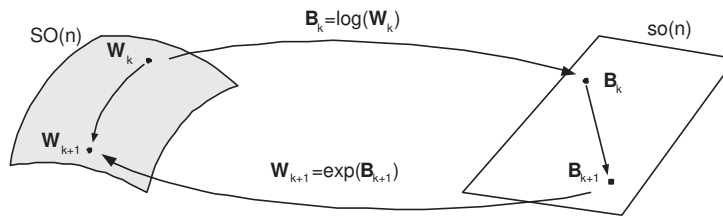
Outra forma de se otimizar a matriz de separação levando-se em conta a ortogonalidade é introduzir na função custo um termo de penalidade ligado a esta restrição. Ainda assim, é possível que o desempenho do algoritmo fique comprometida [114]. O ideal, portanto, seria obter um método para otimização de uma função contraste que

restringa o espaço de busca ao espaço de matrizes ortogonais sem lançar mão destes métodos.

A área da matemática conhecida como geometria diferencial está na origem de algumas ferramentas bastante conhecidas na área de separação de fontes. Um bom exemplo disso é o algoritmo do gradiente natural, comentado anteriormente, que foi desenvolvido tendo como base os conceitos encontrados neste campo da matemática.

Em [114], o autor propõe um método que explora certas peculiaridades do espaço de busca, de maneira semelhante à abordagem do gradiente natural.

Considere o diagrama apresentado na figura 3.4.  $SO(n)$  representa o espaço de matrizes ortogonais de ordem  $n$ . Sabe-se que o conjunto das matrizes ortogonais formam um *grupo sob a operação de multiplicação*, no qual ainda é possível definir um gradiente, o que o configura como um *grupo de Lie* [114].



**Figura 3.4:** Mapeamento entre o grupo de matrizes ortogonais  $SO(n)$  e o grupo de matrizes anti-simétricas  $so(n)$ .

Ao se definir o gradiente neste espaço podemos aplicar diretamente o método definido em (3.38) para realizar a adaptação da matriz de separação. No entanto, como o conjunto de matrizes ortogonais forma um grupo sob a operação de *multiplicação*, não podemos garantir que em todas as iterações do método a matriz resultante seja ortogonal, já que estamos essencialmente somando, e não multiplicando, matrizes.

A solução para este impasse consiste em se considerar o mapeamento biunívoco existente entre matrizes ortogonais e matrizes anti-simétricas, dado pela função  $\exp(\cdot)$ . O conjunto de matrizes anti-simétricas apresenta propriedades interessantes que não são verificadas para o grupo  $SO(n)$ , como o fato de ser fechado sob a operação de adição e multiplicação por escalares, o que o configura como um *espaço vetorial*, denotado por  $so(n)$ .

Além disso, ainda é possível definir o gradiente da função no espaço  $so(n)$ . Dessa forma, efetuando a busca utilizando o método do gradiente neste espaço, garantimos, ao realizar o mapeamento inverso para o grupo  $SO(n)$ , que as matrizes *sempre* serão

ortogonais. A figura (Figura 3.4) ilustra tal idéia, e uma possibilidade para o método de busca é descrita no quadro 3.

1. Inicie em  $\mathbf{0} \in so(n)$ , o que equivale a iniciar em  $\mathbf{I} \in SO(n) = \exp(\mathbf{0})$
2. Desloque-se em  $so(n)$  até  $\mathbf{B} \in so(n)$ , de acordo com a direção apontada pelo gradiente, definido como:

$$\nabla_{\mathbf{B}} J(W) = \left( \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \right) \mathbf{W}^T - \mathbf{W} \left( \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \right)^T$$

3. Use o mapeamento  $\exp(\cdot)$  para voltar a  $SO(n)$ , i.e.,  $\mathbf{R} = \exp(\mathbf{B})$
4. Calcule a nova matriz ortogonal  $\mathbf{W}_{k+1} = \mathbf{R} \cdot \mathbf{W}_k \in SO(n)$

**Quadro 3:** Algoritmo considerando a álgebra de Lie associada ao grupo de matrizes ortogonais

Outras possíveis extensões deste algoritmo são comentadas em [115] visando a melhora na sua velocidade de convergência.

## 3.4 Sumário

Neste capítulo abordamos alguns dos principais critérios e algoritmos para a separação cega de fontes no contexto de misturas lineares instantâneas.

Vimos que, embora tenham nascido a partir de formulações diferentes, os critérios, em geral, se relacionam diretamente à abordagem de minimização da informação mútua entre os sinais, destacando-a como referência sólida para a caracterização da condição de independência entre os sinais estimados. No entanto, em grande parte das abordagens práticas, a função custo é normalmente aproximada por medidas que envolvem, explícita ou implicitamente, estatísticas de ordem superior dos dados observados.

Além disso, vimos alguns dos algoritmos desenvolvidos visando a implementação da ICA. Alguns deles destacam-se pelo caráter pioneiro no uso de determinados critérios, sendo que outros merecem destaque pela inovação preponderantemente relacionada ao método de otimização, como o método do gradiente relativo/natural e o método baseado em álgebra de Lie.

# 4

## Misturas Lineares Convolutivas

O modelo de mistura linear instantânea tem sido empregado com sucesso em diferentes aplicações práticas, o que pode ser confirmado em diversos trabalhos publicados na literatura [82, 66]. No entanto, em algumas situações onde o processo de mistura apresenta caráter convolutivo, o modelo instantâneo provê uma estimativa grosseira do que ocorre na prática. Exemplos típicos ocorrem no processamento de sinais de áudio, captados por meio de microfones em um ambiente com reverberação [99], e também em sistemas de comunicação sem fio [30]. Portanto, um novo modelo deve ser adotado nestes casos.

Neste capítulo abordaremos o problema de separação de fonte no contexto de misturas lineares com memória, dito modelo de mistura convolutiva. Serão discutidas algumas abordagens usuais para o problema e apresentaremos um novo método para separação dos sinais baseado em filtros de erro de predição.

O método propõe que a separação dos sinais seja realizada em duas etapas. Em um primeiro momento, o problema convolutivo é reduzido a uma mistura linear instantânea por meio de filtros de erro de predição não-lineares para que, numa segunda etapa, métodos como os descritos no capítulo anterior possam efetivamente recuperar os sinais das fontes. Em particular, nossa proposta consiste em se empregar filtros baseados em sistemas nebulosos, especialmente adequados para lidar com cenários



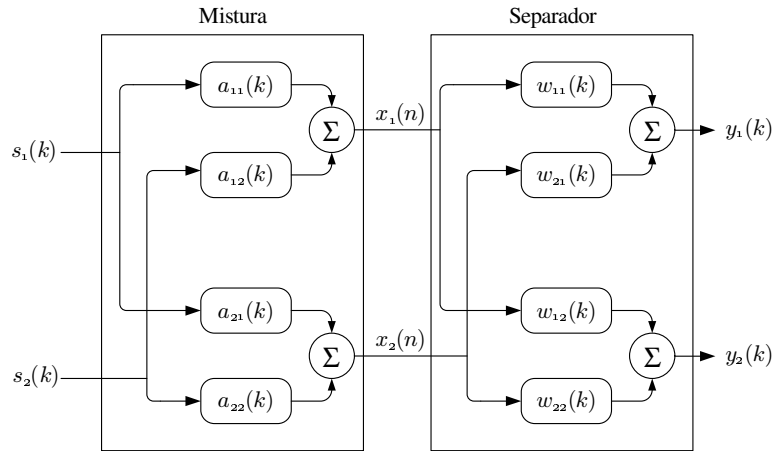
nos quais as fontes apresentam distribuições discretas.

## 4.1 Mistura Linear Convolutiva

No modelo de mistura linear convolutiva, diferentemente daquele associado a misturas instantâneas, é considerado o efeito da propagação do sinal através do meio. Isto é modelado através de função de transferência entre cada par fonte - sensor, conforme ilustrado na figura 4.1, de maneira que o sinal recebido é uma composição de sinais filtrados, podendo ser expresso por:

$$\begin{aligned} x_i(k) &= \sum_j \sum_{\tau} a_{ij}(\tau) s_j(k - \tau) \\ &= \sum_j a_{ij}(k) * s_j(k) \end{aligned} \quad , \quad (4.1)$$

para  $i = 1, \dots, N$ , onde  $a_{ij}(k)$  denota a resposta ao impulso correspondente à propagação entre a fonte  $i$  e o sensor  $j$  e o símbolo  $*$  representa a operação de convolução. A separação, neste caso, pode ser obtida através de um conjunto de filtros  $w_{ij}(k)$  de forma que a estimativa das fontes também resulta de um processo de filtragem.



**Figura 4.1:** Modelo de mistura convolutiva considerando duas fontes e dois sensores.

Note que o modelo de mistura, assim como o sistema separador, são essencialmente filtros com múltiplas entradas e múltiplas saídas (MIMO - *Multiple Input - Multiple Output*), de maneira que é usual utilizar a notação matricial para representar os sinais envolvidos. Assim, as observações podem ser descritas através da convolução

entre as entradas e uma função de transferência matricial, representada através de sua transformada  $z$ :

$$\mathbf{A}[z] = \mathbf{A}(0) + \mathbf{A}(1)z^{-1} + \cdots + \mathbf{A}(L-1)z^{-L+1}. \quad (4.2)$$

A notação introduzida em (4.2) é uma extensão direta do conceito de filtros lineares do caso SISO (no caso, com resposta ao impulso finita), substituindo-se os escalares por matrizes. Assim, as observações e os sinais estimados podem ser expressos através da convolução entre os sinais e a função de transferência dos filtros, ou seja:

$$\mathbf{x}(k) = \sum_l \mathbf{A}(l)\mathbf{s}(k-l), \quad (4.3)$$

$$\mathbf{y}(k) = \sum_l \mathbf{W}(l)\mathbf{x}(k-l). \quad (4.4)$$

Alternativamente, é possível definir estruturas recursivas para a separação dos sinais, a exemplo da estrutura proposta por Heráult, Jutten e Ans no caso de misturas instantâneas [76].

#### 4.1.1 Inversibilidade de estruturas MIMO

Uma questão de extrema relevância ao tratar a separação de misturas convolutivas diz respeito às condições sob as quais é possível inverter perfeitamente o processo de mistura por meio de uma estrutura linear. Matematicamente, queremos determinar sob quais condições é possível obter um filtro  $\mathbf{W}[z]$  tal que

$$\mathbf{W}[z]\mathbf{A}[z] = \mathbf{I}. \quad (4.5)$$

Esta condição é conhecida no contexto de comunicações como *condição zero-forcing*, e  $\mathbf{W}[z]$  é denominado *equalizador zero-forcing* [28].

No contexto de filtros MIMO, a condição para inversibilidade do sistema é definida através do posto da matriz  $\mathbf{A}[z]$  [78][28]:

**Teorema 4.1.1** (Inversibilidade de estruturas MIMO). *Seja  $\mathbf{A}[z]$  uma matriz polinomial conforme (4.2). Existirá uma matriz polinomial  $\mathbf{W}[z]$  tal que  $\mathbf{W}[z]\mathbf{A}[z] = \mathbf{I}$  se e somente se o posto de  $\mathbf{A}[z]$  for completo, para qualquer valor de  $z$  na circunferência de raio unitário,  $|z| = 1$ .*

O teorema assegura a existência de uma estrutura linear capaz de inverter a mistura. No entanto, não garante que a estrutura possua resposta ao impulso finita, ou FIR (*Finite Impulse Response*), tampouco que a resposta será causal. Vejamos alguns exemplos.

**Exemplo 4.1.1** (Equalizador zero-forcing linear e não-linear). *Considere inicialmente o sistema definido por:*

$$\mathbf{A}[z] = \begin{bmatrix} 1 & 2z^{-1} \\ -1 & 3 \end{bmatrix}. \quad (4.6)$$

*Pode-se observar que  $\det \mathbf{A}[z] = 3 + 2z^{-1}$  é diferente de zero na circunferência de raio unitário, mostrando que a matriz possui posto completo. Portanto podemos obter*

$$\mathbf{W}[z] = \frac{1}{3 + 2z^{-1}} \begin{bmatrix} 3 & -2z^{-1} \\ 1 & 1 \end{bmatrix}, \quad (4.7)$$

*onde observa-se que o filtro obtido é anti-causal. Agora, considere um segundo sistema*

$$\mathbf{A}[z] = \begin{bmatrix} 3 & 1z^{-1} \\ 3z^{-1} & 1 \end{bmatrix}, \quad (4.8)$$

*para o qual verifica-se prontamente que  $\det \mathbf{A}[z] = 3 - 3z^{-2}$ . Neste segundo caso, o determinante se anula para  $z = \pm 1$ , indicando que neste cenário não existe estrutura linear capaz de levar à condição zero-forcing.*

**Exemplo 4.1.2** (Equalizador FIR). *Considere outro exemplo de sistema MIMO, definido por sistema definido por:*

$$\mathbf{A}[z] = \begin{bmatrix} z^{-1} & -1 - 3z^{-1} \\ 1 - z^{-1} & -1 + z^{-1} \\ z^{-1} & -1 \end{bmatrix}. \quad (4.9)$$

*Neste caso pode-se verificar que*

$$\mathbf{W}[z] = \frac{1}{6} \begin{bmatrix} -3 - 2z^{-1} + 3z^{-2} & 6 + 12z^{-1} + 9z^{-2} & -3 + 5z^{-1} + 6z^{-2} \\ -3 + z^{-2} & 6z^{-1} + 3z^{-2} & -3 + 3z^{-1} - 2z^{-2} \end{bmatrix} \quad (4.10)$$

representa o equalizador zero-forcing com resposta ao impulso finita.

Este último exemplo evidencia um fato não observado no contexto de filtros lineares SISO: em alguns casos, é possível inverter um sistema MIMO-FIR com outra estrutura de mesma natureza. De fato, isso ocorre para os casos enunciados no seguinte teorema, que corresponde à *Identidade de Bezout* [78]:

**Teorema 4.1.2.** *Seja  $A[z]$  uma matriz polinomial retangular, tal que o posto de  $A[z]$  seja completo para qualquer  $z$  incluindo  $\infty$ . Então, existe uma matriz polinomial*

$$\mathbf{W}[z] = \sum_{l=0}^K \mathbf{W}(l) z^{-l}, \quad (4.11)$$

*com  $K$  finito, tal que  $\mathbf{W}[z]\mathbf{A}[z] = \mathbf{I}$ .*

Um fato observado em matrizes polinomais com mais linhas do que colunas, o que representaria uma mistura com mais sensores do que fontes, é que, a menos de alguns casos patológicos, a condição expressa pelo teorema 4.1.2 é quase sempre satisfeita [72]. O mesmo, no entanto, não ocorre para o caso de matrizes quadradas, para as quais o conjunto de matrizes que atendem a tais condições é restrito a um conjunto de canais denominados unimodulares [28].

### 4.1.2 Separação dos sinais através da independência

Da mesma forma que no caso de mistura instantânea, foi provado que a independência entre as estimativas das fontes garante a separação da mistura [138]. Entretanto, é necessário ressaltar que o sentido dado à independência estatística entre os sinais no contexto de misturas convolutivas é ligeiramente diferente daquele utilizado no caso de misturas instantâneas. Considere o seguinte exemplo:

**Exemplo 4.1.3.** *Sejam duas fontes,  $s_1(k)$  e  $s_2(k)$ , independentes entre si, cujas amostras são independentes e identicamente distribuídas (i.i.d.). Considere uma resposta global do processo de mistura e sistema de separação  $\mathbf{G}[z] = \mathbf{W}[z]\mathbf{A}[z]$  tal que:*

$$\mathbf{G}[z] = \begin{bmatrix} 1 & z^{-1} \\ 0 & 1 \end{bmatrix}. \quad (4.12)$$

Neste caso, os sinais na saída do sistema separador será dada por:

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} s_1(k) + s_2(k-1) \\ s_2(k) \end{bmatrix}. \quad (4.13)$$

Note que os sinais  $y_1(k)$  e  $y_2(k)$  são independentes entre si, entretanto o sinal  $y_1(k)$  ainda representa uma mistura das fontes.

No caso de misturas instantâneas, a independência entre os sinais estimados  $y_1(k)$  e  $y_2(k)$  é garantida através da independência estatística entre variáveis aleatórias referentes ao instante  $k$ . No entanto, a presença de elementos de memória no modelo de mistura, conforme visto no exemplo, faz com que seja necessário verificar a independência entre os processos aleatórios, i.e.,  $y_1(k)$  e  $y_2(k-l)$  devem ser independentes para todo  $k$  e  $l$ .

Portanto, se utilizarmos a informação mútua para verificar a independência entre variáveis aleatórias, a função custo a ser otimizada no caso convolutivo deve tomar a forma de

$$J = \sum_m I(y_1(k), y_2(k-m)), \quad (4.14)$$

garantindo que se a função custo for minimizada os processos aleatórios são independentes entre si.

De maneira similar ao caso de misturas instantâneas, a separação no caso convolutivo também apresenta ambigüidade quanto à permutação das fontes. A diferença, no entanto, se revela no que se refere à ambigüidade de amplitude, que no caso convolutivo dá lugar a uma indeterminação de filtragem. Em outras palavras, se garantirmos que os sinais  $y_1(k)$  e  $y_2(k)$  são independentes, verifica-se que  $\hat{y}_1(k) = h_1(k) * y_1(k)$  e  $\hat{y}_2(k) = h_2(k) * y_2(k)$ , para quaisquer filtros  $h_1(k)$  e  $h_2(k)$  inversíveis, também serão independentes entre si.

Os métodos desenvolvidos no contexto de misturas convolutivas também se utilizam dos critérios enunciados no capítulo 3. Uma das primeiras propostas para separação dos sinais baseada na idéia de independência dos sinais é encontrada em [76], onde uma extensão do algoritmo de Heráult, Jutten e Ans é proposta para a separação de misturas convolutivas. A estrutura de separação utilizada segue a proposta inicial, mostrada na figura 3.3, substituindo-se os coeficientes da estrutura recursiva por filtros lineares. A função custo utilizada segue a versão original descrita pela equação (3.10). Posteriormente em [127], uma modificação na função custo é proposta, de maneira que os filtros de separação são adaptados visando minimizar o cumulante

conjunto dos sinais estimados.

Outros métodos inicialmente desenvolvidos para lidar com o caso de misturas instantâneas também foram modificados para lidar com sistemas MIMO, como é o caso do algoritmo do gradiente natural [7, 6], o algoritmo Infomax [130, 131], bem como o algoritmo FastICA [128].

### 4.1.3 Separação dos sinais no domínio da frequência

Embora a separação dos sinais seja garantida empregando-se a idéia de independência entre os sinais, a forma particular do processo de mistura, e conseqüentemente da estrutura de separação, permite que o problema seja abordado de uma outra maneira.

A idéia é explorar o fato de que as observações são resultado de convoluções temporais entre as fontes e os filtros de mistura, e tratar os dados no domínio freqüencial [122, 91]. Os sinais observados no domínio temporal  $x_i(k)$ , amostrados a uma freqüência  $f_s$ , podem ser convertidos em uma seqüência de representações freqüenciais  $\bar{x}_i(f, \tau)$  através de uma transformada de Fourier de curta duração, ou STFT (*Short-Time Fourier Transform*), com número finito de pontos:

$$\bar{x}_i(f, \tau) = \sum_{l=-\frac{L}{2}}^{\frac{L}{2}+1} x_i(\tau + l) w(l) e^{-j2\pi fl}, \quad (4.15)$$

onde  $f \in \{0, \frac{1}{L}f_s, \dots, \frac{L-1}{L}f_s\}$  representam as freqüências,  $w(k)$  uma função de janelamento (como a janela de Hanning), e  $\tau$  representa um novo índice temporal.

Assim, o sinal observado pode ser aproximadamente representado como resultado de uma mistura instantânea em cada freqüência:

$$\bar{x}_i(f, \tau) = \sum_j \bar{a}_{ij}(f) \bar{s}_j(f, \tau) \quad (4.16)$$

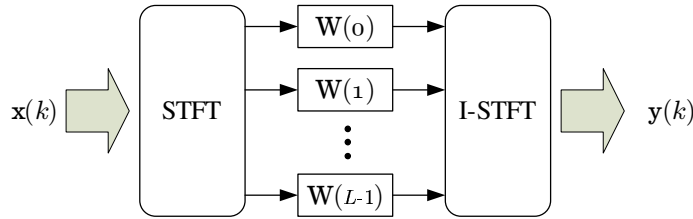
onde  $\bar{a}_{ij}(k)$  é a resposta em freqüência entre a fonte  $j$  e o sensor  $i$ , e  $\bar{s}_j(f, \tau)$  é a STFT da fonte  $j$ . Finalmente, podemos organizar todas as observações no formato matricial, resultando em:

$$\bar{\mathbf{x}}(f, \tau) = \bar{\mathbf{A}}(f) \bar{\mathbf{s}}(f, \tau) \quad (4.17)$$

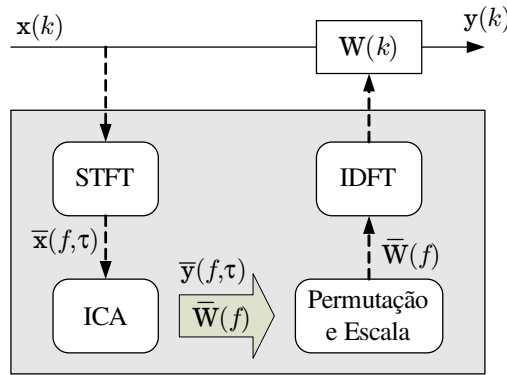
Dessa forma, o problema de separação de misturas convolutivas é convertido em um conjunto de problemas equivalentes de misturas instantâneas, para os quais métodos

baseados nos princípios descritos no capítulo 3 podem ser empregados. No entanto, é necessário ressaltar que a passagem para o domínio freqüencial transforma os dados reais em valores complexos, e alguns ajustes devem ser feitos aos algoritmos apresentados para lidar com isto. Por exemplo, em [122] uma versão modificada do algoritmo Infomax é proposta para realizar a separação de sinais complexos. É possível encontrar versões do algoritmo FastICA para dados complexos [19][48]. Já o algoritmo JADE [23] foi originalmente proposto para sinais complexos, de forma que pode ser diretamente aplicado ao problema de separação no domínio da freqüência.

As figuras 4.2 e 4.3 ilustram duas possíveis estratégias que exploram este princípio. Na figura 4.2 a separação é realizada no domínio freqüencial e as fontes recuperadas são obtidas no tempo através da transformada inversa. Já na figura 4.3 apenas o projeto dos filtros de separação é efetuado no domínio da freqüência, mas a separação é realizada no domínio temporal.



**Figura 4.2:** Método de separação de misturas convolutivas no domínio da freqüência.



**Figura 4.3:** Outra possível estratégia para separação de fontes utilizando o domínio da freqüência.

Como a separação é realizada de forma independente nas diferentes freqüências, cada solução obtida pode apresentar uma permutação entre os sinais e uma escala ar-

bitrária, i.e.:

$$\mathbf{y}(f, \tau) = \mathbf{P}(f)\Lambda(f)\mathbf{s}(f, \tau). \quad (4.18)$$

A indeterminação quanto à amplitude dos sinais dá origem à ambigüidade de filtragem dos sinais, comentada anteriormente. Em geral, a exemplo do que ocorre no caso de misturas instantâneas, esta ambigüidade só será resolvida se informações adicionais forem consideradas. Um exemplo disto ocorre em aplicações de comunicações digitais, onde os sinais envolvidos possuem uma distribuição conhecida e, na maioria das vezes, pode-se considerar que as amostras das fontes são i.i.d..

Mesmo na impossibilidade de se obter exatamente o sinal das fontes, é usual, especialmente em aplicações de sinais de áudio e voz, utilizar o *princípio de mínima distorção* [96, 95] para determinação dos ganhos dos filtros e amenização do problema de filtragem dos sinais. Segundo esse princípio, o filtro de separação deve ser escolhido de forma que o sinal observado sofra um mínimo de distorção possível. Em outras palavras, o princípio sugere que a separação seja realizada através de uma estrutura que influa o mínimo possível na forma dos sinais observados.

Matematicamente, o princípio de mínima distorção pode ser traduzido na escolha da estrutura de separação tal que  $E \{[\mathbf{y}(k) - \mathbf{x}(k)]^2\}$  seja mínimo. É possível mostrar que a solução que minimiza o erro quadrático entre as estimativas dos sinais e as observações possui a seguinte propriedade:

$$\Lambda(f) = \text{diag} \bar{\mathbf{A}}(f), \quad (4.19)$$

o que é obtido impondo a condição  $\text{diag} \mathbf{W}^{-1}(f) = \mathbf{I}$  sobre as matrizes de separação no domínio freqüencial.

Com isso, vemos que a ambigüidade de amplitude não chega a ser um grande empecilho na recuperação dos sinais. Entretanto, a ambigüidade quanto à ordem dos sinais recuperados torna-se de grande relevância na abordagem freqüencial.

Caso as permutações não forem consistentes ao longo das freqüências, a reconstrução dos sinais no domínio do tempo resultará em componentes que são composições de diferentes fontes em faixas de freqüência distintas. Assim, torna-se imperativo o uso de métodos para o alinhamento das diferentes componentes freqüenciais de uma mesma fonte antes de retornar ao domínio temporal.

Algumas propostas para a solução da indeterminação de permutação exploram características particulares das matrizes de mistura e separação. Um exemplo seria considerar que a transição entre as matrizes de mistura/separação no domínio freqüencial



seja suave. Assim, é possível corrigir as permutações avaliando a variação dos coeficientes das matrizes entre duas frequências consecutivas, i.e., para cada frequência  $f_k$  determina-se a permutação que minimiza a distância entre a matriz de mistura estimada  $\bar{\mathbf{A}}(f_k)$  da frequência atual e  $\bar{\mathbf{A}}(f_{k-1})$  da banda anterior [15].

Outro método presente na literatura se utiliza de informação sobre a estrutura dos filtros de separação, considerando a geometria dos sensores para obter a direção de chegada dos sinais (DOA - *Direction of Arrival*). Como, a princípio, as fontes encontram-se em locais diferentes, é possível agrupar as diferentes componentes frequenciais dos sinais de acordo com a direção de chegada estimada através dos coeficientes dos filtros de separação e, dessa forma, realinhar as componentes de maneira correta [83, 70].

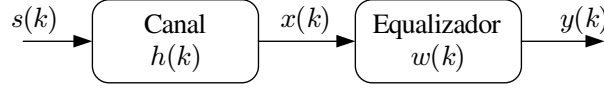
Além destas abordagens, encontram-se na literatura métodos para a solução o problema de permutação utilizando informação espectral dos sinais estimados. Alguns métodos, efetivos principalmente na recuperação de sons naturais, utilizam-se do fato que neste tipo de sinal as componentes espectrais em torno de uma mesma frequência apresentam correlação. Desta maneira, o problema de permutação pode ser resolvido verificando-se a correlação entre as componentes obtidas em frequências adjacentes [69, 121]. Esta abordagem, aliada à informação espacial das fontes, pode prover métodos mais eficientes para a reordenação das componentes espectrais [120].

## 4.2 Uma proposta para separação de sinais baseada em predição

Conforme comentado anteriormente, na ausência de informações adicionais sobre a natureza das fontes, os métodos de separação de misturas convolutivas recuperarão as fontes a menos de uma ambigüidade quanto à filtragem. No entanto, no caso particular de fontes com amostras i.i.d., como no caso de comunicações digitais, a indeterminação quanto à filtragem pode ser solucionada garantindo que os sinais recuperados também apresentem tal propriedade.

De fato, ao assumir que os sinais são i.i.d., o problema de separação de fontes toma a forma de um problema de desconvolução cega, semelhante ao problema de equalização de canais de comunicação. Considere a figura 4.4, onde um sinal i.i.d. não-gaussiano  $s(k)$  é transmitido através de um canal com resposta ao impulso  $h(k)$ . O sinal recebido  $x(k)$  será resultado da convolução  $x(k) = h(k) * s(k)$ , e o papel do equalizador é recuperar, da melhor maneira possível, o sinal transmitido. Assim como em BSS, o

termo cego explicita o fato de que o equalizador deve ser obtido sem conhecimento prévio do canal, tampouco de amostras do sinal transmitido.



**Figura 4.4:** Caso particular de mistura convolutiva - Sistema SISO.

As técnicas de equalização cega, da mesma forma que os métodos de ICA, fazem uso de estatísticas de ordem superior para compor critérios que levem à recuperação dos sinais. Um exemplo é o conhecido critério do módulo constante [56] (CM - *Constant Modulus*), que propõe que o ajuste do equalizador seja realizado de maneira a minimizar

$$J_{CM}(y(k)) = E\{|y(k)|^2 - R_2\}^2 \quad (4.20)$$

onde a constante  $R_2$  depende das estatísticas do sinal transmitido. Note a grande semelhança entre (4.20) e a função custo da técnica de PCA não-linear (3.36), este sendo apenas um exemplo de estreita ligação entre os métodos de BSS e os de equalização cega [13, 14, 60].

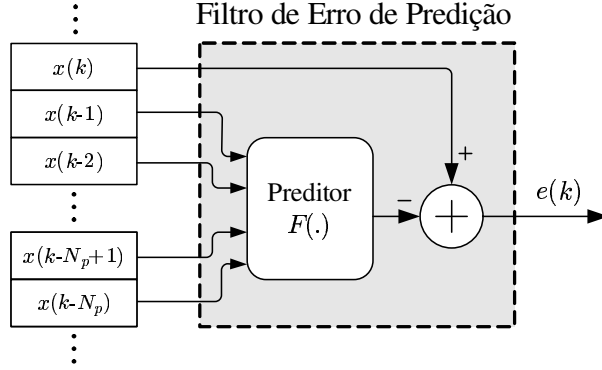
Uma outra abordagem desenvolvida no contexto de equalização cega, e de particular interesse para nossos propósitos, consiste na utilização de preditores na tarefa de equalização [29, 42, 52]

Em linhas gerais, a tarefa de predição é estimar valores futuros a partir de amostras atuais e passadas de um determinado sinal. A figura 4.5 ilustra esta idéia, onde o preditor, caracterizado por uma função  $F$ , utiliza  $N_p$  valores passados de  $x$  para estimar a amostra atual  $x(n)$ . O erro de estimação é denominado *erro de predição*, enquanto que a estrutura mostrada na figura 4.5 é conhecida como *filtro de erro de predição*.

A aplicabilidade de filtros de erro de predição ao problema de equalização pode ser melhor compreendida ao considerarmos o mapeamento entrada-saída do canal, dado por:

$$\begin{aligned} x(k) &= \sum_{l=0}^{L-1} h(l)s(k-l) \\ &= h(0)s(k) + h(1)s(k-1) + \cdots + h(L-1)s(k-L+1) \end{aligned} \quad (4.21)$$

onde  $h(l)$  representa a resposta ao impulso do canal e  $L$  o seu comprimento. Se utilizarmos um preditor que considere  $N_p$  amostras passadas para estimar a amostra



**Figura 4.5:** Filtro de erro de predição SISO.

$x(k)$  do sinal recebido, obteremos um mapeamento da seguinte forma:

$$\begin{aligned} y(k) &= F[x(k-1), x(k-2), \dots, x(k-N_p)] \\ &= P[s(k-1), s(k-2), \dots, s(k-N_p-L+1)] \end{aligned} \quad (4.22)$$

Note que a saída do preditor será função do sinal transmitido, exceto pela amostra  $s(k)$ , que, por hipótese, é independente das demais amostras transmitidas. Dessa forma, dado que uma estrutura com certo grau de flexibilidade seja utilizada como preditor, o erro de predição tenderá a ser proporcional ao sinal  $s(k)$ , a única informação à qual o preditor não tem acesso [52]. Em outras palavras, caso seja possível projetar tal estrutura, o filtro de erro de predição eliminará quaisquer redundâncias existentes nas amostras do sinal observado, provendo em sua saída apenas o sinal  $h(0)s(k)$ .

Se considerarmos que tanto o canal como o equalizador possuem estruturas lineares FIR, veremos que será impossível obter tal estrutura, uma vez que obter na saída do equalizador um sinal proporcional a  $s(k)$  é equivalente a inverter, a menos de um ganho, um filtro FIR com outro de mesma natureza.

Porém, conforme vimos na seção 4.1.1, esta restrição é de certa forma relaxada ao considerarmos estruturas MIMO. Utilizando o teorema 4.1.2, em [58] os autores, considerando apenas o caso em que há mais sensores do que fontes, mostram que a sequência observada  $\mathbf{x}(k)$  é equivalente a um processo auto-regressivo, com processo de inovação dado por  $\mathbf{A}(0)\mathbf{s}(k)$ . Dessa forma, considerando um filtro de erro de predição, na forma de

$$\mathbf{W}[z] = \mathbf{I} + \sum_{l=1}^{N_p} \mathbf{W}(l) z^{-l}, \quad (4.23)$$

é possível recuperar um sinal que representa uma mistura instantânea de  $\mathbf{s}(\mathbf{k})$ .

Assim, uma vez obtido o processo de inovação, a separação efetiva das fontes pode ser efetuada com qualquer algoritmo proposto para o contexto de misturas instantâneas, como os apresentados no capítulo 3. A mesma abordagem é seguida em [44, 94], lembrando que todos consideram a propriedade observada em sistemas que representam cenários com mais sensores do que fontes. Além disso, assume-se que as condições enunciadas no teorema 4.1.2 são respeitadas, o que exclui canais como os apresentados no exemplo 4.1.1, para os quais não é possível inverter a mistura com uma estrutura MIMO-FIR.

Tendo em vista as limitações impostas pela predição linear, buscamos uma solução utilizando filtros de erro de predição mais flexíveis, empregando a mesma idéia delineada para o caso de equalização de sistemas SISO. Considerando que as fontes possuem distribuições discretas, e que as amostras transmitidas são i.i.d., a partir de (4.3), podemos escrever o sinal observado pelo sensor  $i$  como

$$x_i(k) = \mathbf{a}_i(0)\mathbf{s}(k) + \mathbf{a}_i(1)\mathbf{s}(k-1) + \cdots + \mathbf{a}_i(L-1)\mathbf{s}(k-L+1), \quad (4.24)$$

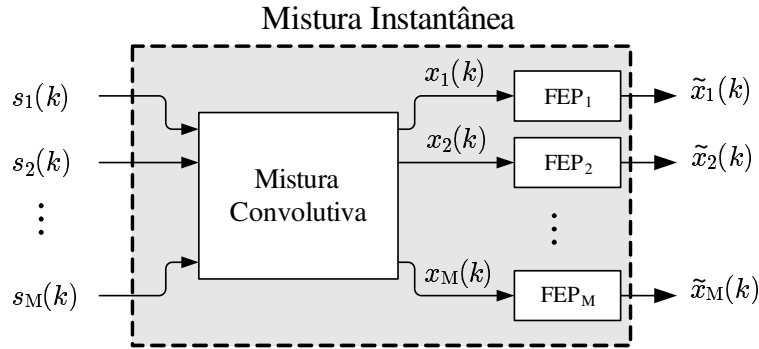
onde  $\mathbf{a}_i(k)$  denota a linha  $i$  da matriz  $\mathbf{A}(k)$ . Note que o sinal observado possui a mesma forma de (4.21), e, de maneira análoga, ao aplicar um filtro de predição não-linear aos sinais  $x_i(k)$ , obtemos, em uma condição ideal, sinais de erro  $e_i(k) = \mathbf{a}_i(0)\mathbf{s}(k)$ . Dessa forma, o vetor  $\mathbf{e}(k) = [e_1(k), e_2(k), \dots, e_M(k)]^T$  será dado por

$$\mathbf{e}(k) = \tilde{\mathbf{x}}(k) = \mathbf{A}(0) \cdot \mathbf{s}(k), \quad (4.25)$$

ou seja, o vetor de observações será dado por um processo de mistura instantânea definido apenas pela matriz  $\mathbf{A}(0)$ . Assim, a partir de  $\mathbf{e}(k)$  podemos extrair os sinais das fontes utilizando, por exemplo, o algoritmo FastICA, descrito no capítulo anterior. A figura 4.6 ilustra o método proposto, considerando um conjunto de  $N = M$  fontes.

### 4.3 Solução através de filtros fuzzy

A escolha da estrutura é de grande importância para o sucesso do método, uma vez que é necessário um certo grau de flexibilidade para que seja possível eliminar as redundâncias existentes no sinal observado e, assim, eliminar o caráter convolutivo da mistura. Nossa escolha, seguindo resultados de trabalhos anteriores [52], recaiu sobre a estrutura nebulosa, ou *fuzzy*, apresentada a seguir.



**Figura 4.6:** Proposta de utilização de filtros de erro de predição não-lineares no problema de misturas convolutivas.

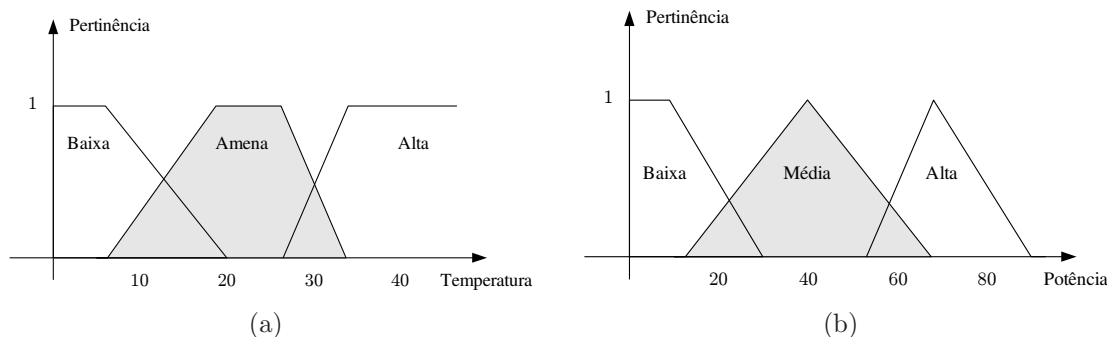
### 4.3.1 Filtros Fuzzy

Os filtros nebulosos, ou *fuzzy*, são estruturas não-lineares dotadas de capacidade de aproximação universal, isto é, da capacidade de aproximar com precisão arbitrária qualquer mapeamento não-linear contínuo definido sobre uma região compacta do domínio [135, 98].

Em sua essência, o filtro nebuloso pode ser compreendido como uma estrutura que realiza o mapeamento não-linear através de regras que relacionam conjuntos definidos no espaço de entrada a conjuntos no espaço de saída. Estes conjuntos, denominados *conjuntos nebulosos*, são definidos através de uma *função de pertinência*, que expressa o grau de pertinência de um determinado elemento a um determinado conjunto, em contraposição à teoria clássica, na qual um elemento ou pertence ( $\in$ ) ou não pertence ( $\notin$ ) a um conjunto.

Por exemplo, imagine que desejamos estabelecer conjuntos de temperaturas que consideramos baixas, amenas e altas. Segundo a teoria clássica, devemos estabelecer limites precisos, a partir dos quais um determinado valor de temperatura passa a uma outra categoria. No entanto, ao definir um grau de pertinência para cada um dos grupos, um mesmo valor de temperatura, como 17 graus, pode ter pertinência de 0,7 ao conjunto de temperaturas amenas, e pertinência 0,2 ao conjunto de temperaturas baixas. A figura 4.7a mostra um exemplo de definição dos conjuntos através das funções de pertinência.

As regras que relacionam os conjuntos de entrada aos de saída são definidas na forma *Se X então Y*, onde X e Y representam os conjuntos de entrada e saída, respectivamente. Suponha que os conjuntos definidos anteriormente representem os conjuntos



**Figura 4.7:** Exemplo de definição de conjuntos fuzzy. (a) Conjuntos relacionados à temperatura do ambiente; (b) Conjuntos relacionados à potência de um aquecedor.

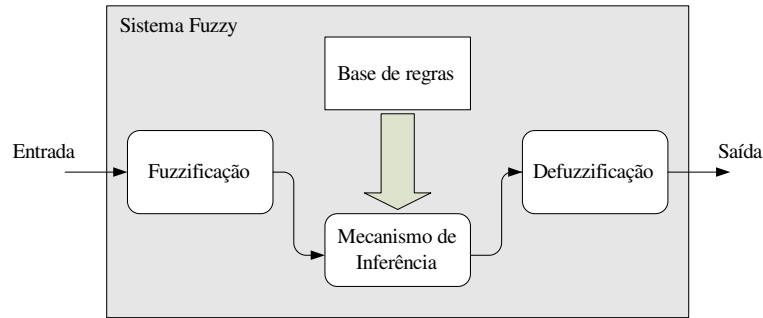
de entrada de um controle de temperatura ambiente, e que os conjuntos de saída deste controlador sejam relativos à potência de um aquecedor (potência baixa, média e alta, conforme ilustra a figura 4.7b). Uma possibilidade para as regras que compõem o controlador são

1. Se *temperatura baixa* então *potência alta*
2. Se *temperatura amena* então *potência média*
3. Se *temperatura alta* então *potência baixa*

Até o momento, descrevemos a parte central do sistema fuzzy, definindo os conjuntos nos espaços de entrada e saída, bem como as regras que associam tais conjuntos. No entanto, a fim de se implementar um controle de temperatura, as variáveis de entrada e saída devem representar valores numéricos definidos de forma rígida, ou *crisp*. No exemplo do controle de temperatura, o sinal de entrada representa um valor de temperatura, e sua saída deve representar um valor de potência. Vejamos então como um sistema fuzzy opera com tais valores, com o auxílio da figura 4.8, onde um diagrama em blocos do sistema é apresentado.

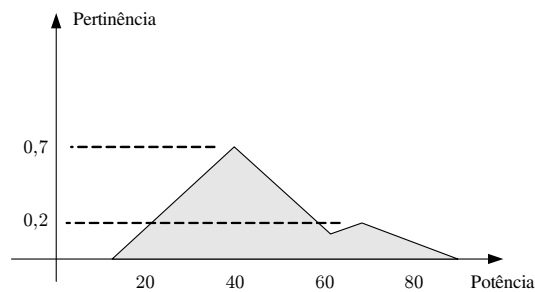
Como dissemos, a entrada do sistema será um valor numérico da temperatura ambiente. A primeira etapa do filtro consiste em se transformar tal valor numérico em *valores de pertinência* a cada um dos conjuntos definidos no espaço de entrada, processo chamado *fuzzificação*. Suponhamos que a temperatura seja  $17^{\circ}\text{C}$  e que este valor tenha grau de pertinência 0,2 em relação ao conjunto de temperaturas baixas e 0,7 em relação ao de temperaturas amenas.

Através destes dados, o controlador buscará em seu conjunto de regras aquelas as-



**Figura 4.8:** Diagrama esquemático de um sistema fuzzy.

sociadas aos conjuntos de entradas *temperatura baixa* e *temperatura amena*, o que corresponde às duas primeiras regras descritas anteriormente. Neste ponto, como duas regras são ativadas pela entrada, é necessário estabelecer a forma como estas serão combinadas para gerar o conjunto de saída correspondente, papel executado pelo *mecanismo de inferência*. Uma solução seria ponderar o conjunto de saída associado a cada uma das regras com os valores da pertinência obtidos nos conjuntos de entrada, e combinar os conjuntos de saída através da união dos mesmos. Note que a união dos conjuntos fuzzy pode ser definida através da operação max entre as duas funções de pertinência, conforme ilustra a figura 4.9.



**Figura 4.9:** Conjunto fuzzy de saída do controlador.

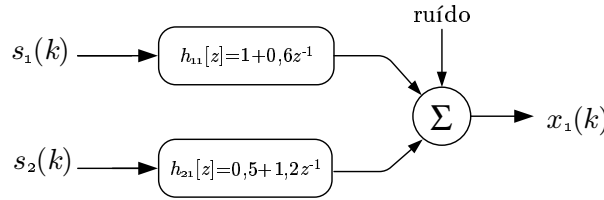
O passo final, uma vez obtido o conjunto de saída correspondente, é realizar a *defuzzificação* do conjunto obtido, retornando na saída do controlador um valor para a potência. Diferentes abordagens podem ser empregadas neste processo, como atribuir à saída o valor de maior pertinência do grupo, o centro de massa etc. A escolha de um tipo particular de inferência e *defuzzificação* altera a forma do mapeamento entrada-saída obtido.

### 4.3.2 Estrutura fuzzy aplicada ao problema de predição

No exemplo apresentado, a escolha das regras foi feita por meio do senso comum, o que evidencia uma característica marcante deste tipo de dispositivo: a possibilidade de introduzir, de maneira intuitiva, informação *a priori* sobre o problema no filtro.

Entretanto, em muitas situações, só é possível dispor de um conjunto de dados de treinamento, como no problema de projeto do preditor. Neste contexto, o sistema fuzzy pode ser definido, apenas estruturalmente, através da escolha das formas das funções de pertinência e dos mecanismos de inferência e defuzzificação. Isto resultará em um mapeamento genérico, com parâmetros a serem ajustados posteriormente de acordo com os dados disponíveis para treinamento [135].

Consideremos um exemplo no qual o sinal de duas fontes independentes transmitindo símbolos  $\{+1, -1\}$  chegam a um sensor através de dois canais de comprimento 2, conforme ilustrado na figura 4.10.

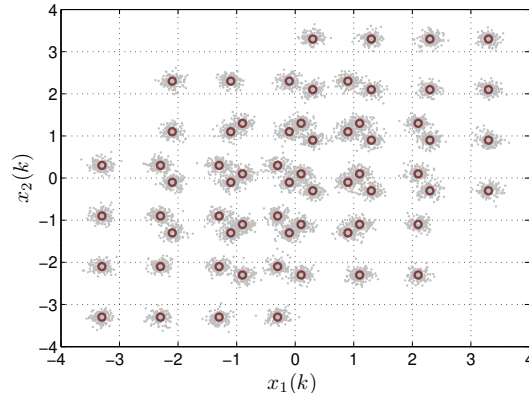


**Figura 4.10:** Exemplo de um sinal  $x_1(k)$  que representa a mistura convolutiva de duas fontes, e considera-se a presença de ruído aditivo.

Suponha que desejássemos obter um preditor fuzzy que utilizasse apenas duas amostras do sinal recebido,  $x_1(k)$  e  $x_1(k-1)$ . No contexto de comunicações digitais, devido ao fato de as amostras das fontes pertencerem a um alfabeto finito e o processo de mistura envolver um número finito de amostras passadas do sinal (ligado ao comprimento  $L$  do canal), o sinal observado, na ausência de ruído, também pertencerá a um alfabeto finito. Portanto, o vetor de entrada  $\mathbf{x}_p(k) = [x_1(k), x_1(k-1)]^T$  do filtro fuzzy será composto de um conjunto finito de possíveis combinações de amostras do sinal recebido, denominados *estados do canal*.

A figura 4.11 mostra o gráfico contendo as amostras  $x_1(k)$  e  $x_1(k-1)$ , onde é possível distinguir 64 “nuvens” de dados, cada uma delas associadas a um estado de canal (denotados pelo símbolo  $\circ$ ). Note que neste exemplo consideramos a presença de ruído gaussiano aditivo nas observações, e por este motivo observamos um aglomerado de dados ao invés de um único ponto em cada estado de canal.





**Figura 4.11:** Diagrama com amostras do sinal recebido, onde são indicados os estados do canal.

De acordo com os princípios discutidos previamente sobre o sistema nebuloso, é necessário definir os conjuntos fuzzy no espaço de entrada do filtro. Dada a característica particular do vetor de entrada em nosso problema de predição, parece natural que os conjuntos sejam definidos em consonância com a idéia de “*valores em torno do centro  $\mathbf{c}_j$* ”, onde  $\mathbf{c}_j$  denota um dos estados do canal. Isto pode ser implementado utilizando uma função de pertinência gaussiana<sup>1</sup>, e regulando sua dispersão para acomodar da melhor maneira possível os dados *em torno de  $\mathbf{c}_j$* .

Além disso, devemos associar cada uma das possibilidades de entrada do preditor a uma regra que represente a seguinte idéia: *Se o vetor de entrada estiver em torno de  $\mathbf{c}_j$ , então a saída do preditor deverá estar em torno de  $w_j$* , onde o valor  $w_j$  deverá ser ajustado para representar a melhor estimativa para  $x(k + 1)$ .

Note, que ao fazer tais considerações, definimos o filtro estruturalmente. Os parâmetros, e, conseqüentemente, a base de regras deverão ser ajustados de acordo com os dados de treinamento. Em nosso trabalho, utilizamos funções de pertinência gaussianas, um operador união (ou lógico) definido pela soma de funções de pertinência, um operador intersecção (e lógico) definido pelo produto e defuzzificação realizada através do cálculo do centro de massa. A estrutura nebulosa fruto destas escolhas apresenta

---

<sup>1</sup>A escolha de uma função de pertinência gaussiana está relacionada à distribuição do ruído aditivo presente nas observações, que é usualmente modelada pela distribuição normal.

um mapeamento entre entrada e saída dado por [53]:

$$y(k) = \frac{\sum_{j=1}^{N_r} w_j \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{N_r} \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}, \quad (4.26)$$

onde  $N_r$  denota o número de regras e  $\mathbf{x}_p(k)$  representa o vetor de entrada do filtro, enquanto que  $w_l$ ,  $\mathbf{c}_j$  e  $\sigma^2$  representam os três conjuntos de parâmetros do filtro: os pesos de saída, os centros das funções de pertinência e a variância das funções de pertinência, respectivamente [52].

O treinamento a partir dos dados pode, portanto, ser dividido em duas etapas distintas. A definição dos centros e variâncias das funções de pertinência dependem exclusivamente dos vetores de entrada do filtro. A determinação de seus valores pode ser feita através de um processo de *clusterização* dos dados de entrada, que identifica os centros dos agrupamentos formados e a dispersão dos dados em torno deles.

Uma vez definidos os centros e dispersões, nota-se que o mapeamento (4.26) pode ser reescrito da seguinte forma:

$$y(k) = \sum_{j=1}^{N_r} w_j \psi_j(k), \quad (4.27)$$

onde

$$\psi_j(k) = \frac{\exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}{\sum_{l=1}^{N_r} \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_l\|^2}{2\sigma^2}\right)}, \quad (4.28)$$

ou seja, o ajuste dos pesos se torna um problema *linear nos parâmetros*, o qual pode ser resolvido através de uma infinidade de técnicas.

### 4.3.3 Preditor de Mínimo Erro Quadrático Médio e o Preditor Fuzzy

A opção pela estrutura não-linear baseada em sistemas Fuzzy não foi ao acaso. Devido às características particulares do problema em mãos, podemos verificar que o preditor de mínimo erro quadrático médio possui a mesma estrutura do filtro fuzzy mostrada em (4.26). A seguir mostramos, seguindo os passos delineados em [52], que ambos

são equivalentes.

Sabe-se que o estimador de mínimo erro quadrático de  $x(k)$  a partir do vetor  $\mathbf{x}_p(k) = [x(k-1), x(k-2), \dots, x(k-N_p)]^T$  de  $N_p$  amostras passadas de  $x(k)$  é dado pela esperança condicional [80]

$$\hat{x}(k) = E \{x(k) | \mathbf{x}_p(k)\}, \quad (4.29)$$

isto é

$$E \{x | \mathbf{x}_p\} = \int_{-\infty}^{\infty} x \cdot p(x | \mathbf{x}_p) dx, \quad (4.30)$$

onde  $p(x | \mathbf{x}_p)$  é a pdf condicional de  $x$  em relação a  $\mathbf{x}_p$ . Os índices temporais foram omitidos apenas para não sobrecarregar a notação.

Utilizando o teorema de Bayes, podemos escrever a pdf condicional como

$$p(x | \mathbf{x}_p) = \frac{p(x, \mathbf{x}_p)}{p(\mathbf{x}_p)}, \quad (4.31)$$

onde  $p(x, \mathbf{x}_p)$  é a pdf conjunta entre  $x$  e  $\mathbf{x}_p$  e  $p(\mathbf{x}_p)$  é a pdf marginal de  $\mathbf{x}_p$ .

Vejamos inicialmente o cálculo da distribuição de  $\mathbf{x}_p$ . Utilizando o teorema da probabilidade total [109], podemos expressar a distribuição de  $\mathbf{x}_p$  como

$$p(\mathbf{x}_p) = \sum_{j=1}^{N_\Omega} p(\mathbf{x}_p | \mathbf{c}_j) \cdot P(\mathbf{c}_j), \quad (4.32)$$

sendo que  $\mathbf{c}_j$  são os estados do canal de dimensão  $N_p$ , igual à dimensão dos vetores de entrada do filtro, e  $N_\Omega$  denota o número de possibilidades para os estados  $\mathbf{c}_j$ . Como usualmente considera-se um ruído aditivo gaussiano nas medidas,  $p(\mathbf{x}_p | \mathbf{c}_j)$  é uma distribuição condicional Gaussiana com média igual a  $\mathbf{c}_j$ , ou seja:

$$p(\mathbf{x}_p | \mathbf{c}_j) = (2\pi\sigma^2)^{\frac{-(N_p)}{2}} \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_j\|^2}{2\sigma^2}\right). \quad (4.33)$$

Considerando que todas as fontes transmitem com um mesmo alfabeto de cardinalidade  $S$ , pode-se verificar que o número de possibilidades para o vetor  $\mathbf{c}_j$  de dimensão  $N_p$  é dado por:

$$N_\Omega = S^{(N_p-1)N + \sum_{i=1}^N L_{c_i}}, \quad (4.34)$$

onde  $N$  denota o número de fontes e  $L_{c_i}$  denota o comprimento do canal ligando a fonte  $i$  e a observação. Além disso, é comum assumir que os símbolos transmitidos são equiprováveis, o que torna os estados do canal também equiprováveis, com probabilidade  $P(\mathbf{c}_j) = 1/N_\Omega$ . Assim, a pdf de  $\mathbf{x}_p$  assume a forma de

$$p(\mathbf{x}_p) = \frac{(2\pi\sigma^2)^{-\frac{(N_p)}{2}}}{N_\Omega} \sum_{j=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_j\|^2}{2\sigma^2}\right). \quad (4.35)$$

Resta agora o cálculo da distribuição conjunta  $p(x, \mathbf{x}_p)$ , que pode ser efetuado através do teorema da probabilidade total:

$$\begin{aligned} p(x, \mathbf{x}_p) &= \sum_{j=1}^{N_\Omega+} p(x, \mathbf{x}_p | \mathbf{c}_j^+) \cdot P(\mathbf{c}_j^+) \\ &= \sum_{j=1}^{N_\Omega+} p(x | \mathbf{x}_p, \mathbf{c}_j^+) \cdot p(\mathbf{x}_p | \mathbf{c}_j^+) \cdot P(\mathbf{c}_j^+), \\ &= \sum_{j=1}^{N_\Omega+} p(x | \mathbf{c}_j^+) \cdot p(\mathbf{x}_p | \mathbf{c}_j^+) \cdot P(\mathbf{c}_j^+) \end{aligned} \quad (4.36)$$

onde  $\mathbf{c}_j^+$  é o estado de dimensão  $N_p + 1$  do canal (mesma dimensão do vetor  $[x, \mathbf{x}_p]^T$ ), e  $N_\Omega^+$  o número de possíveis estados desta dimensão, i.e.:

$$N_\Omega^+ = S^{(N_p)N + \sum_{i=1}^N L_{c_i}}. \quad (4.37)$$

De acordo com (4.33), a distribuição condicional de  $x$  dado o estado  $\mathbf{c}_j^+$  pode ser escrita como

$$p(x | \mathbf{c}_j^+) = (2\pi\sigma^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{|x - c_{j,0}^+|^2}{2\sigma^2}\right), \quad (4.38)$$

onde  $c_{j,0}^+$  denota o primeiro elemento do vetor  $\mathbf{c}_j^+$ .

Por sua vez,  $p(\mathbf{x}_p | \mathbf{c}_j^+)$  é expresso por

$$p(\mathbf{x}_p | \mathbf{c}_j^+) = (2\pi\sigma^2)^{-\frac{N_p}{2}} \cdot \exp\left(-\frac{\|\mathbf{x}_p - \hat{\mathbf{c}}_j^+\|^2}{2\sigma^2}\right), \quad (4.39)$$

onde  $\hat{\mathbf{c}}_j^+$  corresponde ao vetor  $[c_{j,1}^+ \ c_{j,2}^+ \ \cdots \ c_{j,N_p}^+]^T$ .

Novamente, sabendo que os estados são equiprováveis, isto é,  $P(\mathbf{c}_j^+) = 1/N_\Omega^+$ , substituindo (4.38) e (4.39) em (4.36), temos:

$$p(x, \mathbf{x}_p) = \frac{(2\pi\sigma^2)^{-\frac{(N_p+1)}{2}}}{N_\Omega^+} \sum_{j=1}^{N_\Omega^+} \exp\left(-\frac{|x - c_{j,0}^+|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\mathbf{x}_p - \hat{\mathbf{c}}_j^+\|^2}{2\sigma^2}\right). \quad (4.40)$$

Finalmente, de posse destes resultados, resta determinar a distribuição condicional de  $x$  dado o vetor  $\mathbf{x}_p$ . Substituindo (4.35) e (4.40) em (4.31), obtemos:

$$p(x|\mathbf{x}_p) = \frac{(2\pi\sigma^2)^{-\frac{1}{2}} \sum_{j=1}^{N_\Omega^+} \exp\left(-\frac{|x - c_{j,0}^+|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\mathbf{x}_p - \hat{\mathbf{c}}_j^+\|^2}{2\sigma^2}\right)}{S^N \sum_{j=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_j\|^2}{2\sigma^2}\right)}. \quad (4.41)$$

Conhecendo a distribuição condicional, podemos agora calcular o mapeamento do preditor de mínimo erro quadrático médio substituindo (4.41) em (4.30):

$$\begin{aligned} \hat{x} &= \int_{-\infty}^{\infty} x \cdot p(x|\mathbf{x}_p) dx \\ &= \frac{\sum_{j=1}^{N_\Omega^+} \exp\left(-\frac{\|\mathbf{x}_p - \hat{\mathbf{c}}_j^+\|^2}{2\sigma^2}\right) \cdot \int_{-\infty}^{\infty} x \cdot (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{|x - c_{j,0}^+|^2}{2\sigma^2}\right) dx}{S^N \cdot \sum_{j=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_j\|^2}{2\sigma^2}\right)} \\ &= \frac{\sum_{j=1}^{N_\Omega^+} \frac{c_{j,0}^+}{S^N} \cdot \exp\left(-\frac{\|\mathbf{x}_p - \hat{\mathbf{c}}_j^+\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_j\|^2}{2\sigma^2}\right)} \end{aligned} \quad (4.42)$$

Como  $\hat{\mathbf{c}}_j^+ = \mathbf{c}_j$ , obtemos a fórmula final para o estimador de mínimo erro quadrático:

$$\hat{x}(k) = \frac{\sum_{j=1}^{N_\Omega} w_j \cdot \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}, \quad (4.43)$$

onde os pesos  $w_j$  são definidos por:

$$w_j = \frac{1}{S^N} \sum_{n=1}^{S^N} c_{n,0}^+ \quad (4.44)$$

Comparando o resultado obtido em (4.42) com (4.26) notamos que o mapeamento provido pelo filtro fuzzy é, exatamente, o preditor de mínimo erro quadrático. Desta forma, a escolha pela estrutura fuzzy é justificada, já que esta será a melhor estrutura para minimizar o erro de predição quadrático.

#### 4.3.4 Algoritmo de Clusterização - K-means + Iterated Local Search

Conforme vimos, a primeira etapa do projeto do preditor baseado em um sistema nebuloso depende da estimação correta dos centros dos agrupamentos no espaço de entrada do filtro. Este processo é obtido empregando-se um algoritmo de clusterização não-supervisionada, como o clássico algoritmo *K-means*.

Considere um conjunto de vetores de observados  $\mathcal{C} = \{\mathbf{x}_p(1), \mathbf{x}_p(2), \dots, \mathbf{x}_p(m)\}$  com dimensão  $N_p$ . Considerando que todas as fontes utilizam o mesmo alfabeto na transmissão serão formados  $N_\Omega$  agrupamentos, indicados pelos conjuntos  $\mathcal{C}_i$ , conforme vimos anteriormente.

O objetivo da etapa de clusterização é, portanto, obter estimativas dos centros  $\mathbf{c}_j$  de maneira que a distância euclidiana quadrática entre os vetores de  $\mathcal{C}$  e o centro do cluster ao qual pertencem seja minimizada, i.e., minimizando a função custo de soma dos quadrados (SQ)

$$J_{SQ} = \sum_{j=1}^{N_\Omega} \sum_{i \in \mathcal{C}_j} \|\mathbf{x}_p(i) - \mathbf{c}_j\|^2 \quad (4.45)$$

onde  $\|\cdot\|$  representa a norma euclidiana.

É necessário dizer que o método de clusterização utilizado, embora não conte com nenhum sinal piloto, depende da informação sobre o número de clusters a ser encontrado. Isto, em nosso problema, requer o conhecimento prévio ou estimação do comprimento dos canais, de forma que seja possível estimar o valor de  $N_\Omega$  em (4.34).

A função custo  $J_{SQ}$  não é convexa e sua minimização constitui um problema complicado de otimização. Um dos métodos heurísticos desenvolvidos para resolvê-lo é

o algoritmo *K-means* [32], que busca minimizar  $J_{SQ}$  de forma iterativa a partir de um conjunto inicial dos centros dos agrupamentos, ditos centróides. Uma das versões presentes na literatura, denominada *Optimal K-means* [32], utilizada em nosso trabalho, é apresentada no quadro 4. A diferença desta versão em relação ao algoritmo original está na escolha do passo de aprendizado, feita neste caso de maneira adaptativa, de forma a melhorar as características de convergência do algoritmo. Note que o algoritmo, além de ajustar os centros dos clusters, também obtém o valor da dispersão de cada um dos agrupamentos, um dos parâmetros do filtro fuzzy.

Após a convergência, ainda é possível realizar um pequeno ajuste ao valor obtido para os centros. Considerando que  $\mathbf{c}_j$  deve estar localizado exatamente no centro de cada agrupamento, podemos atualizar seu valor baseado no particionamento obtido, de acordo com

$$\mathbf{c}_j = \frac{1}{m_j} \sum_{i \in \mathcal{C}_j} \mathbf{x}(i) \quad (4.52)$$

onde  $m_j$  denota o número de padrões associados ao cluster  $j$ .

O algoritmo, apesar de eficaz quando os centros iniciais forem escolhidos de forma adequada, não garante que a melhor solução seja encontrada no caso mais geral devido à função custo ser não-convexa. Portanto, seria interessante obter algum método menos sensível à inicialização, com maior capacidade de exploração do espaço de busca.

A nossa opção foi utilizar um paradigma de otimização, denominado *Iterated Local Search*, ou ILS, que não define propriamente um algoritmo, e sim uma metodologia para utilização mais eficiente de algoritmos de busca já existentes. A idéia central da ILS consiste no seguinte: utiliza-se um algoritmo de busca local, como *K-means*, para encontrar uma solução, possivelmente um mínimo local da função custo. Em seguida, aplica-se uma pequena perturbação (uma mutação) à solução obtida, e o resultado é utilizado como inicialização de um novo procedimento de busca, utilizando-se o mesmo algoritmo. Caso esta nova busca resulte em uma solução que reduza ainda mais a função custo, ela é mantida, e o procedimento é repetido, até que um critério de parada seja atingido. Os passos são descritos no quadro 5.

A utilização conjunta de uma ferramenta de busca local, como o *K-means*, com um procedimento como o ILS, confere ao método de clusterização, ao qual denominaremos *K-means/ILS*, uma grande capacidade de exploração do espaço de busca, e torna o procedimento de busca dos centros menos sensível à inicialização.

Para nossos propósitos, optamos pelas seguintes escolhas para implementação do

1. Inicializar aleatoriamente um conjunto de centróides  $\mathbf{c}_j(0)$  e um conjunto de variâncias  $v_j(k)$ . As variâncias devem ser inicializadas com valores próximos a zero;
2. Enquanto o critério de parada não for satisfeito:
  - (a) Determinar a qual cluster pertence o vetor atual  $\mathbf{x}_p(k)$ :

$$j' = \arg \min_j \{v_j(k) \|\mathbf{x}_p(k) - \mathbf{c}_j(k)\|^2\}, \quad 1 \leq j \leq N_\Omega \quad (4.46)$$

onde  $\mathbf{c}_j(k)$  e  $v_j(k)$  correspondem, respectivamente, ao centróide e à variância do cluster  $j$  na iteração  $k$ ;

- (b) Atualizar o passo de adaptação de acordo com

$$\alpha_k = \frac{\ln(N_\Omega) - \varepsilon(k)}{\ln(N_\Omega)} \quad (4.47)$$

onde  $\varepsilon(k)$  é uma medida de entropia das variâncias dos clusters, dada por

$$\varepsilon(k) = \sum_{j=1}^{N_\Omega} -v_{j,norm}(k) \ln(v_{j,norm}(k)) \quad (4.48)$$

e  $v_{j,norm}(k)$  é definido como

$$v_{j,norm}(k) = \frac{v_j(k)}{\sum_{i=1}^{N_\Omega} v_i(k)} \quad (4.49)$$

- (c) Atualizar os valores do centróide e da variância do cluster  $j'$ :

$$\mathbf{c}_{j'}(k+1) = \mathbf{c}_{j'}(k) + \alpha_k (\mathbf{c}_j(k) - \mathbf{c}_{j'}(k)) \quad (4.50)$$

$$v_{j'}(k+1) = \xi v_{j'}(k) + (1 - \xi) \|\mathbf{c}_j(k) - \mathbf{c}_{j'}(k)\|^2 \quad (4.51)$$

onde  $\xi$  é uma constante de esquecimento ligeiramente menor que 1.

#### Quadro 4: Algoritmo *Optimal K-means*

algoritmo de clusterização *K-means*/ILS:

- **Inicialização:** Considerando um conjunto de  $m$  vetores de dimensão  $N_p$  como entrada de nosso preditor, inicializamos os centros  $\mathbf{c}_j(0)$  com vetores escolhidos aleatoriamente dentre os  $m$  disponíveis;



1. Gerar uma inicialização aleatória  $\mathbf{c}_{inicial}$  para o algoritmo de busca local;
2.  $\mathbf{c}^* = \text{Busca\_local}(\mathbf{c}_{inicial})$ ;
3. Repetir enquanto o critério de parada não for atendido:
  - (a)  $\mathbf{c}' = \text{Perturbação}(\mathbf{c}^*)$ ;
  - (b)  $\mathbf{c}'^* = \text{Busca\_local}(\mathbf{c}')$ ; Se o custo  $J(\mathbf{c}'^*) < J(\mathbf{c}^*)$ , então  $\mathbf{c}^* = \mathbf{c}'^*$

**Quadro 5:** Princípio da meta-heurística *Iterated Local Search*

- **Busca Local:** A busca local é executada pelo algoritmo *Optimal K-means*, apresentado no quadro 4;
- **Mutação:** Uma característica de nosso problema de clusterização é que os estados de entrada do preditor são equiprováveis. Isto faz com que o número de vetores associados a um mesmo cluster tenda a ser igual para todos os agrupamentos. No entanto, em alguns casos, o algoritmo *K-means* converge para uma solução na qual um centro está associado a mais de um agrupamento de dados, enquanto outros estão associados a um mesmo cluster. Assim, o procedimento adotado para realizar a mutação foi o seguinte:
  - Verificar o número médio de vetores associados a cada um dos clusters, o qual deve ser próximo de  $m/N_\Omega$ .
  - Caso alguns dos clusters apresentem uma diferença muito grande em relação ao valor esperado, deve-se identificar o centro associado ao cluster com o menor número de vetores e substituí-lo pelo centro associado ao cluster com a maior quantidade de vetores, e aplicar uma pequena perturbação, i.e.:

$$\mathbf{c}_{novo} = \mathbf{c}_{maior} + \beta \quad (4.53)$$

onde  $\mathbf{c}_{novo}$  representa o novo centro,  $\mathbf{c}_{maior}$  o centro associado ao cluster com maior número de vetores, e  $\beta$  denota uma variável aleatória gaussiana com média nula e variância pequena.

- **Critério de parada:** Novamente, observando-se que o número de vetores em cada um dos clusters deve ser aproximadamente igual, utilizamos o seguinte critério de parada

$$\sum_{i=1}^{N_\Omega} \frac{m_i}{m} \log \left( \frac{m_i}{m} \right) > \alpha \log (N_\Omega) \quad (4.54)$$

onde  $m_i$  representa o número de vetores associados ao cluster  $i$ , e  $\alpha$  denota

uma constante pouco menor que 1. Note que o critério mede uma forma de “entropia” da distribuição dos vetores, que será sempre inferior a  $\log(N_\Omega)$ .

### 4.3.5 Adaptação dos pesos de saída da estrutura fuzzy

Conforme dito anteriormente, uma vez definidos os centros dos agrupamentos de dados, o mapeamento entre a entrada e saída do filtro fuzzy torna-se linear nos parâmetros. Em nossas simulações utilizamos o algoritmo RLS para realizar a adaptação.

Reescrevendo a equação de saída do filtro fuzzy (4.27) temos

$$y(n) = \mathbf{w}_{fuzzy}^T \Psi(k) \quad (4.55)$$

onde  $\mathbf{w}_{fuzzy} = [w_1, w_2, \dots, w_{N_\Omega}]^T$  denota o vetor contendo os pesos de saída da estrutura, e  $\Psi(k) = [\psi_1(k), \psi_2(k), \dots, \psi_{N_\Omega}(k)]^T$ , onde

$$\psi_j(k) = \frac{\exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_j\|^2}{2\sigma^2}\right)}{\sum_{l=1}^{N_\Omega} \exp\left(-\frac{\|\mathbf{x}_p(k) - \mathbf{c}_l\|^2}{2\sigma^2}\right)} \quad (4.56)$$

O algoritmo utilizado é resumido no quadro 6.

1. Inicializar os vetor  $\mathbf{w}_{fuzzy}$  com valores próximos a zero, e  $\mathbf{P}$  com uma matriz diagonal;
2. enquanto não convergir, executar as seguintes etapas de adaptação:

$$\begin{aligned} e &= x(k) - \mathbf{w}_{fuzzy}^T \Psi(k) \\ \mathbf{g} &= \mathbf{P} \Psi(k) [\lambda + \Psi(k)^T \mathbf{P} \Psi(k)]^{-1} \\ \mathbf{P} &= \lambda^{-1} \mathbf{P} - \mathbf{g} \Psi(k)^T \lambda^{-1} \mathbf{P} \\ \mathbf{w}_{fuzzy} &= \mathbf{w}_{fuzzy} + e \mathbf{g} \end{aligned} \quad (4.57)$$

onde  $\lambda$  representa o fator de esquecimento do algoritmo.

**Quadro 6:** Algoritmo RLS utilizado para adaptação do vetor de pesos  $\mathbf{w}_{fuzzy}$ .

## 4.4 Resumo da proposta

Após a exposição dos conceitos relacionados à proposta, resumimos aqui as etapas que compõem o método proposto neste capítulo:

1. Definir o número de amostras a serem consideradas no processo de predição;
2. A partir de um conjunto de amostras do sinal, determinar os centros e variâncias utilizando o algoritmo K-means/ILS;
3. Com os centros e variâncias definidos, realizar a adaptação dos pesos em (4.27), utilizando o algoritmo RLS descrito no quadro 6;
4. Com os parâmetros ajustados, executar o processo de predição e obter os erros de predição  $e_i(k)$ , formando um vetor de observações  $\mathbf{e}(k)$ ;
5. Aplicar qualquer método de separação de mistura linear, como FastICA, ao vetor  $\mathbf{e}(k)$ , obtendo assim as estimativas das fontes.

## 4.5 Resultados de Simulações

De forma a ilustrar e avaliar o método proposto, definimos uma métrica baseada no erro quadrático médio entre o sinal obtido após o uso do algoritmo FastICA e o sinal real das fontes

$$MSE_i = \frac{1}{K} \sum_{n=1}^K \{[y_i(n) - s_j(n)]^2\} \quad (4.58)$$

onde  $K$  é o número de amostras disponíveis. A comparação é realizada após normalização da variância dos dados e levando-se em consideração possíveis permutações nos sinais recuperados. Nos cenários simulados a seguir foram utilizados preditores com 2 entradas apenas, as amostras  $x_i(k-1)$  e  $x_i(k-2)$ .

Como base de comparação, utilizamos uma estrutura linear para estimar os sinais das fontes. O sinal é estimado a partir de amostras dos sinais provenientes de todos os sensores, ou seja, uma estrutura MISO (*Multiple-Input Single-Output*). A saída

do estimador linear é dada por

$$\begin{aligned}\hat{s}_i(k) &= \mathbf{w}_i^T(0) \mathbf{x}(k) + \cdots + \mathbf{w}_i^T(N_p - 1) \mathbf{x}(k - N_p + 1) \\ &= \begin{bmatrix} \mathbf{w}_i^T(0) & \cdots & \mathbf{w}_i^T(N_p - 1) \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \vdots \\ \mathbf{x}(k - N_p + 1) \end{bmatrix} \\ &= \bar{\mathbf{w}}_i^T \bar{\mathbf{x}}(k)\end{aligned}\quad (4.59)$$

onde  $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_M(k)]^T$  representa o vetor contendo os sinais de todos os sensores.

O vetor de pesos  $\bar{\mathbf{w}}_i^T$  que minimiza o erro quadrático médio  $E\{(\hat{s}_i(k) - s_i(k - d))^2\}$  é dado pela solução de Wiener [61]:

$$\bar{\mathbf{w}}_i = \bar{\mathbf{R}}_i^{-1} \bar{\mathbf{p}}_i \quad (4.60)$$

onde

$$\begin{aligned}\bar{\mathbf{R}}_i &= E\{\bar{\mathbf{x}}(k) \bar{\mathbf{x}}^T(k)\} \\ \bar{\mathbf{p}}_i &= E\{\bar{\mathbf{x}}(k) s_i(k - d)\}\end{aligned}\quad (4.61)$$

Para cada possível atraso  $d$  temos uma solução de Wiener correspondente, que pode apresentar maior ou menor erro residual dependendo do processo de mistura. Nos ensaios, a solução de Wiener foi estimada a partir dos dados observados, considerando todo o conjunto de observações disponíveis, avaliando-se qual o melhor atraso para cada uma das fontes. Dessa forma, garantimos que as técnicas de separação baseadas em estruturas lineares apresentarão desempenho, na melhor das hipóteses, semelhantes ao do estimador linear.

## Primeiro cenário

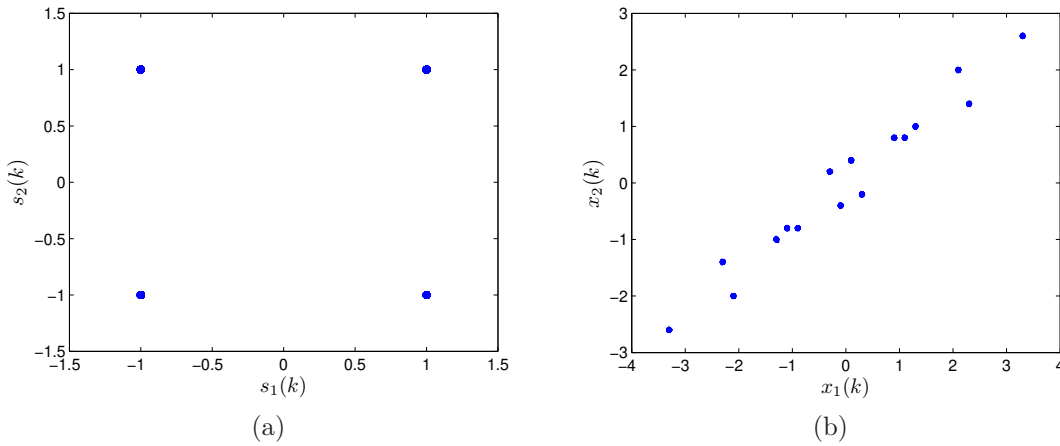
O primeiro cenário de simulação consiste de uma mistura com  $L = 2$ , sendo as matrizes de mistura dadas por:

$$\mathbf{A}(0) = \begin{bmatrix} 1 & 0,5 \\ 0,8 & 0,6 \end{bmatrix}, \mathbf{A}(1) = \begin{bmatrix} 0,6 & 1,2 \\ 0,3 & 0,9 \end{bmatrix} \quad (4.62)$$

e variância do ruído  $\sigma^2 = 10^{-4}$ .

É interessante comentar que a matriz polinomial  $\mathbf{A}[z] = \mathbf{A}(0) + \mathbf{A}(1)z^{-1}$  apresenta um determinante que depende do valor de  $z$ . Isto mostra que o posto da matriz  $\mathbf{A}[z]$  pode se anular para alguns valores de  $z$ , e dessa forma não é passível de inversão perfeita através de uma estrutura MIMO-FIR, conforme estabelecido pelo teorema 4.1.2.

A figura 4.12(a) mostra a distribuição original das fontes, dois sinais independentes com amostras i.i.d., pertencentes ao alfabeto  $\{+1, -1\}$ . Ao serem misturadas através do sistema definido pelas duas matrizes, obtém-se o vetor de sinais  $\mathbf{x}(k)$ , cujos elementos têm a distribuição conjunta mostrada na figura 4.12(b). Foram realizados 20 experimentos no total, considerando-se 10000 amostras em cada experimento.

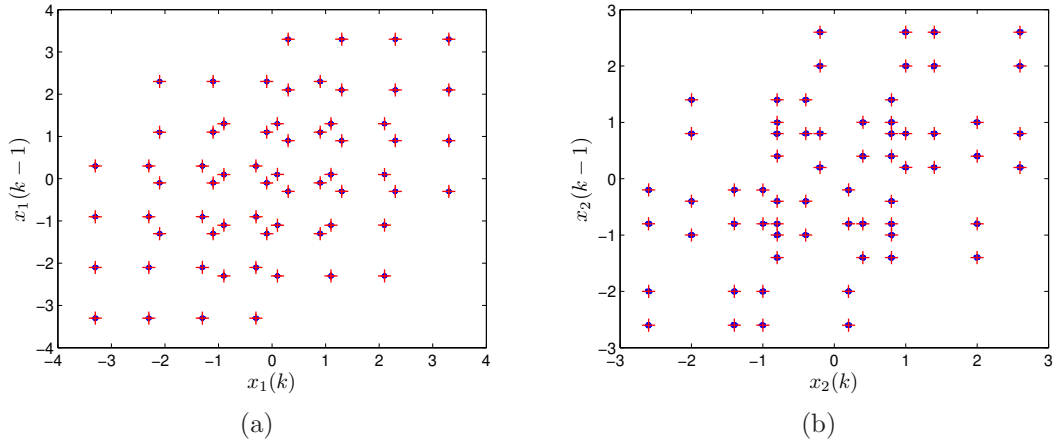


**Figura 4.12:** Distribuição (a) das fontes; e (b) dos dados após o processo de mistura

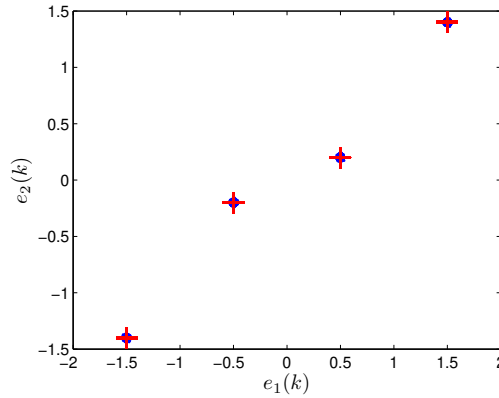
A primeira etapa do treinamento do filtro fuzzy, e crucial para a obtenção da estrutura ótima de filtragem, consiste na clusterização dos vetores de entrada. Neste caso, utilizando (4.34), podemos verificar que existem 64 estados a serem estimados. Nas figuras 4.13(a) e 4.13(b) são mostradas as distribuições dos dados nas entradas dos preditores, bem como os centros obtidos. Neste caso, nota-se que os centros foram encontrados de maneira satisfatória.

Após a adaptação dos centros, passamos ao ajuste dos pesos de saída, feito em nosso exemplo com o algoritmo RLS, com fator de esquecimento  $\lambda = 1$ . Os filtros resultantes foram utilizados na predição dos sinais  $x_1(k)$  e  $x_2(k)$ , de forma que, na figura 4.14, mostramos a distribuição dos erros de predição, onde observa-se que a distribuição resultante equivale à de uma mistura linear instantânea das fontes.

Comparamos o erro definido em (4.58) em 2 situações: na primeira delas (caso ideal),



**Figura 4.13:** Distribuição dos dados nas entradas dos 2 preditores, destacando com um + a posição dos centros encontrados pelo algoritmo de clusterização



**Figura 4.14:** Distribuição conjunta dos erros de predição da estrutura fuzzy; (+) denota a distribuição esperada para a mistura instantânea correspondente.

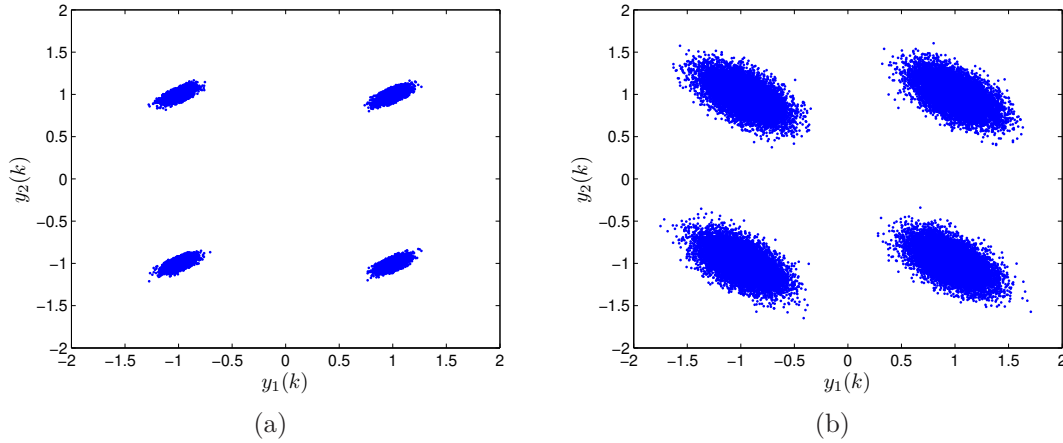
os estados do canal, assim como a variância do ruído, são conhecidos *a priori*; na segunda situação (caso adaptado), os estados do canal são estimados através do algoritmo de clusterização *K-means/ILS*. Os resultados médios obtidos podem ser observados na tabela 4.1, onde incluímos também o resultado para a estrutura linear com 20 entradas.

Observa-se que em ambas situações o método obteve bons resultados. A diferença de desempenho entre o caso ideal e o adaptado se deve a uma diferença mínima entre

**Tabela 4.1:** Resultados para o primeiro cenário

	$MSE_1$	$MSE_2$
Caso Ideal	$1 \times 10^{-4}$	$9,95 \times 10^{-5}$
Caso Adaptado	$5,67 \times 10^{-3}$	$4,9 \times 10^{-3}$
Estimador linear	$13,95 \times 10^{-3}$	$8,56 \times 10^{-3}$

a solução obtida pelo algoritmo K-means/ILS e os centros reais, muito embora o resultado final ainda seja satisfatório. No entanto, a abordagem de filtragem linear, como era de se esperar, apresentou um erro pelo menos duas vezes maior, mesmo contando com um maior número de entradas. Na figura 4.15 mostramos a distribuição conjunta dos sinais recuperados utilizando o algoritmo FastICA.



**Figura 4.15:** Distribuição conjunta dos sinais recuperados (a) a partir do erro de predição da estrutura fuzzy e (b) da estrutura linear.

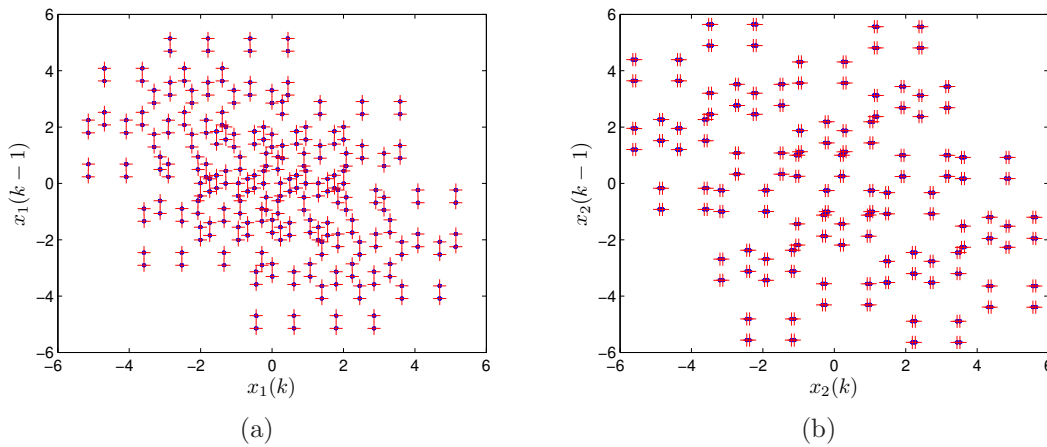
## Segundo cenário

No segundo cenário simulado considerou-se uma mistura com  $L = 3$ , descrito através das matrizes

$$\mathbf{A}(0) = \begin{bmatrix} 1,12 & 0,53 \\ 0,625 & 0,04 \end{bmatrix}, \mathbf{A}_1 = \begin{bmatrix} -1,57 & 0,92 \\ 1,0625 & -2,32 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0,78 & 0,225 \\ -0,375 & 1,22 \end{bmatrix} \quad (4.63)$$

e ruído com variância  $\sigma^2 = 10^{-4}$ .

Da mesma forma que no exemplo anterior, o sistema formado por tais matrizes não admite inversão perfeita por meio de uma estrutura FIR. Os mesmos passos obtidos no primeiro cenário foram seguidos. Na figura 4.16, apresentamos o resultado obtido na etapa de clusterização. Neste caso, o número de estados a serem estimados passa para 256. Nota-se que, a despeito do grande número de agrupamentos, o algoritmo é capaz de estimar com acuidade os centros. Na tabela 4.3 mostramos os resultados de 20 simulações considerando o segundo cenário.



**Figura 4.16:** Distribuição dos dados nas entradas dos 2 preditores, destacando com um + a posição dos centros encontrados pelo algoritmo de clusterização.

**Tabela 4.2:** Resultados para o segundo cenário

	$MSE_1$	$MSE_2$
Caso Ideal	$1 \times 10^{-4}$	$1 \times 10^{-4}$
Caso Adaptado	$8,18 \times 10^{-3}$	$6,69 \times 10^{-3}$
Estimador linear	$28,56 \times 10^{-3}$	$10,55 \times 10^{-3}$

Observa-se novamente uma pequena diferença de desempenho entre o caso de centros adaptados pelo *K-means*/ILS e com conhecimento *a priori* dos estados do canal, enquanto que o caso linear apresenta um patamar de erro acima dos atingidos pela estrutura fuzzy.

Porém, a partir deste exemplo, começa a ficar claro uma das principais dificuldades enfrentadas por este método. Devido à natureza combinatória das possibilidades de vetores de entrada dos preditores, o número de centros, e conseqüentemente de parâmetros, cresce de maneira exponencial com o comprimento dos canais.



### Terceiro cenário

Em um último cenário, consideramos uma mistura definida pelas matrizes utilizadas no exemplo 4.1.1:

$$\mathbf{A}(0) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{A}(1) = \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix}. \quad (4.64)$$

com ruído aditivo de variância  $\sigma^2 = 10^{-2}$ . É importante lembrar que para este tipo de mistura não é possível encontrar uma estrutura linear que a inverta perfeitamente.

Numa primeira rodada de simulações percebemos que a abordagem utilizando preditores não-lineares SISO não conseguia realizar a separação dos sinais. A explicação é revelada observando-se o sinal captado por cada um dos sensores:

$$\begin{aligned} x_1(k) &= 3s_1(k) + 1s_2(k-1) \\ x_2(k) &= 1s_2(k) + 3s_1(k-1) \end{aligned} \quad (4.65)$$

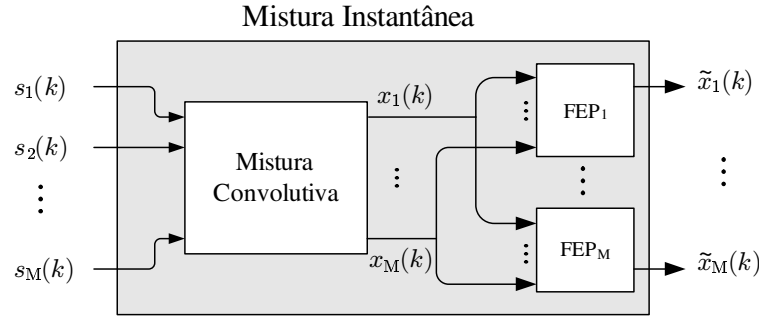
Note que os próprios sinais  $x_1(k)$  e  $x_2(k)$  são compostos por amostras i.i.d., devido à forma particular da mistura. Isto significa que as amostras  $x_1(k-1), x_1(k-2), \dots$ , não carregam nenhuma informação acerca da amostra atual  $x_1(k)$ , o mesmo ocorrendo para o sinal  $x_2(k)$ . Assim, ao tentarmos projetar um preditor SISO para o valor  $x_i(k)$ , obteremos uma estrutura que, segundo (4.30), fornecerá como estimativa o valor  $E\{x_i(k)\} = 0$ . Com isso, a saída do filtro de erro de predição será exatamente igual ao sinal observado, não eliminando o caráter convolutivo da mistura.

Para contornar este problema, modificamos ligeiramente a abordagem utilizada até então empregando preditores não-lineares cujas entradas são amostras passadas dos sinais provenientes de todos os sensores disponíveis, ou seja, o vetor de entrada do preditor fuzzy passa a ser

$$\mathbf{x}_p(k) = [x_1(k-1), \dots, x_1(k-N_p), x_2(k-1), \dots, x_2(k-N_p), \dots, x_{N_f}(k-N_p)]^T, \quad (4.66)$$

transformando o preditor em um sistema MISO, a exemplo do estimador linear em (4.59). A modificação proposta é ilustrada na figura 4.17.

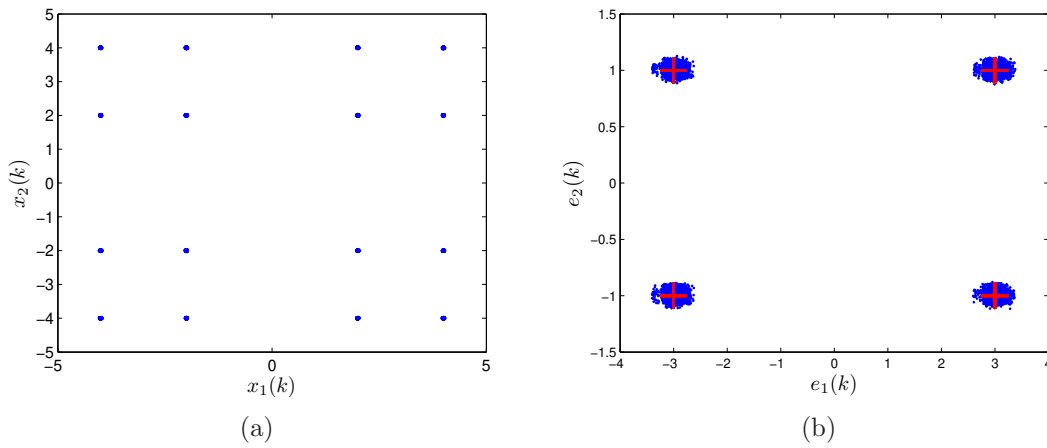
Um ponto a ser comentado sobre a modificação diz respeito ao número de possíveis vetores na entrada do preditor. O aumento na dimensão do vetor  $\mathbf{x}_p(k)$ , em nosso cenário, não acarreta um aumento no número de possíveis estados a serem considerados pelo filtro fuzzy. Como os canais ligando cada par fonte-sensor possuem o mesmo



**Figura 4.17:** Modificação da proposta original, na qual os filtros de erro de predição utilizam como entrada amostras passadas dos sinais provenientes de todos sensores.

comprimento, a cada possível valor da saída  $x_1(k)$  teremos um valor correspondente em  $x_2(k)$  pois os sinais foram gerados pelas mesmas fontes. Isto, no entanto, não significa que a complexidade computacional permanece a mesma, tendo em vista que o processo de clusterização será realizado para vetores de maior dimensão.

Aplicando esta nova estrutura ao problema de separação, obtivemos nas saídas do filtro de erro de predição a distribuição mostrada na figura 4.18.



**Figura 4.18:** Distribuição (a) dos dados observados; e (b) dos dados após o filtro de erro de predição, indicando com + o resultado da mistura instantânea equivalente esperada

**Tabela 4.3:** Resultados para o terceiro cenário

	$MSE_1$	$MSE_2$
Caso Ideal	$1,19 \times 10^{-5}$	$9,92 \times 10^{-4}$
Caso Adaptado	$2,7 \times 10^{-3}$	$11,4 \times 10^{-3}$
Estimador linear	$14,4 \times 10^{-3}$	$133,5 \times 10^{-3}$

## 4.6 Discussão

A metodologia proposta neste capítulo mostrou-se eficiente na remoção do caráter convolutivo da mistura nos cenários simulados. Nos dois primeiros casos, ao utilizar apenas filtros de erro de predição SISO e considerando-se apenas duas entradas, foi possível reduzir o problema convolutivo a um de mistura instantânea, mostrando que estruturalmente o filtro de erro de predição fuzzy foi capaz de inverter uma estrutura MIMO-FIR. O filtro linear, nos casos mencionados, apesar de contar com um número significativamente maior de entradas, não foi capaz de reduzir o erro aos mesmos níveis obtidos com a estrutura nebulosa.

A modificação proposta no último cenário confere à estrutura do filtro nebuloso uma maior flexibilidade, tornando possível, por exemplo, lidar com canais cuja inversa linear é inexistente. Além disso, a estrutura modificada do preditor também representa o estimador de mínimo erro quadrático.

No entanto, é necessário comentar que, ao utilizarmos um filtro de erro de predição progressivo, assumimos tacitamente que a matriz correspondente à mistura linear instantânea também será inversível, pois somente assim seria possível recuperar as fontes de maneira acurada. Na impossibilidade de se contar com esta propriedade, poderíamos lançar mão de estruturas de predição regressiva, buscando eliminar a redundância dos sinais e prover um sistema de mistura equivalente à matriz de mistura correspondente ao máximo atraso  $A(L - 1)$ . Ou ainda, pode-se empregar uma cascata de preditores progressivos e regressivos de forma a ajustar qual o atraso a ser selecionado. Esta mesma estratégia já foi empregada com sucesso no problema de equalização cega de canais SISO [52].

A principal dificuldade enfrentada pela abordagem proposta reside em seu custo computacional crescente. A natureza combinatória das possibilidades de estados de canal faz com que o número de parâmetros a serem ajustados, como os centros e pesos da saída do filtro, cresçam de maneira exponencial com o comprimento dos canais, dificultando a implementação do método em tempo real. Uma alternativa para isso seria considerar outras estruturas não-lineares, como filtros de Volterra, ou redes neu-

rais, que, embora não possuam estruturalmente a forma do preditor ótimo, também apresentam a característica de aproximação universal. Além disso, tais estruturas podem viabilizar o uso deste método em cenários nos quais as fontes não possuem distribuição discreta.

## 4.7 Sumário

Neste capítulo abordamos o problema de separação de fontes no contexto de misturas convolutivas. Conforme vimos, neste cenário, as ferramentas para solução se dividem em dois grandes grupos: separação no domínio freqüencial e separação no domínio temporal.

No primeiro grupo, os métodos realizam a separação dos sinais transformados no domínio da freqüência, onde a mistura convolutiva torna-se um conjunto de misturas instantâneas. Isto permite que os sinais sejam separados com técnicas bem estabelecidas, de maneira independente nas diferentes freqüências. Isto, no entanto, faz com que o problema de permutação das fontes seja revisto, pois permutações não coerentes ao longo das freqüências analisadas incorrerão em sinais recuperados compostos por componentes espectrais de diferentes fontes.

Já no segundo grupo, o critério de independência é utilizado diretamente na adaptação dos pesos da estrutura de separação. As soluções obtidas ainda estão sujeitas a ambigüidades de permutação dos sinais; no entanto, a ambigüidade quanto à amplitude dá lugar a uma indeterminação quanto a um processo de filtragem. Caso informações adicionais sobre as fontes estejam disponíveis, como é o caso de sinais de comunicação digital, é possível recuperar exatamente o sinal das fontes.

Neste contexto de sinais de comunicação digital, propusemos um método para separação dos sinais baseado em estruturas não-lineares de predição. A escolha por uma estrutura em particular, os filtros fuzzy, mostrou-se adequada, haja vista sua equivalência com o preditor ótimo de mínimo erro quadrático. Utilizando tais filtros de predição, foi possível eliminar as redundâncias contidas no sinal recebido, fornecendo como erro de predição um sinal que representa uma combinação linear de amostras das fontes em um único instante temporal  $k$ .

Em outras palavras, a abordagem proposta emprega os filtros de erro de predição com o intuito de reduzir o problema a um de mistura instantânea, para o qual as técnicas vistas no capítulo 3 podem ser aplicadas diretamente. No entanto, deve-se ressaltar que a principal dificuldade imposta por esta abordagem se refere ao

seu custo computacional elevado. O algoritmo de clusterização proposto mostrou-se bastante robusto, conseguindo encontrar os centros de maneira adequada nos cenários propostos. Porém, sua complexidade está diretamente relacionada ao número de centros a serem estimados, que podem crescer exponencialmente com o número de coeficientes dos canais.

A despeito de sua complexidade, a metodologia proposta, conforme visto através das simulações apresentadas, mostrou-se eficaz, atuando em cenários nos quais as estruturas MIMO-FIR não seriam capazes de inverter perfeitamente o processo de mistura, e mesmo no caso em que a mistura não é passível de inversão por estruturas lineares.

# 5

## Misturas Não-Lineares

O sucesso das diferentes abordagens desenvolvidas para a separação de fontes em misturas lineares levou à consideração de outros tipos de misturas, de maneira a cobrir um maior número de cenários encontrados na prática. Neste contexto, uma extensão natural dos trabalhos em BSS seria considerar modelos não-lineares para o processo de mistura.

Neste capítulo, discutiremos sobre o problema de BSS quando o processo de mistura é modelado por um sistema não-linear. Veremos algumas propostas que buscaram soluções para o caso geral de misturas não-lineares, e comentaremos sobre um modelo particular, denominado modelo de mistura com não-linearidade posterior, ou PNL (*Post-Nonlinear*), para o qual a ICA pode ser empregada para separação das fontes.

Por fim, apresentaremos uma proposta para separação de sinais em uma mistura PNL que utiliza uma ferramenta de otimização baseada em redes imunológicas artificiais. Sua principal característica é a grande capacidade de busca global. Aliado ao método de otimização empregamos um algoritmo para estimação da entropia dos sinais recuperados, grandeza necessária para a avaliação da função custo a ser otimizada, baseado em estatísticas de ordem, que, a despeito de sua simplicidade, fornece valores acurados para o processo de otimização.

## 5.1 Misturas não-lineares

Considerar modelos com misturas não-lineares pode ser visto como uma extensão natural dos trabalhos em BSS que, em sua grande maioria, lidam apenas com o caso linear. Em sua formulação mais geral, o modelo de mistura não-linear instantânea (sem memória) pode ser expresso através de um mapeamento  $\mathbf{F} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  tal que

$$\mathbf{x}(k) = \mathbf{F}(\mathbf{s}(k)). \quad (5.1)$$

Buscamos, portanto, um mapeamento, que neste caso será necessariamente não-linear, ou seja:

$$\mathbf{y}(k) = \mathbf{G}(\mathbf{x}(k)), \quad (5.2)$$

tal que  $\mathbf{y}(k)$  seja uma estimativa precisa das fontes. Note que, em contraste com o caso linear, a escolha da estrutura de separação  $\mathbf{G}$  representa uma tarefa de considerável dificuldade. Como, em geral, funções parametrizadas são utilizadas, é provável que não seja possível inverter completamente o mapeamento  $F(\cdot)$ .

### 5.1.1 Técnicas para o caso geral

Uma das primeiras abordagens sugeridas para a separação de misturas não-lineares consistiu na utilização de mapas auto-organizáveis [107], um paradigma de rede neural cujo aprendizado é feito de maneira competitiva e não-supervisionada.

Um mapa auto-organizável [59] realiza um mapeamento entre o espaço de entrada, constituído pelos dados observados, e o espaço de saída, onde são definidos nós com localizações pré-estabelecidas, normalmente distribuídos uniformemente sobre uma região retangular em duas dimensões. A cada um desses nós associa-se um vetor no espaço de entrada, que será adaptado de forma que a distribuição destes vetores tenda a se assimilar à distribuição efetiva dos dados observados<sup>1</sup>.

Uma vez que o mapa tenha sido treinado, teremos um conjunto de vetores  $\mathbf{c}_j$  no espaço de entrada, cada um associado a um nó no espaço de saída  $\alpha_j$ . A partir deste ponto, a cada vetor observado  $\mathbf{x}$  verifica-se qual o vetor  $\mathbf{c}_j$  mais próximo (vetor vencedor), e a saída será dada pelo nó  $\alpha_j$  correspondente. Este procedimento, no entanto, leva a uma quantização nos valores de saída, a qual pode ser atenuada através da realização de uma interpolação no mapa.

---

<sup>1</sup>De certa forma, este mesmo princípio é utilizado nos filtros fuzzy, descrito no capítulo 4.

Note que, se a distribuição dos vetores  $\mathbf{c}_j$  seguir a distribuição dos dados observados, os nós no espaço de saída serão escolhidos de forma equiprovável. Dessa forma, os sinais de saída apresentarão distribuição uniforme e componentes independentes entre si. Esta razão motivou o uso de mapas auto-organizáveis em problemas nos quais as fontes possuem distribuição uniforme, uma vez que o mapeamento provido pelo mapa tende a gerar em suas saídas sinais uniformes e independentes entre si. No entanto, a abordagem exibe restrições, sendo a mais severa a possibilidade de lidar somente com cenários nos quais as fontes possuem distribuições que se assemelham a uma distribuição uniforme.

Outra proposta para lidar com misturas não-lineares é apresentada por Valpola, em [132], tomando como base a teoria de treinamento bayesiano de redes neurais. A metodologia proposta é desenvolvida tendo em vista a identificação de modelos não-lineares que representem os dados observados da melhor maneira possível. Em seu trabalho, Valpola considerou que a mistura não-linear é modelada de maneira adequada por uma rede neural MLP (*Multi-Layer Perceptron*) e que as fontes podem ser modeladas através de uma mistura de densidades gaussianas. O ajuste da estrutura não-linear é efetuado de forma a obter não um conjunto de parâmetros para a rede neural, mas sim suas distribuições de probabilidade *a posteriori* após a observação dos dados. Embora o método seja capaz de modelar uma ampla gama de processos de mistura não-lineares, não se garante que o modelo obtido equivalha ao processo real.

De maneira semelhante, em [4], Almeida propõe o uso de redes neurais MLP para realizar a separação de misturas não-lineares, utilizando um algoritmo denominado MISEP, que pode ser visto como uma extensão do algoritmo Infomax. Apesar de realizar a separação dos sinais em alguns casos, novamente não há garantias de que isto sempre ocorrerá.

### 5.1.2 ICA não-linear

Tendo em mente o sucesso obtido no caso linear, uma tentativa para solucionar o problema de mistura não-linear seria considerar uma extensão direta da ICA. Sob a hipótese de que as fontes são independentes entre si, a estrutura de separação deveria ser ajustada a fim de recuperar em suas saídas sinais também independentes entre si. Esta abordagem recebeu o nome de ICA não-linear (*Nonlinear ICA*).

Considere o mapeamento  $\mathcal{H}(\mathbf{s}) = \mathbf{G}(\mathbf{F}(\mathbf{s}))$ , resultado da composição entre o mapea-



mento da mistura e do sistema separador tal que

$$\mathcal{H}(\mathbf{s}) = [h_1(s_{\rho(i)}), \dots, h_n(s_{\rho(i)})], \quad (5.3)$$

onde  $h_i$  denota uma função não-linear arbitrária e  $\rho$  representa um operador de permutação sobre o conjunto  $\{1, 2, \dots, n\}$ . Este tipo particular de mapeamento, conhecido como *mapeamento trivial* [75], transforma um vetor de componentes independentes em outro com a mesma propriedade, uma vez que cada uma das componentes de  $\mathcal{H}(\mathbf{s})$  representa uma função não-linear de apenas um componente de  $\mathbf{s}$ .

Portanto, obter o sistema separador  $\mathbf{G}$  tal que o mapeamento resultante entre as fontes e os sinais recuperados seja trivial configura-se como uma possível solução da ICA não-linear. Note que, neste caso, ainda há a ambigüidade de permutação das fontes, sendo ainda possível existir uma distorção não-linear dos sinais originais, devido às funções  $h_i$ .

Embora extremamente atraente, a aplicação da ICA no caso não-linear geral leva a resultados que não se restringem aos mapeamentos triviais, e desta forma não garante que os sinais recuperados representam os sinais das fontes. Um exemplo desta condição, extraído de [75], é mostrado a seguir.

**Exemplo 5.1.1.** *Considere duas fontes  $s_1$  e  $s_2$  independentes entre si, com as seguintes distribuições:  $p_{s_1}(s_1) = s_1 \exp(-s_1^2/2)$ , i.e., uma distribuição Rayleigh, e  $p_{s_2}(s_2) = 2/\pi$ , uma distribuição uniforme no intervalo  $s_2 \in [0, \pi/2)$ . A probabilidade conjunta das variáveis será expressa pelo produto das densidades marginais, ou seja:*

$$p_{s_1 s_2}(s_1, s_2) = \begin{cases} \frac{2}{\pi} s_1 \exp\left(\frac{-s_1^2}{2}\right), & s_2 \in [0, \pi/2); \\ 0, & s_2 \in (-\infty, 0) \text{ ou } s_2 \in [\pi/2, +\infty). \end{cases} \quad (5.4)$$

*Considere agora o seguinte mapeamento  $\mathcal{H}$  entre as fontes e suas estimativas*

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathcal{H}(\mathbf{s}) = \begin{bmatrix} s_1 \cos(s_2) \\ s_1 \sin(s_2) \end{bmatrix}. \quad (5.5)$$

*Note que os elementos de  $\mathbf{y}$  são misturas dos elementos de  $\mathbf{s}$ , uma vez que o jacobiano desta transformação não é diagonal.*

*Avaliemos agora a distribuição conjunta do vetor  $\mathbf{y}$ , que se relaciona com a distri-*

buição de  $\mathbf{s}$  conforme a seguinte equação [109]:

$$p_{y_1 y_2}(y_1, y_2) = \frac{p_{s_1 s_2}(s_1, s_2)}{|\det \mathbf{J}_{\mathcal{H}}(\mathbf{y})|}, \quad (5.6)$$

onde  $\mathbf{J}_{\mathcal{H}}(\mathbf{y})$  corresponde ao jacobiano do mapeamento  $\mathcal{H}(\cdot)^2$ .

Neste caso, temos que o jacobiano da transformação não-linear será expresso por

$$\begin{aligned} |\det \mathbf{J}_{\mathcal{H}}(\mathbf{y})| &= \left| \det \begin{pmatrix} \cos(s_2) & \sin(s_2) \\ s_1 \sin(s_2) & -s_1 \cos(s_2) \end{pmatrix} \right| = \\ &= |-s_1(\cos^2(s_2) + \sin^2(s_2))| = |s_1|. \end{aligned} \quad (5.7)$$

Substituindo (5.4) e (5.7) em (5.6), e observando-se que  $s_1^2 = y_1^2 + y_2^2$ , chegamos a

$$p_{y_1 y_2}(y_1, y_2) = \frac{2}{\pi} \exp\left(-\frac{(y_1^2 + y_2^2)}{2}\right) = \left(\sqrt{\frac{2}{\pi}} \exp\left(-\frac{y_1^2}{2}\right)\right) \left(\sqrt{\frac{2}{\pi}} \exp\left(-\frac{y_2^2}{2}\right)\right). \quad (5.8)$$

Note que a densidade conjunta de  $\mathbf{y}$  é dada por um produto de funções exclusivas de  $y_1$  e  $y_2$ , ou seja, os elementos de  $\mathbf{y}$  são estatisticamente independentes entre si, embora eles ainda sejam uma mistura de  $s_1$  e  $s_2$ .

O exemplo 5.1.1 ilustra que é possível obter mapeamentos não-triviais que ainda preservem a independência entre os componentes de um vetor. Este resultado havia sido observado inicialmente por Darmois em 1951, no contexto de análise por fatores não-lineares [126]. No contexto de separação de fontes, a questão sobre a unicidade de soluções para o caso não-linear foi tratada em [68], onde os autores descrevem um procedimento para construir famílias de soluções não-triviais.

Pode-se dizer que a dificuldade em se obter uma solução única através da ICA reside justamente na flexibilidade das funções não-lineares. Como neste caso não temos um sinal piloto para guiar a adaptação do sistema separador, vimos que é possível recuperar a independência entre os sinais e mesmo assim não garantir que estes correspondam às fontes. Deste modo, as abordagens encontradas na literatura no que se refere a modelos não-lineares de mistura buscam, de uma forma ou de outra, diminuir esse grau de flexibilidade dos mapeamentos<sup>3</sup> [74].

---

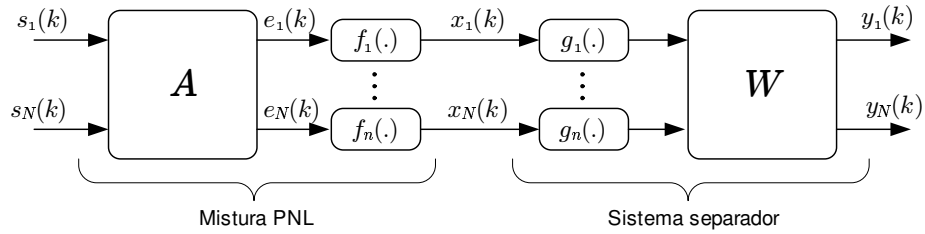
<sup>2</sup>Note que, no intervalo delimitado por  $\mathbf{s}$ , este mapeamento é inversível, o que permite obter o jacobiano em função  $\mathbf{y}$ .

<sup>3</sup>É interessante notar que no caso de misturas convolutivas realizamos justamente o contrário, aumentando o grau de flexibilidade da estrutura de separação. No entanto, a diferença era que

Uma possível abordagem seria restringir o problema de forma estrutural, concentrando-se em modelos de misturas não-lineares menos abrangentes, mas nos quais as ferramentas desenvolvidas para ICA ainda fossem efetivas. Em um primeiro momento foram considerados problemas nos quais a mistura era modelada por mapeamentos suaves, como os gerados por meio de uma rede neural MLP (*Multi-Layer Perceptron*) [137, 4]. No entanto, mesmo nestas condições ainda é possível obter soluções indesejadas [93], indicando que outras condições devam ser consideradas. Nesta linha, destaca-se o modelo introduzido por Taleb e Jutten, denominado *mistura com não-linearidade posterior*, ou PNL (*Post-Nonlinear*).

## 5.2 Mistura *Post-Nonlinear*

O modelo de mistura PNL, introduzido por Taleb e Jutten [125], consiste de uma seção de mistura linear instantânea seguida de uma seção de não-linearidades aplicadas a cada um dos sinais linearmente misturados. O sistema PNL é ilustrado na figura 5.1.



**Figura 5.1:** Sistema com não-linearidade posterior.

Este tipo de modelo é particularmente útil para modelar cenários nos quais os sensores apresentam características não-lineares, como por exemplo em aplicações de comunicações, onde estágios amplificadores na recepção podem distorcer os sinais recebidos. Matematicamente, temos que os sinais observados  $\mathbf{x}$  podem ser expressos por:

$$\mathbf{x} = \mathbf{f}(\mathbf{A}\mathbf{s}), \quad (5.9)$$

onde  $\mathbf{f}(\mathbf{e}) = [f_1(e_1), f_2(e_2), \dots, f_N(e_N)]$ .

---

naquele caso havia um sinal de referência para o treinamento da estrutura e, portanto, a dificuldade encontrada era restrita ao método de otimização.

Dada a forma da mistura, a escolha natural para o sistema separador consiste em uma estrutura similar à da mistura, empregando um conjunto de não-linearidades  $g_i(\cdot)$  e uma matriz de separação  $\mathbf{W}$ , conforme mostrado na figura 5.1. A saída, portanto, pode ser expressa por:

$$\mathbf{y} = \mathbf{W}\mathbf{g}(\mathbf{x}), \quad (5.10)$$

onde  $\mathbf{g}(\mathbf{x}) = [g_1(x_1), g_2(x_2), \dots, g_N(x_N)]$ .

### 5.2.1 Separabilidade de misturas PNL

Conforme havíamos comentado, o grande interesse em se considerar modelos mais restritos seria obter misturas não-lineares às quais a idéia subjacente à ICA ainda fosse efetiva para a recuperação das fontes. Isto, de fato, ocorre para o modelo de mistura PNL, e as condições para a separabilidade deste modelo são enunciadas através do seguinte teorema [126, 2]:

**Teorema 5.2.1** (Separabilidade do Modelo PNL). *Considere o modelo PNL apresentado na figura 5.1, e o seguinte conjunto de hipóteses:*

- *A matriz  $\mathbf{A}$  é inversível e, de fato, mistura as fontes, ou seja, há ao menos dois elementos não-nulos em cada coluna e linha desta matriz;*
- *As funções  $\mathbf{f}(\cdot)$  e  $\mathbf{g}(\cdot)$  são monotônicas; conseqüentemente, a composição  $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$  também será monotônica;*
- *Há, no máximo, uma fonte gaussiana;*
- *A função densidade de probabilidade conjunta das fontes é diferenciável e sua derivada é contínua em todo o seu suporte.*

*Nessas condições, se os elementos de  $\mathbf{y}$  forem independentes entre si, então os elementos do vetor  $\mathbf{h}(\cdot)$  corresponderão a funções lineares e  $\mathbf{WA} = \mathbf{\Lambda P}$ .*

### 5.2.2 Minimização da informação mútua

A exemplo do que ocorre no caso linear, o critério de minimização da informação mútua também ocupa lugar de destaque no contexto de misturas *post-nonlinear*. Relembrando a equação (3.6), a informação mútua entre os elementos do vetor  $\mathbf{y}$  é dada por

$$I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y}), \quad (5.11)$$

onde  $H(y_i)$  representa a entropia de cada um dos elementos de  $\mathbf{y}$ .

Considerando a estrutura de separação da figura 5.1, podemos expressar a entropia conjunta  $H(\mathbf{y})$  em termos da entropia de  $\mathbf{x}$ :

$$H(\mathbf{y}) = H(\mathbf{x}) + E \{ \log |J| \}, \quad (5.12)$$

onde  $J$  representa o *Jacobiano* da transformação não-linear entre o vetor de observações e as saídas do sistema separador. Aplicando este resultado à equação (5.11), é possível obter [126]

$$I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{x}) - \log |\det \mathbf{W}| - E \left\{ \log \prod_i |g'_i(x_i)| \right\} \quad (5.13)$$

onde  $g'_i(\cdot)$  denota a derivada de  $g_i(\cdot)$ . É importante ressaltar que esta relação é válida somente se as não-linearidades  $g_i(\cdot)$  forem inversíveis.

Em [126] é proposto um algoritmo baseado na otimização de (5.13) a partir do método do gradiente. Avaliando-se o gradiente de (5.13) em relação ao parâmetros de  $\mathbf{W}$  obtém-se

$$\frac{\partial I(\mathbf{y})}{\partial \mathbf{W}} = -E \{ \Psi(\mathbf{y}) \mathbf{e}^T \} - (\mathbf{W}^T)^{-1}, \quad (5.14)$$

onde  $\Psi(\mathbf{y}) = [\psi_{y_1}(y_1), \dots, \psi_{y_N}(y_N)]$ , para  $\psi_{y_i}(y_i) = (p_{y_i}(y_i)' / p_{y_i}(y_i))$ , conhecidas também como *score functions*.

Considerando funções não-lineares parametrizadas, denotadas por  $g_i(\theta_i, x_i)$ ,  $i = 1, \dots, N$ , onde  $\theta_i$  correspondem ao conjunto de parâmetros da função, o gradiente da informação mútua em relação a  $\theta_i$  será dado por:

$$\frac{\partial I(\mathbf{y})}{\partial \theta_i} = -E \left\{ \frac{\partial \log |g'_i(\theta_i, x_i)|}{\partial \theta_i} \right\} - E \left\{ \left( \sum_{k=1}^N \psi_{y_k}(y_k) w_{ki} \right) \frac{\partial g_i(\theta_i, x_i)}{\partial \theta_i} \right\}, \quad (5.15)$$

onde  $w_{ki}$  corresponde ao elemento  $(k, i)$  da matriz  $\mathbf{W}$ .

Nota-se com isso que a adaptação dos parâmetros está diretamente ligada à estimativas das *score functions*, que por sua vez se relacionam com estimativas da pdf dos sinais  $y_i$ . Conforme comentado no capítulo 3, uma solução usual no caso de misturas lineares consiste em aproximar as distribuições através de expansões em séries que levam em consideração estatísticas de ordem superior dos sinais observados. Porém, se por um lado as estimativas geradas por meio de expansões em série reduzem consideravelmente o custo computacional, por outro lado os erros obtidos nesta estimação podem comprometer severamente a qualidade das soluções no caso não-linear [126, 3].

### 5.2.3 Abordagem de gaussianização

A forma peculiar da mistura com não-linearidade posterior, com uma divisão clara entre a sessão linear e não-linear do processo de mistura, sugere que a adaptação possa ser efetuada de maneira separada nos dois blocos. No entanto, isto não é possível considerando-se apenas a independência entre as saídas do sistema.

A técnica de gaussianização [123][142] baseia-se na hipótese de que os sinais à saída da seção linear apresentam uma distribuição gaussiana. Esta hipótese apóia-se no teorema central do limite, já que os sinais em questão são combinações lineares de sinais independentes entre si e, portanto, devem tender a uma distribuição gaussiana.

Como a seção não-linear da mistura é aplicada a estes sinais, as observações certamente apresentarão distribuição distinta da normal. Este fato abre caminho para uma divisão no processo de adaptação do sistema separador, tornando o ajuste das não-linearidades  $g_i(\cdot)$  independente do ajuste da matriz de separação.

Em um primeiro momento, as funções  $g_i(\cdot)$  são ajustadas de forma a recuperar sinais com distribuição gaussiana, o que pode ser feito da seguinte forma

$$g_i(\cdot) = F_{gauss}^{-1}(F_{x_i}(\cdot)) \quad (5.16)$$

onde  $F_{x_i}(x_i)$  denota a distribuição cumulativa de  $x_i$  e  $F_{gauss}^{-1}(\cdot)$  a inversa da distribuição cumulativa de uma variável aleatória gaussiana normalizada, i.e., a função quantil.

A partir dos dados transformados, espera-se que as fontes serão recuperadas utilizando algum método de ICA desenvolvido para misturas lineares, como os descritos na seção 3.

Os resultados da separação utilizando esta abordagem dependem fortemente da validade da hipótese de gaussianidade dos sinais, o que ocorre quase sempre para um número elevado de fontes na mistura. No entanto, nas condições nas quais isto não é verificado, a distorção residual nos dados transformados pode ser considerável, inviabilizando a separação com técnicas usuais de ICA.

### 5.2.4 Outras abordagens

Seguindo na mesma linha de adaptar separadamente as seções lineares e não-linear do sistema de separação, em [93] é proposta uma abordagem geométrica para o pro-

blema. Conforme visto no exemplo 2.2.1, com a presença de duas fontes com distribuição uniforme, a distribuição conjunta dos dados observados na saída da mistura linear toma a forma de um paralelogramo. No entanto, ao passar os dados através de uma não-linearidade, a forma geométrica da distribuição conjunta será alterada. Dessa forma, o método geométrico ajusta o conjunto de não-linearidades de maneira a recuperar uma forma geométrica semelhante ao paralelogramo. O método, no entanto, é bastante restrito, pois lida somente com cenários com duas fontes, e além disso elas devem necessariamente apresentar distribuições com suporte finito.

Outra extensão obtida para o caso específico de misturas post-nonlinear é proposto em [71], onde os autores utilizam as mesmas idéias delineadas por Valpola, mas restringindo-se ao modelo de misturas com não-linearidade posterior.

Em [1] um critério alternativo à informação mútua entre sinais é proposto, baseado na minimização de uma função denominada *medida de dependência quadrática*. A função, a exemplo da informação mútua, se anula se e somente se as variáveis envolvidas forem independentes entre si. O uso do critério no problema de separação de fontes PNL alivia de certa forma o problema de estimação da informação mútua, já que não envolve diretamente a estimação das *score functions*. Por outro lado, o critério depende da escolha de *kernels* específicos, e observa-se em alguns casos que a superfície da função custo apresenta mínimos locais.

Seguindo uma outra linha, em [118] e [117], técnicas baseadas em heurísticas, como algoritmos genéticos e *simulated annealing* são propostas. A motivação para o trabalho é evitar problemas de convergência para mínimos locais, observados em algoritmos utilizando a informação do gradiente [93, 1], por meio de algoritmos com reconhecida capacidade de busca multimodal. A função custo adotada no trabalho utilizou séries de Gram-Charlier para aproximação do valor da informação mútua. No entanto, conforme comentado anteriormente, o uso de tais aproximações podem levar a estimativas grosseiras da função custo, o que levaria inevitavelmente à convergência para soluções sub-ótimas.

### 5.3 Proposta para Separação de misturas PNL

Um dos problemas enfrentados por técnicas de otimização baseadas no gradiente é a possibilidade de convergência para soluções sub-ótimas. Este problema, embora presente também no caso de misturas lineares, torna-se mais pronunciado no caso de misturas não-lineares.

A convergência para soluções sub-ótimas, no entanto, não decorre apenas do uso de algoritmos baseados na informação do gradiente, mas também, conforme comentamos, de estimativas pouco precisas da função custo ou suas derivadas, como as *score functions*.

Uma solução para o primeiro problema consiste em empregar algoritmos de busca mais eficientes, como os algoritmos evolutivos, que possuem uma grande capacidade de busca multimodal. Em adição aos métodos empregados em [118] e [117], podemos destacar os algoritmos imunológicos [41], cuja robustez e capacidade de busca global têm se mostrado superior às dos algoritmos genético comuns [11].

No entanto, a ferramenta de busca só será capaz de realizar o ajuste correto do sistema separador se a estimativa da função custo for feita de maneira precisa. Neste sentido, métodos mais adequados para a estimação da informação mútua são necessários.

A nossa proposta busca, portanto, contornar estes dois problemas relacionados ao modelo de misturas PNL. Em resumo, a proposta é definida por:

- Uma ferramenta de otimização capaz de adaptar os parâmetros do sistema separador, evitando a convergência para soluções sub-ótimas. Nossa opção foi utilizar um algoritmo evolutivo denominado *opt-aiNet* (*Optimization version of an Artificial Immune Network*) [40], um algoritmo imunológico cuja aplicação em problemas típicos de processamento de sinais vem sendo bastante eficaz [12, 11];
- Um método de estimação para a função custo, relacionado à informação mútua entre os sinais, que basicamente depende da estimação das entropias marginais dos sinais na saída do sistema separador. Para este fim, empregamos um método baseado em estatísticas de ordem [110, 51], que oferece um bom compromisso entre acuidade e complexidade computacional;

### 5.3.1 Estimação de entropia baseada em estatísticas de ordem

Se observarmos a equação (5.13), que descreve a informação mútua entre as saídas do sistema separador, veremos que o valor da entropia conjunta  $H(\mathbf{x})$  dos sinais observados não depende dos parâmetros a serem ajustados, e portanto pode ser desconsiderado no processo de adaptação. Dessa forma, a estimação da função custo depende de estimativas acuradas dos valores de  $H(y_i)$ , que podem ser obtidas através de métodos baseados em estatísticas de ordem (*Order Statistics*).



Considere um conjunto de variáveis aleatórias  $Y_1, Y_2, \dots, Y_T$ . A cada realização  $y_1, y_2, \dots, y_T$  deste conjunto, o valor  $m = \max y_i$  representará também a realização de uma variável aleatória, que representaremos por  $Y_{(T:T)}$ . De maneira equivalente, poderíamos considerar o valor mínimo em cada realização, o que por sua vez representa a realização de uma outra variável aleatória. Generalizando este conceito, chegamos a um conjunto de variáveis aleatórias ordenadas

$$Y_{(1:T)} \leq Y_{(2:T)} \leq \dots \leq Y_{(T:T)}, \quad (5.17)$$

onde  $Y_{(t:T)}$  corresponde à  $t$ -ésima *estatística de ordem*, e representa a variável aleatória associada ao  $t$ -ésimo menor valor das realizações de  $Y_1, Y_2, \dots, Y_T$ .

Note que o conceito de estatística de ordem também pode ser definido para uma única variável aleatória. De fato, cada realização de um conjunto de  $T$  variáveis aleatórias com distribuições de probabilidade idênticas pode ser vista como  $T$  realizações de uma única variável aleatória.

O interesse em se empregar o conceito de estatística de ordem na estimação da entropia reside em sua ligação com a função quantil, que corresponde ao inverso da função cumulativa de probabilidade. A relação entre as estatísticas de ordem e a função quantil de uma variável aleatória  $Y$ , representada por  $Q_Y(\cdot)$ , é dada por [110]:

$$E\{Y_{(t:T)}\} = Q_Y\left(\frac{t}{T+1}\right), \quad t = 1, \dots, T. \quad (5.18)$$

Por sua vez, a entropia de uma variável aleatória pode ser definida em termos da função quantil, e expressa por

$$H(Y) = \int_0^1 \log Q'_Y(u) du. \quad (5.19)$$

onde  $Q'_Y$  denota a derivada da função quantil. Empregando uma aproximação numérica para implementar a integral em (5.19) obtemos

$$H(Y) = \sum_{l=2}^L \log \left[ \frac{Q_Y(u_l) - Q_Y(u_{l-1})}{u_l - u_{l-1}} \right] \frac{u_l - u_{l-1}}{u_L - u_1}, \quad (5.20)$$

onde  $\{u_1, \dots, u_L\}$  é um conjunto de números crescentes no intervalo  $[0, 1]^4$ . Portanto, empregando as estatísticas de ordem podemos obter informação sobre a função quan-

---

<sup>4</sup>Consideramos também que  $u_i - u_{i-1}$  é constante para todo  $i$ .

til, que por sua vez pode ser utilizada para calcular a entropia da variável aleatória.

No caso em que há apenas uma realização de  $T$  amostras da variável aleatória  $Y$ , a estimação da função quantil pode ser feita, de maneira aproximada, com base na realização correspondente das estatísticas de ordem, ou seja

$$Y_{(t:T)} \approx Q_Y\left(\frac{t}{T+1}\right), \quad t = 1, \dots, T. \quad (5.21)$$

Para valores da função quantil que não pertencem ao conjunto  $\frac{t}{T+1}$  é possível realizar uma interpolação linear, gerando um estimador de entropia simples e eficaz.

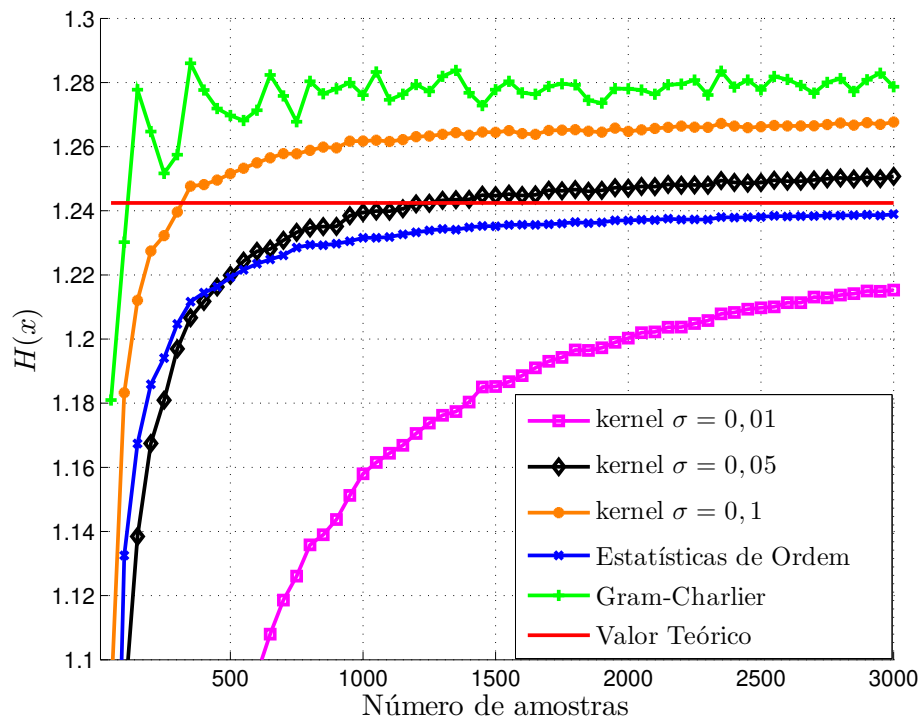
Na figura 5.2 mostramos o comportamento do estimador baseado em estatística de ordem para uma variável aleatória uniforme no intervalo  $[-\sqrt{3}, \sqrt{3}]$  [49]. Na figura também são ilustrados os comportamentos de estimadores baseados na estimação por *kernels* gaussianos com diferentes valores de desvio padrão [30], e a estimação utilizando a expansão de Gram-Charlier para a pdf da variável aleatória [36].

Nota-se que o desempenho do método baseado de estatísticas de ordem se assemelha ao do método empregando *kernels*, tanto em acuidade como em sensibilidade quanto ao número de amostras consideradas. O método baseado na expansão de Gram-Charlier, por outro lado, necessita de um número menor de dados para realizar a estimativa de maneira razoável, porém o erro obtido com este estimador é visivelmente maior.

### 5.3.2 Algoritmo de otimização

Tendo definido o critério a ser utilizado e a forma como a função custo será estimada, deve-se definir a ferramenta de otimização a ser empregada. A escolha do método de otimização é de extrema importância quando lidamos com o problema de misturas não-lineares. Os métodos tradicionais empregados no contexto de separação são baseados no gradiente da função custo e apresentam bons resultados. Porém, no caso de misturas não-lineares, a função custo a ser minimizada pode apresentar mínimos locais, e assim levar os algoritmos tradicionais a resultados não satisfatórios. Os algoritmos evolutivos, por outro lado, são particularmente apropriados para lidar com este tipo de problema.

Em oposição aos algoritmos derivados do gradiente, um algoritmo evolutivo é um algoritmo populacional, de forma que cada indivíduo da população representa uma possível solução para o problema de otimização. A idéia, portanto, é evoluir esta



**Figura 5.2:** Estimação da entropia de uma variável aleatória uniforme.

população de indivíduos de maneira que as soluções sejam gradativamente melhoradas e, ao final do processo, tenhamos a solução global para o problema.

Em [11], um amplo estudo sobre a aplicabilidade destas ferramentas a problemas típicos de processamento de sinal foi conduzido. Os resultados apresentados indicam que o emprego de algoritmos evolutivos em cenários nos quais o problema é inerentemente multimodal pode resultar em ganhos significativos. Motivados por tais observações, adotamos em nossa proposta o algoritmo *opt-aiNet* [40, 11], que pode ser definido como uma ferramenta de busca que agrega idéias relacionadas ao funcionamento de nosso sistema imunológico.

Nosso sistema imunológico apresenta uma capacidade inata de responder a um grande número de microorganismos, e também a capacidade de se adaptar para lidar com novos *antígenos*. A capacidade de adaptação a novos tipos de antígenos está relacionada às noções de *seleção clonal* e *maturação de afinidade*.

A essência da seleção clonal é a seguinte: quando há o reconhecimento positivo de um antígeno, a célula bem sucedida passa por um processo de clonagem, gerando

um conjunto de cópias capazes de gerar anticorpos apropriados para o antígeno. O reconhecimento positivo está ligado a uma certa *afinidade* entre os receptores presentes na célula e as características do antígeno.

A característica de adaptação do sistema se revela durante o processo de clonagem, onde são observadas mutações nos receptores das células geradas inversamente proporcionais à afinidade entre a célula e o antígeno. Portanto, podemos entender este mecanismo como um processo de *maturação de afinidade* entre os receptores e o antígeno. Em conjunto com esse mecanismo principal, atuam ainda mecanismos capazes de selecionar apenas os melhores produtos dessas mutações e um engenhoso esquema capaz de introduzir material significativamente novo na população de células de defesa.

Dessa forma, podemos ver que a adaptação ocorrida no sistema imunológico é essencialmente um processo de otimização das células, com o objetivo de melhorar a afinidade entre os seus receptores e as características do antígeno. Traduzindo estas idéias para o contexto de otimização de uma função custo qualquer, podemos realizar os seguintes paralelos:

- cada possível solução do problema corresponde a um indivíduo da população de células. No caso de separação de misturas PNL, cada solução é composta de um conjunto de parâmetros que definem as não-linearidades e a matriz de separação;
- a função custo a ser otimizada representa a medida de afinidade (ou *fitness*) entre as células e os antígenos.

Assim, organizando as idéias relacionadas aos mecanismos de adaptação de nosso sistema imunológico, propôs-se o algoritmo opt-aiNet, descrito resumidamente no quadro 7.

A etapa de controle de diversidade é de extrema importância para o bom desempenho do algoritmo. Através deste mecanismo, eliminam-se possíveis elementos semelhantes entre si, o que poderia resultar em convergência sub-ótima, e são introduzidos novos elementos na população. Note que este mecanismo também atua como um controle do tamanho da população, ajustando-a de acordo com a dificuldade do problema.

De maneira mais ampla podemos enxergar o algoritmo opt-aiNet atuando em dois níveis diferentes. Em um primeiro nível, definido pelo passo 2, ocorre uma busca local através dos ajustes baseados no valor do *fitness* de cada indivíduo. Em um segundo nível, há o processo de busca global, que decorre dos mecanismos de eliminação de indivíduos próximos entre si e inserção de novas células na população.

1. **Inicialização:** Criar aleatoriamente os indivíduos da população;
2. **Busca local:** Enquanto não se atingir o critério de parada:
  - (a) **Expansão clonal:** Determine o fitness de cada um dos indivíduos. Gere um conjunto de  $N_c$  clones (cópias exatas) de cada um dos indivíduos.
  - (b) **Maturação de afinidade:** Aplique a mutação a cada clone com uma taxa inversamente proporcional ao fitness do indivíduo que o originou, de acordo com

$$c' = c + \alpha N(0, 1), \text{ com } \alpha = \beta^{-1} \exp(-f^*) \quad (5.22)$$

onde  $c'$  e  $c$  representam os indivíduos após a mutação e o original, respectivamente.  $\beta$  é um parâmetro que controla a mutação,  $N(0, 1)$  uma variável aleatória gaussiana padronizada de mesma dimensão do vetor de parâmetros, e  $f^*$  é o fitness do indivíduo. Para cada conjunto de clones, selecione aquele com maior fitness e calcule o fitness médio da população restante.

- (c) **Convergência local:** se a média do fitness da população não variar significativamente entre as iterações, prossiga; caso contrário retorne ao passo 2;
3. **Controle de diversidade:** determine a afinidade (proximidade, medida através da distância euclidiana) entre os indivíduos da população; Elimine todos indivíduos, exceto um deles, cujas afinidades estejam abaixo de um valor  $\sigma_s$  pré-definido.
4. Introduza um certo número de novos indivíduos e retorne ao passo 2.

**Quadro 7:** Resumo do algoritmo opt-aiNet.

É necessário, no entanto, ressaltar um pequeno detalhe na implementação do algoritmo de otimização. Em sua forma original, o algoritmo opt-aiNet assume que a função de fitness, ou a função custo do problema, define sempre um problema de maximização. Como para nossos propósitos consideramos a minimização da informação mútua (ou uma quantidade diretamente relacionada à informação mútua), a função de fitness em nosso problema assume a forma:

$$J_{fitness} = \frac{1}{\sum_i H(y_i) - \log |\det \mathbf{W}| - E \left\{ \log \prod_i |g'_i(x_i)| \right\}}. \quad (5.23)$$

### 5.3.3 Primeiros Resultados

O método proposto foi colocado a prova diante de dois cenários. No primeiro, duas fontes com distribuição uniforme no intervalo  $[-1, +1]$  foram misturadas através de um sistema definido por:

$$\mathbf{A} = \begin{bmatrix} 1 & 0,6 \\ 0,5 & 1 \end{bmatrix} \text{ e } \begin{cases} f_1(e_1) = \tanh(2e_1) \\ f_2(e_2) = 2\sqrt[5]{e_2} \end{cases}. \quad (5.24)$$

O segundo cenário é composto de três fontes com distribuição uniforme, misturadas através do seguinte sistema:

$$\mathbf{A} = \begin{bmatrix} 1 & 0,6 & 0,5 \\ 0,5 & 1 & 0,4 \\ 0,4 & 0,6 & 1 \end{bmatrix} \text{ e } \begin{cases} f_1(e_1) = 2\sqrt[3]{e_1} \\ f_2(e_2) = 2\sqrt[3]{e_2} \\ f_3(e_3) = 2\sqrt[3]{e_3} \end{cases}. \quad (5.25)$$

Em ambos os casos, o sistema separador consistiu de uma matriz  $\mathbf{W}$  e funções não-lineares do tipo  $g_i(x) = a_i x_i^5 + b_i x_i^3 + c_i x_i$ . De maneira a respeitar as condições de separabilidade do modelo PNL [2], a composição  $g_i(f_i(\cdot))$  deve ser uma função monotônica. Assim, os coeficientes dos polinômios  $g_i(\cdot)$  foram restringidos ao conjunto  $\Re^+$ .

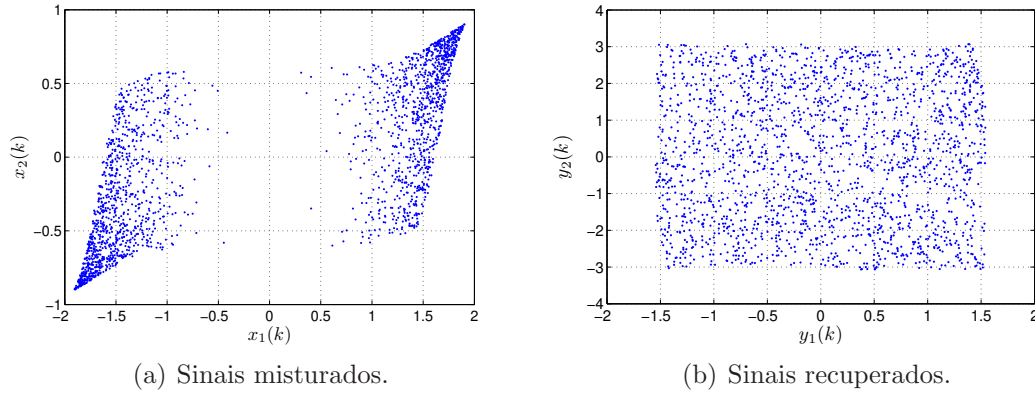
Desta maneira, cada indivíduo da população é representado por um vetor de números reais, compreendido de:

- Uma seção que contém os elementos da matriz de separação  $\mathbf{W}$ ;
- Um certo número de seções contendo os coeficientes de cada um dos polinômios  $g_i(\cdot)$ .

Os parâmetros do algoritmo evolutivo utilizados foram:  $N_c = 7$ ,  $\beta = 60$  e  $\sigma_s = 2$ . Foram utilizadas cerca de 2000 amostras dos sinais observados e considerou-se um número máximo de 10000 gerações para o algoritmo evolutivo.

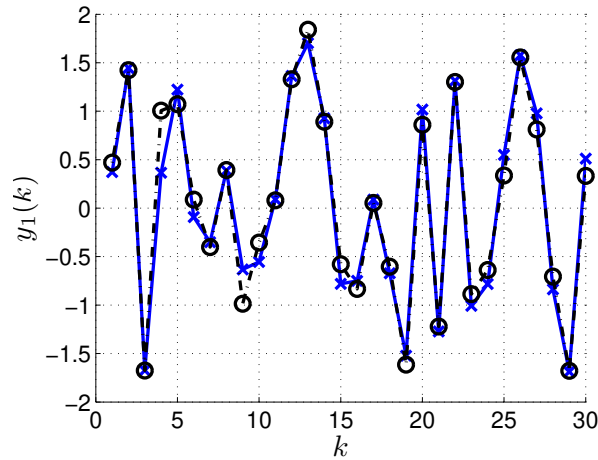
A figura 5.3 apresenta a distribuição conjunta dos dados observados (misturados) e após a separação para o primeiro cenário. Nota-se que a abordagem proposta foi capaz de recuperar a distribuição das fontes e verificou-se que o erro médio quadrático entre os sinais estimados e as fontes (após normalização da variância) era da ordem de  $1 \times 10^{-3}$ , mostrando a eficiência da técnica.

Para o segundo cenário, os resultados foram similares. Na impossibilidade de visuali-



**Figura 5.3:** Primeiro cenário.

zar a distribuição conjunta das fontes, é apresentado na figura 5.4 o sinal recuperado de uma fonte, sobreposta ao sinal original. Visualmente é possível verificar que as não-linearidades foram aproximadamente invertidas, e, na tabela 5.1, mostramos os erros médios para as estimativas das fontes.



**Figura 5.4:** Segundo cenário - sinal recuperado correspondente a uma das fontes: (—) sinal original; (· — ·) sinal estimado

## 5.4 Uma solução modificada

Os resultados obtidos com a abordagem proposta foram bastante positivos, tendo sido recuperados os sinais originais nos cenários simulados, com 2 e 3 fontes. Porém,

**Tabela 5.1:** Segundo cenário - Erro quadrático médio.

EQM ( $\times 10^{-3}$ )	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$
opt-aiNet	33,5	11	15

uma crítica ao método proposto seria o grande esforço computacional demandado para problemas de dimensões maiores do que as consideradas. Como a solução do problema é obtida por meio da evolução de uma população de indivíduos, um aumento na dimensão do problema de separação leva a um aumento considerável da carga computacional exigida.

Nas simulações apresentadas, vemos que o número de parâmetros relacionados à seção não-linear da estrutura de separação cresce linearmente com o número de fontes presentes na mistura. Porém, o número de parâmetros da matriz de separação  $W$  cresce de maneira quadrática, o que representa um aumento considerável do espaço de busca dos parâmetros e, conseqüentemente, do custo computacional da abordagem.

Uma solução para isso seria obter uma função custo para o ajuste das não-linearidades que não dependesse dos parâmetros da matriz  $\mathbf{W}$ <sup>5</sup>. Dessa forma, uma vez feito o ajuste das não-linearidades, a separação das fontes poderia ser completada por meio de métodos eficientes para o caso linear, como o algoritmo FastICA.

Motivados por esta idéia, buscamos uma maneira de realizar a adaptação das seções linear e não-linear separadamente, mas mantendo uma função custo “global” que levasse em conta a informação mútua entre as saídas do sistema separador.

A idéia, em essência, é simples. Ao invés de adaptarmos todos os parâmetros da estrutura de separação, cada indivíduo representará apenas um conjunto de não-linearidades. A seção linear do sistema separador será calculada a partir dos sinais transformados por estas não-linearidades utilizando-se o algoritmo FastICA, escolhido por sua grande velocidade de convergência e baixo custo computacional. Desse modo estamos de certa forma obtendo qual seria a melhor matriz de separação para aquele conjunto específico de não-linearidades e, com base nos sinais recuperados, é possível calcular o *fitness* associado ao indivíduo.

Em suma, o processo de avaliação do fitness passa a ser:

1. Para um dado indivíduo (conjunto de parâmetros das funções não-lineares), calcule as saídas da seção não-linear  $\mathbf{e}$ ;

---

<sup>5</sup>Note que esta mesma idéia está presente nas abordagens de gaussianização e geométrica, comentadas anteriormente.



2. Aplique o algoritmo FastICA nos sinais  $\mathbf{e}$ , obtendo, assim, uma matriz  $\mathbf{W}$  que representa a “melhor” matriz de separação para o dado conjunto de não-linearidades;
3. Avalie o *fitness* do indivíduo utilizando a expressão (5.23), calculada com as saídas fornecidas pelo FastICA.

Esta simples modificação tem como consequência uma drástica diminuição do espaço de busca do algoritmo evolutivo. Evidentemente, se por um lado há uma redução no espaço de busca, por outro há um aumento na complexidade computacional por iteração, haja vista que o algoritmo FastICA agora é executado a cada avaliação do *fitness* dos indivíduos.

### 5.4.1 Avaliação da solução modificada

De modo a avaliar os ganhos obtidos com a modificação proposta, realizamos simulações para um dos cenários já utilizados anteriormente. Consideramos novamente 3 fontes uniformemente distribuídas no intervalo  $[-1, 1]$ , misturadas pelo sistema definido em (5.25). Comparamos o desempenho da técnica proposta anteriormente, a qual nos referenciaremos como opt-aiNet, com a nova estratégia, denominada opt-aiNet/FastICA, utilizando o conjunto de parâmetros mostrado na tabela 5.2.

**Tabela 5.2:** Parâmetros dos algoritmos opt-aiNet e opt-aiNet/FastICA

	$N_c$	$\beta$	$\sigma_s$
opt-aiNet	7	60	3
opt-aiNet/FastICA	5	50	3

Realizamos um conjunto de 20 simulações para cada método, definindo um critério de parada baseado no número máximo de iterações dos algoritmos: 15000 para o opt-aiNet, e 6000 para o opt-aiNet/FastICA. Mesmo com uma margem maior para o opt-aiNet, observamos que este método não convergiu para soluções satisfatórias em 5 das 20 simulações executadas. Isto indica que o critério de parada do algoritmo o limitou a tal ponto de não permitir que a solução adequada para o problema fosse atingida. Entretanto, como nosso intuito era obter uma comparação no tempo de convergência dos algoritmos, estas simulações mal sucedidas não foram levadas em consideração.

A tabela 5.3 apresenta alguns tempos relativos à convergência dos dois algoritmos<sup>6</sup>.

---

<sup>6</sup>As simulações foram realizadas em um computador Athlon64 3000, com 1GB de memória.

**Tabela 5.3:** Velocidade de convergência.

	opt-aiNet	opt-aiNet/FastICA
Velocidade de convergência (iteração)	10500	1400
Tempo médio por iteração (ms)	80	182
Tempo médio de convergência (min)	14	4,5

**Tabela 5.4:** Erro quadrático médio.

EQM ( $\times 10^{-3}$ )	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$
opt-aiNet	7,33	9,50	6,45
opt-aiNet/FastICA	2,19	1,58	1,73

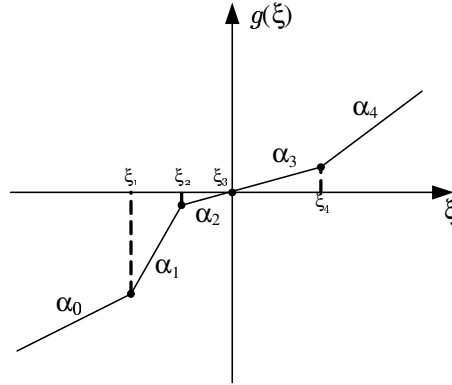
Conforme esperado, a introdução do FastICA no laço de adaptação realmente implica um aumento no tempo médio de cada iteração. No entanto, a redução do espaço de busca diminuiu consideravelmente o número de iterações necessário para convergência do algoritmo, resultando, de um modo geral, em um tempo de convergência menor que o da opt-aiNet.

A tabela 5.4 mostra os valores dos erros quadráticos médios entre as fontes e as estimativas para ambas as técnicas. Nota-se que, devido à presença de um mecanismo de busca extremamente adaptado ao problema, a técnica opt-aiNet/FastICA provê melhores resultados. Evidentemente, seria possível que a opt-aiNet obtivesse resultados mais expressivos, às custas, porém, de um aumento no número de iterações do algoritmo.

### 5.4.2 Estrutura linear por partes

Os resultados obtidos com o novo método mostraram-se promissores, indicando que a inclusão de uma ferramenta de busca altamente especializada no processo de adaptação do sistema separador pode trazer grandes benefícios. Porém, conforme pode-se notar, a escolha das funções não-lineares presentes na mistura e da forma das funções  $g_i(\cdot)$  foi feita de maneira que a inversão perfeita do sistema fosse possível, o que pode não ocorrer em situações práticas. Com o intuito de tornar a abordagem mais flexível, substituímos as funções polinomiais  $g_i(\cdot)$  por estruturas lineares por partes, conforme a ilustrada na figura 5.5.

A estrutura linear por partes, considerando-se  $N_{sec}$  seções, é definida pelo seguinte conjunto de parâmetros:



**Figura 5.5:** Estrutura linear por partes utilizada.

- $(N_{sec})$  coeficientes angulares para cada uma das seções. A monotonicidade do mapeamento é garantida restringindo-se os valores dos coeficientes angulares  $\alpha_i$  a assumirem a mesma polaridade.
- $(N_{sec} - 1)$  valores que definem a largura de cada uma das seções  $\xi_i$ . Os intervalos constituem então conjuntos semi-abertos  $(-\infty, \xi_0], (\xi_0, \xi_1], \dots, (\xi_{N_{sec}}, \infty)$
- Um valor  $\beta_0$  que define o coeficiente linear da primeira seção. Note que os demais coeficientes lineares podem ser obtidos a partir dos coeficientes angulares das seções.

Note, portanto, que há um aumento no número de parâmetros a serem ajustados em relação aos ensaios anteriores. Porém, ainda assim, este número cresce linearmente com a quantidade de fontes.

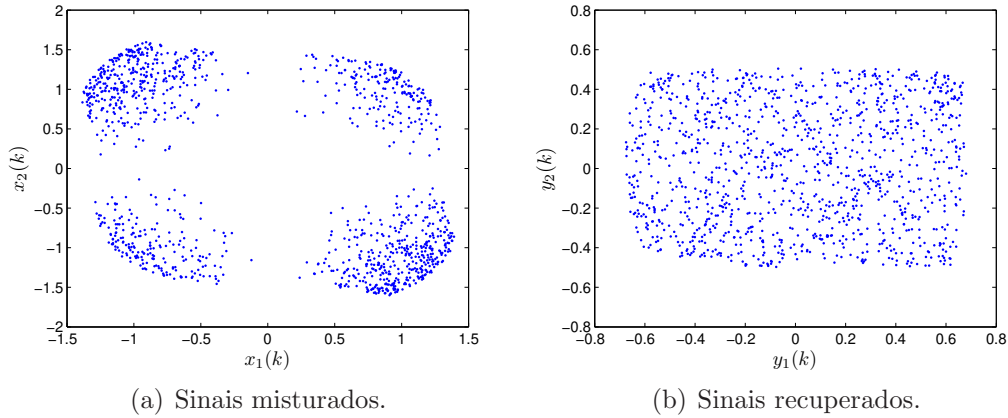
O cálculo da função custo, em especial o termo que depende das derivadas da função não-linear  $g_i(\cdot)$ , deve ser ligeiramente alterado. Como o cálculo da derivada não é definido nos pontos de transição da estrutura linear por partes, o valor da derivada da função não-linear para um certo valor de entrada é dada pelo coeficiente angular do intervalo ao qual pertence a amostra.

Foram realizadas 10 simulações para uma configuração semelhante à do ensaio anterior, alterando-se as funções  $g_i(\cdot)$  por estruturas lineares por partes e gerando aleatoriamente a matriz de mistura. Neste caso, optamos por estruturas com 10 seções lineares, para garantir que a inversa das funções não-lineares fosse razoavelmente aproximada. No entanto, os parâmetros do algoritmo evolutivo passaram a ser  $N_c = 20$ ,  $\beta = 60$  e  $\sigma_s = 5$ . O aumento no número de clones se deve ao fato do número de parâmetros a ser ajustado ser substancialmente maior que nos casos anteriores.

**Tabela 5.5:** Média dos EQMs para estrutura linear por partes

EQM ( $\times 10^{-3}$ )	$\hat{s}_1$	$\hat{s}_2$
opt-aiNet/FastICA	5,398	6,612

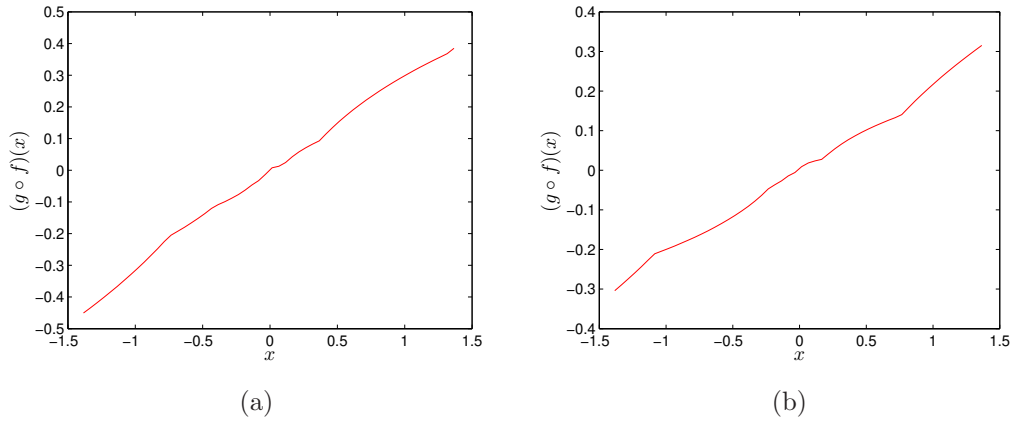
Os resultados de erro quadrático médio, após normalização da variância dos sinais estimados e das fontes reais, são apresentados na tabela 5.5. Novamente, vemos que o algoritmo é capaz de realizar a estimação relativamente precisa dos sinais das fontes, indicando que a estrutura linear por partes aproximou com relativa acuidade a inversa das funções  $f_i$ . Na figura 5.7 apresentamos a composição das funções não-lineares  $f_i$  e as funções lineares por partes  $g_i$  obtidas em uma das simulações, onde se nota que o mapeamento resultante é próximo a um mapeamento linear. Na figura 5.6 comparamos a distribuição dos sinais observados e dos sinais recuperados com as mesmas funções que deram origem à figura 5.7.

**Figura 5.6:** Distribuição dos sinais (a) observados e (b) recuperados.

A exemplo do que havia sido observado anteriormente, a técnica foi capaz de realizar o ajuste do sistema separador de maneira satisfatória. A estrutura não-linear utilizada foi capaz de prover um mapeamento inverso aproximado para as distorções presentes na mistura, mostrando que o método é bastante flexível.

## 5.5 Sumário

Neste capítulo, discutimos sobre a separação cega de fontes em cenários com misturas não-lineares. No caso mais geral, a abordagem desenvolvida para os modelos lineares



**Figura 5.7:** Composição das funções  $f_i$  e  $g_i$  para as duas fontes.

não é mais válida, e algumas propostas distintas foram concebidas para tentar separar as fontes. No entanto, ao contrário do que se observa no caso linear, não existem garantias de que seja possível recuperá-las.

Na impossibilidade de se obter uma solução geral para o problema, considerou-se um modelo mais restrito, mas para o qual a idéia de independência entre os sinais ainda é válida para o projeto da estrutura de separação. O modelo em questão, denominado mistura PNL, é composto por uma seção linear e um conjunto de funções não-lineares, que são aplicadas individualmente a cada um dos sinais linearmente misturados. Dada a forma peculiar do modelo, o sistema separador possui uma estrutura semelhante, constituindo-se também de uma matriz  $\mathbf{W}$  e um conjunto de não-linearidades  $g_i(\cdot)$ .

Uma das dificuldades relacionadas a este tipo de mistura reside na determinação dos parâmetros do sistema separador, que, diferentemente do caso linear, é mais sensível a erros de estimação da função custo ou de sua derivada. Além disso, verifica-se que a utilização de técnicas de otimização baseadas no gradiente leva, por vezes, a soluções sub-ótimas.

Por este motivo, propusemos uma nova metodologia baseada em uma ferramenta de busca robusta, denominada opt-aiNet, cuja principal característica é sua grande capacidade de busca multimodal. A função custo, neste caso representada pela informação mútua entre as saídas do sistema separador, necessita de estimativas da entropia marginal de cada um dos sinais recuperados, de forma que empregamos um estimador baseado em estatísticas de ordem para realizar esta tarefa.

Os resultados desta primeira proposta mostraram-se promissores, atestando a capaci-

dade de busca do algoritmo evolutivo, e também a acuidade do estimador de entropia escolhido. No entanto, numa tentativa de se reduzir a complexidade computacional do método, propusemos uma nova modificação, que introduz uma ferramenta de busca altamente especializada no laço de adaptação. Desta forma, somente as estruturas não-lineares são adaptadas pelo algoritmo evolutivo, enquanto que a seção linear é obtida por meio do algoritmo FastICA, o que mostrou-se bastante efetivo no que se refere à diminuição do tempo de execução do algoritmo.

Finalmente, testamos a metodologia na adaptação de uma estrutura não-linear mais flexível, visando aumentar a gama de cenários nos quais a abordagem proposta seria capaz de separar os sinais. Apesar de um aumento no número de parâmetros a serem ajustados, obtivemos resultados semelhantes aos observados anteriormente para uma estrutura mais restrita, recuperando os sinais das fontes com grande acuidade.

# 6

## Conclusões e Perspectivas

No presente trabalho, abordamos o problema de separação cega de fontes, buscando comentar nossas contribuições originais para os casos de mistura convolutiva e não-linear.

No capítulo 2 apresentamos um panorama geral sobre as técnicas desenvolvidas para separação de fontes no contexto de misturas instantâneas. Vimos que, sob algumas hipóteses plausíveis, é possível recuperar os sinais a partir da análise por componentes independentes. E caso outras hipóteses sejam consideradas, métodos mais específicos podem ser desenvolvidos.

No capítulo 3 abordamos especificamente a técnica de ICA, apresentando os principais critérios e algoritmos desenvolvidos. Como uma decorrência desse estudo mais aprofundado sobre BSS foi gerado um artigo em duas partes, que apresentamos como anexo a esse capítulo.

Já no capítulo 4, consideramos o problema de misturas convolutivas, para o qual propusemos um método baseado em filtros de erro de predição não-lineares. A abordagem, aplicável a sinais com distribuição discreta, emprega os filtros não-lineares de maneira a reduzir redundâncias nos sinais observados e, dessa forma, transforma o problema em uma mistura instantânea equivalente.

No contexto do capítulo 4, o uso de filtros nebulosos é particularmente adequado, haja vista a equivalência entre sua estrutura e a do estimador ótimo de mínimo erro quadrático. O treinamento dos parâmetros é feito por meio de uma técnica de clusterização, que denominamos *K-means*/ILS, e o algoritmo RLS. Conforme pudemos observar através das simulações, a conjunção destes dois métodos foi bastante eficaz na determinação dos parâmetros da estrutura não-linear.

No capítulo 5 passamos ao problema de separação de misturas não-lineares, onde também pudemos contribuir propondo uma nova técnica para efetuar o ajuste do sistema separador. A abordagem, aplicável no contexto de misturas com não-linearidade posterior, alia o poder de busca de um algoritmo evolutivo a um método bastante eficaz para estimação da função custo a ser otimizada. Com isso, obtivemos uma ferramenta extremamente flexível para separação dos sinais, com uma estrutura não-linear capaz de inverter de maneira adequada o processo de mistura.

Como perspectivas de continuidade deste trabalho temos, principalmente, propostas que reduzam a complexidade computacional dos métodos expostos. No caso de misturas convolutivas, vislumbramos o uso de outras estruturas não-lineares, como os filtros de volterra e redes neurais, a fim de atenuar o problema do número de parâmetros a ser ajustado. Além disso, tais estruturas podem viabilizar o uso deste método em cenários nos quais as fontes não possuem distribuição discreta. Isto porque, embora não possuam estruturalmente a forma do preditor ótimo, as estruturas mencionadas apresentam a capacidade de aproximação universal, e espera-se que sejam flexíveis o bastante para realizar a mesma tarefa que o filtro nebuloso desempenha no caso de fontes discretas.

Uma outra perspectiva diz respeito à aplicação dos métodos desenvolvidos no problema de separação de fontes em misturas convolutivas PNL, que pode ser visto como uma extensão natural do modelo de mistura PNL.



## Trabalhos publicados durante o doutorado

Listamos abaixo os trabalhos publicados e aceitos para publicação durante o período do doutorado.

### Artigos de conferência:

1. R. R. F. Attux, R. Suyama, R. Ferrari, C. Junqueira, R. Krummenauer, P. Larzabal, A. Lopes, *A Clustering-based Method for DOA Estimation in Wireless Communications*, Aceito para apresentação no 15<sup>th</sup> European Signal Processing Conference (EUSIPCO'07), Poznań, Polônia, setembro de 2007
2. Ricardo Suyama, Murilo B. Loiola, Romis R. de F. Attux, Cynthia Junqueira, *Separação de Fontes Usando o Filtro de Kalman Estendido*, Aceito para apresentação no XXV Simpósio Brasileiro de Telecomunicações (SBrT'2007), Recife, Brasil, setembro de 2007
3. Tiago M. Dias, Ricardo Suyama, Leonardo T. Duarte, Romis R. de F. Attux, João Marcos T. Romano, *Blind Source Separation of Post-Nonlinear Mixtures Using Evolutionary Computation a Local Search Procedure*, Aceito para apresentação no XXV Simpósio Brasileiro de Telecomunicações (SBrT'2007), Recife, Brasil, setembro de 2007
4. Fabiano S. Chaves, Fábio F. G. Fernandes, Romis R. F. Attux, R. Suyama, Renato R. Lopes, João Marcos T. Romano, *Une analyse du problème de l'égalisation basée sur la théorie des jeux*, Aceito para apresentação no XXI Colóquio GRETSI, Troyes, França, Setembro de 2007
5. D. Zanatta Filho, R. R. Lopes, R. Ferrari, M. B. Loiola, R. Suyama, G. Simões, C. Wada, J. M. T. Romano, B. Dortsch e J. Rius i Riu. *The Capacity of Binders for MIMO Digital Subscriber Lines*, VI International Telecommunications Symposium (ITS'2006), Fortaleza, Brasil, setembro de 2006
6. Helder Knidel, Rafael Ferrari, Leonardo T. Duarte, Ricardo Suyama, Romis R. F. Attux, Leandro N. de Castro, Fernando J. Von Zuben, João M. T. Romano, *A Hybrid Unsupervised Clustering Algorithm for Channel Equalization*, IEEE International Workshop on Machine Learning for Signal Processing (MLSP'2006), Maynooth, Irlanda, setembro de 2006
7. Cynthia Junqueira, Fabrício O. de França, Romis R. F. Attux, Ricardo Suyama, Leandro N. de Castro, Fernando J. Von Zuben, João Marcos T. Romano, *A*

*Proposal for Blind FIR Equalization of Time-Varying Channels*, IEEE International Workshop on Machine Learning for Signal Processing (MLSP'2005), Mystic, Connecticut, EUA, setembro de 2005

8. Clodoaldo A. M. Lima, Cynthia Junqueira, Ricardo Suyama, Fernando J. Von Zuben, João M. T. Romano. *Least-Squares Support Vector Machines for DOA Estimation: A Step-by-Step Description and Sensitivity Analysis*, IEEE International Joint Conference on Neural Networks (IJCNN'2005), Vancouver, BC, Canadá, julho de 2005
9. Cynthia Junqueira, Ricardo Suyama, Fernando J. Von Zuben, João Marcos T. Romano. *An evolutive algorithm for Blind Adaptive Beamforming in GPS Applications*, IEEE International Workshop on Machine Learning for Signal Processing (MLSP'2004), São Luís, Brasil, setembro-outubro de 2004

#### Artigos de revista:

10. Ricardo Suyama, Leonardo T. Duarte, Rafael Ferrari, Leandro E. P. Rangel, Romis R. de F. Attux, Charles C. Cavalcante, Fernando J. Von Zuben, João M. T. Romano. *A Nonlinear-Prediction Approach to the Blind Separation of Convolutional Mixtures*, EURASIP Journal on Applied Signal Processing, 2007
11. Aline Neves, Romis R. de F. Attux, Ricardo Suyama, Maria D. Miranda, João M. T. Romano. *Sobre Critérios para Equalização Não-Supervisionada*, Sba: Controle & Automação Sociedade Brasileira de Automatica, scielo, 2006, 17, pp. 278-299
12. Romis R. de F. Attux, Aline Neves, Leonardo T. Duarte, Ricardo Suyama, Cynthia Junqueira, Leandro E. P. Rangel, Tiago M. Dias, João M. T. Romano, *On the Relationships between Blind Equalization and Blind Source Separation - Part I: Foundations*, aceito para publicação na Revista da Sociedade Brasileira de Telecomunicações
13. Romis R. de F. Attux, Aline Neves, Leonardo T. Duarte, Ricardo Suyama, Cynthia Junqueira, Leandro E. P. Rangel, Tiago M. Dias, João M. T. Romano, *On the Relationships between Blind Equalization and Blind Source Separation - Part II: Relationships*, aceito para publicação na Revista da Sociedade Brasileira de Telecomunicações

#### Capítulos de livro:

14. Ricardo Suyama, Leonardo T. Duarte, Aline Neves, Rafael Ferrari, Romis R. F. Attux, Charles C. Cavalcante, Cynthia Junqueira, João M. T. Romano. *Unsupervised Signal Processing: Concepts, Applications and Trends*, em *Telecommunications: Advances and Trends in Transmission, Networking and Applications*, University of Fortaleza Press, 2006, pp. 29-55, ISBN: 85-98876-18-6
15. Leonardo Tomazeli Duarte, Ricardo Suyama, Romis R. de F. Attux, Fernando J. Von Zuben, and João Marcos T. Romano. *Blind Source Separation of Post-Nonlinear Mixtures Using Evolutionary Computation and Order Statistics*, em *Independent Component Analysis and Blind Signal Separation*, Lecture Notes in Computer Science, volume 3889/2006, Springer Berlin / Heidelberg, 2006, pp. 66-73, ISBN: 3-540-32630-8

**Pedidos de Patente:**

16. B. Dorschty, D. Zanatta Filho, R. Suyama, M. Loiola, R. Ferrari, R. Lopes e J. M. T. Romano, *Method for discrete bit allocation for precoded DMT-based xDSL systems*, patente européia pendente, P22391WO, 04 de setembro de 2006.

## Referências Bibliográficas

- [1] S. Achard. *Mesures de Dépendance pour la séparation aveugle de sources: Application aux mélanges post non linéaires*. PhD thesis, Université Joseph Fourier, 2003. citado na(s) página(s): 92
- [2] S. Achard and C. Jutten. Identifiability of post-nonlinear mixtures. *IEEE Signal Processing Letters*, 12(5):423–426, May 2005. citado na(s) página(s): 89, 99
- [3] S. Achard and D.-T. Pham. Blind source separation in post nonlinear mixtures. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2001)*, 2001. citado na(s) página(s): 90
- [4] L. B. Almeida. Linear and nonlinear ica based on mutual information. In *Proc. IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)*, pages 117–122, 2000. citado na(s) página(s): 85, 88
- [5] S.-I. Amari. "natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. citado na(s) página(s): 39
- [6] S.-I. Amari, A. Cichocki, and H. H. Yang. *Unsupervised Adaptive Filtering Volume I: Blind Source Separation*, chapter Blind Signal Separation and Extraction: Neural and Information Theoretic Approaches, pages 63–138. John Wiley & Sons, 2000. citado na(s) página(s): 51
- [7] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang. Multichannel blind deconvolution and equalization using the natural gradient. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication (SPAWC'97)*, pages 101–104, 1997. citado na(s) página(s): 51
- [8] S. Arberet, R. Gribonval, and F. Bimbot. *Independent Component Analysis and Blind Signal Separation*, volume 3889/2006 of *Lecture Notes in Computer*

- Science*, chapter A Robust Method to Count and Locate Audio Sources in a Stereophonic Linear Instantaneous Mixture. Springer, 2006. citado na(s) página(s): 17
- [9] S. Arberet, R. Gribonval, and F. Bimbot. A robust method to count and locate audio sources in a stereophonic linear anechoic mixture. In *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process (ICASSP'07)*, 2007. citado na(s) página(s): 17
- [10] Barry Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12:35–50, Jul 1992. citado na(s) página(s): 2
- [11] R. R. F. Attux. *Novos Paradigmas para equalização e identificação de canais baseados em estruturas não-lineares e algoritmos evolutivos*. PhD thesis, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, 2005. citado na(s) página(s): 93, 96
- [12] R. R. F. Attux, M. B. Loiola, R. Suyama, L. N. de Castro, F. J. Von Zuben, and J. M. T. Romano. Blind search for optimal wiener equalizers using an artificial immune network model. *EURASIP Journal of Applied Signal Processing*, 2003(8):740–747, 2003. citado na(s) página(s): 93
- [13] R. R. F. Attux, A. Neves, L. T. Duarte, R. Suyama, C. Junqueira, L. E. P. Rangel, T. M. Dias, and J. M. T. Romano. On the relationships between blind equalization and blind source separation - part i: Foundations. *Revista da Sociedade Brasileira de Telecomunicações*, 2007. citado na(s) página(s): 31, 55
- [14] R. R. F. Attux, A. Neves, L. T. Duarte, R. Suyama, C. Junqueira, L. E. P. Rangel, T. M. Dias, and J. M. T. Romano. On the relationships between blind equalization and blind source separation - part ii: Relationships. *Revista da Sociedade Brasileira de Telecomunicações*, 2007. citado na(s) página(s): 31, 55
- [15] W. Baumann, B.-U. Köhler, D. Kolossa, and R. Orglmeister. Real time separation of convolutive mixtures. In *Proc. International Conference on Independent Component Analysis and Signal Separation (ICA2001)*, pages 65–69, 2001. citado na(s) página(s): 54
- [16] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995. citado na(s) página(s): 29, 30, 34

- [17] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, Feb. 1997. citado na(s) página(s): 15
- [18] A. Benveniste, M. Goursat, and G. Ruget. Identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications. *IEEE Transactions on Automatic Control*, AC-25(3):385–399, June 1980. citado na(s) página(s): 31
- [19] E. Bingham and A. Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex-valued signals. *International Journal of Neural Systems*, 10, 2000. citado na(s) página(s): 52
- [20] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353–2362, November 2001. citado na(s) página(s): 17
- [21] H. Buchner, R. Aichner, and W. Kellermann. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. *IEEE Transactions on Speech and Audio Processing*, 13(1):120–134, Janeiro 2005. citado na(s) página(s): 2
- [22] V. Calhoun, T. Adali, L. Hansen, J. Larsen, and J. Pekar. ICA of Functional MRI Data: an Overview. In *Proc. 4th international symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003. citado na(s) página(s): 3
- [23] J. F. Cardoso. Blind beamforming for non-gaussian signals. *IEE Proceedings*, 140(6):362–370, December 1993. citado na(s) página(s): 24, 25, 52
- [24] J.-F. Cardoso. Infomax and maximum likelihood of blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997. citado na(s) página(s): 12, 30
- [25] J.-F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, Outubro 1998. citado na(s) página(s): 30
- [26] J.-F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, December 1996. citado na(s) página(s): 37, 38

- [27] Jean-François Cardoso and Antoine Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164, 1996. citado na(s) página(s): 25
- [28] R. T. Causey. *Blind Multiuser Detection Based on Second-Order Statistics*. PhD thesis, School of Electrical and Computer Engineering - Georgia Institute of Technology, Atlanta, Georgia, 1999. citado na(s) página(s): 47, 49
- [29] C. C. Cavalcante. Predição neural e estimação de função de densidade de probabilidade aplicadas à equalização cega. Master's thesis, Universidade Federal do Ceará, Fortaleza, Ceará, Fevereiro 2001. citado na(s) página(s): 55
- [30] C. C. Cavalcante. *Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário*. PhD thesis, Faculdade de Engenharia Elétrica e de Computação - FEEC - UNICAMP, Campinas, April 2004. citado na(s) página(s): 3, 12, 32, 45, 95
- [31] W. Chen, T. Nemoto, T. Kobayashi, Y. Honda, T. Saito, and E. Kasuya. Foetal ecg extraction from maternal body surface measurement using independent component analysis. In *Proc. 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2001.*, volume 2, pages 1990–1993, 25-28 Oct. 2001. citado na(s) página(s): 3
- [32] C. Chinrungrueng and C. H. Seguin. Optimal adaptive k-means algorithm with dynamic adjustment of learning rate. *IEEE Transactions on Neural Networks*, 6(1):157–169, 1995. citado na(s) página(s): 68
- [33] S. Choi, A. Cichocki, H. M. Park, and S.-Y. Lee. Blind source separation and independent component analysis: A review. *Neural Information Processing - Letters and Reviews*, 6(1):1–57, Jan 2005. citado na(s) página(s): 11
- [34] A. Cichocki and S.-I. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002. citado na(s) página(s): 3, 15, 16
- [35] A. Cichocki and R. Unbehauen. Robust neural networks with on-line learning for blind identification and blind separation of sources. *IEEE Transaction on Circuits and Systems - I: Fundamental Theory and Applications*, 43:894–906, 1996. citado na(s) página(s): 34
- [36] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, Abril 1994. citado na(s) página(s): 4, 10, 11, 24, 25, 95



- [37] J. X. Cong. Historical Development of Central Limit Theorem (CLT). Technical report, Rice University, Fevereiro 2003. Disponível on-line em <http://www.stat.rice.edu/blairc/seminar/Files/julieTalk.pdf>. citado na(s) página(s): 26
- [38] T. M Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991. citado na(s) página(s): 22
- [39] S. A. Cruces-Alvarez, A. Cichock, and S.-I. Amari. From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms, and stability. *IEEE Transactions on Neural Networks*, 15(4):859–873, July 2004. citado na(s) página(s): 28
- [40] L. N. de Castro and J. Timmis. An artificial immune network for multimodal function optimization. In *Proc. International Conference on Evolutionary Computation (ICEC'02)*, pages 699–674, Honolulu, Hawaii, 2002. citado na(s) página(s): 93, 96
- [41] L. N. de Castro and F. J. Von Zuben. Learning and Optimization Using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251, June 2002. citado na(s) página(s): 93
- [42] Mauricio Sol de Castro. Soluções adaptativas para equalização autodidata multicanal. Master's thesis, Faculdade de Engenharia Eletrica e de Computação - Universidade Estadual de Campinas, 2002. citado na(s) página(s): 55
- [43] L. de Lathauwer, B. de Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, 47(5):567–572, May 2000. citado na(s) página(s): 3
- [44] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: A second-order stable algorithm for the general case. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, 47(7):1056–1071, 2000. citado na(s) página(s): 57
- [45] Y. Deville. A unified stability analysis of the hérault-jutten source separation neural network. *Signal Processing*, 51(3):229–233, 1996. citado na(s) página(s): 34
- [46] K.I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks*. John Wiley & Sons, 1996. citado na(s) página(s): 36



- [47] S. Dodel, J. M. Herrmann, and T. Geisel. Localization of brain activity – Blind separation of fMRI data. *Neurocomputing*, 32:701–708, 2000. citado na(s) página(s): 3
- [48] S. C. Douglas. Fixed-point fastica algorithms for the blind separation of complex-valued signal mixtures. In *Proc. Asilomar Conference on Signals, Systems and Computers (ASILOMAR2005)*, pages 1320–1325, 2005. citado na(s) página(s): 52
- [49] L. T. Duarte. Um estudo sobre separação cega de fontes e contribuições ao caso de misturas não-lineares. Master’s thesis, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, 2006. citado na(s) página(s): 95
- [50] D. P. W. Ellis. *Prediction-driven computational auditory scene analysis*. PhD thesis, Massachusetts Institute of Technology, 1996. citado na(s) página(s): 2
- [51] J. Even. *Contribution a la separaration de sources à l’aide de statistiques d’ordre*. PhD thesis, Université Joseph-Fourier Grenoble, 2003. citado na(s) página(s): 93
- [52] R. Ferrari. Equalização de Canais de Comunicação baseada em Filtros Fuzzy. Master’s thesis, Faculdade de Engenharia Elétrica e de Computação - FEEC - UNICAMP, Campinas, February 2005. citado na(s) página(s): 55, 56, 57, 63, 80
- [53] R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, L. N. de Castro, F. J. Von Zuben, and J. M. T. Romano. Unsupervised Channel Equalization Using Fuzzy Prediction-Error Filters. In *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP03)*, Toulouse, França, 2003. citado na(s) página(s): 63
- [54] I. K. Fodor and C. Kamath. Using independent component analysis to separate signals in climate data. In *Proc. Independent component analysis, Wavelets, and Neural Networks - SPIE Aerosense*, Orlando, 2003. citado na(s) página(s): 3
- [55] P. G. Georgiev, F. Theis, and A. Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Transactions on Neural Networks*, 16(4):992–996, July 2005. citado na(s) página(s): 17
- [56] D. N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transactions on Communications*, COM-28(11):1867–1875, 1980. citado na(s) página(s): 55

- [57] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, 1996. citado na(s) página(s): 12, 28
- [58] A. Gorokhov and P. Loubaton. Blind Identification of MIMO-FIR Systems: A Generalized Linear Prediction Approach. *Signal Processing*, 1999(73):105–124, 1999. citado na(s) página(s): 56
- [59] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1998. citado na(s) página(s): 84
- [60] S. Haykin. *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*, volume 2. Wiley, 2000. citado na(s) página(s): 37, 55
- [61] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 4 edition, 2001. citado na(s) página(s): 73
- [62] Z. He and A. Cichocki. *Independent Component Analysis and Blind Signal Separation*, chapter K-EVD Clustering and Its Applications to Sparse Component Analysis, pages 90–97. Springer, 2006. citado na(s) página(s): 17
- [63] M. T. Heath. *Scientific Computing: An Introductory Survey*. McGraw-Hill, 2002. citado na(s) página(s): 33, 40
- [64] J. Herault, C. Jutten, and B. Ans. Detection de grandeurs primitive dans un message composite par une architecture de calcul neuromimetique en apprentissage non supervise. In *Proc. 10th GRETSI Symposium on Signal and Image Processing*, pages 1017–1022, Nice, France, May 1985. citado na(s) página(s): 3, 4
- [65] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999. citado na(s) página(s): 33, 40
- [66] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley Interscience, 2001. citado na(s) página(s): 15, 26, 29, 32, 33, 36, 37, 45
- [67] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997. citado na(s) página(s): 40

- [68] A. Hyvärinen and P. Pajunen. Nonlinear independent componentes analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999. citado na(s) página(s): 87
- [69] S. Ikeda and N. Murata. An approach to blind source separation of speech signals. In *Proc. ICANN'98*, pages 1855–1865, 1998. citado na(s) página(s): 54
- [70] M. Z. Ikram and D. R. Morgan. A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP2002)*, pages 881–884, 2002. citado na(s) página(s): 54
- [71] A. Ilin and A. Honkela. Post-nonlinear independent component analysis by variational bayesian learning. In *Proc. Fifth International Conference (ICA2004)*, pages 766–773, 2004. citado na(s) página(s): 92
- [72] Y. Inouye and R.-W. Liu. A System Theoretic Foundation for Blind Equalization of an FIR MIMO Channel System. *IEEE Transactions on Circuits and Systems*, 49(4):425–436, Abril 2002. citado na(s) página(s): 49
- [73] C.J. James and O.J. Gibson. Temporally constrained ica: an application to artifact rejection in electromagnetic brain signal analysis. *IEEE Transactions on Biomedical Engineering*, 50(9):1108–1116, Sept. 2003. citado na(s) página(s): 3
- [74] C. Jutten, M. Babaie-Zadeh, and Shahram Hosseini. Three easy ways for separating nonlinear mixtures? *Signal Processing*, 84(2):217–229, Fevereiro 2004. citado na(s) página(s): 87
- [75] C. Jutten and J. Karhunen. Advances in nonlinear blind source separation. In *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 245–256, 2003. citado na(s) página(s): 86
- [76] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and J. Caelen. Blind separation of sources: an algorithm for separation of convolutive mixtures. In *Proc. International Signal Processing Workshop on Higher Order Statistics*, 1991. citado na(s) página(s): 47, 50
- [77] A. M. Kagan, Y. V. Linnik, and C. R. Rao. *Characterization Problems in Mathematics Statistics*. John Wiley & Sons, 1973. citado na(s) página(s): 10
- [78] T. Kailath. *Linear Systems*. Prentice Hall, 1980. citado na(s) página(s): 47, 49

- [79] S. T. Kaplan. Principal and independent component analysis for seismic data. Master's thesis, University of British Columbia, Canada, 2003. citado na(s) página(s): 3
- [80] S. M. Kay. *Fundamentals of Statistical Signal Processing*. Prentice Hall, 1993. citado na(s) página(s): 19, 64
- [81] K. H. Knuth. A bayesian approach to source separation. In *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA '99)*, pages 283–288, Aussois, France, January 1999. citado na(s) página(s): 19
- [82] E. Kofidis. Blind Source Separation: Fundamentals and Recent Advances. *Mini-curso no XIX Simpósio Brasileiro de Telecomunicações (SBrT2001)*, September 2001. citado na(s) página(s): 1, 12, 24, 29, 31, 45
- [83] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura. Evaluation of frequency-domain blind signal separation using directivity pattern under reverberant conditions. In *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP'00)*, pages 3140–3143, 2000. citado na(s) página(s): 54
- [84] L. De Lathauwer, B. De Moor, and J. Vanderwalle. Ica algorithms for 3 sources and 2 sensors. In *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics (HOS'99)*, pages 116–120, Caesarea, Israel, June 1999. citado na(s) página(s): 25
- [85] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:708–791, 1999. citado na(s) página(s): 16
- [86] J. Lee, K.L. Park, and K.J. Lee. Temporally constrained ica-based foetal ecg separation. *Electronics Letters*, 41(21):1158–1160, 13 Oct. 2005. citado na(s) página(s): 3
- [87] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, 11(2):417–441, 1999. citado na(s) página(s): 35
- [88] Y. Li, A. Cichocki, and S.-I. Amari. Sparse component analysis for blind source separation with less sensors than sources. In *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 89–94, Kyoto, Japan, 2003. citado na(s) página(s): 17

- [89] Y. Li and J. Wang. Sequential blind extraction of instantaneously mixed sources. *IEEE Transactions on Signal Processing*, 50(5):997–1006, May 2002. citado na(s) página(s): 28
- [90] S. Makino, S. Araki, R. Mukai, and H. Sawada. Audio source separation based on independent component analysis. In *Proc. International Symposium on Circuits and Systems (ISCAS2004)*, pages 668–671, May 2004. citado na(s) página(s): 2
- [91] S. Makino, H. Sawada, R. Mukai, and S. Araki. Blind source separation of convolutive mixtures of speech in frequency domain. *IEICE Transaction of Electronics, Communications and Computer Sciences*, E88-A(7):1640–1655, 2005. citado na(s) página(s): 51
- [92] M. Malcangi. *Computer Music Modeling and Retrieval*, chapter Source Separation and Beat Tracking: A System Approach to the Development of a Robust Audio-to-Score System, pages 71–82. Springer, 2005. citado na(s) página(s): 2
- [93] M. Babaiezadeh Malmiri. *On blind source separation in convolutive and nonlinear mixtures*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, France, September 2002. citado na(s) página(s): 88, 91, 92
- [94] A. Mansour, C. Jutten, and P. Loubaton. Adaptive subspace algorithm for blind separation of independent sources in convolutive mixture. *IEEE Transactions on Signal Processing*, 48(2):583–586, 2000. citado na(s) página(s): 57
- [95] K. Matsuoka. Minimal distortion principle for blind source separation. In *Proc. 41st SICE Annual Conference (SICE'2002)*, volume 4, pages 2138–2143, 2002. citado na(s) página(s): 53
- [96] K. Matsuoka and S. Nakashima. Minimal distortion principle for blind source separation. In *Proc. International Conference on Independent Component Analysis and Signal Separation (ICA2001)*, pages 722–727, 2001. citado na(s) página(s): 53
- [97] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995. citado na(s) página(s): 15
- [98] J. M. Mendel. Fuzzy logic systems for engineering: a tutorial. *Proceedings of the IEEE*, 83(3):345–377, 1995. citado na(s) página(s): 58

- [99] N. Mitianoudis and M. E. Davies. Audio source separation of convolutive mixtures. *IEEE Transactions on Speech and Audio Processing*, 11(5):489 – 497, Setembro 2003. citado na(s) página(s): 2, 45
- [100] A. Mohammad-Djafari. A bayesian approach to source separation. In *Proc. 19th International Workshop on Bayesian and Maximum Entropy methods (MaxEnt 1999)*, pages 1–26, Boise, Idaho, USA, 1999. citado na(s) página(s): 19
- [101] J.-P. Nadal and N. Parga. Nonlinear neurons in the low-noise limit: a factorial code maximises information transfer. *Network: Computation in Neural Systems*, 5(4):565–581, November 1994. citado na(s) página(s): 29
- [102] A. O. Neves, R. R. F.l Attux, R. Suyama, M. D. Miranda, and J. M. T. Romano. Sobre critérios para equalização não-supervisionada. *Sba: Controle & Automação Sociedade Brasileira de Automática*, 17(3):278–299, 2006. citado na(s) página(s): 3, 37
- [103] C. Nikias and A. P. Petropulu. *Higher-Order Spectra Analysis*. Prentice Hall PTR, 1993. citado na(s) página(s): 24
- [104] P. D. O’Grady, B. A. Pearlmutter, and S. T. Rickard. Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology (IJIST)*, 15:18–33, 2005. citado na(s) página(s): 17
- [105] E. Oja and M. D. Plumbley. Blind separation of positive sources by globally convergent gradient search. *Neural Computation*, 16(9):1811–1825, September 2004. citado na(s) página(s): 16
- [106] P. Paatero and U. Tapper. Positive Matrix Factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994. citado na(s) página(s): 16
- [107] P. Pajunen. Nonlinear independent component analysis by self-organizing maps. In *Proc. International Conference on Neural Information Processing (ICONIP’96)*, pages 1207–1210, Hong Kong, China, Setembro 1996. citado na(s) página(s): 84
- [108] P. Pajunen and J. Karhunen. Least-squares methods for blind source separation based on nonlinear pca. *International Journal of Neural Systems*, 8(5-6):601–612, Outubro-Dezembro 1997. citado na(s) página(s): 37



- [109] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill International, 3 edition, 1991. citado na(s) página(s): 9, 11, 12, 13, 19, 24, 26, 27, 64, 87
- [110] D. T. Pham. Blind separation of instantaneous mixture of sources based on order statistics. *IEEE Transactions on Signal Processing*, 2(48):363–375, February 2000. citado na(s) página(s): 93, 94
- [111] D.T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, 2001. citado na(s) página(s): 15
- [112] M. Plumbley. Conditions for nonnegative independent component analysis. *IEEE Signal Processing Letters*, 9(6):177–180, June 2002. citado na(s) página(s): 16
- [113] M. D. Plumbley. Algorithms for nonnegative independent component analysis. *IEEE Transactions on Neural Networks*, 14(3):534–543, 2003. citado na(s) página(s): 16
- [114] M. D. Plumbley. *Independent Component Analysis and Blind Signal Separation*, proc. fifth international conference, ica 2004, granada, spain, september 22-24, 2004 Lie Group Methods for Optimization with Orthgonality Constraints, pages 1245–1252. Number 3195 in Lecture Notes in Computer Sciences. Springer-Verlag, 2004. citado na(s) página(s): 41, 42
- [115] M. D. Plumbley. Geometrical methods for non-negative ica: Manifolds, lie groups and toral subalgebras. *Neurocomputing*, 67:161–197, August 2005. citado na(s) página(s): 43
- [116] M. D. Plumbley, S. A. Abdallah, J. P. Bello, M. E. Davies, G. Mont, and M. B. Sandler. Automatic music transcription and audio source separation. *Cybernetics and Systems*, 33(6):603 – 627, 2002. citado na(s) página(s): 2
- [117] F. Rojas, C. G. Puntonet, M. Rodríguez-Álvarez, I. Rojas, and R. Martín-Clemente. Blind source separation in post-nonlinear mixtures using competitive learning, simulated annealing, and a genetic algorithm. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 34(4):407–416, 2004. citado na(s) página(s): 92, 93
- [118] F. Rojas, I. Rojas, R. Martín Clemente, and C. G. Puntonet. Nonlinear blind source separation using genetic algorithms. In *Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation (ICA*

- 2001), pages 400–405, San Diego, USA, December 2001. citado na(s) página(s): 92, 93
- [119] S. Sanei, S. Constantinides, C. C. Took, and B. Song. Incorporating frequency warping into sparse component analysis. In *Proc. 13th European Signal Processing Conference (EUSIPCO 2005)*, Turkey, 2005. citado na(s) página(s): 17
  - [120] H. Sawada, R. Mukai, S. Araki, and S. Makino. A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Transaction on Speech and Audio Processing*, 12(5):530–538, 2004. citado na(s) página(s): 54
  - [121] C. Serviere and D.-T. Pham. *Independent Component Analysis and Blind Signal Separation*, chapter A Novel Method for Permutation Correction in Frequency-Domain in Blind Separation of Speech Mixtures, pages 807–815. Springer, 2004. citado na(s) página(s): 54
  - [122] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22:21–34, 1998. citado na(s) página(s): 51, 52
  - [123] J.i Solé-Casals, M. Babaie-Zadeh, C. Jutten, and D.-T. Pham. Improving algorithm speed in pml mixture separation and wiener system inversion. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japão, 2003. citado na(s) página(s): 91
  - [124] E. Sorouchyari. Blind separation of sources, part III: Stability analysis. *Signal Processing*, 24:21–29, 1991. citado na(s) página(s): 34
  - [125] A. Taleb and C. Jutten. Nonlinear source separation: The post-nonlinear mixtures. In *Proc. ESANN*, Bruges, Bélgica, 1997. citado na(s) página(s): 88
  - [126] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing*, 47(10):2807–2820, Outubro 1999. citado na(s) página(s): 4, 87, 89, 90
  - [127] H.-L. Nguyen Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal Processing*, 45(2):209–229, 1995. citado na(s) página(s): 50
  - [128] J. Thomas, Y. Deville, and S. Hosseini. Time-domain fast fixed-point algorithms for convolutive ica. *IEEE Signal Processing Letters*, 13(4):228 – 231, 2006. citado na(s) página(s): 51



- [129] L. Tong, Ruey wen Liu, and V. C. Soon. Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 38(5):499–509, Maio 1991. citado na(s) página(s): 15
- [130] K. Torkkola. Blind separation of convolved sources based on information maximization. In *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pages 1–10, Kyoto, Japan, 1996. citado na(s) página(s): 51
- [131] K. Torkkola. *Unsupervised Adaptive Filtering Volume I: Blind Source Separation*, chapter Blind Separation of Delayed and Convolved Sources, pages 321–375. John Wiley & Sons, 2000. citado na(s) página(s): 51
- [132] H. Valpola. Nonlinear independent component analysis using ensemble learning: Theory. In *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 251–256, Finlândia, 2000. citado na(s) página(s): 85
- [133] H. Valpola and P. Pajunen. Fast algorithms for bayesian independent component analysis. In *Proc. Second International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 233–237, Helsinki, Finland 2000. citado na(s) página(s): 19
- [134] M. van der Baan. PP/PS Wavefield separation by independent component analysis. *Geophysical Journal International*, 166:339–348, 2006. citado na(s) página(s): 3
- [135] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6):1414–1427, 1992. citado na(s) página(s): 58, 61
- [136] C. S. Wong, D. Obradovic, and N. Madhu. (ICA) for Blind Equalization of Frequency Selective Channels. In *IEEE 13th Workshop on Neural Networks for Signal Processing (NNSP'03)*, pages 419–428, 2003. citado na(s) página(s): 3
- [137] H. H. Yang, S.-I. Amari, and A. Cichocki. Information-theoretic approach to blind separation of sources in non-linear mixture. *Signal Processing*, 64:291–300, 1998. citado na(s) página(s): 88
- [138] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Transactions on Signal Processing*, 42(8):2158–2168, Agosto 1994. citado na(s) página(s): 4, 49

- [139] V. Zarzoso and A. K. Nandi. Improving mimo channel equalization with independent component analysis. In *Proc. 6th IMA International Conference on Mathematics in Signal Processing*, pages 223–226, Cirencester, UK, December 2004. citado na(s) página(s): 3
- [140] Y. Zhanga, X. Shia, and C. H. Chenb. A gaussian mixture model for underdetermined independent component analysis. *Signal Processing*, 86(7):1538–1549, July 2006. citado na(s) página(s): 19
- [141] W. Zhou, J. Zhou, H. Zhao, and L. Ju. Removing eye movement and power line artifacts from the eeg based on ica. In *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005*, pages 6017–6020, 2005. citado na(s) página(s): 3
- [142] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller. Blind separation of post-nonlinear mixtures usins linearizing transformations and temporal decorrelation. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japão, 2003. citado na(s) página(s): 91
- [143] A. Ziehe and K. Muller. TDSEP – an efficient algorithm for blind separation using time structure. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, pages 675–680, Suécia, 1998. citado na(s) página(s): 15