

FERNANDO ALVES DE LIMA MORETO

**ANÁLISE DE COMPONENTES INDEPENDENTES APLICADA À
SEPARAÇÃO DE SINAIS DE ÁUDIO**

Dissertação apresentada ao Programa de Pós-Graduação do curso de Engenharia Elétrica da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica

São Paulo

2008

FERNANDO ALVES DE LIMA MORETO

**ANÁLISE DE COMPONENTES INDEPENDENTES APLICADA À
SEPARAÇÃO DE SINAIS DE ÁUDIO -
POR MEIO DE BUSCA DE PROJEÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação do curso de Engenharia Elétrica da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica

Área de concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr Miguel Arjona Ramírez

São Paulo

2008

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, de abril de 2008.

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Moreto, Fernando Alves de Lima

Análise de componentes independentes aplicada à separação de sinais de áudio / F. A. L. Moreto. -- ed.rev. --São Paulo, 2008.

p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

**1. Processamento de sinais acústicos 2. Análise multivariada
3. Teoria da informação I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II. t.**

Dedicatória

À minha esposa, Priscila Ranieri Moreto, por compreender o motivo de minha ausência em muitos momentos e, principalmente, pelo seu amor e incentivo incondicional.

Agradecimentos

Primeiramente, agradeço a Deus por todas as oportunidades proporcionadas, que tornaram possível a realização deste projeto.

Ao meu orientador, Dr. Miguel Arjona Ramírez, pela orientação, incentivo e preocupação despendidas com muita dedicação à realização desta pesquisa. Sua colaboração permitiu não só o crescimento acadêmico, mas, sim, um valioso desenvolvimento pessoal.

A todos os professores que acreditaram, colaboraram e propiciaram para que, eu chegasse até esse momento, em especial aos professores que participaram da minha banca de qualificação, Dra. Cinthia Itiki, Dr. Magno Teófilo Madeira e Dr. Cristiano Magalhães Panazio.

Aos meus pais, meu irmão e minha avó que sempre estiveram ao meu lado, ajudando, apoiando e acompanhando em todos os momentos da minha vida.

A todos os familiares e amigos que contribuíram de alguma forma ou estiveram presentes em todos os momentos.

Aos amigos que compartilharam o árduo processo de desenvolvimento científico. Em especial aos amigos, Edmila Montezani e Roberto Nery da Fonseca, pela união, perseverança e a amizade que nos uniu desde o início do mestrado.

Aos diretores da VTB Consultoria e Treinamento que incentivaram e contribuíram para a conclusão deste projeto.

Resumo

Este trabalho estuda o modelo de análise em componentes independentes (ICA) para misturas instantâneas, aplicado na separação de sinais de áudio. Três algoritmos de separação de misturas instantâneas são avaliados: FastICA, PP (*Projection Pursuit*) e PearsonICA; possuindo dois princípios básicos em comum: as fontes devem ser independentes estatisticamente e não-Gaussianas. Para analisar a capacidade de separação dos algoritmos foram realizados dois grupos de experimentos. No primeiro grupo foram geradas misturas instantâneas, sinteticamente, a partir de sinais de áudio pré-definidos. Além disso, foram geradas misturas instantâneas a partir de sinais com características específicas, também geradas sinteticamente, para avaliar o comportamento dos algoritmos em situações específicas. Para o segundo grupo foram geradas misturas convolutivas no laboratório de acústica do LPS.

Foi proposto o algoritmo PP, baseado no método de Busca de Projeções comumente usado em sistemas de exploração e classificação, para separação de múltiplas fontes como alternativa ao modelo ICA. Embora o método PP proposto possa ser utilizado para separação de fontes, ele não pode ser considerado um método ICA e não é garantida a extração das fontes. Finalmente, os experimentos validam os algoritmos estudados.

Palavras-Chave: Processamento de sinais, Separação Cega de Fontes, Análise em Componentes Independentes, *Cocktail Party*, Busca de Projeção, Estatística de Ordem Superior.

Abstract

This work studies Independent Component Analysis (ICA) for instantaneous mixtures, applied to audio signal (source) separation. Three instantaneous mixture separation algorithms are considered: FastICA, PP (Projection Pursuit) and PearsonICA, presenting two common basic principles: sources must be statistically independent and non-Gaussian. In order to analyze each algorithm separation capability, two groups of experiments were carried out. In the first group, instantaneous mixtures were generated synthetically from predefined audio signals. Moreover, instantaneous mixtures were generated from specific signal generated with special features, synthetically, enabling the behavior analysis of the algorithms. In the second group, convolutive mixtures were probed in the acoustics laboratory of LPS at EPUSP.

The PP algorithm is proposed, based on the Projection Pursuit technique usually applied in exploratory and clustering environments, for separation of multiple sources as an alternative to conventional ICA. Although the PP algorithm proposed could be applied to separate sources, it couldn't be considered an ICA method, and source extraction isn't guaranteed. Finally, experiments validate the studied algorithms.

Key words: Signal Processing, Blind Source Separation, Independent Component Analysis, Cocktail Party, Projection Pursuit, Higher Order Statistics.

Lista de Figuras

Figura 1.1 Esquema do Problema <i>Cocktail Party</i>	4
Figura 1.2: Sinais originais das falas (fontes) do filme Harry Potter.....	5
Figura 1.3: Sinais das misturas das falas do filme Harry Potter.....	5
Figura 1.4: Sinais estimados das falas do filme Harry Potter.....	6
Figura 1.5: Diagrama esquemático do problema de separação cega de fontes.....	6
Figura 3.1: Arquitetura de Héroult-Jutten para separação de fontes.....	31
Figura 3.2: Regiões no plano (β_1, β_2) para várias distribuições de Pearson. (De E.S. Pearson, Seminars, Princeton University, 1960).....	43
Figura 5.1: Componentes independentes extraídas aplicando FastICA baseado na curtose....	53
Figura 5.2: Histograma das fontes do Experimento 1: Leitura de trecho do texto em uma revista científica e trecho de uma música clássica.....	55
Figura 5.3: Histograma das fontes do Experimento 2: um sinal senoidal, um sinal periódico com descontinuidades, um sinal dente de serra e um ruído impulsivo.....	57
Figura 5.4: Primeiras 200 amostras, normalizadas, das fontes originais (azul) do Experimento 3 (apenas uma fonte gaussiana) e suas respectivas estimativas (verde), normalizadas, usando o método simétrico e a função custo gaussiano do algoritmo FastICA.....	58
Figura 5.5: Primeiras 200 amostras, normalizadas, das fontes originais (azul) do experimento 4 (duas fonte gaussiana) e suas respectivas estimativas (verde), normalizadas, usando o método simétrico e a função custo gaussiano do algoritmo FastICA.....	58
Figura 5.6: Esquema de realimentação para melhorar a convergência do algoritmo FastICA.	60
Figura 5.7: Gráfico do erro de convergência do algoritmo FastICA (método simétrico) para o Experimento 2, com critério de parada $\varepsilon = 1 \times 10^{-5}$	60
Figura 5.8: Primeiras 200 amostras, normalizadas, das fontes originais do experimento 2 e suas respectivas estimativas, normalizadas, usando o método simétrico e a função custo Tanh do algoritmo FastICA com estimativa inicial W_0 definida pela equação (5.8).....	61

Figura 5.9: Primeiras 200 amostras, normalizadas, das fontes do Experimento 2 e suas respectivas estimativas, normalizadas, usando o método simétrico e a função custo Tanh do algoritmo FastICA após a repetição de uma vez do algoritmo com o novo W_0 igual ao W obtido na execução anterior.....	61
Figura 5.10: Histogramas das fontes originais do Experimento 5: as 3 primeiras fontes são os sinais do filme Harry Potter, usado na Seção 1.1; as fontes 4 e 5 são do Experimento 1 e as fontes 6 e 7 são dois trechos de música do gênero rock.....	62
Figura 5.11: Configuração da sala acústica no Experimento 6.	64
Figura 5.12: Função coerência entre as estimativas das fontes do algoritmo PearsonICA no Experimento 6.....	65
Figura 5.13: Configuração da sala acústica nos Experimentos 7 e 8, à esquerda e Experimento 9, à direita.	65
Figura 5.14: Configuração da sala acústica no Experimento 10 (esquerda) e Experimento 11 (Direita)	66
Figura 5.15: Função coerência entre as fontes originais usadas nos experimentos 10 e 11.....	66
Figura 5.16: Função coerência entre as misturas e entre as estimativas das fontes feita pelo algoritmo PearsonICA no Experimento 11.	67

Lista de tabelas

Tabela 3.1: Algoritmo do gradiente usando a curtose.....	35
Tabela 3.2: Algoritmo FastICA usando curtose.....	36
Tabela 3.3: Novo passo 3 do algoritmo FastICA para múltiplas fontes.....	37
Tabela 3.4: Algoritmo do gradiente estocástico para maximização da não gaussianidade.....	39
Tabela 3.5: Algoritmo FastICA usando negentropia.....	41
Tabela 3.6: Algoritmo PearsonICA.....	44
Tabela 4.1: Algoritmo de busca de projeção pelo gradiente ascendente.....	47
Tabela 5.1: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 1.....	55
Tabela 5.2: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 2.....	56
Tabela 5.3: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 3.....	59
Tabela 5.4: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 4.....	59
Tabela 5.5: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 5.....	63

Abreviaturas

ASR	<i>Automatic Speech Recognition</i> (Reconhecimento automático de fala)
BSS	<i>Blind Source Separation</i> (Separação Cega de Fontes)
CRB	<i>Cramér-Rao Bound</i> (Limite de Cramér-Rao)
EPUSP	Escola Politécnica - Universidade de São Paulo
EVD	<i>Eigenvalue Decomposition</i> (Decomposição em autovalores)
FA	<i>Factor Anaysis</i> (Análise de Fatores)
GSO	<i>Gram-Schmidt Orthogonalization</i>
HOS	<i>Higher Order Statistics</i> (Estatísticas de Ordem Superior)
ICA	<i>Independent Component Analysis</i> (Análise em Componentes Independentes)
LPS	Laboratórios de Processamento de Sinais - Departamento de Engenharia de Sistemas Eletrônicos da USP
PCA	<i>Principal Component Analysis</i> (Análise em Componentes Principais)
<i>pdf, fdp</i>	<i>Probability Density Function</i> (Função Densidade de Probabilidade)
PP	<i>Projection Pursuit</i> (Busca de projeções)
SAR	<i>Source-to-Artifact Ratio</i> (Relação Fonte-Artefato)
SDR	<i>Source-to-Distortion Ratio</i> (Relação Fonte-Distorção)
SIR	<i>Source-to-Interference Ratio</i> (Relação Fonte-Interferência)
SNR	<i>Source-toNoise Ratio</i> (Relação Fonte-Ruído)
SONAR	<i>Sound Navigation Ranging</i>
SOS	<i>Second Order Statistics</i> (Estatística de segunda ordem)

Simbologia

A	Maiúscula para matrizes constantes
\hat{A}	Estimativa da matriz A
A^T	Transposta da matriz A
A^H	Hermitiana (conjugada transposta) da matriz A
A^{-1}	Inversa da matriz A
C_{xx}	Matriz de covariância
C_{xy}	Matriz de covariância cruzada
\hat{C}_{xx}	Estimativa da matriz de covariância
\hat{C}_{xy}	Estimativa da matriz de covariância cruzada
$\text{coh}_{xy}(j\omega)$	Função coerência
$E\{\cdot\}$	Valor esperado
$f_x(x)$	Função densidade de probabilidade da variável aleatória x
$f_{xy}(x, y)$	Função densidade de probabilidade conjunta das variáveis aleatórias x e y
$f(\cdot), g(\cdot)$	Função contínua
$H(\cdot)$	Entropia da variável aleatória
$h(\cdot)$	Entropia diferencial
$J(\cdot)$	Negentropia
$I_M(x_1, \dots, x_n)$	Informação mútua
J	Jacobiano
$I(\cdot)$	Função custo
m_x	Média da variável aleatória x
A	Matriz de mistura
\tilde{A}, W	Matriz de mistura ortogonal
$p(x)$	Probabilidade da variável aleatória x
\mathbf{s}	Vetor de fontes
s, s_i	Sinal de uma fonte
R_{xx}	Matriz de autocorrelação

R_{xy}	Matriz de correlação cruzada
\hat{R}_{xx}	Estimativa da matriz de autocorrelação
\hat{R}_{xy}	Estimativa da matriz de correlação cruzada
$S_{xx}(j\omega)$	Densidade espectral de potência
$S_{xy}(j\omega)$	Densidade espectral cruzada de potência
u	Minúscula itálica para variáveis aleatórias
\mathbf{u}	Minúscula itálica e negrito para vetor de variável aleatória
\mathbf{v}	Vetor de ruído
x, x_i	Sinal de uma mistura
\mathbf{x}	Vetor de misturas
\mathbf{x}^T	Transposto do vetor \mathbf{x}
$\hat{\mathbf{y}}$	Vetor da estimativa das fontes
\mathbf{W}	Estimativa da matriz de mistura ($\hat{\mathbf{A}}$)
\mathbf{z}	Vetor de misturas branqueado
μ'_k	Momento de ordem k
μ_k	Momento central de ordem k
$\boldsymbol{\mu}_k$	Vetor de momentos centrais de ordem k
κ_k	Cumulante de ordem k
$\hat{\mu}_k$	Estimativa do momento de ordem k , ou momento amostral de ordem k
$\delta(p_x, p_y)$	Divergência de <i>Kullback-Leibler</i>
$\frac{\partial g(x)}{\partial x}$	Gradiente de $g(x)$ em relação à x

Sumário

Resumo	i
Abstract	ii
Lista de Figuras	iii
Lista de tabelas	v
Abreviaturas	vi
Simbologia	vii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos.....	7
1.3 Metodologia.....	7
1.4 Estrutura do trabalho	8
2 Conceitos para separação cega de sinais	10
2.1 Introdução à separação de sinais	10
2.2 Momentos e cumulantes	11
2.3 Independência Estatística e Correlação	16
2.3.1 Entropia e Informação mútua	16
2.4 Gaussianidade (Normalidade)	18
2.4.1 Curtose.....	20
2.4.2 Negentropia	21
2.5 Fontes gaussianas	24
3 Métodos para Separação de Fontes	26
3.1 Pré-Processamento	26
3.1.1 PCA - Análise em Componentes Principais	29
3.2 Algoritmo de Héroult Jutten	31
3.3 ICA pela curtose	32
3.3.1 Algoritmo de ponto fixo usando a curtose.....	35
3.4 ICA pela negentropia.....	38
3.4.1 Algoritmo rápido de ponto fixo usando negentropia.....	40
3.5 PearsonICA - Curvas de Pearson	42
4 Busca de projeções (<i>Projection Pursuit</i>)	45
4.1 Introdução à busca de projeções.....	45

4.2	Separação de fontes por meio de busca de projeção	46
5	Experimentos e análise dos algoritmos	50
5.1	Método de medição	50
5.2	Experimentos	52
5.2.1	Visão geral	52
5.2.2	Misturas instantâneas	54
5.2.3	Misturas convolutivas	63
6	Conclusões	68
6.1	Principais contribuições	68
6.2	Sugestões para pesquisa futura	69
	Apêndices	70
A.	Limite de Cramér-Rao	70
B.	Coeficiente de correlação e função coerência	71
C.	Conteúdo do CD-ROM	72
	Referências	73

1 Introdução

Em muitas situações práticas é necessário o uso de sensores para coletar informações, nas quais geralmente há o problema de os sinais fornecidos pelos sensores serem misturas dos sinais (fontes) desejados. Além disso, em geral, não há como observar as fontes diretamente, nem é conhecida a forma como as fontes foram misturadas. Esse problema é conhecido como o problema da Separação Cega de Fontes (BSS – *Blind Source Separation*). O termo separação cega de fontes se deve ao fato de não haver, ou haver com escassez, conhecimento com relação às fontes e aos sinais misturados. Por esse fato, a BSS tem sido alvo de muitos trabalhos em diversas áreas de conhecimento como, por exemplo, bioengenharia, telecomunicações, economia, processamento de sinais e imagens, entre outras.

Podem-se encontrar duas situações bem distintas no problema da BSS. A primeira situação corresponde àquela na qual se tem misturas instantâneas, formadas por uma combinação linear das fontes e, na segunda situação, há misturas convolutivas, na qual cada fonte aparece convoluída com a resposta impulsiva do trajeto que intermedia a sua origem e o ponto de mistura.

Um dos métodos mais difundidos para BSS é a Análise em Componentes Independentes (ICA). O presente trabalho concentra-se no estudo de algumas das principais técnicas ICA para misturas instantâneas e na sua aplicação em separação de sinais de áudio.

A ICA é um método usado para encontrar fontes a partir de um conjunto de dados multidimensional, e é caracterizada por procurar fontes que são estatisticamente independentes e não-gaussianas. Neste capítulo é feita uma breve introdução ao problema de separação cega de fontes, sendo também apresentada a estrutura do trabalho.

1.1 Motivação

Uma técnica inovadora para separação cega de fontes foi proposta, em 1985, por J. Héroult, C. Jutten, e B. Ans [67]. Essa técnica foi denominada por eles como análise em componentes independentes (inicialmente INCA, a qual foi muito pouco utilizada, e posteriormente ICA - *Independent Component Analysis*), devido à sua similaridade com a técnica de análise em componentes principais (PCA – *Principal Component Analysis*). A proposta deles era aplicar a ICA na separação de sinais neurológicos, na qual as únicas

considerações assumidas foram a de independência estatística e a não-gaussianidade das fontes [1, 2, 4, 10].

Atualmente, a ICA é um dos métodos mais aplicados à separação de fontes. Geralmente baseadas nos mesmos princípios, as diferentes metodologias ICA existentes podem ser usadas para resolver o problema da BSS.

Um problema bem conhecido de separação cega de fontes é o *Cocktail Party*, no qual, por exemplo, várias pessoas estão em uma sala conversando ao mesmo tempo e uma música toca ao fundo. O problema consiste, então, em aplicar um método de separação para encontrar o sinal de voz de cada uma das pessoas e da música, a partir dos sinais misturados obtidos pelos sensores.

O problema *Cocktail Party* foi, primeiramente, mencionado em 1953 por Cherry [3], no contexto da capacidade humana de reconhecimento de mensagens de voz. Desde então, muitos autores têm buscado novas formas de representar esse problema [8, 5, 6].

Inicialmente, a análise em componentes independentes foi criada para lidar com problemas semelhantes ao problema do *Cocktail Party*. Entretanto, existem vários estudos em diversas áreas do conhecimento com aplicações muito interessantes como, por exemplo, telecomunicações [73, 92-97], biomedicina [85-91, 100], processamento de voz [99, 105, 106], processamento de áudio e imagem [76, 102-110, 100] e análises financeiras [113, 115].

Para o problema *Cocktail Party*, considere uma situação em que há três pessoas em uma sala conversando simultaneamente, existindo também três microfones em locais distintos, captando três misturas diferentes das fontes. Sejam $x_1(t)$, $x_2(t)$ e $x_3(t)$ os sinais das misturas gravados pelos microfones em função do tempo e $s_1(t)$, $s_2(t)$ e $s_3(t)$ os sinais das fontes. No caso de misturas instantâneas, matematicamente as misturas são combinações lineares das fontes e podem ser representadas pelo sistema de equações (1.1).

$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)\end{aligned}\tag{1.1}$$

Ou, na forma matricial, como na equação (1.2).

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s}\tag{1.2}$$

em que, para o caso geral, \mathbf{x} é o vetor das misturas, de dimensão m , \mathbf{A} é uma matriz $m \times n$ a ser estimada das constantes de mistura, cuja dimensão n representa o número de fontes e m o número de sensores, e \mathbf{s} é o vetor das fontes independentes, de dimensão n , a serem estimadas.

A matriz de misturas depende da disposição geográfica dos sensores em relação às fontes e do ganho de cada sensor (microfone). No contexto de separação de fontes os valores observados de \mathbf{x} correspondem à sequência de realizações dos sinais de tempo discreto $x_i(t)$, para $t=1,2,\dots$, logo, as componentes $s_i(t)$ são chamadas de sinais das fontes. É considerada somente a estrutura espacial dos dados, ou seja, todas as informações relevantes estão presentes nas distribuições das fontes, que são tratadas como sinais estacionários. Porém, existem outros métodos ICA que se baseiam em alguma suposição referente à estrutura temporal dos sinais, por exemplo, autocovariância [120]. Esses modelos não serão tratados no presente trabalho.

Pode-se observar que, se a matriz de mistura fosse conhecida, o problema se resumiria à solução do sistema de equações. No entanto, os valores da matriz devem ser estimados; o que torna o problema mais complicado e instigante.

Considerando o mesmo número de fontes e sensores, o problema ICA pode ser proposto da seguinte forma: deseja-se encontrar uma estimativa, $\hat{\mathbf{A}}$ $n \times n$, para a matriz de misturas \mathbf{A} e, conseqüentemente, encontrar $\hat{\mathbf{s}}$, usando a matriz inversa da estimativa da matriz de misturas, $\hat{\mathbf{A}}^{-1}$.

Entretanto, se o número de sensores for menor que o número de fontes, ter-se-ia um sistema de equações incompleto e, portanto, haveria perdas na estimativa das fontes ou até mesmo não seria possível a estimativa. Nesse caso, é possível estimar as fontes usando a pseudo-inversa da estimativa da matriz de misturas. Porém, muitas vezes, são necessárias técnicas mais sofisticadas de estimativa [12]. Para o caso em que número de sensores é maior que o número de fontes tem-se um sistema com mais equações do que incógnitas e a solução é trivial.

O ruído será tratado como sendo uma das fontes no modelo ICA. No entanto, o modelo ICA também pode ser representado com um ruído aditivo, como mostra a equação (1.3).

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} + \mathbf{v} \quad (1.3)$$

em que \mathbf{v} é o ruído aditivo. Porém, nesse caso, o problema se torna mais complicado e esse modelo também não será tratado no presente trabalho.

Veja, por exemplo, a Figura 1.1, que representa o problema do *Cocktail Party* usando três fontes e três sensores, gerando três misturas.

No seguinte exemplo ilustrativo, considere as falas do filme *Harry Potter and the Sorcerer's Stone*. Todas as falas são truncadas, com duração de quatro segundos [128]:

- s_1 - “There will be no foolish wand waving or silly incantations” - [Professor Severus Snape]
- s_2 - “Clearly fame isn't”- [Professor Severus Snape]
- s_3 - “I think we had a bad influence on her.” - [Ron Weasley]

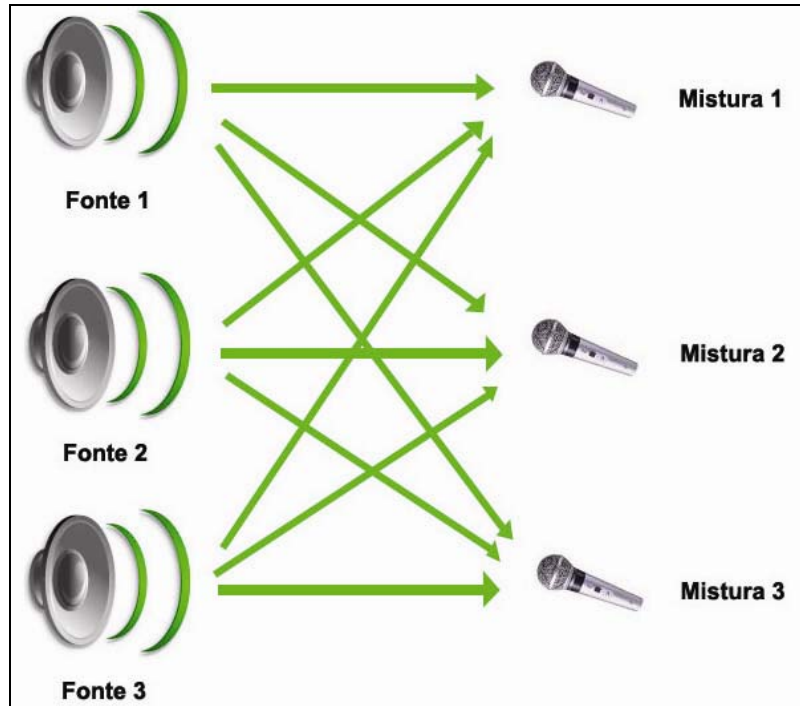


Figura 1.1 Esquema do Problema *Cocktail Party*.

Observe na Figura 1.2 os três sinais das falas s_1 , s_2 e s_3 respectivamente. Na Figura 1.3 estão representadas as misturas obtidas a partir das fontes da Figura 1.2, usando a matriz de misturas, gerada aleatoriamente no Matlab, na equação (1.4).

$$\begin{bmatrix} 1.40 & 0.63 & -0.91 \\ 1.68 & -0.19 & -1 \\ -0.15 & 0.06 & 1.12 \end{bmatrix} \quad (1.4)$$

Os sinais estimados das fontes representados na Figura 1.4 foram obtidos a partir dos sinais misturados da Figura 1.3, e da estimativa da matriz inversa de misturas, \hat{A}^{-1} . O algoritmo usado na estimativa foi o FastICA, que será detalhado nas seções 3.3.1 e 3.4.1.

É possível observar que os sinais estimados podem estar permutados em relação à ordem dos sinais originais, Além disso, a sua energia também poderá ser diferente; porém o sinal audível será muito próximo do original.

Dessa forma, o problema da ICA propõe determinar \hat{A} , a menos de um fator de escala e de permutação que satisfaça o modelo da equação (1.2). Então, a partir de sua inversa, determinar as fontes, assumindo que exista a inversa, como mostra a equação (1.5).

$$\hat{\mathbf{s}} = \mathbf{A}^{-1}\mathbf{x} \quad (1.5)$$

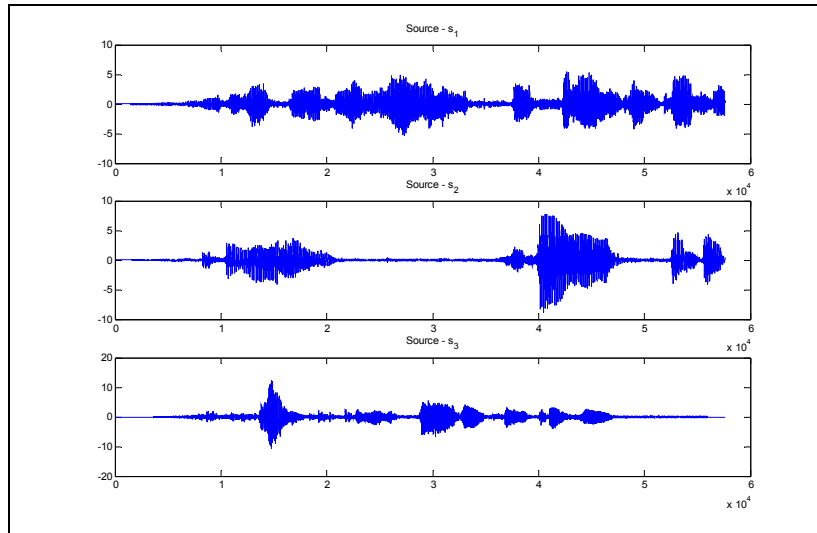


Figura 1.2: Sinais originais das falas (fontes) do filme Harry Potter.

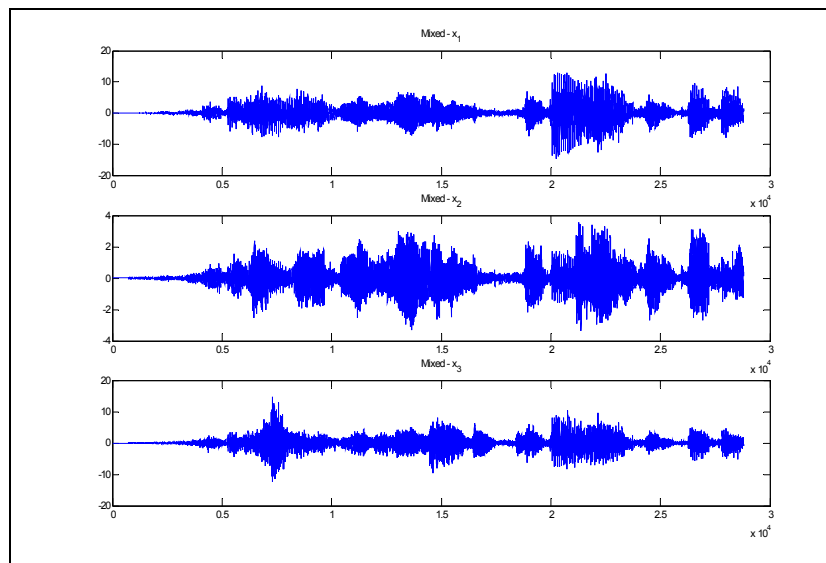


Figura 1.3: Sinais das misturas das falas do filme Harry Potter.

Na prática, \hat{A}^{-1} é estimada diretamente e, no presente trabalho, a estimativa da matriz de mistura será denominada simplesmente por W . Para simplificar os cálculos, consideraremos que todas as variáveis têm média nula e variância unitária, exceto quando especificado. Além disso, deve-se notar que W , ou sua inversa, geralmente não pode ser obtida diretamente em função do conjunto de dados, devendo ser calculada a partir de métodos baseados em função custo, que também é conhecida por função contraste ou

objetivo. Dessa forma, W pode ser obtida a partir da maximização ou minimização dessa função custo. A Figura 1.5 mostra um diagrama esquemático do problema de separação cega de fontes. Cabe salientar que a matriz de separação é obtida diretamente não havendo necessidade de a matriz de mistura ser quadrada. Será considerado somente o caso em que o número de sensores é igual ou superior ao número de fontes.

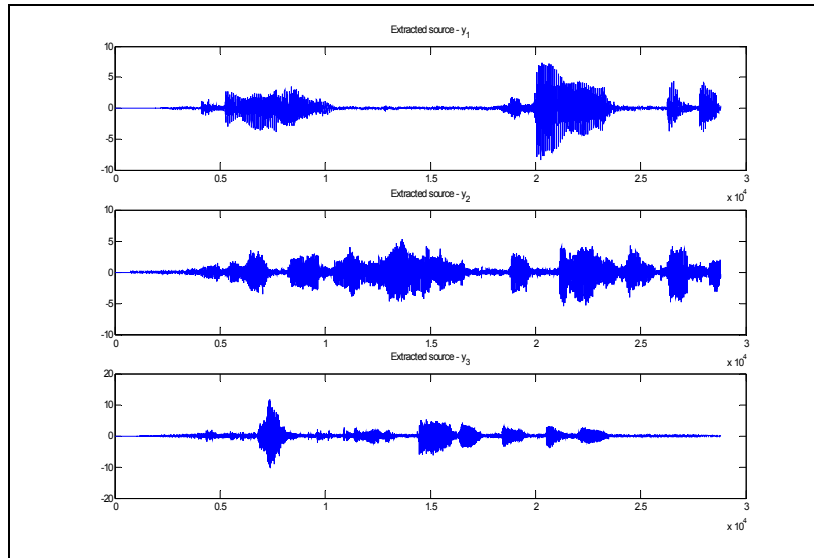


Figura 1.4: Sinais estimados das falas do filme Harry Potter.

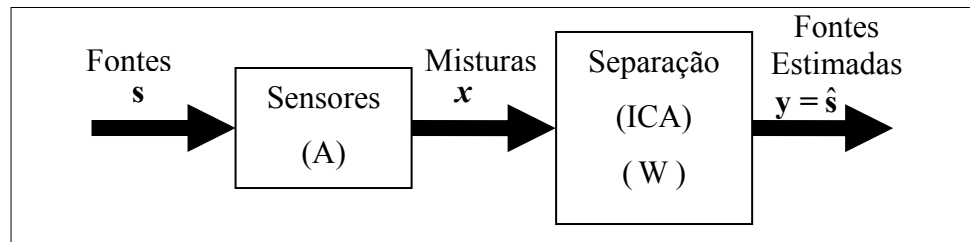


Figura 1.5: Diagrama esquemático do problema de separação cega de fontes

Recentemente, foram realizadas comparações entre alguns algoritmos ICA e BSS. Por exemplo, [73] compara oito métodos de separação de fontes aplicados na separação de sinais de rádio-comunicação. Em [75], é feita uma comparação entre 5 métodos para ICA usando três conjuntos de dados distintos, com objetivo de avaliar tanto a separação de fontes, quanto de encontrar projeções de interesse para propósito de visualização.

Vincent *et al* [76] apresentam uma discussão a respeito da análise de desempenho de algoritmos de separação de sinais de áudio. No presente trabalho, é feita a comparação de três algoritmos de separação cega de fontes, aplicados à separação de sinais de áudio.

1.2 Objetivos

O principal objetivo desse trabalho é fazer uma comparação de diversos métodos ICA para misturas instantâneas e aplicá-los à separação cega de sinais de áudio, considerando todos os aspectos importantes da análise estatística, intrínseca aos métodos existentes.

Sabe-se que existem duas situações bem distintas no problema da separação de fontes, a de misturas instantâneas e a de misturas convolutivas (do “mundo real”). Alguns artigos que tratam da aplicação da ICA em misturas instantâneas sugerem aplicar os algoritmos em sinais de áudio. Para isso o presente trabalho deseja comparar e analisar alguns desses algoritmos aplicados à separação de sinais de áudio em ambos os casos.

Serão estudadas as aplicações de alguns dos principais algoritmos (descritos nos capítulos 3 e 4), na separação de misturas instantâneas e o comportamento desses algoritmos na separação de sinais convolutivos, obtidos sob certas condições controladas em laboratório.

Também será avaliado o método ICA baseado na maximização da curtose, o qual faz uso de estatística de ordem superior, e o método ICA baseado na negentropia, que é derivado do conceito de informação mútua. O método da Busca de Projeção, o qual é baseado na maximização de um índice de interesse, é proposto para separação de múltiplas fontes. Nesse caso, o índice de interesse pode ser, por exemplo, a curtose ou a negentropia. Outro método a ser avaliado é um método baseado na estimativa da distribuição da fonte com o auxílio das famílias de curvas de Pearson.

O presente trabalho visa a apresentar análises experimentais para prover comparações entre os métodos, além de propiciar a verificação da aplicabilidade de cada método em casos particulares e suas limitações, bem como de avaliar o potencial de aplicação da análise em componentes independentes em situações práticas.

1.3 Metodologia

Os modelos apresentados no trabalho são validados por meio de experimentos práticos com o auxílio do software Matlab da MathWorks, usando diversos sinais de áudio ou sintéticos. Os experimentos estão divididos em duas etapas, detalhadas a seguir.

A primeira etapa é realizada por meio de sinais gravados separadamente como, por exemplo, trechos musicais, trechos de sinais de fala, sinais aleatórios ou gerados por alguma função para forçar situações particulares ou extremas. Esses sinais são misturados, aplicando-

se uma matriz de mistura gerada aleatoriamente, visando identificar a influência de diversos fatores no desempenho dos algoritmos ICA como, por exemplo, a distribuição das fontes.

Na segunda etapa, realizada por meio de simulações práticas, são obtidos os sinais já misturados de sensores (microfones) em uma simulação do problema *Cocktail Party*, realizado no Laboratório de Processamento de Sinais e Acústica Prof. Normonds Alens do LPS, na EPUSP, permitindo, dessa forma, que seja avaliado o comportamento dos algoritmos em situações reais, com misturas feitas por sensores reais distintos.

Foi possível realizar a comparação entre os sinais estimados e os sinais originais, permitindo avaliar a qualidade da separação, quando as fontes originais estavam disponíveis. Por exemplo, a relação Sinal-Interferência (SIR), apresentada na Seção 5.1, pode ser usada para fazer essa comparação. Foram usados gráficos e índices, sempre que possível e necessário, com objetivo de elucidar os experimentos para melhor visualização e análise dos resultados. Todos os modelos apresentados no presente trabalho estão tratados no campo dos números reais e sem a adição de ruídos, com exceção dos problemas na qual o ruído também seja considerado uma fonte. Somente os casos cujo número de sensores é igual ou superior ao número de fontes são estudados.

As dimensões reais da sala acústica (isolada) em metros são 2,60x2,95x2,95 (PxLxA). Além disso, a atenuação média da sala é 30dB e o tempo de reverberação (RT60) está próximo a 0,6s. Os equipamentos usados no laboratório foram: par de caixas Roland DS 5, interface de captura de áudio de 24 bits (firewire) FA101-Edirol (Frequência de amostragem usada: 44.1 kHz), microfone dinâmico AKG D880 M, microfone tipo condensador M-Audio Nova, microfone dinâmico Shure SM58 e software para gravação Cakewalk SONAR.

1.4 Estrutura do trabalho

O presente capítulo visa a apresentar uma breve introdução ao assunto, bem como a motivação para realização desse estudo. Além disso, são apresentados o escopo e a metodologia utilizada no desenvolvimento da pesquisa.

O Capítulo 2 traz uma breve introdução teórica ao tema de separação cega de sinais, tratando de aspectos importantes para sua análise como independência estatística e normalidade. É apresentada também uma seção sobre momentos, pois muitos algoritmos ICA são baseados em estatísticas de ordem superior. *A priori*, como nenhuma informação sobre a distribuição das fontes é conhecida, estimativas de momentos e cumulantes podem ser uma

das formas de caracterizarmos, de forma amostral, a sua distribuição. Dessa forma, as seções do Capítulo 2, com as quais o leitor já esteja familiarizado, podem ser dispensadas.

Primeiramente, no Capítulo 3, é apresentada uma técnica de pré-processamento de sinais, muito utilizada em métodos de separação de fontes. Então, é apresentado o algoritmo de Héroult-Jutten, que é um dos precursores da ICA; em seguida são apresentados os métodos ICA, baseados na curtose e na negentropia, bem como seus respectivos algoritmos baseado em iterações de ponto fixo, FastICA, o último algoritmo apresentado nesse capítulo, PearsonICA, baseia-se na estimativa da distribuição das fontes, com base nas curvas de Pearson. No Capítulo 4 é proposto o algoritmo PP (*Projection Pursuit*) para separação de múltiplas fontes, baseado na técnica de busca de projeção (*Projection Pursuit*) e sua relação com a ICA.

Os experimentos que demonstram e validam as previsões teóricas são apresentados no Capítulo 5.

Finalmente, no Capítulo 6 são apresentadas as conclusões e as sugestões de trabalho futuro com base nos resultados da pesquisa.

2 Conceitos para separação cega de sinais

Este capítulo é dedicado a uma revisão de alguns dos conceitos mais importantes para compreensão da análise em componentes independentes. Portanto, caso o leitor já esteja familiarizado com os conceitos presentes nas seções seguintes, poderá dispensá-los sem prejuízo para o entendimento dos algoritmos apresentados no capítulo 3.

Na Seção 2.1 é apresentada uma introdução à separação de sinais, na qual são expostos alguns dos conceitos mais importantes dentro desse contexto. A Seção 2.2 introduz os conceitos básicos de momentos e cumulantes, pois muitos algoritmos ICA são baseados em estatísticas de ordem superior.

Em seguida é introduzido o conceito de independência estatística e gaussianidade (normalidade), nas Seções 2.3 e 2.4 respectivamente, que são condições necessárias e suficientes para resolver o problema da separação cega de sinais.

Por fim, na Seção 2.5 é apresentado o comportamento de fontes gaussianas no modelo ICA, deixando clara a razão das fontes não poderem ser gaussianas.

2.1 Introdução à separação de sinais

A separação cega de sinais, ou BSS, consiste basicamente em recuperar sinais ocultos, ou simplesmente fontes da observação de alguns sinais misturados. Tipicamente, são obtidos de um conjunto de sensores, nos quais, cada um representa uma combinação diferente das fontes. ICA é um método específico aplicado em BSS, e nesse estudo serão consideradas várias técnicas ICA aplicadas à BSS, nas quais serão considerados seus aspectos estatísticos.

Quando dois sinais são misturados podem-se considerar alguns aspectos, os quais podem ser usados nos algoritmos para auxiliar na separação das fontes. Pode-se considerar como os mais importantes, independência e gaussianidade (normalidade). Também existem algoritmos que são baseados na complexidade de Kolmogorov do sinal [13, 15]. A complexidade de Kolmogorov é uma teoria da informação e da aleatoriedade, que trata da quantidade de informação de objetos individuais; refere-se a um conceito pontual de aleatoriedade, ao invés de uma aleatoriedade média como o faz a teoria das probabilidades [137].

Esses aspectos são suficientes para estabelecermos os princípios básicos para recuperação de fontes, a partir de um conjunto de misturas. Em cada aplicação, a linha de raciocínio será a mesma, ou seja, os sinais extraídos das misturas são mais independentes,

estatisticamente, que suas respectivas fontes e são mais distantes da gaussiana que suas misturas (teorema central do limite). Além disso, a complexidade de Kolmogorov de uma mistura sempre será maior que a complexidade de seus sinais constituintes. Dessa maneira, esses aspectos podem ser usados para estimarmos os sinais das fontes [13].

De fato, somente assumindo a condição de independência das fontes, o problema é bem definido e o modelo representado na equação (1.2) pode ser estimado se, e somente se, existir no máximo uma fonte gaussiana, como é demonstrado em [1]. Na Seção 2.5 é ilustrado por que não é possível separar mais de uma fonte gaussiana.

Muitos algoritmos são baseados em estatísticas de ordem superior, que podem ser usadas para caracterizar a distribuição estatística das fontes e, de certa forma, podem ser usadas para verificar a gaussianidade das fontes. Na Seção 2.2 são apresentados os conceitos de momentos e cumulantes, que são comumente utilizados na caracterização da distribuição.

A condição de independência é mais restritiva que a condição de correlação. Uma forma de verificar a dependência de variáveis aleatórias é a informação mútua, a qual será apresentada na Seção 2.3.1.

2.2 Momentos e cumulantes

Como nenhuma informação é conhecida acerca das distribuições das fontes ou da matriz de mistura, é possível usar estatísticas baseadas em momentos para caracterizar as distribuições dos sinais estimados. Nesta seção estão apresentados alguns conceitos sobre momentos, que são muito importantes em ICA.

A função densidade de probabilidade (fdp - *pdf* - *probability density function*) de uma variável aleatória pode ser caracterizada em termos de seus momentos. O k -ésimo momento da variável aleatória x com fdp $f_x(x)$ é definido por:

$$\mu'_k = E\{x^k\} = \int_{-\infty}^{\infty} f_x(x)x^k dx \quad (2.1)$$

Vale salientar que existem distribuições com todos os seus momentos possuindo valor infinito e, além disso, há o inconveniente de, mesmo conhecendo os momentos, não ser possível caracterizar a fdp unicamente. Porém, felizmente na prática, para a maioria das distribuições, todos os momentos possuem valor finito e o conhecimento deles é equivalente a conhecer a fdp, conforme Hyvärinen *et al* [12] *apud* Nandi [29].

Os momentos mais importantes são os de primeira ao de quarta ordem. O primeiro momento, μ'_1 , é também denominado de média e aqui será representado por \bar{x} . Pode-se também definir o conceito de momento central como sendo:

$$\mu_k = E\{(x - \bar{x})^k\} = \int_{-\infty}^{\infty} f_x(x)(x - \bar{x})^k dx \quad (2.2)$$

Mostra-se também que [27]

$$\mu_k = \sum_{r=0}^k \binom{k}{r} (-\bar{x})^{k-r} \mu'_r \quad (2.3)$$

O segundo momento central é também denominado de variância e sua raiz quadrada é o desvio padrão, chamado de σ , ou seja, $\mu_2 = \sigma^2$. É fácil provar que $E\{x^2\} = \sigma^2 + \bar{x}^2$.

O momento central de terceira ordem, definido na equação (2.4), também é conhecido como *skewness* e pode ser usado como uma medida de assimetria.

$$\mu_3 = E\{(x - \bar{x})^3\} = \int_{-\infty}^{\infty} f_x(x)(x - \bar{x})^3 dx \quad (2.4)$$

Logo, o momento de quarta ordem e o momento central de quarta ordem são $E\{x^4\}$ e $E\{(x - \bar{x})^4\}$, respectivamente. No entanto, a estatística de quarta ordem, denominada curtose, é utilizada em substituição ao momento central de quarta ordem, devido à facilidade proporcionada nos cálculos. A seguir, será apresentado o conceito de cumulantes, no qual a curtose é baseada. A Seção 2.4.1 é dedicada à curtose e suas propriedades; importantes na análise em componentes principais.

Analogamente, para um vetor aleatório \mathbf{x} de dimensão n , tem-se:

$$\boldsymbol{\mu}'_k = E\{\mathbf{x}^k\} = \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x})\mathbf{x}^k d\mathbf{x} \quad (2.5)$$

Para $k=1$ o vetor média $m_{\mathbf{x}} = \boldsymbol{\mu}'_1$, na qual cada elemento será dado por (2.6), e $f_{x_i}(x_i)$ é a densidade marginal do i -ésimo elemento x_i do vetor \mathbf{x} .

$$m_{x_i} = E\{x_i\} = \int_{-\infty}^{\infty} f_{x_i}(x_i)x_i dx_i \quad (2.6)$$

Freqüentemente, a função densidade de probabilidade é desconhecida. Como afirmamos anteriormente, podemos usar os momentos para caracterizar a distribuição. Nesse caso, devemos fazer uma estimativa dos valores esperados dadas as amostras, que são as únicas informações disponíveis, denominada de momento amostral. Seja, então, o conjunto de N amostras (x_1, \dots, x_N) de uma variável aleatória x . A estimativa do valor esperado de x será dada por:

$$\hat{\mu}'_k = \frac{1}{N} \sum_{j=1}^N x_j^k \quad (2.7)$$

Similarmente,

$$\hat{\mu}_k = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^k \quad (2.8)$$

Outro conjunto muito importante de momentos são as correlações entre os pares de elementos do vetor \mathbf{x} , dado pelo momento de segunda ordem definido em (2.9).

$$r_{ij} = E\{x_i x_j\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{x_i, x_j}(x_i, x_j) x_i x_j dx_j dx_i \quad (2.9)$$

Convém notar ser possível expressar as correlações na forma matricial, que é dada pela matriz de autocorrelação definida na equação (2.10). A matriz de autocorrelação possui várias propriedades importantes: simétrica e positiva semi-definida.

$$\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\} \quad (2.10)$$

O momento central correspondente à matriz de autocorrelação é chamado de autocovariância e é definido pela matriz de autocovariância, como na equação (2.11).

$$\mathbf{C}_{xx} = E\{(\mathbf{x} - m_x)(\mathbf{x} - m_x)^T\} \quad (2.11)$$

É fácil verificar que

$$\mathbf{R}_{xx} = \mathbf{C}_{xx} + m_x m_x^T \quad (2.12)$$

Observe que, se o vetor média $m_x = \mathbf{0}$, for nulo, as matrizes de autocorrelação e autocovariância serão as mesmas. Para tornar o vetor média nulo, basta subtraí-lo do vetor de dados original. Como será visto mais adiante, essa é uma prática conveniente nos passos de pré-processamento.

Agora, estendendo ao caso de dois vetores aleatórios \mathbf{x} e \mathbf{y} , em termos de sua densidade conjunta, obtém-se a equação (2.13).

$$\mu'_{xy} = E\{\mathbf{x}\mathbf{y}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x}\mathbf{y} f_{xy}(\mathbf{x}, \mathbf{y}) d\mathbf{x}d\mathbf{y} \quad (2.13)$$

em que $f_{xy}(\mathbf{x}, \mathbf{y})$ é a densidade conjunta \mathbf{x} e \mathbf{y} . As integrais são calculadas sobre todos os elementos de \mathbf{x} e \mathbf{y} .

Esses resultados podem ser facilmente estendidos para o caso de r variáveis aleatórias.

Os valores esperados conjuntos de duas variáveis mais usados são: a matriz de correlação cruzada de \mathbf{x} e \mathbf{y} e a matriz de covariância cruzada de \mathbf{x} e \mathbf{y} .

$$\mathbf{R}_{xy} = E\{\mathbf{x}\mathbf{y}^T\} \quad (2.14)$$

$$C_{xy} = E\{(\mathbf{x} - m_x)(\mathbf{y} - m_y)^T\} \quad (2.15)$$

Nesse caso, convém notar também que as dimensões dos vetores \mathbf{x} e \mathbf{y} podem ser diferentes. Então, as matrizes de correlação cruzada e covariância cruzada não são necessariamente quadradas, e não são simétricas, em geral.

Se forem conhecidos somente os pares de amostras $(x_1, y_1), \dots, (x_N, y_N)$ dos vetores aleatórios \mathbf{x} e \mathbf{y} , em vez da sua densidade conjunta de probabilidade $f_{xy}(x, y)$, pode-se estimar a correlação de \mathbf{x} e \mathbf{y} pela equação (2.16).

$$\hat{R}_{xy} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{y}_j^T \quad (2.16)$$

Similarmente,

$$\hat{C}_{xy} = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}_j - m_x)(\mathbf{y}_j - m_y)^T \quad (2.17)$$

Ou utilizando as médias amostrais

$$\hat{C}_{xy} = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}_j - \hat{m}_x)(\mathbf{y}_j - \hat{m}_y)^T \quad (2.18)$$

Facilmente, obtêm-se as expressões de \hat{R}_{xx} e \hat{C}_{xx} .

Dessa forma, os momentos amostrais constituem um conjunto de constantes descritivas de uma distribuição, muito útil na determinação de suas propriedades e, em certas circunstâncias, para especificá-la. No entanto, esse não é o único e nem o melhor conjunto de constantes para esse propósito. Outro conjunto de constantes baseadas em cumulantes possui propriedades que são muito úteis do ponto de vista teórico.

Seja x uma variável aleatória real contínua de média nula e fdp $f_x(x)$. Define-se a primeira função característica de x , $\varphi(\omega)$, como sendo a transformação de Fourier da fdp $f_x(x)$ [22].

$$\varphi(\omega) = E\{e^{j\omega x}\} = \int_{-\infty}^{\infty} e^{j\omega x} f_x(x) dx \quad (2.19)$$

Na qual $j = \sqrt{-1}$ e ω é a variável transformada correspondente a x . Substituindo $j\omega$ por s , obtém-se:

$$\Phi(s) = \int_{-\infty}^{\infty} e^{sx} f_x(x) dx \quad (2.20)$$

A expansão de $\Phi(s)$ em série de Taylor resulta em:

$$\Phi(s) = \sum_{k=0}^{\infty} \frac{\mu'_k s^k}{k!} \quad (2.21)$$

Os cumulantes $\kappa_1, \kappa_2, \dots, \kappa_r$ são definidos, formalmente, pela identidade em s dada pela equação (2.22) [27].

$$e^{\sum_{k=1}^{\infty} \frac{\kappa_k s^k}{k!}} = \sum_{k=0}^{\infty} \frac{\mu'_k s^k}{k!} \quad (2.22)$$

Deve-se notar que não existe κ_0 . Das equações (2.22) e (2.21) pode-se escrever a equação (2.23).

$$e^{\sum_{k=1}^{\infty} \frac{\kappa_k s^k}{k!}} = \Phi(s) \quad (2.23)$$

Desse modo, enquanto o k -ésimo momento μ'_k é o coeficiente de $\frac{s^k}{k!}$ em $\Phi(s)$, o k -ésimo cumulante κ_k é o coeficiente de $\frac{s^k}{k!}$ em $\ln(\Phi(s))$, se existir a expansão em série de Taylor. A função $\Psi(s) = \ln(\Phi(s))$ é denominada de função característica de segunda ordem de x ou função geradora de cumulantes de x [27].

O k -ésimo cumulante pode ser obtido pela derivada da função característica de segunda ordem de x tomada na origem, conforme representado na equação (2.24)

$$\kappa_k = \left. \frac{d^k \Psi(s)}{ds^k} \right|_{s=0} \quad (2.24)$$

Assim, após considerável esforço, podem-se obter as relações entre momentos e cumulantes, representadas pelas equações (2.25). O cumulante de quarta ordem é também conhecido como curtose [27].

$$\begin{aligned} \kappa_1 &= \mu'_1 \\ \kappa_2 &= \mu'_2 - (\mu'_1)^2 \\ \kappa_3 &= \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3 \\ \kappa_4 &= \mu'_4 - 3(\mu'_2)^2 - 4\mu'_3\mu'_1 + 12\mu'_2(\mu'_1)^2 - 6(\mu'_1)^4 \end{aligned} \quad (2.25)$$

Para média nula, as equações (2.25) resultam respectivamente nas seguintes equações (2.26):

$$\begin{aligned}
\kappa_1 &= 0 \\
\kappa_2 &= \mu_2 \\
\kappa_3 &= \mu_3 \\
\kappa_4 &= \mu_4 - 3(\mu_2)^2
\end{aligned} \tag{2.26}$$

2.3 Independência Estatística e Correlação

Duas variáveis aleatórias x e y têm fdp's $f_x(x)$ e $f_y(y)$, respectivamente, as quais são também denominadas fdp's marginais da fdp conjunta $f_{xy}(x, y)$. Elas são independentes se, e somente se, a identidade (2.27) for satisfeita para todos os valores de x e y .

$$f_{xy}(x, y) = f_x(x) \cdot f_y(y) \tag{2.27}$$

Se duas variáveis são independentes, é possível obter a sua distribuição conjunta simplesmente pela multiplicação de suas fdp's marginais, como sugere a equação (2.27). Essa condição implica que a informação contida em uma variável não fornece nenhuma informação sobre a outra variável. Pode-se facilmente expandir esse resultado para n variáveis aleatórias. Consulte [22] para mais referências.

Se duas variáveis x e y forem independentes é fácil provar que isso implica em:

$$E\{x^k y^r\} = E\{x^k\}E\{y^r\} \tag{2.28}$$

O momento cruzado de segunda ordem da fdp conjunta de \mathbf{x} e \mathbf{y} , $E\{\mathbf{xy}^T\}$ é também conhecido como correlação cruzada de \mathbf{x} e \mathbf{y} . Note que se as variáveis não são correlacionadas, então é suficiente que $R_{xy} = E\{\mathbf{xy}^T\} = E\{\mathbf{x}\}E\{\mathbf{y}^T\}$, ou seja, $C_{xy} = E\left\{(\mathbf{x} - m_x)(\mathbf{y} - m_y)^T\right\} = 0$. Em contrapartida, para que as variáveis sejam independentes, é necessário a equação (2.28) ser satisfeita para quaisquer valores positivos de k e r . Portanto, a independência é uma propriedade muito mais restritiva que a correlação. Por exemplo, se duas variáveis são independentes, elas serão, necessariamente, não correlacionadas; porém se duas variáveis são não correlacionadas, nada se pode afirmar quanto sua independência.

2.3.1 Entropia e Informação mútua

Informação mútua é uma medida da informação que um membro de um conjunto de variáveis aleatórias contém em relação às outras variáveis do conjunto, ou seja, fornece uma indicação da dependência entre as variáveis. Com isso, observa-se que tal medida de

informação é uma importante candidata à função custo para o problema de otimização do modelo ICA. Pode-se analisar a Informação mútua fazendo uso da definição de entropia de uma variável aleatória ou interpretá-la como distâncias.

No presente trabalho, apenas uma breve introdução no assunto será abordada, que ajudará no desenvolvimento do tema e na interpretação dos resultados. Em [12], um capítulo é dedicado ao assunto e, em [17], é possível obter maiores detalhes, além de outras referências.

Entropia é uma medida da incerteza de uma variável aleatória, ou seja, é a medida da quantidade de informação que pode ser codificada por uma variável aleatória. Esse é um conceito básico da teoria da informação, definido pela equação (2.29).

$$H(x) = -\sum_i p(x_i) \log_2 p(x_i) \quad (2.29)$$

em que $p(x_i)$ representa a probabilidade de cada valor possível de x , o logaritmo é na base 2 e a entropia é medida em bits. No contexto de comunicações, a entropia foi primeiramente definida por Shannon em [47]. Para exemplos práticos, veja também [30].

Informação mútua pode ser interpretada tomando a entropia como medida de comprimento de código, ou seja, comprimento médio das palavras usadas para representar a informação. A expressão (2.30) representa a informação mútua I_M , entre as n variáveis aleatórias x_i , na qual o vetor \mathbf{x} contém todos x_i .

$$I_M(x_1, x_2, \dots, x_n) = \sum_{i=1}^n H(x_i) - H(\mathbf{x}) \quad (2.30)$$

Por exemplo, codificando todas as variáveis juntas, pode-se obter uma redução no comprimento do código em contraste, se as codificassem separadamente. Porém, se as variáveis forem independentes, é possível codificar separadamente, sem incremento no comprimento do código.

Também é possível considerar o conceito de entropia definido acima para variáveis discretas e estendê-lo a variáveis contínuas, como definido na equação (2.31), chamada de entropia diferencial.

$$h(x) = -\int_S f(x) \log f(x) dx \quad (2.31)$$

em que S é o suporte de x , cujas fdp's $f(x) > 0$. Porém, apesar de também estar relacionada com o comprimento do código e ser semelhante ao caso discreto em muitos aspectos, há diferenças importantes entre as duas e deve-se ter cautela ao aplicar essa definição [17].

De forma alternativa, pode-se usar a função conhecida como divergência de *Kullback-Leibler* ou entropia relativa para interpretar a informação mútua. Essa função pode ser considerada como uma distância entre duas densidades de probabilidades, apesar de não ser uma distância real, pois não é simétrica e, dessa forma, não satisfaz a desigualdade triangular. Entretanto, é fácil verificar que ela nunca será negativa e será nula se, e somente se, as distribuições forem iguais [17]. Sejam duas fdp's p_x e p_y , então se define a divergência de *Kullback-Leibler* pela equação (2.32).

$$\delta(p_x, p_y) = \int p_x(u) \log \frac{p_x(u)}{p_y(u)} du \quad (2.32)$$

Para aplicar a divergência de *Kullback-Leibler*, considere que, se as variáveis aleatórias $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$ forem independentes, é possível fatorar sua fdp conjunta de acordo com a definição de independência. Dessa forma, pode-se medir o grau de dependência de x_i como sendo a divergência de *Kullback-Leibler* entre a fdp real $p_x(u)$ e a fdp fatorada p_y , dada pela equação (2.33)

$$p_y(u) = \prod_{i=1}^n p_{x_i}(u_i) \quad (2.33)$$

em que p_{x_i} são as fdp's marginais de x_i . Assim, a partir de (2.32) e (2.33) obtém-se (2.34) para a determinação da informação mútua, ou seja, dependência da variável \mathbf{x} .

$$\delta(p_x) = \int p_x(u) \log \frac{p_x(u)}{\prod p_{x_i}(u_i)} du \quad (2.34)$$

Sabendo que a divergência de *Kullback-Leibler* é tal que $\delta(p_x, p_y) \geq 0$, a partir da equação (2.34), fica claro que a informação mútua será nula se, e somente se, as variáveis forem mutuamente independentes, ou seja, $p_y(u) = p_x(u)$. Caso contrário, a divergência de *Kullback-Leibler* será estritamente positiva [1].

2.4 Gaussianidade (Normalidade)

Anteriormente foi dito que o problema só pode ser resolvido se, e somente se, as fontes não forem gaussianas, ou seja, não possuírem distribuição normal. Para explicar esse fato é possível recorrer ao teorema central do limite.

O teorema central do limite garante que, se for tomado a soma de variáveis aleatórias independentes e identicamente distribuídas, a função densidade de probabilidade dessa soma

tenderá a uma distribuição gaussiana quando o número de variáveis tende a infinito [22]. Como pode ser observado a seguir.

Seja y a soma de uma seqüência de n variáveis aleatórias identicamente distribuídas e independentes x_i . Considerando a forma normalizada de y , visto que sua média e variância podem crescer consideravelmente com $n \rightarrow \infty$, pode-se provar que $f(y)$ tende a uma distribuição normal, com média nula e variância unitária, como mostra a equação (2.35).

$$\lim_{n \rightarrow \infty} f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (2.35)$$

De maneira intuitiva é possível dizer que as fontes são mais distantes da gaussiana que as misturas e, dessa forma, ao localizar máximos locais de uma medida de não-gaussianidade, estão sendo encontradas as componentes independentes. Em outras palavras, é possível criar uma sistemática, na qual é realizada a localização das componentes em que é maximizada a não-gaussianidade.

Para clarificar essa idéia, seja o vetor de misturas \mathbf{x} distribuído de acordo com o modelo ICA da equação (1.2), em que A é a matriz de misturas, \mathbf{x} é o vetor de sinais misturados e \mathbf{s} é o vetor das fontes independentes. Para encontrar uma fonte é preciso considerar uma combinação linear de x_i , conforme indicado na equação (2.36).

$$y = \mathbf{b}^T \mathbf{x} = \sum_i b_i x_i \quad (2.36)$$

em que \mathbf{b} é um vetor a ser determinado. Das equações (1.2) e (2.36) pode-se escrever:

$$y = \mathbf{b}^T A \mathbf{s} \quad (2.37)$$

Dessa forma, y é uma combinação linear de s_i com coeficientes dados por $\mathbf{b}^T A$. Indicando esses coeficientes por \mathbf{w} , obtém-se, então, a equação (2.38):

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{w}^T \mathbf{s} = \sum_i w_i s_i \quad (2.38)$$

Se \mathbf{b} fosse uma das linhas de A^{-1} , essa combinação linear, equação (2.36), seria igual a uma das componentes independentes. Nesse caso o \mathbf{w} correspondente seria de tal forma que um de seus elementos é 1 e os demais são nulos.

O problema, agora, resume-se a determinar os coeficientes de \mathbf{b} , de tal modo que seja igual a uma das linhas de A^{-1} . Na prática não é possível determinar \mathbf{b} exatamente, pois não há nenhum conhecimento da matriz A . Mas pode-se determinar um estimador que gera uma boa aproximação. Veja um exemplo em [12].

A questão fundamental agora, considerando o teorema central do limite, é que a soma, mesmo que de apenas duas variáveis aleatórias, é usualmente mais próxima da gaussiana que as variáveis originais. Da equação (2.38), pode-se ver que é possível variar tanto \mathbf{b} quanto \mathbf{w} , pela definição. Assim, variando \mathbf{b} e verificando a distribuição de $\mathbf{b}^T \mathbf{x}$, a componente independente será encontrada exatamente quando y se tornar o mais distante da gaussiana possível.

Em vista do exposto, é possível formular a seguinte proposição: $y = \mathbf{b}^T \mathbf{x} = \mathbf{w}^T \mathbf{s}$ é igual a uma das componentes independentes quando maximizada a não-gaussianidade de $\mathbf{b}^T \mathbf{x}$. Isso é equivalente a dizer que, quando maximizada a não-gaussianidade de $\mathbf{b}^T \mathbf{x}$, \mathbf{w} está na mesma direção de uma das componentes e, portanto, um de seus elementos vale 1 e os demais elementos são nulos. Logo, a componente correspondente é ortogonal às demais componentes.

A não-gaussianidade pode ser medida de diversas formas. As seções seguintes tratarão de duas delas: a curtose e a negentropia.

2.4.1 Curtose

Curtose é uma estatística de quarta ordem que, para o caso de média nula, é definida pela equação (2.39).

$$curt(x) = E\{x^4\} - 3E\{x^2\}^2 \quad (2.39)$$

A curtose é muito utilizada em diversas áreas de aplicação, devido à sua simplicidade e propriedades. Além de ser baseada em cumulantes, também compartilha de suas propriedades [27, 28].

Também é comum o uso de uma versão normalizada da curtose, como definida na equação (2.40), na qual dividimos a curtose definida em (2.39) pelo quadrado da variância.

$$C(x) = \frac{E\{x^4\}}{E\{x^2\}^2} - 3 \quad (2.40)$$

A curtose normalizada é nula para distribuições gaussianas e, na prática, é não-nula para distribuições não gaussianas. Dessa forma, ela pode ser usada como uma medida de normalidade. O valor da curtose normalizada pode ser tanto positivo como negativo, mas deve-se ter cautela ao trabalhar com curtoses normalizadas positivas, já que não há limite para os valores e pode existir curtose infinita. No entanto, para curtose normalizada negativa, o menor valor possível é -2, quando a variância for unitária. No caso de a curtose normalizada

ser positiva, a distribuição é comumente chamada de super-gaussiana ou leptocúrtica, negativa de sub-gaussiana ou platicúrtica e, no caso de curtose normalizada nula, mesocúrtica.

Para dados nos quais a variância é unitária, ou seja, $E\{x^2\} = 1$, a expressão da curtose normalizada se resume a equação (2.41).

$$curt(x) = C(x) = E\{x^4\} - 3 \quad (2.41)$$

Sejam duas variáveis aleatórias x_1 e x_2 estatisticamente independentes e α um escalar qualquer, assim valem as duas propriedades descritas pelas equações abaixo.

$$curt(x_1 + x_2) = curt(x_1) + curt(x_2) \quad (2.42)$$

e

$$curt(\alpha x) = \alpha^4 curt(x) \quad (2.43)$$

Uma limitação na utilização da curtose normalizada como medida de normalidade se deve ao fato dela ser sensível à *outliers*, também designados por observações anormais ou extremas ou aberrantes. Os *outliers* podem gerar resultados falsos, pois fazem com que o valor da curtose seja alterado consideravelmente. Um simples *outlier* pode ter uma grande influência em métodos estatísticos convencionais, pois estatísticas baseadas em cumulantes favorecem consideravelmente valores na cauda da distribuição [84]. No entanto, a simplicidade e facilidade computacional da curtose normalizada a tornam uma medida atrativa para aplicação em diversos algoritmos.

2.4.2 Negentropia

A negentropia é uma medida quantitativa baseada na entropia diferencial. A entropia tem como base a quantidade de informação que uma variável aleatória fornece; e quanto mais aleatória, imprevisível e desestruturada for, maior a entropia. Seja um vetor aleatório \mathbf{x} com fdp $f_x(x)$ entropia diferencial é definida, de acordo com a equação (2.44).

$$h(\mathbf{x}) = -\int f_x(x) \log f_x(x) dx \quad (2.44)$$

Um resultado muito importante da teoria da informação é que a distribuição mais imprevisível, menos estruturada, ou seja, com maior entropia, é a distribuição gaussiana em relação a todas as outras de mesma variância [17]. Portanto, pode-se considerar este advento para o uso da entropia como uma medida de não-gaussianidade. Para obter uma medida que seja nula para variáveis gaussianas e sempre não-negativa, usa-se freqüentemente uma versão normalizada da entropia diferencial, chamada de negentropia, definida a seguir.

$$J(\mathbf{x}) = H(\mathbf{x}_{gauss}) - H(\mathbf{x}) \quad (2.45)$$

em que \mathbf{x}_{gauss} é um vetor gaussiano aleatório de mesma matriz de correlação e covariância que o vetor \mathbf{x} .

A vantagem de se utilizar a negentropia como medida de não-gaussianidade consiste no fato de ela ser bem justificada pela teoria estatística e é, de certa forma, um estimador ótimo da não-gaussianidade. O maior problema com relação à negentropia é a sua complexidade computacional, pois exige o conhecimento ou a estimativa (provavelmente não-paramétrica) da fpd. Entretanto, é possível obter aproximações razoáveis que podem ser aplicadas na prática [12].

Uma aproximação para entropia pode ser o método da expansão polinomial da densidade de probabilidade de uma variável aleatória. Pode-se assumir que distribuição da variável é muito próxima da gaussiana; sendo assim, é possível fazer uma aproximação semelhante à expansão de Taylor. Geralmente, duas aproximações são usadas para isso: a expansão de Edgeworth e a expansão de Gram-Charlier [44, 21, 1, 12]. Como a determinação da densidade de probabilidade se torna computacionalmente complicada, esses métodos permitem a aproximação da densidade de probabilidade em função de seus cumulantes, possibilitando a estimativa da entropia. Dessa forma, como é mostrado em [12], é possível medir a não gaussianidade de uma variável aleatória normalizada usando a negentropia a partir da aproximação, descrita pela equação (2.46).

$$J(x) \approx \frac{1}{12} \left(E\{x^3\}^2 + \frac{1}{4} curt(x)^2 \right) \quad (2.46)$$

Uma abordagem mais detalhada sobre a expansão da densidade de probabilidade pode ser encontrada em [27, 45].

Fica claro que essa aproximação pode ser precária, pois ela depende da boa aproximação dos cumulantes, ou seja, estatísticas baseadas em cumulantes são muito sensíveis a *outliers*, podendo depender apenas de alguns valores errôneos, além de favorecer, consideravelmente, valores na cauda da distribuição. Assim sendo, a estimativa da entropia também se torna sensível a *outliers*. Hyvärinen em [46] propôs um novo método de aproximação mais robusta para entropia, baseado em uma aproximação do método da máxima entropia.

Antes de prosseguir a discussão, será apresentado brevemente o método da máxima entropia. Seja a informação disponível em uma densidade $f_x(\bullet)$ de uma variável aleatória, descrita pela forma da equação (2.47).

$$\begin{aligned} f_x(x) &\geq 0 \\ \int_S f_x(x) &= 1 \\ \int_S f_x(x) F_i(x) dx &= c_i \quad 1 \leq i \leq n \end{aligned} \quad (2.47)$$

em que $F_i(x)$ é uma função, em geral, não polinomiais, pois usando funções polinomiais os mesmo problemas enfrentados pela aproximação baseada em cumulantes permaneceram presentes.

A questão principal aqui é maximizar a entropia $H(x)$ que satisfaça as restrições impostas em (2.47). Um resultado muito importante da teoria da informação mostra que as fdp's que satisfazem essa condição têm a forma descrita pela equação (2.48). Veja em [17] a demonstração desse resultado.

$$f(x) = A e^{\sum_{i=1}^n a_i F_i(x)} \quad (2.48)$$

em que A e a_i são constantes, de tal forma que $f(x)$ satisfaz as restrições impostas nas equações (2.47).

Baseado nesse resultado, Hyvärinen [46] derivou uma aproximação da entropia semelhante à expansão de Edgeworth e à expansão de Gram-Charlier, obtendo a expressão em (2.49).

$$J(x) \approx \frac{1}{2} \sum_{i=1}^n E\{F_i(x)\}^2 \quad (2.49)$$

Embora essa aproximação não seja precisa, a equação (2.49) pode ser usada como uma medida de não gaussianidade, no sentido de obter o valor mínimo, quando x tiver uma distribuição gaussiana [12, 46].

A escolha das funções não polinomiais deve ser feita de acordo com os 3 critérios estabelecidos em [12] e compiladas abaixo:

1. Na prática, a estimativa de $E\{F_i(x)\}$ não deve ser estatisticamente difícil; em particular ela não deve ser muito sensível a *outliers*.

2. O método da entropia máxima assume que a função $f(x)$, em (2.48), é integrável. Dessa forma, a função $F_i(x)$ não deve crescer rapidamente, o que permitiria a integração de $f(x)$.
3. A função $F_i(x)$ deve considerar aspectos importantes da distribuição de x , os quais forem relevantes à determinação da entropia.

Em um caso especial, pode-se considerar duas funções: F_1 ímpar e F_2 par, por exemplo, $F_1(x) = x^3$ e $F_2(x) = x^4$, que gera uma aproximação a partir de uma generalização da aproximação pelos cumulantes de ordem superior. Porém, se em vez de usar essas funções polinomiais, fossem utilizadas funções não quadráticas, obter-se-ia a aproximação da equação (2.50).

$$J(x) \approx k_1 (E\{F_1(x)\})^2 + k_2 (E\{F_2(x)\} - E\{F_2(v)\})^2 \quad (2.50)$$

em que k_1 e k_2 são constantes positivas e v é uma variável gaussiana de média nula e variância unitária [12].

2.5 Fontes gaussianas

Intuitivamente pode-se dizer que as distribuições gaussianas são mais simples. Seus cumulantes de ordem superior são nulos; porém, as informações contidas nesses cumulantes são essenciais na estimativa do modelo ICA. Desse modo, é fundamentalmente impossível realizar a estimativa se as fontes tiverem distribuição gaussiana.

O branqueamento, que será visto na Seção 3.1, ajuda a compreender por que sinais de fontes gaussianas não podem ser estimados pelo modelo ICA. O branqueamento é fundamentalmente uma transformação, na qual restringe a estimativa ao espaço de matrizes ortogonais.

Primeiramente, considere a transformação de variáveis aleatórias. Sejam \mathbf{x} e \mathbf{y} dois vetores aleatórios de dimensão n , relacionados pelas seguintes transformações.

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{y} \\ g^{-1}(\mathbf{y}) &= \mathbf{x} \end{aligned} \quad (2.51)$$

Pode ser demonstrado que a fdp de \mathbf{y} é obtida a partir da fdp de \mathbf{x} , de acordo com a equação (2.52).

$$f_y(y) = \frac{f_x(x)}{J(x, y)} = \frac{f_x(g^{-1}(y))}{\det\left(\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}\right)} \quad (2.52)$$

em que $J(x, y)$ é o determinante da matriz Jacobiana da transformação.

Considere duas componentes independentes gaussianas \mathbf{s}_1 e \mathbf{s}_2 , cuja fdp conjunta é descrita pela equação (2.53).

$$f_{\mathbf{s}_1, \mathbf{s}_2}(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2\pi} e^{\left(-\frac{\mathbf{s}_1^2 + \mathbf{s}_2^2}{2}\right)} = \frac{1}{2\pi} e^{\left(-\frac{\|\mathbf{s}\|^2}{2}\right)} \quad (2.53)$$

Assumindo, por exemplo, pelo pré-processamento que a matriz de mistura A , na equação (1.2), é ortonormal. Usando a transformação de fdp e o fato da matriz ortogonal ser unitária, ou seja, $A^{-1} = A^T$, logo, a fdp de \mathbf{x} será dada por:

$$f_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2\pi} e^{\left(-\frac{\|A^T \mathbf{x}\|^2}{2}\right)} |\det(A^T)| = \frac{1}{2\pi} e^{\left(-\frac{\|\mathbf{x}\|^2}{2}\right)} \quad (2.54)$$

Portanto, pode-se concluir que uma matriz ortogonal não altera a fdp das fontes, quando elas forem gaussianas, pois ela desaparece após a transformação. Em outras palavras, as distribuições das misturas e das fontes originais são iguais. Isso quer dizer que não há como fazer inferência alguma sobre as fontes a partir das misturas, visto que a matriz de mistura não alteraria a sua forma nessa condição. O fato da matriz de mistura ortogonal não poder ser estimada para variáveis gaussianas está relacionado à propriedade que variáveis aleatórias gaussianas não-correlacionadas são necessariamente independentes. Assim, fica provado que fontes gaussianas não podem ser extraídas pelo modelo ICA.

Será visto mais adiante que, se existir somente uma fonte gaussiana, ainda é possível estimá-la usando o modelo ICA.

Note também, que estão sendo descartadas todas as informações contidas na estrutura temporal dos sinais e considerando-se que todas as informações relevantes, e suficientes, para estimativa das fontes estão presentes em suas distribuições. Existe também uma classe de métodos ICA, que se baseiam na estrutura temporal dos sinais. Nesse caso, a suposição de não gaussianidade é substituída por alguma suposição relativa à estrutura temporal, por exemplo, a autocovariância [120, 132], então, não há restrição às fontes serem não-gaussianas.

3 Métodos para Separação de Fontes

Muitos métodos de separação cega de fontes têm sido propostos; alguns para misturas instantâneas, ou seja, combinação linear de fontes, e outros para casos de fontes que sofreram convolução. No presente trabalho, somente os métodos para combinação linear de fontes, misturas instantâneas, serão tratados. Observa-se que muitos deles possuem vários pontos em comum, sendo alguns já abordados anteriormente como, por exemplo, independência e gaussianidade. Outro tópico muito importante, comumente encontrado em vários métodos de separação, é o pré-processamento.

O pré-processamento pode tornar o problema da estimativa do modelo ICA consideravelmente simplificado. Portanto, será feita uma breve introdução na Seção 3.1 a respeito de técnicas de pré-processamento, cujo objetivo principal é facilitar a análise e o desenvolvimento teórico, além de diminuir a complexidade computacional dos algoritmos.

Na Seção 3.2 será apresentada a primeira aproximação para solução do problema de separação de fontes, o algoritmo de Héroult-Jutten, na qual as fontes são obtidas a partir de um sistema de rede neural adaptativa, com objetivo de obter suas saídas independentes. Embora essa seja uma das soluções pioneiras, não será considerada em detalhes.

Nas Seções 3.3 e 3.4 serão apresentadas as bases dos algoritmos ICA pela curtose e pela negentropia, respectivamente, bem como a implementação de seus respectivos algoritmos rápidos baseados em iterações de ponto fixo, FastICA.

Finalmente, na Seção 3.5, é apresentado o algoritmo PearsonICA, baseado nas curvas de Pearson para a determinação da função custo (contraste).

3.1 Pré-Processamento

A técnica de pré-processamento dos sinais das misturas, introduzida nesta seção, será muito útil para aplicação nos diversos métodos avaliados no presente trabalho.

Uma técnica muito utilizada no pré-processamento de ICA é o branqueamento (*whitening*). De uma forma geral, um sinal é dito branco se a sua densidade espectral de potência for constante [33]. Em BSS, um vetor aleatório $\mathbf{z} = (z_1 \dots z_n)^T$ de média nula é dito branco, se os seus elementos z_i não são correlacionados e possuem a mesma variância. Por

conveniência os vetores, no presente trabalho, serão normalizados para terem variância unitária. Em termos de matriz de covariância, significa que $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, na qual \mathbf{I} é a matriz identidade. O branqueamento é largamente utilizado em ICA, Busca de Projeção (PP – *Projection Pursuit*) e em outras técnicas de análise multivariada [112, 15, 19, 101].

O branqueamento é sempre possível e um método muito utilizado é a PCA, descrita na Seção 3.1.1, seguida de uma normalização para que os vetores tenham variância unitária. A PCA pode ser obtida, por exemplo, pela decomposição em valores singulares (SVD – *Singular Value Decomposition*) do conjunto de vetores de mistura.

Considere a transformação linear de branqueamento $\mathbf{z} = \mathbf{B}\mathbf{x}$. Seja a matriz $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$, $n \times m$ o conjunto de vetores de m misturas com n valores, na qual $\text{posto}(\mathbf{X}) = \min(n, m) = m$, então, a decomposição de \mathbf{X} em valores singulares pode ser expressa como descrito na equação (3.1).

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (3.1)$$

em que $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)$ é uma matriz $n \times m$ e $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)$ é uma matriz $m \times m$, cujas colunas são denominadas vetores singulares à esquerda e vetores singulares à direita, respectivamente. \mathbf{S} é uma matriz diagonal $m \times m$ de m valores singulares ordenados, na qual cada valor singular é igual a σ_i , em que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$ para $i = 1, 2, \dots, m$. Cada σ_i é um valor singular referente a um vetor singular, ou autovetor, em \mathbf{U} e \mathbf{V} , sendo os vetores singulares de \mathbf{U} e \mathbf{V} ortonormais. Tendo em vista que os vetores em \mathbf{U} são combinações das colunas de vetores em \mathbf{X} , de tal forma que todas as colunas em \mathbf{U} são mutuamente ortogonais, é suficiente usar somente a matriz \mathbf{U} para os novos sinais de mistura [32, 50].

Note que, por convenção, os vetores singulares obtidos a partir da SVD possuem módulo unitário [32]. No entanto, precisa-se de vetores com variância unitária. Dessa forma, é preciso aplicar um fator de escala em cada um dos vetores para normalizá-los.

A partir do exposto, podem-se representar os novos sinais branqueados em função dos novos sinais misturados em \mathbf{U} . Logo, a expressão (3.2) descreve os vetores dos sinais branqueados \mathbf{z}_i em função dos vetores \mathbf{u}_i , em que \mathbf{u}_i representa os vetores de \mathbf{U} e $E\{\mathbf{u}_i^2\}$ é a variância de \mathbf{u}_i para $i = 1 \dots m$.

$$\mathbf{z}_i = \frac{\mathbf{u}_i}{E\{\mathbf{u}_i^2\}} \quad (3.2)$$

Como \mathbf{U} é uma combinação linear de \mathbf{X} , \mathbf{z} também será. Tendo que o modelo ICA é $\mathbf{x} = \mathbf{A}\mathbf{s}$ e considerando a transformação linear $\mathbf{z} = \mathbf{B}\mathbf{x}$, pode-se escrever o vetor branqueado em função das fontes pela equação (3.3).

$$\mathbf{z} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s} \quad (3.3)$$

Como \mathbf{z} é um vetor branqueado, é fácil verificar que $\tilde{\mathbf{A}}$ é ortogonal, em que $\tilde{\mathbf{A}} = \mathbf{B}\mathbf{A}$, veja a Equação (3.6). Cabe observar também que, apesar do branqueamento ou da não-correlação estar relacionada com independência, o branqueamento não resolve o problema da ICA. A não-correlação é mais fraca que a independência e não é uma condição suficiente para garantir a independência, como visto anteriormente. Conseqüentemente, não é suficiente para estimar o modelo ICA.

A fim de demonstrar isso, considere a transformação ortogonal \mathbf{C} de \mathbf{z} .

$$\mathbf{y} = \mathbf{C}\mathbf{z} \quad (3.4)$$

O vetor \mathbf{z} é branco, portanto $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$. Devido à ortogonalidade de \mathbf{C} , tem-se:

$$E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{C}\mathbf{z}\mathbf{z}^T\mathbf{C}^T\} = \mathbf{C}\mathbf{I}\mathbf{C}^T = \mathbf{I} \quad (3.5)$$

Ou seja, \mathbf{y} também será um sinal branco e, dessa forma, não é possível dizer se as componentes independentes são dadas por \mathbf{y} ou \mathbf{z} , usando somente a propriedade do branqueamento. Em contrapartida, será mostrado agora que o branqueamento é muito útil no pré-processamento da ICA. Se a nova matriz de misturas for ortogonal, como se pode observar na equação (3.6), a busca da matriz de misturas está restrita ao espaço de matrizes ortogonais.

$$E\{\mathbf{z}\mathbf{z}^T\} = \tilde{\mathbf{A}}E\{\mathbf{s}\mathbf{s}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \quad (3.6)$$

Portanto, em vez de estimar n^2 parâmetros, que são os elementos da matriz original \mathbf{A} , é preciso estimar uma matriz ortogonal $\tilde{\mathbf{A}}$, que contém $n(n-1)/2$ graus de liberdade [16]. Observa-se que, desta forma, o trabalho da estimativa do modelo ICA é reduzido consideravelmente.

Similarmente, é possível branquear os sinais a partir da decomposição em autovalores (EVD – *Eigenvalue Decomposition*) da matriz de covariância das misturas [12].

Para mostrar a relação entre a EVD e a SVD, considere $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ a decomposição em valores singulares de \mathbf{X} e a matriz $\mathbf{C}_X = \mathbf{X}^T\mathbf{X}$, então:

$$\mathbf{C}_X = \mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{S}^T\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{V}\mathbf{S}^2\mathbf{V}^T \quad (3.7)$$

Sendo V unitária e S^2 diagonal, logo, VS^2V^T é a EVD de C_X . Portanto, os autovalores de C_X são iguais ao quadrado dos valores singulares de X , ou seja, $\sigma_i = \sqrt{\lambda_i}$, em que σ_i são os valores singulares de X e λ_i são os autovalores de $X^T X$ e as colunas de V são os autovetores de $X^T X$. Similarmente, verifica-se que as colunas de U são os autovetores de XX^T .

Em geral, a SVD é calculada em vez da EVD, pois do ponto de vista numérico, a SVD é preferível à EVD [83, 124].

3.1.1 PCA - Análise em Componentes Principais

A análise em componentes principais é largamente utilizada para uma série de problemas semelhantes, aos quais é aplicada a técnica ICA. A principal diferença entre ICA e PCA é que, no método ICA, obtêm-se sinais não gaussianos independentes e no método PCA, obtêm-se simplesmente sinais não correlacionados, gaussianos ou não. Para o caso da PCA, pode-se observar que, se os sinais são gaussianos, também serão independentes, porém fisicamente, podem não condizer com sinais provindos de fontes reais como, por exemplo, o sinal de voz de uma pessoa.

PCA é uma técnica muito utilizada na redução dimensional para compressão, extração de características e para classificação. No entanto, há aplicações, nas quais ela não é uma solução ótima como, por exemplo, mostra [42]. A ICA também pode ser usada para redução dimensional, além de outras técnicas como a análise de fator (FA – *Factor analysis*) [117-119, 43].

Pode-se considerar que PCA é um caso particular da ICA, pois, na PCA, obtêm-se sinais que não são correlacionados e, na ICA, sinais que são independentes. Ou seja, na PCA, considera-se apenas estatística até segunda ordem, enquanto na ICA são consideradas estatísticas de ordem superior (HOS – *High-Order Statistics*).

Na prática, antes de se aplicar a PCA, os dados são centralizados para que sua média seja nula. Então, as amostras são linearmente transformadas, de forma que a nova dimensão seja menor e a redundância devida à correlação seja eliminada. Dessa maneira, pode-se dizer que houve uma compressão dos dados com perdas.

Semelhante à ICA, a transformação de um conjunto de sinais misturados $\mathbf{x} = (x_1, x_2)^T$ de média nula em um conjunto de sinais $\mathbf{y} = (y_1, y_2)^T$ extraídos pode ser representada pela

matriz W_{PCA} . Portanto, apesar de não haver um modelo para PCA, pode-se derivar a partir de um, como descrito na equação (3.8).

$$\mathbf{y} = W_{PCA} \mathbf{x} \quad (3.8)$$

Agora será apresentada a análise matemática para resolver o problema da PCA. Considere y_1 uma combinação linear dos elementos do vetor \mathbf{x} , em que os elementos do vetor \mathbf{w} são os pesos e y_1 é denominada primeira componente principal de \mathbf{x} , se a variância de y_1 for máxima. Como a variância depende tanto da norma, quanto da orientação do vetor \mathbf{w}_1 , e cresce conforme a norma cresce, na prática, é imposta a restrição de que a norma seja constante, em geral igual a 1.

$$y_1 = \sum_{k=1}^n w_{k1} x_k = \mathbf{w}_1^T \mathbf{x} \quad (3.9)$$

O problema de otimização da PCA se resume em encontrar o vetor \mathbf{w}_1 que maximiza a função custo da PCA, definida na equação (3.10).

$$I_1(\mathbf{w}_1) = E\{y_1^2\} = E\{(\mathbf{w}_1^T \mathbf{x})^2\} = \mathbf{w}_1^T E\{\mathbf{x}\mathbf{x}^T\} \mathbf{w}_1 = \mathbf{w}_1^T C_x \mathbf{w}_1 \quad (3.10)$$

$$\text{Sujeito a } \|\mathbf{w}_1\| = 1$$

em que C_x é a matriz de covariância do vetor \mathbf{x} de média nula e a norma de \mathbf{w}_1 é, em geral, a norma euclidiana. A solução do problema acima é bem conhecida da álgebra linear e é dada em termos dos autovetores de norma unitária $\mathbf{e}_1, \dots, \mathbf{e}_n$ da matriz C_x . A ordem dos autovetores é tal que os correspondentes autovalores satisfazem as seguintes desigualdades $\lambda_1 \geq \dots \geq \lambda_n$. Dessa forma, a solução da PCA para equação (3.10) será $\mathbf{w}_1 = \mathbf{e}_1$. Logo, a primeira componente principal de \mathbf{x} será $y_1 = \mathbf{e}_1^T \mathbf{x}$. A equação (3.10) pode ser generalizada para as componentes principais m-ésimas, com $1 \leq m \leq n$, então a m-ésima componente principal será $y_m = \mathbf{w}_m^T \mathbf{x}$, na qual a norma unitária de \mathbf{w}_m maximiza a variância de y_m , sujeito a restrição de y_m não ser correlacionado com todas as componentes principais encontradas anteriormente, ou seja, $E\{y_m y_k\} = 0, k < m$.

Tendo em vista o exposto anteriormente, não é difícil mostrar que a k-ésima componente principal será $y_k = \mathbf{e}_k^T \mathbf{x}$ [43].

Na PCA é encontrada uma representação dos sinais, de tal forma que eles não sejam correlacionados e suas variâncias sejam máximas. Portanto, como a variância sozinha não

garante que os sinais sejam independentes, a PCA não é suficiente para realizar a separação de fontes.

3.2 Algoritmo de Héroult Jutten

A primeira aproximação para solução do problema de separação cega de fontes foi proposta por Héroult e Jutten [2], que, tentando emular o comportamento do sistema nervoso, propuseram uma arquitetura neural, conforme a representada na Figura 3.1, para o caso 2×2 e um algoritmo adaptativo, o qual denominaram de INCA (*INdependent Component Analysis*), a fim de encontrar seus parâmetros.

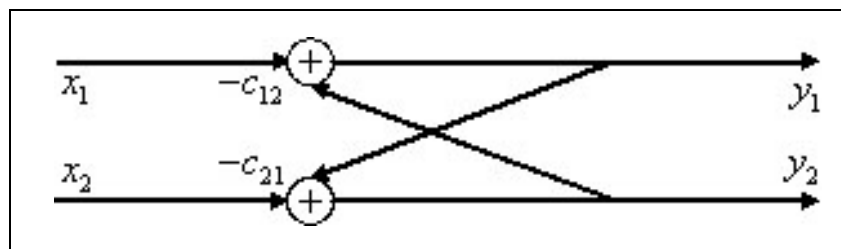


Figura 3.1: Arquitetura de Héroult-Jutten para separação de fontes.

As saídas iniciais são realimentadas no sistema e, então, são recalculadas até que o sistema entre em equilíbrio. A equação que relaciona a entrada e a saída para o sistema de separação de fontes apresentado é dada pela equação (3.11), em que C é a matriz dos coeficientes da rede, de tal forma que $\text{diag}(C) = 0$.

$$\mathbf{y} = \mathbf{x} - C\mathbf{y} \quad (3.11)$$

Pode-se reescrever a equação (3.11), como segue:

$$\mathbf{y} = (\mathbf{I} + C)^{-1} \mathbf{x} \quad (3.12)$$

Dessa forma, a matriz de separação será $\mathbf{W} = (\mathbf{I} + C)^{-1}$, confrontando com o modelo da ICA, na qual se tem $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$. Logo, se $(\mathbf{I} + C) = \mathbf{A}$, então \mathbf{y} será igual a \mathbf{s} .

A solução proposta consistia em adaptar os coeficientes de realimentação do sistema, de tal forma que as saídas do sistema se tornassem independentes. Para tanto, Héroult e Jutten usaram um critério de correlação não linear para independência, conforme a seguinte regra de aprendizado, representada na equação (3.13), para determinação dos elementos da matriz C .

$$\Delta c_{ij} = \mu f(y_i)g(y_j) \quad (3.13)$$

Na qual μ é a taxa de aprendizado e ambas as funções $f(\cdot)$ e $g(\cdot)$ são ímpares, tipicamente $f(y) = y^3$ e $g(y) = \tan^{-1}(y)$, embora $g(y) = y$ e $g(y) = \text{sign}(y)$ também permitam a separação das fontes. A análise das limitações e da estabilidade do algoritmo

proposto está delineada em [65, 66]. Embora a estabilidade local seja mostrada em [66], a convergência global do algoritmo não é garantida.

Dessa forma, se o aprendizado convergir, o lado direito da equação (3.13) deverá ser nulo na média. Ou seja, isso implica que:

$$E\{f(y_i)g(y_j)\} = E\{f(y_i)\}E\{g(y_j)\} = 0, \text{ para } i \neq j \quad (3.14)$$

Logo, com isso, espera-se conseguir a independência das fontes.

Embora esse seja um dos métodos pioneiros, na prática existem várias desvantagens. Por exemplo, se os sinais estiverem mal dimensionados (uma ou mais fontes/sensores podem estar ruins) ou se a matriz de mistura for mal condicionada, o algoritmo pode falhar ou não trabalhar de forma eficiente. Também o número de fontes que o algoritmo pode separar é muito limitado. Cichocki *et al* [70, 69] desenvolveram um algoritmo baseado no algoritmo de Héroult-Jutten, porém com algumas vantagens sobre este.

3.3 ICA pela curtose

A ICA é uma importante ferramenta quando há sinais gerados pela combinação de duas ou mais fontes. Essencialmente, pode-se interpretar como sendo um método que permite extrair informações úteis de um conjunto de dados e, por isso, atualmente existem várias áreas de aplicação que fazem uso desse método.

Como visto anteriormente, não há informação alguma a respeito das fontes ou de como elas foram misturadas; apenas são conhecidos os sinais das misturas que serão utilizados para realizar a estimativa, aplicando uma função custo para otimizar o processo de estimativa, baseando-se nas medidas de não gaussianidade e de independência, observadas no capítulo 2.

Como foi introduzido no capítulo 1, o modelo da ICA para estimativa das fontes pode ser considerado como sendo uma combinação linear de sinais misturados, conforme é representado na equação (3.15), em que W é a estimativa da matriz de mistura A .

$$\mathbf{y} = W^T \mathbf{x} \quad (3.15)$$

A primeira coisa a ser considerada será como a maximização ou minimização da curtose permite encontrar as componentes independentes, conforme foi apresentado no capítulo 2. Para simplificar, considere o modelo para o caso bidimensional da equação $\mathbf{x} = A \cdot \mathbf{s}$, sem perda da generalidade, assumindo que as componentes s_1 e s_2 possuem curtose $curt(s_1)$ e $curt(s_2)$, respectivamente, ambas diferentes de zero.

Dessa forma, deve-se procurar por uma fonte a partir da equação $y = \mathbf{b}^T \mathbf{x}$. Considerando a transformação $\mathbf{q} = A^T \mathbf{b}$, é possível escrever a equação (3.16).

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{b}^T A \mathbf{s} = \mathbf{q}^T \mathbf{s} = \sum_i q_i s_i \quad (3.16)$$

Note que se \mathbf{b} for uma das linhas da inversa da matriz de misturas A , a combinação linear da equação (3.16) será uma das componentes. E nesse caso, implica que, a correspondente \mathbf{q} tenha um elemento não nulo e todos os outros elementos nulos.

A curtose normalizada pode ser usada como uma medida de não gaussianidade, sendo nula para variáveis gaussianas, positiva super-gaussianas e negativa para sub-gaussianas, veja Seção 2.4.1. Aplicando as propriedades da curtose ao caso bidimensional tem-se a equação (3.17).

$$curt(y) = q_1^4 curt(s_1) + q_2^4 curt(s_2) \quad (3.17)$$

Note que, por conveniência, as fontes possuem variância unitária devido ao pré-processamento. Portanto, a variância de y deve ser igual a 1. Isso implica na seguinte restrição:

$$E\{y^2\} = q_1^2 + q_2^2 = 1 \quad (3.18)$$

Ou seja, geometricamente isso significa que o vetor \mathbf{q} está limitado ao círculo unitário no plano bidimensional. O problema da otimização, desta vez, se resume a encontrar o máximo no círculo unitário, para a seguinte função:

$$|curt(y)| = |q_1^4 curt(s_1) + q_2^4 curt(s_2)| \quad (3.19)$$

Dessa forma, a partir da equação (3.19), pode-se verificar que o máximo ocorre no ponto onde exatamente um dos elementos de \mathbf{q} é nulo e o outro é não nulo. Como as condições de contorno restringem ao círculo unitário, o elemento não nulo deve ser igual a +1 ou -1. Este ponto ocorre exatamente quando y é igual à componente independente $\pm s_i$. A prova detalhada desta propriedade pode ser verificada em [12].

Pode-se estender a equação (3.17) para o caso de n fontes. Então, a equação passa ter a seguinte forma:

$$curt(y) = \sum_{i=1}^n q_i^4 curt(s_i) \quad (3.20)$$

Observa-se a partir da equação (3.20), que pelo fato de a curtose de variáveis gaussianas ser nula não é possível realizar a separação de fontes com distribuição gaussiana.

Já considerando o pré-processamento pelo branqueamento será mostrado o motivo principal de sua importância. Considere a procura de combinações lineares $\mathbf{w}^T \mathbf{z}$ para o vetor branqueado \mathbf{z} , as quais maximizam a não-gaussianidade. Como $\mathbf{q} = \tilde{\mathbf{A}}^T \mathbf{w}$, logo:

$$\|\mathbf{q}\|^2 = (\mathbf{w}^T \tilde{\mathbf{A}})(\tilde{\mathbf{A}}^T \mathbf{w}) = \|\mathbf{w}\|^2 \quad (3.21)$$

Sendo assim, restringir \mathbf{q} à esfera unitária é o mesmo que restringir \mathbf{w} à esfera unitária. Portanto, basta maximizar o valor absoluto da curtose da estimativa da fonte \mathbf{y} , ou seja, maximizar o valor da curtose de $\mathbf{w}^T \mathbf{z}$, sob a restrição de $\|\mathbf{w}\| = 1$.

Uma forma de resolver o problema de maximização é implementando o método do gradiente, o qual se inicia com um vetor \mathbf{w} e segue movendo-o na direção em que o valor absoluto da curtose de $y = \mathbf{w}^T \mathbf{z}$ é maximizado. Como o sinal da mistura é branqueado, então $E\{(\mathbf{w}^T \mathbf{z})^2\} = \|\mathbf{w}\|^2$. Portanto, é possível calcular o gradiente do valor absoluto da curtose de $y = \mathbf{w}^T \mathbf{z}$, como segue:

$$\begin{aligned} \frac{\partial |curt(\mathbf{w}^T \mathbf{z})|}{\partial \mathbf{w}} &= \frac{\partial |E\{(\mathbf{w}^T \mathbf{z})^4\} - 3(E\{(\mathbf{w}^T \mathbf{z})^2\})^2|}{\partial \mathbf{w}} \\ &= |4E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} - 12\mathbf{w}\|\mathbf{w}\|^2| \end{aligned} \quad (3.22)$$

Finalmente, resulta a equação (3.23), em que $sign$ é o sinal algébrico estimado da curtose.

$$\frac{\partial |curt(\mathbf{w}^T \mathbf{z})|}{\partial \mathbf{w}} = 4sign(curt(\mathbf{w}^T \mathbf{z})) [E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} - 3\mathbf{w}\|\mathbf{w}\|^2] \quad (3.23)$$

Como se deve procurar \mathbf{w} que esteja dentro da esfera unitária ($\|\mathbf{w}\|^2 = 1$), basta dividi-lo por sua norma. Note que, para sinal branqueado, o segundo termo entre colchetes na equação (3.23) desaparece, pois $E\{(\mathbf{w}^T \mathbf{z})^2\} = \|\mathbf{w}\|^2$; assim pode-se omiti-lo. Tendo em vista o exposto, resulta o seguinte algoritmo:

$$\begin{aligned} \Delta \mathbf{w} &\propto sign(curt(\mathbf{w}^T \mathbf{z})) E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|} \end{aligned} \quad (3.24)$$

Olhando atentamente o método descrito acima, nota-se que, após o branqueamento, é possível interpretar as combinações $\mathbf{w}^T \mathbf{z}$, como sendo projeções em uma reta. Essa

observação será muito importante para o Capítulo 4, na qual é realizada a conexão do método ICA com o método denominado “*Projection Pursuit*”.

Uma versão adaptativa (on-line) do algoritmo pode ser obtida tomando-se os valores instantâneos de \mathbf{z} (Veja a Tabela 3.1). Portanto, cada observação \mathbf{z} pode ser usada uma vez no algoritmo [12].

Apesar da simplicidade deste algoritmo para separação de fontes em esquemas não-estacionários, é possível verificar que a convergência pode ser muito demorada, dependendo da boa escolha do fator de aprendizado.

Delfosse e Lobaton [31] apresentam um método, no qual mostram a convergência do algoritmo do gradiente no problema da maximização da curtose. Expõem que é possível extrair uma fonte de cada vez, obtendo o mínimo local da função custo restringindo à esfera unitária, e aplicando um processo denominado *deflation*, que consiste num processo de projeção ortogonal, para extração das demais fontes.

Tabela 3.1: Algoritmo do gradiente usando a curtose.

<ol style="list-style-type: none"> 1) $\Delta \mathbf{w} \propto \text{sign}(\text{curl}(\mathbf{w}^T \mathbf{z})) \mathbf{z} (\mathbf{w}^T \mathbf{z})^3$ 2) $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\ \mathbf{w}\ }$ 3) Se não convergir, voltar ao passo 1.
--

Verifica-se que como uma fonte é extraída de cada vez, se houver uma fonte gaussiana, ainda será possível extrair essa fonte gaussiana, que será o resíduo do algoritmo, após a extração de todas as outras fontes. É claro que, caso haja mais de uma fonte gaussiana não será possível separá-las, pois, conforme visto na Seção 2.5, não é possível fazer inferência alguma acerca das fontes, caso sejam gaussianas.

3.3.1 Algoritmo de ponto fixo usando a curtose

Na seção anterior foi derivado o algoritmo do gradiente para resolver o problema da ICA pela maximização da curtose. Apesar da vantagem desse método admitir que a entrada \mathbf{z} seja usada uma vez no algoritmo, permitindo uma rápida adaptação em aplicações não estacionárias, sua convergência é lenta e depende da boa escolha da taxa de aprendizado. O algoritmo baseado em iterações de ponto fixo, apresentado nesta seção, pode tornar o processo de aprendizado muito mais rápido.

Um algoritmo eficiente para análise em componentes independentes foi desenvolvido por Aapo Hyvärinen e Erkki Oja, em 1997. Esse algoritmo está implementado no pacote,

denominado FastICA, disponível em [52]. Esta seção é baseada no algoritmo publicado em [14].

Seja $\mathbf{z}(t)$ a seqüência de observações e $\mu(t)$ a seqüência de taxa de aprendizado, considerando a equação (3.23), pode-se escrever o algoritmo de aprendizado on-line, conforme descrito pela equação (3.25).

$$\mathbf{w}(t+1) = \mathbf{w}(t) \pm \mu(t) \left[\mathbf{z}(t) (\mathbf{w}(t)^T \mathbf{z}(t))^3 - 3 \|\mathbf{w}(t)\|^2 \mathbf{w}(t) \right] \quad (3.25)$$

Em que os termos entre colchetes são obtidos a partir do gradiente da $\text{curt}(\mathbf{w}^T \mathbf{z})$, tomados pelos seus valores instantâneos invés de seus valores esperados, na qual o segundo termo é o “termo de esquecimento” para evitar o crescimento ilimitado dos pesos $\mathbf{w}(t)$.

A regra de aprendizado de \mathbf{w} , baseada em iterações de ponto fixo pode ser obtida tomando-se os valores esperados e igualando a mudança em \mathbf{w} , na equação (3.25), a zero. Conforme descrito na equação (3.26).

$$E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} - 3 \|\mathbf{w}\|^2 \mathbf{w} = 0 \quad (3.26)$$

O algoritmo baseado em iterações de ponto fixo derivado do gradiente da curtose consiste em aplicar o algoritmo descrito na Tabela 3.2, na qual o valor esperado é calculado em blocos (por exemplo, blocos de 1000 pontos). A convergência do algoritmo é demonstrada ser cúbica, em [14]. Além disso, por ser baseado em iterações de ponto fixo, não há passo de adaptação nem parâmetros de ajuste, o que torna o algoritmo mais fácil de usar.

Tabela 3.2: Algoritmo FastICA usando curtose.

- 1) Escolher um vetor inicial \mathbf{w}_0 de norma igual a 1 e definir $k = 1$.
- 2) $\mathbf{w}_k = E\{\mathbf{z}(\mathbf{w}_{k-1}^T \mathbf{z})^3\} - 3\mathbf{w}_{k-1}$.
- 3) Dividir \mathbf{w}_k por sua norma.
- 4) Se $|\mathbf{w}_k^T \mathbf{w}_{k-1}|$ não for suficientemente próximo de 1, então definir $k = k + 1$ e voltar ao passo 2. Caso contrário, terminar em \mathbf{w}_k .

Assumindo que o vetor do sinal \mathbf{z} é o sinal das misturas (já branqueado), após a convergência do algoritmo descrito na Tabela 3.2, resulta o vetor \mathbf{w}_k final, que será uma das colunas da matriz ortogonal de mistura $\tilde{\mathbf{A}}$. E finalmente, $\mathbf{w}_k^T \mathbf{z}$ será a estimativa de uma das fontes.

Para fazer a separação de múltiplas fontes, basta aplicar várias vezes o algoritmo acima até encontrar o número de fontes desejadas. No entanto, para assegurar que será encontrada uma componente diferente a cada repetição, é necessário adicionar uma projeção de ortogonalização na repetição.

Tendo que as colunas da matriz de mistura \tilde{A} são ortogonais, devido ao branqueamento, e que a sua inversa é $\tilde{A}^T = W^T$, definimos \bar{A} como sendo a matriz, cujas colunas são as colunas \mathbf{w}_k , previamente encontradas da matriz \tilde{A} . Dessa forma, garante-se que o vetor \mathbf{w} , encontrado no passo atual, seja ortogonal aos vetores encontrados anteriormente; portanto, isso permite que se encontre uma fonte diferente, após cada iteração. Assim, o novo algoritmo com a projeção ortogonal para estimativa de várias fontes pode ser escrito, conforme apresentado na Tabela 3.3.

Tabela 3.3: Novo passo 3 do algoritmo FastICA para múltiplas fontes.

- 1) Escolher um vetor inicial \mathbf{w}_0 de norma igual a 1 e definir $k = 1$.
- 2) $\mathbf{w}_k = E\{\mathbf{z}(\mathbf{w}_{k-1}^T \mathbf{z})^3\} - 3\mathbf{w}_{k-1}$.
- 3) $\mathbf{w}_k = \mathbf{w}_k - \bar{A}\bar{A}^T \mathbf{w}_k$, na qual \bar{A} é a matriz composta pelos \mathbf{w}_k encontrados anteriormente.
- 4) Dividir \mathbf{w}_k por sua norma.
- 5) Se $|\mathbf{w}_k^T \mathbf{w}_{k-1}|$ não for suficientemente próximo de 1, então definir $k = k + 1$ e voltar ao passo 2. Caso contrário, terminar em \mathbf{w}_k .

Existem outras formas de realizar a ortogonalização dos vetores de peso, como é possível notar em [14], a qual também apresenta a ortogonalização simétrica, por exemplo.

O presente algoritmo mostra-se superior na maioria das aplicações em relação ao algoritmo do gradiente, exceto em algumas aplicações em que há necessidade de adaptação rápida, em condições de um esquema dinâmico, no qual o algoritmo do gradiente pode ser preferido. A convergência do presente algoritmo é cúbica e pode ser verificada em [12, 14].

3.4 ICA pela negentropia

Na Seção 2.4.1 foi visto que a curtose não é uma medida robusta com relação à *outliers*, ou seja, há a possibilidade de a curtose depender somente de algumas observações na cauda da distribuição, que podem elevar seu valor. Portanto, uma medida errada ou irrelevante pode nos levar a uma conclusão equivocada. Também foi apresentada na Seção 2.4.2 a entropia / negentropia, que pode ser usada como uma medida de gaussianidade mais robusta que a curtose em relações aos *outliers*. Além disso, foi feita uma breve introdução aos métodos de aproximação da entropia como, por exemplo, o uso de cumulantes e o método da máxima entropia.

A aproximação baseada no método da máxima entropia [46] será usada neste algoritmo ICA, pois, geralmente, mostra-se mais robusta que as aproximações convencionais. Um caso simples para essa aproximação é representado na equação (3.27).

$$J(y_i) \approx k[E\{F(y_i)\} - E\{F(v)\}]^2 \quad (3.27)$$

Em que F é praticamente qualquer função não quadrática; k é uma constante irrelevante [48, 46], a qual será assumida $k=1$; e v é uma variável aleatória gaussiana de média nula e variância unitária. Essa aproximação da negentropia permite definir uma nova função custo para o problema da ICA. Para encontrar uma componente independente, na qual $y_i = \mathbf{w}^T \mathbf{x}$, deve-se maximizar a função J_F , dada pela equação (3.28), em que \mathbf{w} é o vetor peso sujeito à $E\{(\mathbf{w}^T \mathbf{x})^2\} = 1$.

$$J_F(\mathbf{w}) = [E\{F(\mathbf{w}^T \mathbf{x})\} - E\{F(v)\}]^2 \quad (3.28)$$

Como a negentropia é invariante para transformações lineares inversíveis, é óbvio que encontrar uma transformação inversível W que minimiza a informação mútua será equivalente a encontrar direções, nas quais a negentropia é maximizada [1]. Da mesma forma desenvolvida com a curtose, pode-se derivar um algoritmo do gradiente para maximização da negentropia. O sinal é branqueado, portanto $E\{(\mathbf{w}^T \mathbf{z})^2\} = \|\mathbf{w}\|^2 = 1$. Logo, o algoritmo representado pelas equações em (3.29) é obtido.

$$\begin{aligned} \Delta \mathbf{w} &\propto \gamma E\{\mathbf{z}f(\mathbf{w}^T \mathbf{z})\} \\ \mathbf{w} &\leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|} \end{aligned} \quad (3.29)$$

Em que $\gamma = E\{F(\mathbf{w}^T \mathbf{z})\} - E\{F(v)\}$, f é a derivada da função F , usada na aproximação da negentropia e v é uma variável aleatória gaussiana de média nula e variância unitária.

O parâmetro γ , que permite um tipo de auto-adaptação, é facilmente estimado pela equação (3.30). Corresponde ao sinal algébrico da curtose na Tabela 3.1 e pode ser substituído pelo seu sinal algébrico, caso haja conhecimento prévio de distribuição das fontes. Note que ele não altera pontos estacionários do aprendizado, porém seu sinal algébrico afeta a sua estabilidade. Dessa forma, é possível substituir γ por seu sinal algébrico, sem afetar o comportamento do aprendizado.

$$\Delta\gamma \propto (F(\mathbf{w}^T \mathbf{z}) - E\{F(v)\}) - \gamma \quad (3.30)$$

Algumas funções usadas para F e suas respectivas derivadas estão descritas abaixo.

$$F_1 = \frac{1}{a} \ln(\cosh(a \cdot y)), \quad f_1 = \tanh(a \cdot y) \quad (3.31)$$

$$F_2 = -\exp\left(\frac{-y^2}{2}\right), \quad f_2 = y \cdot \exp\left(-\frac{y^2}{2}\right) \quad (3.32)$$

$$F_3 = \frac{y^4}{4}, \quad f_3 = y^3 \quad (3.33)$$

Em que a é uma constante conveniente no intervalo $1 \leq a \leq 2$, em geral $a=1$. As funções F_1 e F_2 apresentam melhores resultados que a função F_3 , que obtém uma aproximação baseada na curtose. Em particular, escolhendo uma função F que não cresça rapidamente, obtém-se um estimador mais robusto.

A Tabela 3.4 sumariza o algoritmo do gradiente estocástico para estimativa de uma fonte. O algoritmo apresentado considera que o sinal da mistura foi previamente branqueado, obtendo-se o sinal \mathbf{z} .

Tabela 3.4: Algoritmo do gradiente estocástico para maximização da não gaussianidade.

1. Escolher um vetor inicial \mathbf{w} de norma unitária e um valor inicial para γ .
2. $\Delta\mathbf{w} \propto \gamma E\{\mathbf{z}f(\mathbf{w}^T \mathbf{z})\}$.
3. Dividir \mathbf{w} por sua norma.
4. Se o sinal algébrico de γ não for conhecido previamente:

$$\Delta\gamma \propto (F(\mathbf{w}^T \mathbf{z}) - E\{F(v)\}) - \gamma.$$
5. Se não convergir voltar ao passo 2.

3.4.1 Algoritmo rápido de ponto fixo usando negentropia

O algoritmo rápido baseado em iterações de ponto fixo usando negentropia, apresentado nesta seção, foi primeiramente publicado por Hyvärinen [48]. Esse algoritmo faz parte do pacote FastICA [52]. Similarmente ao algoritmo baseado na curtose, este é um algoritmo eficiente que procura direções do vetor peso unitário, \mathbf{w} , na qual a projeção $\mathbf{w}^T \mathbf{z}$ maximiza a não gaussianidade.

A partir da equação (3.29), é possível escrever a equação da iteração de ponto fixo, como segue, omitindo γ por causa da normalização de \mathbf{w} :

$$\mathbf{w} \leftarrow E\{\mathbf{z}f(\mathbf{w}^T \mathbf{z})\} \quad (3.34)$$

Entretanto, essa iteração não possui as mesmas propriedades de convergência do algoritmo usando a curtose, portanto, precisa ser modificada. Dessa forma, pode-se adicionar $\alpha \mathbf{w}$ em ambos os lados da equação (3.34), sem modificar a iteração de ponto fixo. Logo, é possível escolher adequadamente um valor para α , no qual o algoritmo convirja rapidamente como o algoritmo baseado em iterações de ponto fixo usando a curtose.

$$(1 + \alpha)\mathbf{w} = E\{\mathbf{z}f(\mathbf{w}^T \mathbf{z})\} + \alpha \mathbf{w} \quad (3.35)$$

O coeficiente α pode ser obtido usando uma aproximação do método de Newton. Embora o método de Newton requeira uma inversão de matriz em cada iteração, será visto a seguir, que após algumas considerações com relação ao problema de separação cega, essa inversão será simplificada.

O máximo de $J_F(\mathbf{w})$ é obtido para certo valor ótimo de $E\{F(\mathbf{w}^T \mathbf{z})\}$. De acordo com as condições de Kuhn-Tucker [55, 53, 54], o ótimo de $E\{F(\mathbf{w}^T \mathbf{z})\}$ sujeito a condição $E\{(\mathbf{w}^T \mathbf{z})^2\} = \|\mathbf{w}\|^2 = 1$ é obtido no ponto que satisfaz a equação (3.36).

$$E\{\mathbf{z}f(\mathbf{w}^T \mathbf{z})\} - \beta \mathbf{w} = 0 \quad (3.36)$$

Em que β é uma constante facilmente calculada, resultando em $\beta = E\{\mathbf{w}_o^T \mathbf{z}f(\mathbf{w}_o^T \mathbf{z})\}$, na qual \mathbf{w}_o é o valor de \mathbf{w} no ponto ótimo.

Para resolver essa equação pelo método de Newton, denomine de G o primeiro membro da equação (3.36). Então, pode-se determinar a matriz Jacobiana de G pela expressão (3.37).

$$JG(\mathbf{w}) = \frac{\partial G}{\partial \mathbf{w}} = E\{\mathbf{z}\mathbf{z}^T f'(\mathbf{w}^T \mathbf{z})\} - \beta \mathbf{I} \quad (3.37)$$

Como os dados estão branqueados, é razoável fazer a seguinte aproximação (Equação (3.38)), para o primeiro termo da expressão (3.37), a fim de simplificar a inversão da matriz.

$$E\{\mathbf{z}\mathbf{z}^T f'(\mathbf{w}^T \mathbf{z})\} \approx E\{\mathbf{z}\mathbf{z}^T\}E\{f'(\mathbf{w}^T \mathbf{z})\} = E\{f'(\mathbf{w}^T \mathbf{z})\}I \quad (3.38)$$

Dessa forma, a matriz Jacobiana se torna diagonal e a inversão pode ser obtida facilmente. Também é possível aproximar β pelo valor atual de \mathbf{w} ao invés de \mathbf{w}_o . Assim, é obtida a seguinte iteração pela aproximação de Newton:

$$\begin{aligned} \mathbf{w}_k &= \mathbf{w}_{k-1} - \frac{[E\{\mathbf{z}f(\mathbf{w}_{k-1}^T \mathbf{z})\} - \beta \mathbf{w}_{k-1}]}{[E\{f'(\mathbf{w}_{k-1}^T \mathbf{z})\} - \beta]} \\ \mathbf{w}_{k-1} &= \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \end{aligned} \quad (3.39)$$

Multiplicando ambos os lados da primeira equação em (3.39) por $\beta - E\{f'(\mathbf{w}^T \mathbf{z})\}$, são obtidas as seguintes iterações de ponto fixo para o algoritmo.

$$\begin{aligned} \mathbf{w}_k &= E\{\mathbf{z}f(\mathbf{w}_{k-1}^T \mathbf{z})\} - E\{f'(\mathbf{w}_{k-1}^T \mathbf{z})\}\mathbf{w}_{k-1} \\ \mathbf{w}_{k-1} &= \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \end{aligned} \quad (3.40)$$

O nome ponto fixo foi mantido da versão do algoritmo usando curtose, no entanto, do ponto de vista da derivação do algoritmo, é preferível o método de Newton à iteração de ponto fixo [48]. O algoritmo FastICA usando a negentropia está resumido na Tabela 3.5.

Como resultado da aproximação feita para o Jacobiano e, conseqüentemente, para a matriz Hessiana sob a condição de branqueamento, o algoritmo FastICA se reduz ao algoritmo do gradiente-descendente convencional com passo fixado; portanto, não requer passo de adaptação [63].

Tabela 3.5: Algoritmo FastICA usando negentropia.

- | |
|---|
| <ol style="list-style-type: none"> 1) Escolher um vetor inicial \mathbf{w}_o de norma igual a 1 e definir $k = 1$. 2) $\mathbf{w}_k = E\{\mathbf{z}f(\mathbf{w}_{k-1}^T \mathbf{z})\} - E\{f'(\mathbf{w}_{k-1}^T \mathbf{z})\}\mathbf{w}_{k-1}$. 3) Dividir \mathbf{w}_k por sua norma. 4) Se não convergir retorna ao passo 2. |
|---|

3.5 PearsonICA - Curvas de Pearson

O algoritmo PearsonICA, abordado nesta seção, foi primeiramente desenvolvido e publicado em [56], por J. Karvanen *et al*, e estendido posteriormente em [57], por J. Karvanen e V. Koivunen. O método consiste em aplicar o sistema de curvas de Pearson para modelar a distribuição das fontes. Esse método pode ser aplicado para uma série de distribuições, que também podem ser assimétricas e, inclusive, ter curtose normalizada nula.

O método proposto em [56] combina duas técnicas bem conhecidas: função de contraste não-linear fixa e aproximação pela máxima verossimilhança. Na aproximação pela máxima verossimilhança, o sistema de curvas de Pearson é usado para modelar a distribuição das fontes.

As curvas de Pearson formam um conjunto de famílias de distribuições paramétricas, proposta por Karl Pearson, no qual cada família pode ser gerada pela solução da equação diferencial (3.41) para a variável aleatória x com fdp $f(x)$, dada pela escolha adequada dos quatro parâmetros a , b_0 , b_1 e b_2 .

$$\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2} \quad (3.41)$$

Para a aproximação da máxima verossimilhança para ICA, a função escore “*score*” da distribuição hipotética da fonte é usada como função custo, na qual pode ser facilmente resolvida pela equação (3.42) para o sistema de Pearson. Como se pode observar, a simplicidade da função “*score*” torna o método bastante atrativo.

$$\varphi(x) = -\frac{f'(x)}{f(x)} = -\frac{(x-a)}{b_0 + b_1x + b_2x^2} \quad (3.42)$$

Derivando a equação da função escore resulta em:

$$\varphi'(x) = -\frac{b_0 + ab_1 + 2axb_2 - x^2b_2}{(b_0 + b_1x + b_2x^2)^2} \quad (3.43)$$

A estimativa dos parâmetros a , b_0 , b_1 , b_2 pode ser obtida pelo método dos momentos, como mostra [27, 58]. As equações das estimativas em função dos momentos centrais estão representadas abaixo, em que $A = 10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3$. Além disso, os momentos teóricos são estimados pelos respectivos momentos amostrais.

$$a = b_1 = -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{A} \quad (3.44)$$

$$b_0 = -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{A} \quad (3.45)$$

$$b_2 = -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{A} \quad (3.46)$$

A Figura 3.2 apresenta vários tipos de distribuições no plano (β_1, β_2) , em que $\beta_1 = \mu_3^2/\mu_2^3$ e $\beta_2 = \mu_4/\mu_2^2$.

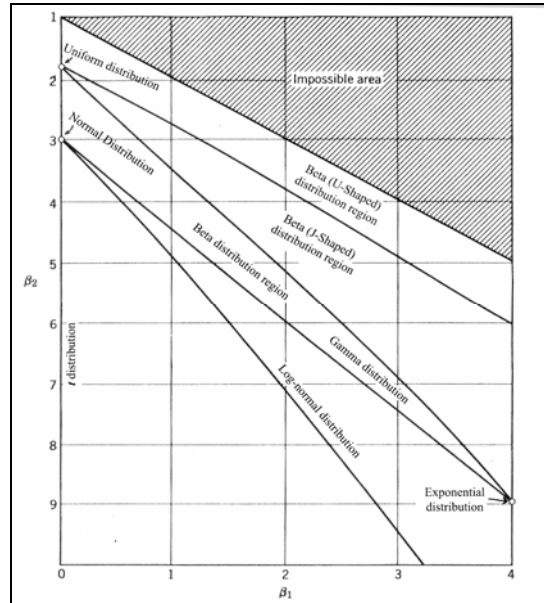


Figura 3.2: Regiões no plano (β_1, β_2) para várias distribuições de Pearson. (De E.S. Pearson, Seminars, Princeton University, 1960).

No método proposto, as distribuições das fontes são estimadas pelas distribuições marginais através do ajuste às curvas de Pearson, utilizando o método descrito acima, que é realizado iterativamente até que o algoritmo convirja.

Qualquer algoritmo ICA que utilize a maximização da verossimilhança de uma função custo pode ser usado [57], como, por exemplo, o algoritmo do gradiente relativo [11]. A equação (3.47) representa o algoritmo do gradiente relativo para determinação da matriz W

$$W_k = W_{k-1} + \mu(I - \varphi(\mathbf{y})\mathbf{y}^T)W_{k-1} \quad (3.47)$$

Em que μ é a taxa de aprendizado. Ou também, por exemplo, é possível usar o algoritmo baseado em iterações de ponto fixo [41, 48]. A equação (3.48) representa a iteração de ponto fixo para o algoritmo na determinação da matriz W

$$W_k = W_{k-1} + D(E\{\varphi(\mathbf{y})\mathbf{y}^T\} - \text{diag}(E\{\varphi(y_i)y_i\}))W_{k-1} \quad (3.48)$$

Na qual, $D = \text{diag}(1/E\{\varphi(y_i)y_i\} - E\{\varphi'(y_i)\})$.

É bem conhecido que o método do momento, usado na estimativa dos parâmetros da curva de Pearson, só é eficiente para distribuições próximas da distribuição normal [58]. Portanto, para contornar esse problema, quando a curtose for próxima de uma distribuição normal, o algoritmo aplica a função custo baseada nas curvas de Pearson. Caso contrário, será usada uma função custo não linear bem conhecida, que pode ser, por exemplo, $\tanh(2y)$.

O algoritmo PearsonICA [56] está sumarizado abaixo.

Tabela 3.6: Algoritmo PearsonICA.

1. Calcular os momentos amostrais $\hat{\mu}_3$ e $\hat{\mu}_4$ para os dados atuais $\mathbf{y}_k = \mathbf{W}_k \mathbf{x}$ e selecionar a função custo: Sistema Pearson ou função não-linear de acordo com a Equação (3.49).
2. Caso o sistema Pearson seja selecionado estimar os seus parâmetros pelo método dos momentos.
3. Calcular a função escore $\varphi(\mathbf{y}_k)$ para o sistema Pearson ou para função não linear
4. Calcular a matriz de separação \mathbf{W}_k usando o algoritmo da equação (3.47) ou (3.48).
5. Se não convergir voltar ao passo 1

A Equação (3.49) mostra o intervalo $(\hat{\mu}_3^2, \hat{\mu}_4)$ para escolha da função custo usado no algoritmo ICA, baseado nas curvas de Pearson e função não-linear, sendo que o limite para todas as distribuições é $\hat{\mu}_4 = \hat{\mu}_3^2 + 1$, e os limites para fonte super-gaussianas e sub-gaussianas, escolhido através de experiências práticas, são respectivamente $\hat{\mu}_4 = 2.6$ e $\hat{\mu}_4 = \hat{\mu}_3^2 + 4$ [56].

$$\text{Sistema Pearson: } \max\{2, 6; \hat{\mu}_3^2 + 1\} \leq \hat{\mu}_4 \leq \hat{\mu}_3^2 + 4$$

$$\text{Tanh: } \begin{cases} \hat{\mu}_4 > \hat{\mu}_3^2 + 4 \\ \hat{\mu}_3^2 + 1 < \hat{\mu}_4 < 2.6 \end{cases} \quad (3.49)$$

4 Busca de projeções (*Projection Pursuit*)

Na Seção 4.1 será apresentada uma breve introdução à técnica de Busca de Projeções (*Projection Pursuit* - PP) e, finalmente, na Seção 4.2 um algoritmo baseado na técnica PP é proposto para separação de múltiplas fontes, bem como uma análise de sua relação com o método ICA.

4.1 Introdução à busca de projeções

Busca de projeções (*Projection Pursuit*), denominada no presente trabalho simplesmente por PP, constitui uma técnica estatística desenvolvida com o objetivo de obter uma visão mais estruturada, utilizando uma dimensão menor, a partir dos dados de uma dimensão maior [18, 19]. A técnica PP foi desenvolvida primeiramente para exploração (*exploratory data analysis*), mas há outras aplicações como, por exemplo, em estimativa de fdp e análise de regressão (Veja [20, 35, 36, 37, 38, 39]). Geralmente a estrutura dos dados observada na dimensão completa será também observada em uma projeção de dimensão menor e cada projeção pode fornecer uma nova introspecção.

Os precursores desta técnica são Friedman e Tukey que, em 1974, publicaram um artigo [18], propondo um algoritmo para análise exploratória de dados. Nesse artigo, propuseram um algoritmo, cuja idéia básica era associar um índice numérico, em uma ou duas dimensões, para cada projeção, de forma a caracterizar o montante de informação presente com o propósito de melhor explorar os dados. Uma vez que não seria prático tentar mapear todas as possíveis projeções, esse índice então é maximizado em relação aos parâmetros de interesse, denominado por eles como *interestingness*, definindo, então, as projeções convenientes. Por exemplo, PCA pode ser considerada um método de PP na qual o parâmetro de interesse é a variância total dos dados projetados, ou seja, a PCA procura pela contribuição de cada componente na variância total dos dados [20, 21].

Sejam $X \in \mathbb{R}^d$ uma projeção de uma variável aleatória $Z \in \mathbb{R}^p$ e A uma matriz $d \times p$ de transformação do $\mathbb{R}^p \rightarrow \mathbb{R}^d$, define-se a projeção linear pela equação (4.1).

$$X = AZ \tag{4.1}$$

Por definição, PP procura por transformações A que maximizam ou minimizam uma determinada função custo ou índice de projeção $I(f_A(X))$, em que I é a função do índice de

projeção, discutida a seguir e, $f_A(X)$ é a fdp de X para uma dada transformação A . Para simplificar a notação será usado simplesmente $I(X)$.

Deve-se notar que não há apenas interesse no extremo global, mas, sim, nos extremos locais. Para compreender melhor essa situação considere, por exemplo, a PCA. Na PCA tem-se uma componente principal a cada máximo local, portanto isso permite analisar a contribuição de cada componente na variância total e não apenas a componente com maior contribuição.

Para o caso exploratório unidimensional, ou seja, $d = 1$, A reduz-se a um vetor linha a^T . Dessa maneira procura-se o vetor a^T , que maximiza/minimiza um determinado índice de projeção, de tal forma que $f_a(X)$ seja relativamente bem estruturada (mais distante possível da gaussiana) [23, 19].

Da mesma maneira que o branqueamento facilita os cálculos no modelo ICA, ele também é muito útil no método PP, permitindo que seja evitada a determinação das variâncias em cada projeção e, portanto, diminuindo o esforço computacional.

Sabe-se que as direções de maior interesse são aquelas nas quais a distribuição seja o distante da gaussian possível [20, 21]. Da teoria da informação, sabe-se que a função densidade de probabilidade mais imprevisível, e que possui máxima entropia, é a gaussiana [17]. Uma forma intuitiva para chegar a essa conclusão é através do teorema central do limite: quanto maior a dimensão, mais próxima da gaussiana é a distribuição; ou seja, menos interessante. Baseando-se nisto, projeções na direção de distribuições mais distantes da gaussiana possível são desejáveis.

Para escolha do índice de projeção, devem-se considerar os seguintes requisitos: a função do índice I deve ser invariante afim; se X e Y forem variáveis aleatórias independentes com variância finita, então, $I(X + Y) \leq \max(I(X), I(Y))$ deve ser satisfeita [20]. Como exemplo de um índice de projeção, veja a equação (4.2).

$$I(X) = |\text{curt}(X)| \quad (4.2)$$

Diferentes índices de projeção, baseados em varias medidas de interesse, aplicados na análise de busca de projeções podem ser encontrados em [18, 19, 20, 21].

4.2 Separação de fontes por meio de busca de projeção

Agora será avaliado como o método de busca de projeções está relacionado com o modelo ICA e um algoritmo para separação de fontes baseado em busca de projeções será

proposto. Anteriormente foi afirmado que as projeções de maior interesse são aquelas que estão na direção de distribuições mais distantes da gaussiana possíveis, e isso é exatamente o que é necessário para estimar o modelo ICA, ou seja, as componentes independentes podem ser encontradas, procurando várias direções que maximizam a não-gaussianidade, com o auxílio de uma medida de normalidade como, por exemplo, a curtose.

Seja $\mathbf{y} = \mathbf{w}^T \mathbf{z}$ um sinal extraído pelo vetor de peso \mathbf{w} , de um conjunto de m misturas transformadas \mathbf{z} . Pode-se observar que rotacionando o vetor \mathbf{w} ao redor da origem, a curtose do sinal extraído será máxima exatamente quando $\mathbf{y} = \mathbf{s}$, e o vetor \mathbf{w} será ortogonal aos eixos projetados [12]. Dessa forma, assumindo que, na prática a maioria das fontes são super-gaussianas e que o conjunto de misturas foi previamente centralizado e branqueado, é possível derivar um algoritmo do gradiente ascendente para encontrar o vetor \mathbf{w} , que maximiza a curtose do sinal estimado [23].

A curtose de \mathbf{y} e o gradiente de sua curtose estão descritos pelas equações (4.3) e (4.4), respectivamente.

$$\text{curt}(\mathbf{y}) = \text{curt}(\mathbf{w}^T \mathbf{z}) = E\{(\mathbf{w}^T \mathbf{z})^4\} - 3 \quad (4.3)$$

$$\frac{\partial \text{curt}(\mathbf{w}^T \mathbf{z})}{\partial \mathbf{w}} \propto E\{\mathbf{z}(\mathbf{w}^T \mathbf{z})^3\} \quad (4.4)$$

Pode-se verificar que o gradiente afeta tanto o comprimento, quanto o ângulo do vetor \mathbf{w} . No entanto, somente o ângulo é importante no algoritmo, pois o seu comprimento altera somente a amplitude do sinal extraído e não altera a sua forma. Portanto, deve-se restringir o vetor \mathbf{w} à norma unitária. A Tabela 4.1 apresenta o algoritmo de busca de projeção.

Tabela 4.1: Algoritmo de busca de projeção pelo gradiente ascendente.

- | |
|--|
| <ol style="list-style-type: none"> 1. Escolher um vetor inicial \mathbf{w} e um valor inicial para o passo de adaptação μ. 2. $\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \cdot E\{\mathbf{z}(\mathbf{w}_n^T \mathbf{z})^3\}$ 3. $\mathbf{w}_{n+1} = \frac{\mathbf{w}_{n+1}}{ \mathbf{w}_{n+1} }$ 4. $\mathbf{w}_n = \mathbf{w}_{n+1}$ 5. Se não convergiu voltar ao passo 2. |
|--|

Considerando apenas os algoritmos ICA convencionais em que uma componente é encontrada de cada vez pela maximização de um índice de interesse, por exemplo, a não-gaussianidade; pode-se considerá-los como sendo um caso particular do método PP, cujas projeções obtidas maximizando esse índice são as fontes desejadas que resolvem o modelo ICA.

De certa forma, considerando o exposto, torna-se possível afirmar que determinados algoritmos ICA podem ser considerados como um caso particular do método PP. De fato, se ao aplicar o algoritmo PP maximizando um índice de interesse, o resultado obtido não respeitar as condições do modelo ICA, o resultado será simplesmente as projeções que maximizam esse índice.

Para separação de múltiplas fontes é necessário remover cada fonte extraída do conjunto de misturas restante aplicando uma ortogonalização, por exemplo, GSO (*Gram-Schmidt Orthogonalization*) e repetir os passos do algoritmo da Tabela 4.1, para que seja encontrada uma fonte diferente a cada repetição. Esse processo pode ser repetido até que todas as fontes sejam extraídas.

A GSO assegura que cada sinal extraído, y_i , seja ortogonal a todas as misturas dos sinais a serem extraídos.

Seja $\mathbf{x}^0 = (x_1^0, \dots, x_m^0)$ o conjunto original de misturas, na qual o número 0 (zero) sobrescrito representa o conjunto original de misturas. Ao obtermos o primeiro vetor, \mathbf{w}_1 , a primeira fonte extraída pode ser expressa pela Equação (4.5).

$$y_1 = \mathbf{w}_1^T \mathbf{x}^0 \quad (4.5)$$

Então, de acordo com a Equação (4.6), é aplicada a GSO para removermos o sinal da fonte, y_1 encontrada, de cada sinal de mistura x_i^0 , garantindo que seja encontrada uma fonte diferente na próxima iteração.

$$x_i^1 = x_i^0 - \frac{E\{y_1 x_i^0\} y_1}{E\{y_1^2\}} \quad (4.6)$$

Resulta então um novo conjunto de misturas x_i^1 num espaço ortogonal à fonte y_1 extraída, ou seja, $E\{x_i^1 y_1\} = 0$ para $i = \{1, \dots, m\}$. Logo, aplicando o algoritmo da Tabela 4.1 ao novo conjunto de misturas \mathbf{x}^1 será extraída uma nova fonte y_2 . O conjunto de misturas \mathbf{x}^1 pode ser agora ortogonalizado para a fonte y_2 e esse processo pode ser repetido até que todas as fontes sejam extraídas.

Portanto, esse procedimento permite a separação de múltiplas fontes, de tal forma que seja garantida a extração de uma fonte diferente a cada repetição da PP. A vantagem desse procedimento é que é possível extrair somente as fontes desejadas.

Embora o desenvolvimento matemático envolvido em ICA e PP seja intimamente relacionado e mais ou menos equivalente no caso básico, é importante enfatizar que existem

diferenças na aplicação, na teoria e também na origem de ambos os métodos. A ICA foi originalmente criada para resolver um problema concreto de separação; ao passo que PP foi originalmente desenvolvido para visualização em análise exploratória. Na ICA, o objetivo é a estimativa das componentes que sejam estatisticamente independentes, embora também possa ser usada em algumas situações para redução de dimensão; já o objetivo da PP é a redução dimensional de um espaço de dimensão elevada para visualização.

Se a estrutura temporal dos sinais e o sistema de mistura forem considerados, a ICA será muito diferente da PP. Como pode ser notado (veja, por exemplo [23]), se ao aplicar o método PP descrito nesta seção, o modelo da ICA não for satisfeito, ou seja, não forem obtidas as componentes independentes estatisticamente, restarão somente como resultado as projeções dos dados em um novo espaço, nas quais não é garantida a extração das fontes.

5 Experimentos e análise dos algoritmos

Neste capítulo, primeiramente é abordado o método de medição usado para comparação entre os algoritmos ICA. O método de medição exposto na Seção 5.1 só pode ser aplicado se as fontes originais forem conhecidas. Quando as fontes não são conhecidas, deve ser usado algum método alternativo para avaliação dos algoritmos, a ser considerado diretamente na seção dos experimentos. Na Seção 5.1, também é introduzida a análise de qualidade da separação em problemas de separação cega de fontes.

Na Seção 5.2.1, é realizado um experimento simples com o algoritmo FastICA (um dos mais utilizados atualmente) para apresentar uma visão geral do problema de separação de sinais de áudio. Nas demais seções serão abordados e analisados diversos experimentos para avaliação e comparação dos algoritmos FastICA, PP e PearsonICA.

Além dos códigos fonte, todos os sinais, utilizados nos experimentos estão disponíveis em [128].

5.1 Método de medição

Para avaliar o desempenho dos algoritmos de separação de fontes foi usado, quando possível, o método desenvolvido em [76], no qual o sinal estimado $\hat{s}(t)$ da fonte $s_i(t)$ é decomposto em uma soma de quatro parcelas, como mostra a equação abaixo.

$$\hat{s}(t) = s_{\text{desejada}}(t) + e_{\text{interf}}(t) + e_{\text{ruído}}(t) + e_{\text{artefato}}(t) \quad (5.1)$$

O sinal $s_{\text{desejada}}(t)$ é a porção referente ao sinal da fonte desejada. A degradação $e_{\text{interf}}(t)$ é causada pela interferência de fontes indesejadas, ou seja, representa a presença de resíduos de outras fontes na fonte sendo estimada. A degradação $e_{\text{ruído}}(t)$ é causada pelo ruído, por exemplo, por fontes não consideradas pelo sistema ou por eco das próprias fontes nas paredes do ambiente. E, finalmente, $e_{\text{artefato}}(t)$ é a degradação produzida por outros artefatos, que pode ser induzida pelo algoritmo de separação de fontes como, por exemplo, “ruído” musical. Em geral, para os experimentos considerados no presente trabalho, $e_{\text{ruído}}$ será nulo.

Essa metodologia só é aplicável quando estiverem disponíveis as fontes originais; caso contrário deve-se utilizar outra metodologia. O sistema denominado *BSS_Eval* foi desenvolvido em Matlab, é de código aberto e está disponível para transferência (*download*), sob a licença GNU em [78]. A partir da decomposição acima são definidas as medidas de desempenho global, expressas em decibéis nas equações abaixo.

SDR (*source-to-distortion ratio* – relação fonte-distorção):

$$SDR = 10 \log_{10} \frac{\|s_{\text{desejada}}\|^2}{\|e_{\text{interf}} + e_{\text{ruído}} + e_{\text{artefato}}\|^2}. \quad (5.2)$$

SIR (*source-to-interferences ratio* – relação fonte-interferência):

$$SIR = 10 \log_{10} \frac{\|s_{\text{desejada}}\|^2}{\|e_{\text{interf}}\|^2}. \quad (5.3)$$

SNR (*source-to-noise ratio* – relação fonte-ruído):

$$SNR = 10 \log_{10} \frac{\|s_{\text{desejada}} + e_{\text{interf}}\|^2}{\|e_{\text{ruído}}\|^2}. \quad (5.4)$$

SAR (*source-to-artifacts ratio* – relação fonte-artefato):

$$SAR = 10 \log_{10} \frac{\|s_{\text{desejada}} + e_{\text{interf}} + e_{\text{ruído}}\|^2}{\|e_{\text{artefato}}\|^2}. \quad (5.5)$$

Essas quatro medidas são baseadas na definição usual de SNR com algumas modificações. A definição de SNR possui o termo $s_{\text{desejada}} + e_{\text{ruído}}$ no numerador objetivando torná-la independente de SIR. Similarmente, a medida SAR é independente de SIR e SNR, pois o numerador possui os termos referentes à interferência e ao ruído.

A avaliação da qualidade de separação de fontes é um problema complexo que depende da área de aplicação. Em alguns casos, não se está interessado em recuperar o sinal mais próximo possível do original, mas em preservar alguma característica de interesse. Por exemplo, quando a recuperação da fonte é usada em um dispositivo de ajuda auditiva, a reconstrução perfeita da fonte original não é tão importante quanto à inteligibilidade. Em aplicações de reconhecimento, por exemplo, é mais importante preservar algumas características acústicas que a inteligibilidade. Além disso, poucos trabalhos têm explorado, com propósito específico, a qualidade na separação em BSS [135].

Para a avaliação de algoritmos BSS diferentes alternativas têm sido usadas, geralmente derivadas de outras áreas do processamento de sinais. Esses métodos podem ser classificados em duas áreas: avaliação subjetiva, na qual é verificada alguma qualidade subjetiva percebida a partir dos sinais estimados [136], ou diferenças entre as formas das fontes estimadas e as fontes originais [56, 136]; e avaliação objetiva, na qual alguma quantidade numérica é associada diretamente à qualidade da separação [76, 131, 135]. Em [135], por exemplo, é usado teste de reconhecimento de fala, como uma forma de avaliar o desempenho de

algoritmos BSS em aplicações de reconhecimento automático de fala (ASR – *Automatic Speech Recognition*).

No presente trabalho a separação de sinais de áudio não é avaliada para uma aplicação específica. Nem se deseja explorar o assunto em detalhe, pois definir padrões e avaliar a qualidade da separação para um objetivo específico não é o foco deste estudo. Porém, deve-se notar que esse é um tópico importante, e necessário, a ser considerado em aplicações específicas.

Dessa forma, sempre que possível, foram usados métodos objetivos; através da relação SIR e do método proposto na Seção 5.2.3, baseado na função coerência. Também foram realizadas avaliações subjetivas para fornecer subsídios e enriquecer as métricas usadas na comparação. As avaliações subjetivas usadas são: a comparação visual entre as fontes estimas e originais; e a inteligibilidade das fontes estimadas comparadas com as originais, considerando acuidades auditivas normais.

5.2 Experimentos

Uma visão geral é apresentada na Seção 5.2.1 e, na seqüência, são apresentados os experimentos realizados, que estão divididos em dois grupos: misturas instantâneas e misturas convolutivas, nas seções 5.2.2 e 5.2.3, respectivamente.

Para simplificar a notação, os sinais das fontes originais serão representados por s_i , em que o índice i representa a ordem, na qual a fonte foi usada no experimento, e as fontes estimadas estão representadas por \hat{s}_i referente à fonte original s_i . Na descrição dos experimentos a seguir, as fontes são citadas na mesma ordem em que elas foram usadas nas análises.

Os algoritmos de separação admitem alguns parâmetros adicionais que permitem ajustes, como por exemplo, o critério de parada na execução, o erro máximo admitido na estimativa da matriz de misturas e o número máximo de iterações permitidas para convergência, são exemplos de parâmetros comumente encontrados nos algoritmos. Todos os experimentos realizados nesta seção foram feitos usando os parâmetros definidos como padrão nos algoritmos, exceto quando especificado.

5.2.1 Visão geral

Para fornecer uma visão geral do problema de separação de fontes, considere, no momento, somente o algoritmo FastICA.

Considere os sinais das falas do filme *Harry Potter* da Figura 1.2. Misturando-se esses sinais (Figura 1.3) e aplicando-se a aproximação “*deflation*” do FastICA, baseada na curtose, são obtidos os sinais representados na Figura 5.1. As componentes (fontes) foram obtidas após 6, 7 e 2 iterações respectivamente, demonstrando que o algoritmo conseguiu convergir rapidamente para a solução do problema. Também é possível verificar que, visualmente, os sinais estimados são muito semelhantes aos sinais de suas respectivas fontes, exceto pelo fato de, eventualmente, estarem permutados e de um fator de escala como, já foi mencionado anteriormente. E de fato o sinal audível resultante da aplicação do algoritmo é muito semelhante ao sinal original.

Como visto anteriormente, o algoritmo FastICA não requer passo de adaptação e, portanto, não é necessária a passagem de nenhum outro parâmetro para a separação. Sendo assim, é suficiente fornecer somente os vetores das misturas para que sejam estimadas as fontes.

Embora o algoritmo FastICA seja o mais difundido, até pouco tempo a sua velocidade, bem como suas características ainda não haviam sido bem fundamentadas, nem comparadas com outros métodos. Têm surgido estudos mais detalhados em aplicações específicas, verificando que em algumas situações o algoritmo FastICA não apresenta um bom desempenho [60, 61, 63, 73].

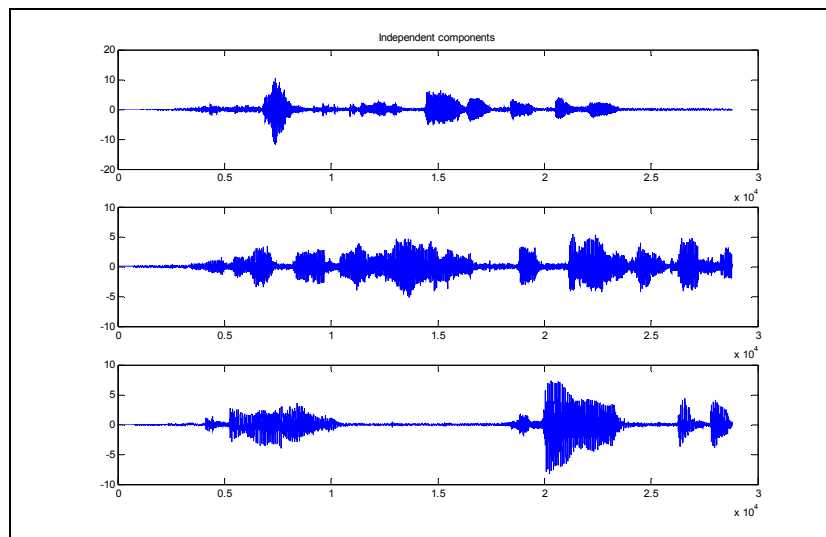


Figura 5.1: Componentes independentes extraídas aplicando FastICA baseado na curtose.

Tichavský *et al.* [60] mostram que o algoritmo FastICA pode ter sua convergência prejudicada, considerando o limite de Cramér-Rao (CRB – *Cramér-Rao Bound*) e, nesse caso, propõem uma adaptação ao FastICA, denominado Smart FastICA, na qual é adicionado um método para escolha da função não-linear usada na estimativa. O CRB é o limite inferior, sob

certas condições (veja Apêndice A), para a variância do estimador, que, no caso da ICA, é a estimativa do parâmetro vetor \mathbf{w} . Pode-se observar também que se o CRB não existir, por exemplo, se as fontes tiverem magnitude limitada ou com distribuição com cauda longa, a variância da estimativa produzida pelo FastICA aproxima-se de zero; caso contrário, com a escolha adequada da função não linear, ela se aproxima assintoticamente do CRB. Entretanto, se o CRB for infinito, o algoritmo pode falhar, pois a convergência se torna muito lenta. Uma variação eficiente do algoritmo FastICA, que satisfaz o CRB é proposta em [61]. Uma avaliação mais detalhada do exposto não faz parte do escopo deste trabalho.

É sabido que sinais de voz, em geral, podem ser modelados pela distribuição laplaciana (exponencial bilateral) ou gama, assim satisfazendo algumas condições necessárias para convergência do algoritmo. Em alguns casos os sinais de voz podem ser modelados por distribuições gaussianas [139], porém não é apropriado para o caso da BSS. Portanto, pode-se supor que o algoritmo FastICA, geralmente, apresenta um bom desempenho para estes sinais, porém não se pode fazer essa suposição para sinais de áudio, de forma geral.

Não há como saber trivialmente se o algoritmo FastICA, ou outros baseados nos mesmos princípios, satisfazem condições necessárias para garantir uma convergência. Portanto, nos experimentos realizados nesta pesquisa, essas condições são assumidas como satisfeitas para aplicação dos algoritmos, sem realmente serem verificadas.

É visada a influência causada sobre os algoritmos ICA devido a vários fatores como, por exemplo, a distribuição estatística das fontes, a influência do ambiente, dos sensores e aproximações usadas com objetivo de buscar melhores alternativas em cada caso.

5.2.2 Misturas instantâneas

Nesta seção estão descritos os principais resultados obtidos nos experimentos realizados com misturas instantâneas, na qual foi utilizado o Matlab para misturar os sinais. Foram realizados cinco experimentos, nos quais foi possível avaliar a separação com o uso da relação SIR, já que em todos os cinco casos, as fontes originais são pré-definidas.

No primeiro experimento (Experimento 1) foram considerados dois sinais de áudio. O primeiro é a leitura de um trecho de texto de uma revista científica, gravado com um microfone comum no PC com o formato de 8 kHz, 16 bits, mono; e o segundo é o trecho de uma música clássica reamostrado para o formato de 8 kHz, 16 bits, mono [128]. Os sinais foram misturados pela matriz da Equação (5.6), gerada aleatoriamente no Matlab.

$$M = \begin{bmatrix} 0.95 & 0.61 \\ 0.23 & 0.49 \end{bmatrix}. \quad (5.6)$$

Pode-se observar na Figura 5.2 que as fdp's dos sinais do Experimento 1 são super-gaussianas. Os resultados mostraram que todos os algoritmos desempenharam bem o seu papel na separação dos sinais, conforme esperado.

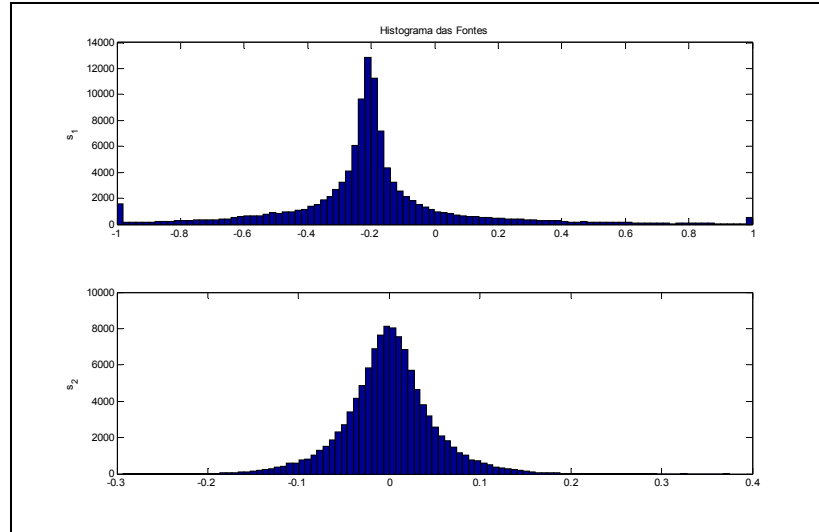


Figura 5.2: Histograma das fontes do Experimento 1: Leitura de trecho do texto em uma revista científica e trecho de uma música clássica.

O valor médio de SIR e o desvio padrão, para 100 repetições, dos três algoritmos avaliados no Experimento 1 estão representados na Tabela 5.1.

Tabela 5.1: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 1.

	FastICA (Simétrica)			PP	PearsonICA
	Pow3	Tanh	Gauss		
\hat{s}_1	42 (1)	50 (1)	52 (5)	42 (4)	50 (1)
\hat{s}_2	40 (1)	51 (2)	56 (7)	28 (8)	51 (2)

Verificou-se que a eficiência global do FastICA usando a separação simétrica foi sutilmente melhor que a separação *deflation*. Em ambas as aproximações, quando aplicada a função custo não-linear Gauss e Tanh na maximização, foram obtidos valores de SIR superiores (Tabela 5.1), significando que a matriz de separação estimada W se aproxima mais de A^{-1} . Isso era esperado, pois as funções não lineares são mais robustas na estimativa da negentropia que funções baseadas em momentos ou cumulantes.

O algoritmo de busca de projeção é baseado no algoritmo do gradiente, portanto, como esperado, a sua convergência depende da boa escolha do passo de adaptação. Quando o passo

de adaptação é muito pequeno ($0 < \mu \ll 1$), o algoritmo demora muito para convergir, por exemplo, o número de iterações necessárias para convergência usando $\mu = 0.01$ é 30 vezes maior, em média, em relação à $\mu = 1$. Além disso, o valor de SIR permanece praticamente constante para passos de adaptação no intervalo analisado, [0.01, 12.51]; Portanto, verificou-se, experimentalmente, que uma boa escolha para o passo de adaptação é $\mu = 1$, pois são necessárias menos iterações para convergência.

Os resultados mostram que outro fator que pode afetar radicalmente o número de iterações necessárias para a convergência do algoritmo de Busca de Projeções é a escolha do vetor de estimativa inicial, \mathbf{w}_0 . Para 100 repetições, usando valores aleatórios para estimativa inicial e $\mu = 1$, o número médio de iterações para as fontes \hat{s}_1 e \hat{s}_2 foram 36 e 43 com desvio padrão de 20 e 25 iterações, respectivamente.

Os resultados obtidos para o algoritmo PearsonICA foram similares aos obtidos pelo FastICA para a função Tanh, pois o algoritmo PearsonICA foi baseado no algoritmo FastICA e no uso da função não-linear Tanh como função custo, além das curvas de Pearson para distribuições próximas a curva de Gauss.

Nos experimentos 2, 3 e 4 foram consideradas quatro fontes geradas sinteticamente no Matlab. Nesses experimentos foi avaliado o comportamento dos algoritmos simulando trechos de sinais com alguma característica específica como periodicidade, descontinuidade e a forma de sua distribuição.

No Experimento 2 foram utilizados: um sinal senoidal, um sinal periódico com descontinuidades, um sinal dente de serra e um ruído impulsivo, veja Figura 5.9; no Experimento 3, o ruído impulsivo foi substituído por um ruído gaussiano. Finalmente, no Experimento 4 o sinal periódico e o ruído impulsivo foram substituídos por sinais gaussianos.

Tabela 5.2: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 2.

	FastICA (Simétrica)			PP	PearsonICA
	Pow3	Tanh	Gauss		
\hat{s}_1	27 (0)	39 (1)	44 (0)	1 (0)	39 (1)
\hat{s}_2	25 (5)	33 (3)	35 (1)	17 (4)	33 (3)
\hat{s}_3	29 (2)	27 (0)	27 (0)	1 (0)	27 (0)
\hat{s}_4	19 (3)	24 (2)	25 (0)	14 (2)	24 (2)

Os histogramas das fontes do Experimento 2 estão representados na Figura 5.3. A matriz da Equação(5.7), gerada aleatoriamente, foi usada nos experimentos 2, 3 e 4 para criar as misturas.

$$M = \begin{bmatrix} 0.82 & 1.19 & -1.60 & -0.81 \\ 0.71 & -1.20 & 0.26 & 0.53 \\ 1.29 & -0.02 & -1.06 & 0.22 \\ 0.67 & -0.16 & 1.42 & -0.92 \end{bmatrix} \quad (5.7)$$

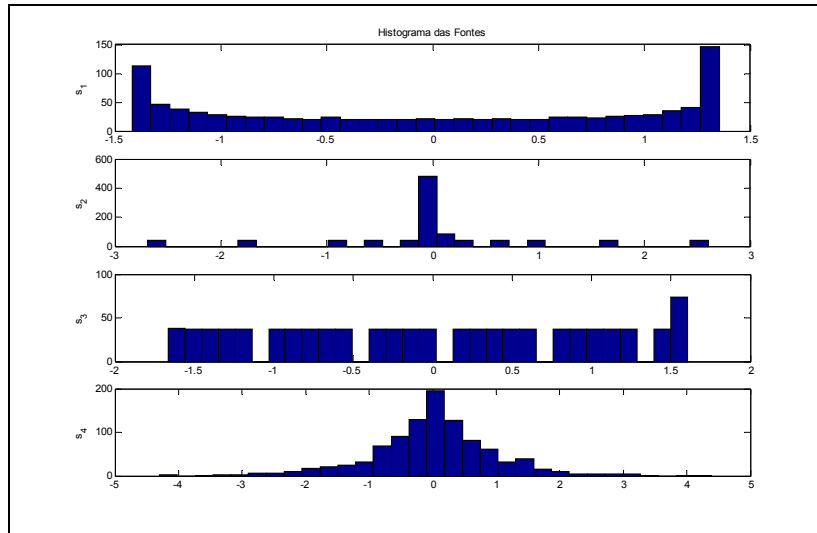


Figura 5.3: Histograma das fontes do Experimento 2: um sinal senoidal, um sinal periódico com descontinuidades, um sinal dente de serra e um ruído impulsivo.

Para o Experimento 2 pode-se verificar que as fontes s_1 e s_3 são sub-gaussianas e as fontes s_2 e s_4 super-gaussianas. No Experimento 3 a fonte s_4 é Gaussiana e no Experimento 4 as fontes s_2 e s_4 são Gaussianas, as demais fontes em ambos os casos são como no Experimento 2.

Verificou-se que os algoritmos são basicamente influenciados pela forma da distribuição das fontes.

O algoritmo FastICA e uso da função custo Pow3 apresentou melhores resultados para as fontes com distribuição sub-gaussianas. No entanto, para as demais funções custo, nada se pode afirmar apesar da separação ser concluída com sucesso.

Verificou-se também que o algoritmo FastICA conseguiu extrair, satisfatoriamente (Figura 5.4), todas as fontes na presença de apenas uma fonte gaussiana, porém não conseguiu extrair todas as fontes, satisfatoriamente (Figura 5.5), na presença de mais de uma fonte gaussiana, como havia sido previsto teoricamente.

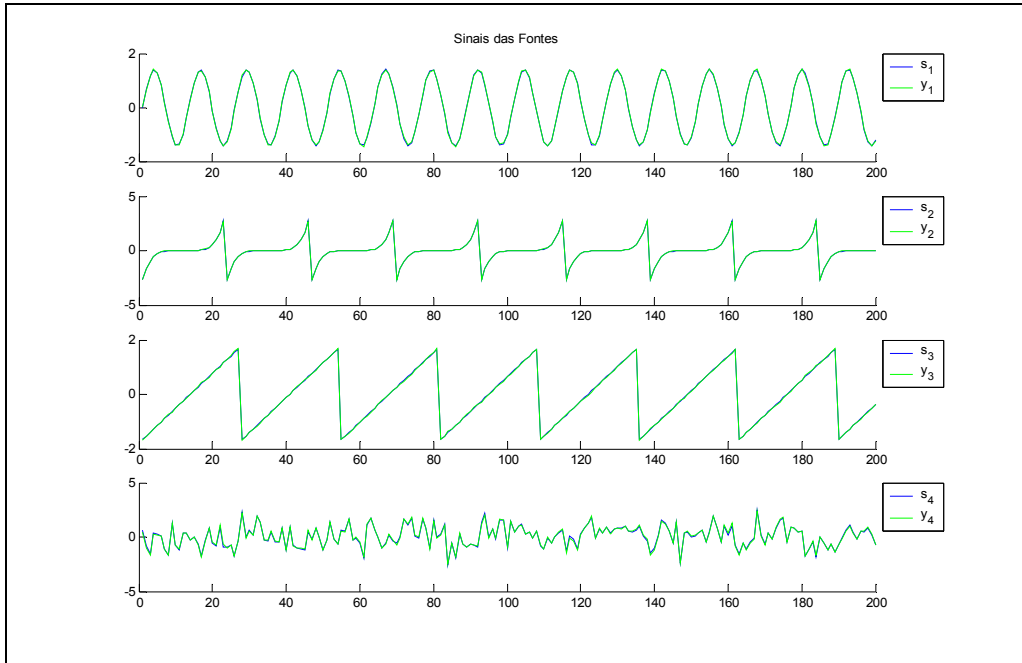


Figura 5.4: Primeiras 200 amostras, normalizadas, das fontes originais (azul) do Experimento 3 (apenas uma fonte gaussiana) e suas respectivas estimativas (verde), normalizadas, usando o método simétrico e a função custo gaussiano do algoritmo FastICA.

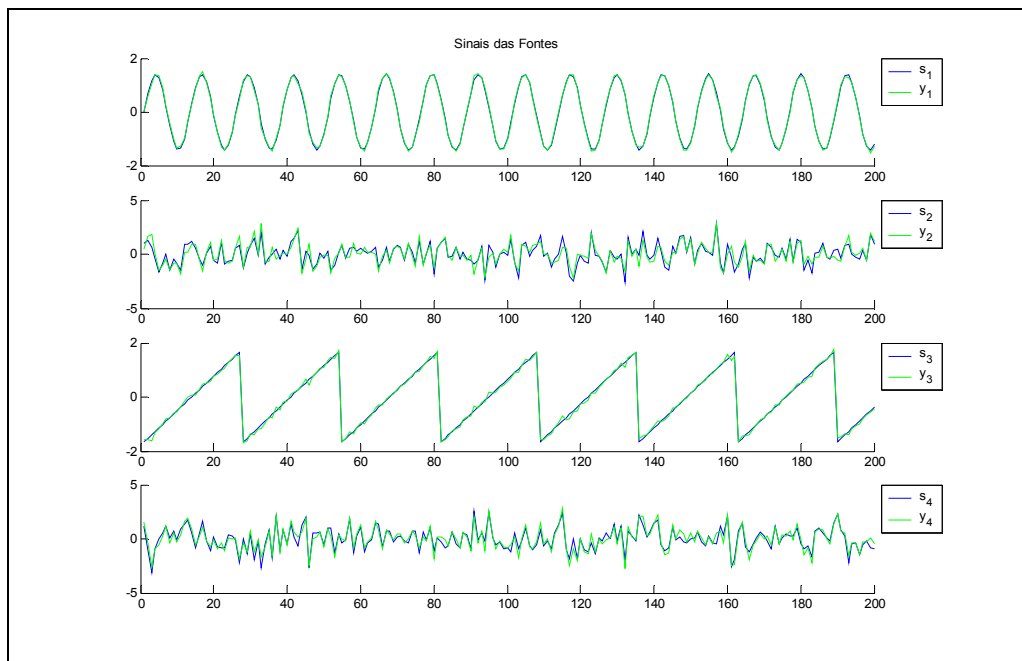


Figura 5.5: Primeiras 200 amostras, normalizadas, das fontes originais (azul) do experimento 4 (duas fonte gaussiana) e suas respectivas estimativas (verde), normalizadas, usando o método simétrico e a função custo gaussiano do algoritmo FastICA.

No Experimento 4, as aplicações da aproximação *deflation* do FastICA, usando as funções custo Pow3 e Tanh, falharam em quase todas as tentativas de separação, ou seja, foi atingido o máximo número de iterações sem obter um erro menor ou igual ao erro máximo

admitido. Isso se deve ao fato de haver duas fontes gaussianas. Verificou-se também que esse comportamento permanece mesmo aumentando o número máximo de iterações como esperado. Apesar de o algoritmo FastICA convergir para as demais funções, os valores de SIR foram muito pequenos para as fontes s_2 e s_4 , veja a Tabela 5.4, ou seja, os valores estimados dessas fontes podem não ser boas aproximações das fontes originais dependendo da aplicação, como pode ser observado na Figura 5.5.

Tabela 5.3: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 3.

	FastICA (Simétrica)			PP	PearsonICA
	Pow3	Tanh	Gauss		
\hat{s}_1	27 (0)	32 (0)	33 (0)	-7 (15)	28 (0)
\hat{s}_2	41 (0)	46 (0)	44 (0)	33 (0)	43 (0)
\hat{s}_3	30 (0)	34 (0)	35 (0)	1 (1)	38 (1)
\hat{s}_4	21 (0)	22 (0)	23 (0)	4 (0)	22 (0)

Tabela 5.4: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 4.

	FastICA (Simétrica)			PP	PearsonICA
	Pow3	Tanh	Gauss		
\hat{s}_1	22 (0)	25 (0)	25 (1)	-5 (4)	32 (0)
\hat{s}_2	8 (0)	8 (0)	8 (1)	1 (0)	11 (0)
\hat{s}_3	22 (0)	23 (0)	23 (1)	-1 (1)	25 (0)
\hat{s}_4	7 (0)	9 (0)	8 (1)	2 (1)	10 (0)

Como havia sido previsto teoricamente, para o algoritmo PearsonICA, houve uma melhora na separação com mais de uma fonte mesocúrtica, no entanto essa melhora é modesta, como pode ser observado na Tabela 5.4.

O algoritmo PP exposto na Seção 4, como presumido, só conseguiu extrair fontes super-gaussianas, veja as tabelas Tabela 5.2, Tabela 5.3 e Tabela 5.4, pois na sua dedução foi assumido que a maioria das fontes, na prática, são super-gaussianas.

Não houve convergência para a estimativa esperada, no Experimento 2, para a função custo Tanh com aproximação simétrica do algoritmo FastICA, usando a estimativa inicial W_0 , definida pela Equação (5.8). Isso pode ocorrer, pois o ponto fixo pode variar com a escolha da estimativa inicial, fazendo com que o algoritmo encontre uma solução diferente da solução esperada. Diminuindo-se o erro máximo permitido (critério de parada) no algoritmo,

pode-se forçar o algoritmo continuar na busca até encontrar a solução esperada. No entanto, é preferível fazer um esquema de realimentação, no qual o algoritmo é executado novamente para as misturas usando a estimativa atual dos pesos W como estimativa inicial da próxima execução, conforme mostra a Figura 5.6. Esse processo pode ser repetido algumas vezes para garantir a convergência do algoritmo.

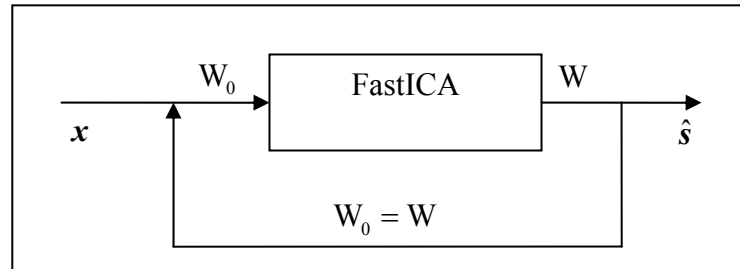


Figura 5.6: Esquema de realimentação para melhorar a convergência do algoritmo FastICA.

O gráfico da Figura 5.7 mostra o erro de convergência do algoritmo FastICA usando critério de parada $\varepsilon = 1 \times 10^{-5}$ (padrão $\varepsilon = 1 \times 10^{-4} = -40dB$) no Experimento 2, na qual nota-se que o algoritmo para a execução na quinta iteração com o valor de ε padrão, veja a Figura 5.8. Também nota-se que após a quinta iteração o ponto fixo se altera, e o algoritmo atende ao novo critério de parada ($\varepsilon = 1 \times 10^{-5}$) somente na iteração 16. Nessa condição a algoritmo conseguiu convergir para a estimativa esperada, separando as fontes corretamente, veja a Figura 5.9. Então, para o Experimento 2, é suficiente repetir o algoritmo somente uma vez.

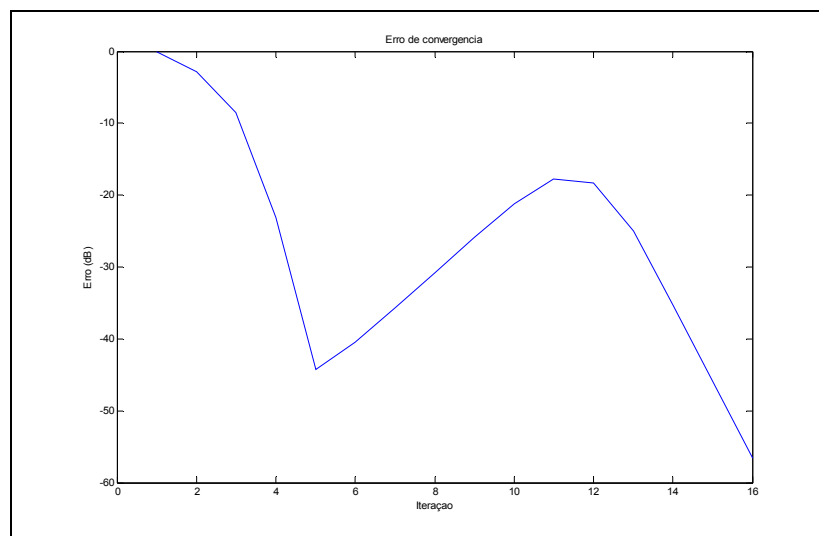


Figura 5.7: Gráfico do erro de convergência do algoritmo FastICA (método simétrico) para o Experimento 2, com critério de parada $\varepsilon = 1 \times 10^{-5}$

As figuras Figura 5.8 e Figura 5.9 apresentam as fontes estimadas do algoritmo FastICA, usando a função custo Tanh e a aproximação simétrica no Experimento 2, para a estimativa

inicial W_0 da Equação (5.8), e após a repetição de uma vez do algoritmo usando o novo W_0 , respectivamente.

$$W_0 = \begin{bmatrix} 0.51 & 0.74 & 0.05 & 1.49 \\ -0.64 & 0.23 & 1.85 & 0.72 \\ -0.82 & -0.001 & 2.17 & -0.77 \\ 0.13 & -0.80 & -1.53 & 0.38 \end{bmatrix} \quad (5.8)$$

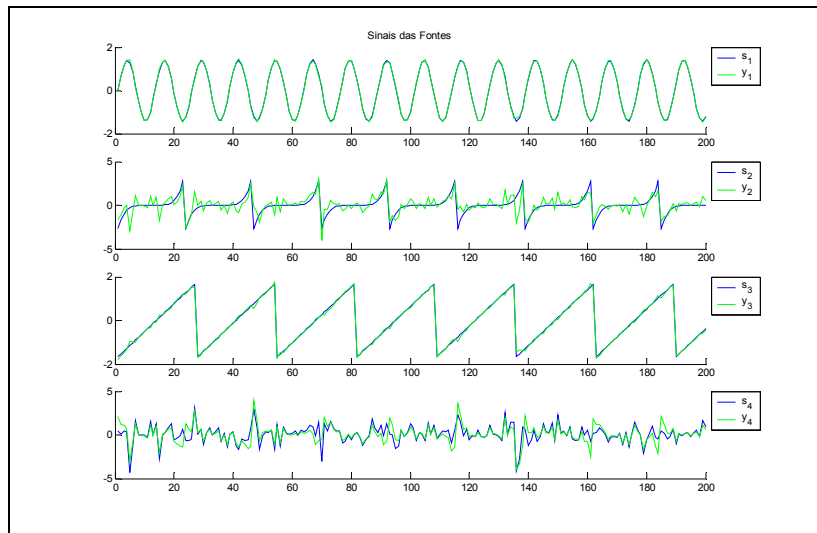


Figura 5.8: Primeiras 200 amostras, normalizadas, das fontes originais do experimento 2 e suas respectivas estimativas, normalizadas, usando o método simétrico e a função custo Tanh do algoritmo FastICA com estimativa inicial W_0 definida pela equação (5.8).

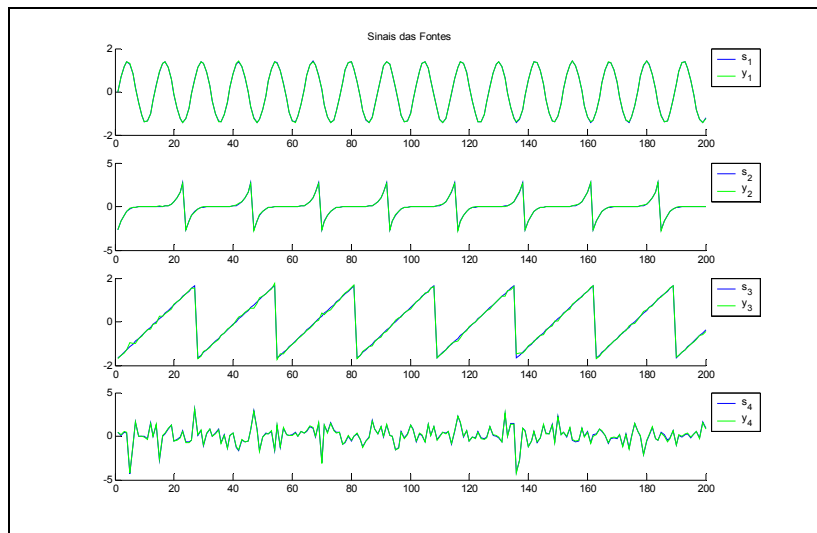


Figura 5.9: Primeiras 200 amostras, normalizadas, das fontes do Experimento 2 e suas respectivas estimativas, normalizadas, usando o método simétrico e a função custo Tanh do algoritmo FastICA após a repetição de uma vez do algoritmo com o novo W_0 igual ao W obtido na execução anterior.

O último experimento do grupo de misturas instantâneas, Experimento 5, visou a avaliar o comportamento dos algoritmos na separação de várias fontes. Foram consideradas sete fontes, sendo todas as fontes sinais de áudio, as quais foram misturadas, usando a matriz de mistura, gerada aleatoriamente da Equação 5.9. As 3 primeiras fontes são os sinais do filme Harry Potter, usados na Seção 1.1; os sinais 4 e 5 são as fontes do Experimento 1; e as fontes 6 e 7 são dois trechos de música do gênero rock.

$$M = \begin{bmatrix} 0.73 & -0.70 & -1.49 & -0.30 & 0.65 & -2.53 & 1.60 \\ 0.33 & -0.15 & 1.43 & 0.75 & 0.68 & 0.48 & -0.75 \\ 1.35 & 0.29 & -0.97 & -0.50 & -0.60 & -0.98 & -0.93 \\ -0.68 & 1.21 & -0.99 & -2.40 & 1.16 & 2.00 & -1.14 \\ 1.14 & 0.25 & 1.38 & 1.40 & 1.11 & -0.24 & 2.17 \\ -3.53 & -0.07 & 1.04 & 1.18 & -0.69 & -1.88 & 0.57 \\ -0.70 & 1.45 & 1.34 & -0.51 & 0.72 & 0.70 & -1.11 \end{bmatrix} \quad (5.9)$$

A partir dos histogramas da Figura 5.10 observa-se que a fonte s_6 é sub-gaussiana e todas as demais fontes são super-gaussianas. Também se pode observar que as fontes s_5 e s_7 são quase gaussianas.

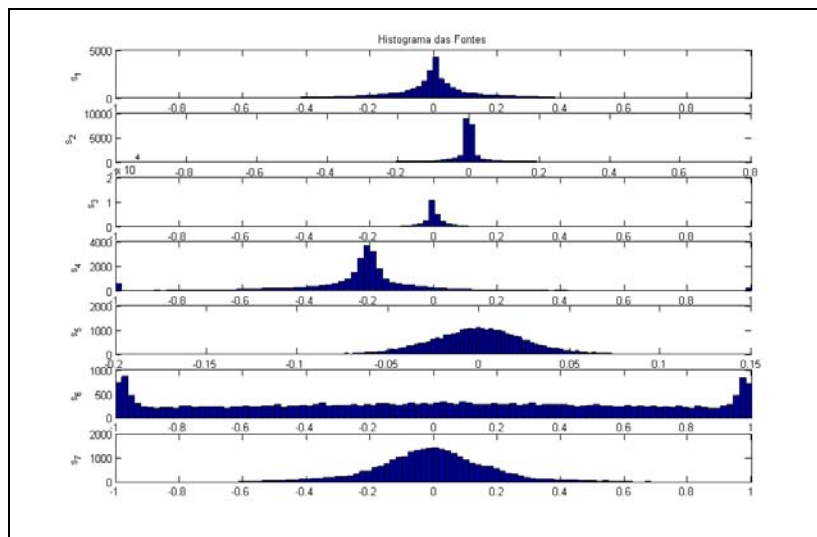


Figura 5.10: Histogramas das fontes originais do Experimento 5: as 3 primeiras fontes são os sinais do filme Harry Potter, usado na Seção 1.1; as fontes 4 e 5 são do Experimento 1 e as fontes 6 e 7 são dois trechos de música do gênero rock.

Verificou-se, que para as distribuições mais próximas de uma gaussiana, foram obtidos valores de SIR inferiores em todos os algoritmos, porém em nenhum dos experimentos houve falha (não houve convergência) na recuperação das fontes. Também se verificou que o algoritmo PP foi capaz de separar a fonte sub-gaussiana, pois todas as outras fontes são super-

gaussianas. Ou seja, o algoritmo PP ainda é capaz de separar todas as fontes na presença de apenas uma fonte gaussiana ou sub-gaussiana.

Os valores de SIR médio, de 100 repetições, dos três algoritmos na separação das fontes do Experimento 5 estão apresentados na Tabela 5.5.

Tabela 5.5: SIR (dB) médio (desvio-padrão) de 100 repetições na separação dos sinais do Experimento 5.

	FastICA (<i>Deflation</i>)			PP	PearsonICA
	Pow3	Tanh	Gauss		
\hat{s}_1	33 (0)	35 (1)	34 (1)	18 (6)	37 (0)
\hat{s}_2	30 (0)	39 (0)	39 (0)	24 (3)	35 (1)
\hat{s}_3	31 (0)	35 (0)	37 (0)	24 (6)	33 (0)
\hat{s}_4	31 (7)	36 (6)	35 (6)	35 (6)	39 (0)
\hat{s}_5	16 (0)	20 (0)	21 (0)	14 (0)	27 (0)
\hat{s}_6	20 (3)	27 (0)	30 (0)	19 (0)	32 (0)
\hat{s}_7	20 (1)	25 (0)	26 (0)	14 (3)	33 (2)

5.2.3 Misturas convolutivas

Todas as misturas dos experimentos realizados nesta seção foram obtidas no laboratório de acústica do LPS na EPUSP. Os experimentos foram divididos em três grupos: no primeiro grupo, há dois sensores e duas fontes; no segundo, há três sensores e duas fontes; e finalmente no último caso há dois sensores e a reprodução de duas fontes gravadas previamente.

O modelo ICA proposto no trabalho não considera que as misturas possam ser convolutivas. Portanto, pode-se supor que os algoritmos não funcionassem. Esses experimentos têm como objetivo avaliar o comportamento dos algoritmos para misturas instantâneas aplicados na separação de misturas convolutivas.

Nos experimentos dos grupos nas quais as fontes originais não estão disponíveis, não há como usar a relação SIR para avaliar a separação das fontes. Logo, para avaliar o resultado da separação é necessário criar um método que avalie os sinais estimados sem a necessidade de conhecer os sinais originais. Em [130] os autores propõem a separação de sinais multi-canal por decorrelação, na qual demonstram que, para sinais não-estacionários, é suficiente usar estatísticas de segunda ordem no critério de separação. Uma forma de avaliar o grau de

dependência linear de duas séries temporais é avaliar a sua correlação, a partir de seus espectros de potência. A função coerência, apresentada no apêndice B, tem esse objetivo.

A função coerência pode ser usada como critério de separação, como realizado em [131]. Tendo em vista essas idéias, tal função foi usada para medir a correlação entre as fontes no domínio da frequência. A função coerência pode não ser suficiente para garantir que os sinais tenham sido separados. No entanto, se os sinais possuírem alguma correlação no domínio da frequência, sabe-se que o algoritmo falhou. Sendo assim, a função coerência é suficiente para provar a falha do algoritmo. A estimativa da função coerência utilizada nos experimentos foi a função do Matlab *cohere*, tendo sido empregado comprimento de 1024 para os cálculos da FFT.

A configuração da sala usada no primeiro experimento do caso convolutivo (Experimento 6) é apresentada na Figura 5.11. Os dois sensores (microfones) são: um Shure e um AKG D880 e as fontes são: uma música instrumental de fundo e uma pessoa do sexo masculino contando de 1 a 10.

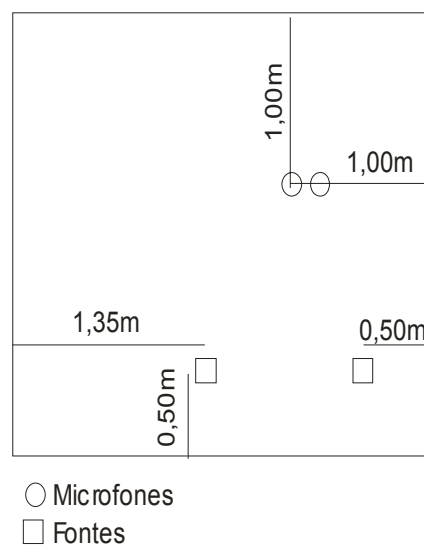


Figura 5.11: Configuração da sala acústica no Experimento 6.

Nos experimentos 7, 8 e 9 foram usados três sensores (microfones): um Shure, um AKG D880 e um M-Audio NOVA. As fontes dos experimentos 7 e 9 são as mesmas do Experimento 6 e as duas fontes do Experimento 8 são: uma música do gênero rock (trecho instrumental) de fundo e uma pessoa do sexo masculino contando de 1 a 10. A configuração da sala, usada nos experimentos 7, 8 e 9 pode ser vista na Figura 5.13.

Nenhum dos algoritmos obteve sucesso na separação das fontes. Em todos os experimentos foram obtidos resultados semelhantes como, por exemplo, apresenta o gráfico da função coerência, na Figura 5.12, relativa aos sinais obtidos na saída do algoritmo

PearsonICA, no Experimento 6. De fato, em nenhum dos experimentos foi obtido um sinal na saída dos algoritmos, na qual houvesse somente uma música instrumental ou somente a contagem.

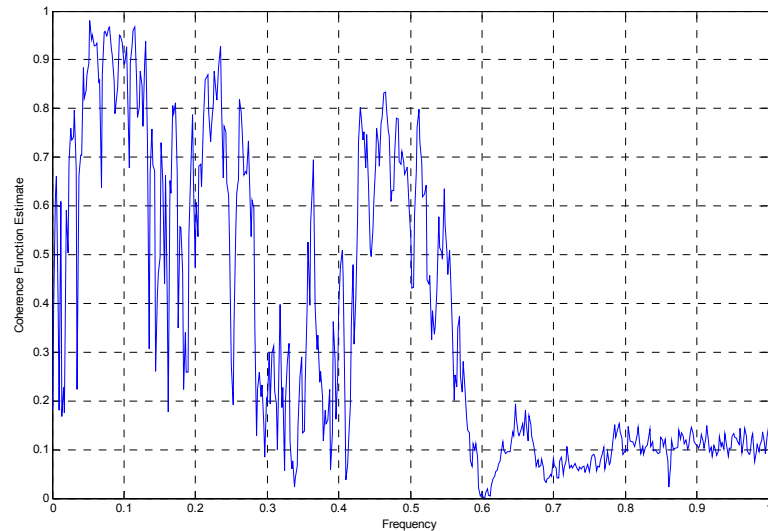


Figura 5.12: Função coerência entre as estimativas das fontes do algoritmo PearsonICA no Experimento 6.

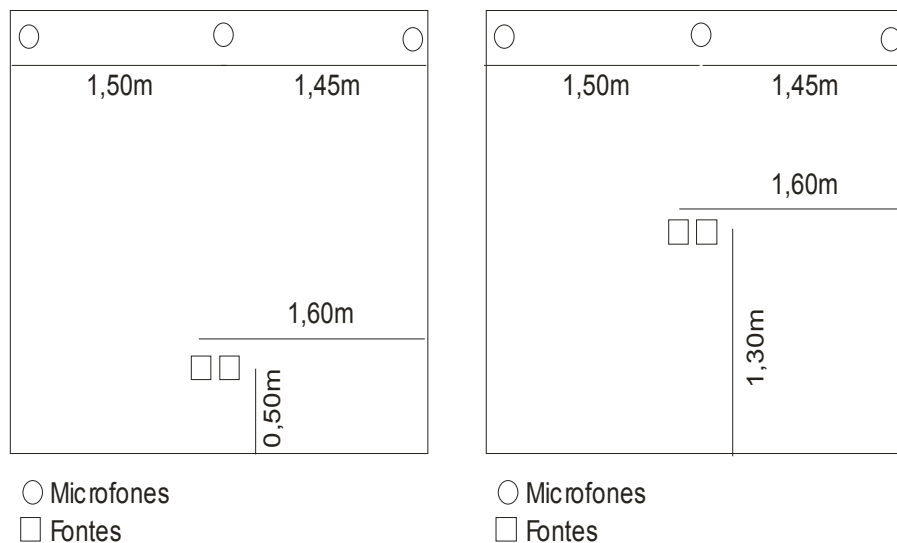


Figura 5.13: Configuração da sala acústica nos Experimentos 7 e 8, à esquerda e Experimento 9, à direita.

Os Experimentos 10 e 11 foram realizados com as fontes originais conhecidas para tornar possível a comparação com as fontes estimadas. Foram usadas duas fontes: a reprodução de uma música instrumental e a reprodução de uma pessoa do sexo masculino contando de 1 a 10. A configuração da sala acústica é apresentada na Figura 5.14.

Nos experimentos 10 e 11 foi possível comparar as fontes estimadas com as fontes originais, utilizando a relação SIR, pois os sinais originais reproduzidos na sala estavam

disponíveis, porém nenhum dos algoritmos conseguiu estimar as fontes originais, apresentando valores de SIR próximo de zero.

Os resultados mostram que, apesar dos algoritmos convergirem para uma solução, os sinais resultantes não correspondem às fontes originais, pois os modelos utilizados na separação consideram somente misturas instantâneas e, portanto para misturas convolutivas é necessário fazer modificações no modelo apresentado no presente trabalho. Ou aplicar outras técnicas, como por exemplo, a deconvolução.

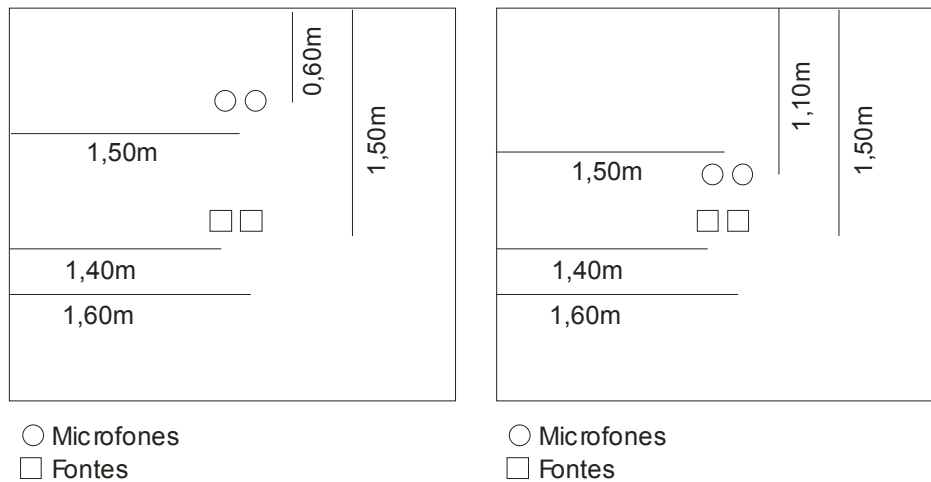


Figura 5.14: Configuração da sala acústica no Experimento 10 (esquerda) e Experimento 11 (Direita)

A Figura 5.15 mostra o gráfico da função coerência entre os sinais das fontes originais usadas nos experimentos 10 e 11, sendo possível observa que não há correlação em nenhuma componente de frequência.

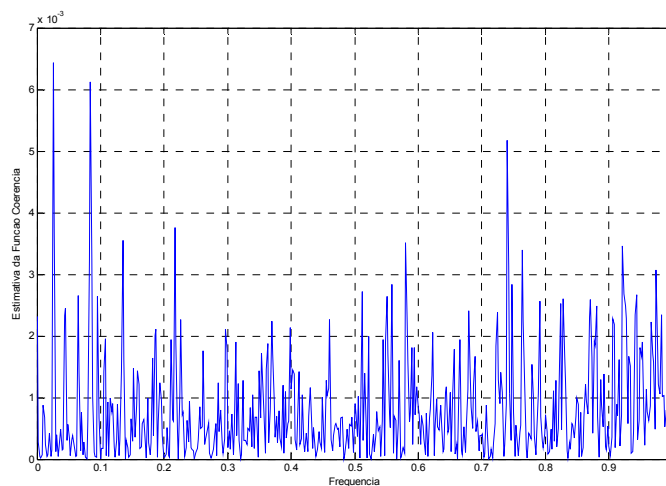


Figura 5.15: Função coerência entre as fontes originais usadas nos experimentos 10 e 11.

Observa-se no gráfico da função coerência entre as misturas do Experimento 11 (esquerda da Figura 5.16), que há correlação em quase todas as componentes de frequência.

Também é possível notar no gráfico da função coerência entre as estimativas obtidas com o PearsonICA (direta da Figura 5.16), que o algoritmo não foi capaz de descorrelacionar os sinais no domínio da frequência e, portanto, o algoritmo falhou na separação das fontes.

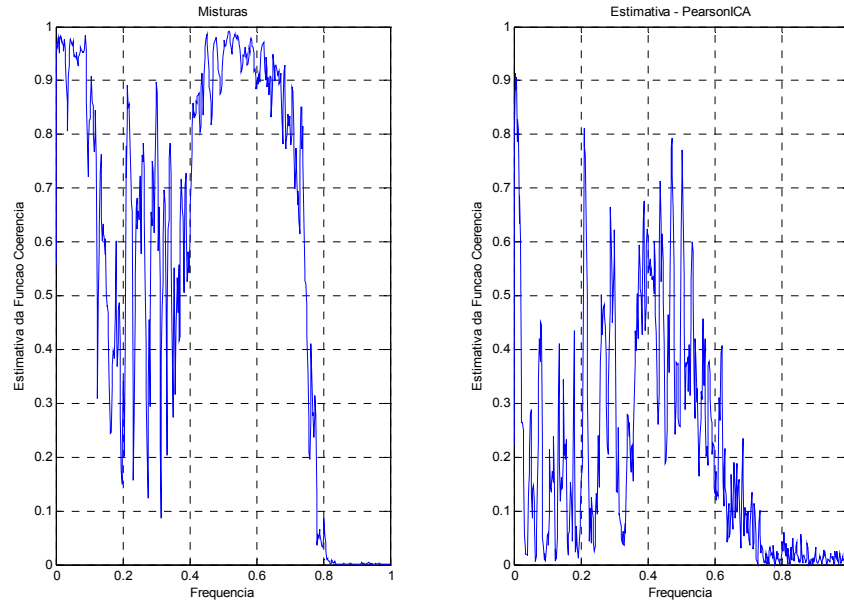


Figura 5.16: Função coerência entre as misturas e entre as estimativas das fontes feita pelo algoritmo PearsonICA no Experimento 11.

6 Conclusões

Nas seções a seguir são apresentadas as principais contribuições da pesquisa realizada no presente trabalho, bem como sugestões para pesquisa futura, com base nos resultados obtidos.

6.1 Principais contribuições

O presente trabalho é focado na separação de misturas instantâneas de sinais para aplicações de áudio. Como alternativa ao método ICA convencional, o algoritmo PP foi proposto para separação de múltiplas fontes. Além disso, foram analisados e comparados os algoritmos FastICA, o algoritmo PP e o PearsonICA.

Foi verificado que o principal fator de influência nas estimativas de todos os algoritmos é a forma da distribuição das fontes. Em geral, outros fatores como escolha de parâmetros iniciais, afetam os algoritmos, principalmente no seu desempenho, não influenciando significativamente as estimativas.

Verificou-se também que fontes com distribuição próximas à gaussiana são mais difíceis de serem estimadas e apresentam resultados inferiores em relação às fontes mais distantes da gaussiana.

Como o algoritmo PP é baseado no algoritmo do gradiente, a principal vantagem é permitir a rápida adaptação em cenários não-estacionários, por exemplo, música e fala. Porém é necessário conhecer alguns aspectos das fontes para decidir se deve ser aplicado o algoritmo gradiente descendente ou ascendente. Por exemplo, é necessário saber se a distribuição das fontes é super-gaussiana ou sub-gaussiana para decidir entre a maximização e minimização da curtose.

Outra conclusão importante é que o método PP proposto não pode ser considerado propriamente como sendo um método ICA ou vice-versa. Embora o método da PP possa ser utilizado para separação de fontes, bem como ICA pode ser usado para redução de dimensão, existem diversas diferenças importantes entre os dois métodos, além de não ser garantida a extração das fontes pelo método da PP. É possível afirmar que o método da PP proposto no presente trabalho decompõe os sinais das misturas de forma muito semelhante aos métodos de ICA, porém isso não garante a extração das fontes, pois a decomposição pode não satisfazer o modelo linear da ICA.

Apesar da tentativa de minimizar alguns fatores, como reverberação e o ruído aditivo, nos experimentos realizados em laboratório, verificou-se que não é possível aplicar nenhum dos algoritmos propostos para separação de misturas convolutivas.

6.2 Sugestões para pesquisa futura

As sugestões apresentadas nesta seção são baseadas nos resultados obtidos na pesquisa do presente trabalho.

Foi verificado que os algoritmos apresentados são excelentes para separação de fontes a partir de misturas instantâneas. No entanto, não são capazes de fazer a separação de misturas convolutivas. Portanto, poderia ser desenvolvida uma pesquisa que ampliasse ou modificasse os modelos apresentados, levando em conta os atrasos e a reverberação, de modo a serem aplicáveis às misturas convolutivas. Por exemplo, o modelo poderia ser analisado no domínio da frequência.

Já que no presente trabalho não foi considerado nenhum tipo de ruído no modelo ICA, poderia ser desenvolvido um modelo, no qual o ruído seja considerado. Por exemplo, pode ser modelado como ruído do sensor, em que é adicionado separadamente em cada sensor; ou como ruído das fontes, em que é adicionado nas componentes independentes (fontes).

Sendo o método PP proposto na pesquisa baseado na curtose, a dedução do algoritmo com base na negentropia poderia ser desenvolvida para obtenção de algoritmo mais robusto, possivelmente, permitindo melhores resultados. Além disso, o algoritmo PP poderia ser modificado para extrair fontes super-gaussianas e sub-gaussianas ao mesmo, por exemplo, poderia ser verificado qual o melhor resultado a maximização ou a minimização durante o processo de separação.

Apêndices

A. Limite de Cramér-Rao

Seja um vetor de parâmetros $\boldsymbol{\theta}$, a ser estimado a partir do vetor de dados \mathbf{x} , cuja densidade de probabilidade é $f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$, usando um estimador não viesado $\hat{\boldsymbol{\theta}}$. O CRB é o limite inferior da variância do estimado $\hat{\boldsymbol{\theta}}$.

Seja a função densidade de probabilidade $f(\mathbf{x}|\boldsymbol{\theta})$ de uma população contínua ou discreta, a função de verossimilhança de uma amostra de n observações independentes é definida pela equação (A.1).

$$L(\boldsymbol{\theta} | x_1, x_2, \dots, x_n) = f(x_1 | \boldsymbol{\theta}) f(x_2 | \boldsymbol{\theta}) \cdots f(x_n | \boldsymbol{\theta}) \quad (\text{A.1})$$

Dado que a função escore é a derivada do logaritmo da função de verossimilhança e a matriz de covariância da função escore é também denominada de matriz da Informação de Fisher. Assim, se a função densidade de probabilidade $f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$ for contínua e se existir a matriz da Informação de Fisher representada na equação (A.2) [77, 60].

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathbb{E} \left\{ \frac{1}{f_{\mathbf{x}|\boldsymbol{\theta}}^2} \frac{\partial \ln f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ln f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \quad (\text{A.2})$$

Supondo que $f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$ é independente de $\boldsymbol{\theta}$, que existe $\frac{\partial \ln f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ para qualquer $\boldsymbol{\theta}$ pertencente a um conjunto aberto e $\mathbb{E} \left\{ \frac{\partial \ln f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} = 0$, então, a inequação (A.3), também conhecida como inequação de Cramér-Rao, será válida [77, 126, 127].

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq \text{CRB}_{\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}}^{-1} \quad (\text{A.3})$$

B. Coeficiente de correlação e função coerência

O coeficiente de correlação de Pearson é definido pela equação (B.1) e indica o grau de dependência linear entre as variáveis x e y .

$$\rho_{xy} = \frac{\text{cov}_{xy}}{\sigma_x \sigma_y} \quad (\text{B.1})$$

Em que $\text{cov}_{xy} = E\{xy\} - E\{x\}E\{y\}$ é a covariância de x e y , $\sigma_x^2 = E\{x^2\}$ e $\sigma_y^2 = E\{y^2\}$ são as variâncias de x e y , respectivamente. O intervalo possível para o coeficiente de correlação é $-1 \leq \rho_{xy} \leq 1$. Valores negativos indicam correlação linear negativa entre as variáveis, ou seja, conforme x cresce, y decresce linearmente e para valores positivos, conforme x cresce, y também cresce linearmente. Obviamente que quanto mais próximo o coeficiente de correlação estiver de zero, menos correlacionadas são as variáveis. O quadrado do coeficiente de correlação, também conhecido como coeficiente de regressão r , é uma medida muito importante, pois indica a porção de variação que é previsível em relação à outra variável. O intervalo possível para o coeficiente de regressão é $0 \leq r^2 \leq 1$, em que $r = \rho_{xy}$.

Uma forma de avaliar a dependência linear de duas séries temporais é avaliar a dependência entre seus espectros de potência. A função coerência está relacionada com a correlação cruzada em termo das densidades espectrais de potência e densidade espectral cruzada de potência, dessa forma, serve como indicador da dependência linear entre as séries.

A função coerência é uma função complexa da frequência comumente representada graficamente pela magnitude e fase separadamente. A equação (B.2) representa a magnitude da função coerência.

$$\text{coh}_{xy}(j\omega) = \frac{|S_{xy}(j\omega)|^2}{S_{xx}(j\omega)S_{yy}(j\omega)} \quad (\text{B.2})$$

Em que S_{xy} é o espectro cruzado de potência entre as variáveis x e y , S_x e S_y são os espectros de potência das variáveis x e y , respectivamente.

A magnitude da função coerência é uma função real entre zero e um, que mede a correlação entre as variáveis x e y , em cada frequência ω .

C. Conteúdo do CD-ROM

O CD-ROM contém os algoritmos e os sinais referentes aos experimentos, além do artigo publicado no IWT 2007 (International Workshop on Telecommunications – 2007). Também estão disponíveis outros sinais, utilizados na dissertação. Os arquivos estão divididos em pastas, conforme descrito a seguir.

- **Algoritmos:** Contém os códigos fonte dos algoritmos utilizados na dissertação. Os arquivos foram testados no Matlab 6.5 e 7.4.
- **Artigo:** Contém o artigo publicado no IWT-2007 International Workshop on Telecommunications – 2007.
- **Experimentos:** Contém os arquivos referentes aos sinais utilizados nos experimentos.
- **Vários:** Contém outros sinais utilizados no desenvolvimento da dissertação e na pesquisa.

Referências

- [1] Comon, P., “*Independent Component Analysis – a new concept?*” *Signal Processing*, 36:287-314, 1994.
- [2] Jutten, C. and Herault, J., “*Blind Separation of Sources, part I: An adaptative algorithm based on neuromimetic architecture*”. *Signal Processing*, 24:1-10, 1991.
- [3] Cherry, E. C., “*Some Experiments on the Recognition of Speech, with One and with Two Ears*”, *The Journal of the Acoustical Society of America*, Vol. 25, Issue 5, pp. 975-979, Sep. 1953.
- [4] Sejnowski, T. J., “http://www.cnl.salk.edu/~tewon/ica_cnl.html”, Aug. 1998.
- [5] Ebata, M., “*Spatial unmasking and attention related to the cocktail party problem*”, *Acoustical Science and Technology*, Vol. 24, No. 5, Special issue on Spatial hearing, pp.208-219, 2003.
- [6] Lee, T.-W., Ziehe, Orglmeister, A., Sejnowski, R. T., “*Combining time-delayed decorrelation and ICA: towards solving the cocktail party problem*”, *Acoustics, Speech and Signal Processing, Proceedings of the 1998 IEEE International Conference on*, Vol. 2, pp 1249-1252, Seattle, WA, USA, May 1998.
- [7] Oja, E., Karhunen, J., Wang, L., and Vigario, R., “*Principal and independent components in neural networks—Recent developments*”, *Proc. VII Italian Wkshp. Neural Nets WIRN’95*, Vietri sul Mare, Italy, May 1995.
- [8] Haykin, S. and Chen, Z., “*The Cocktail Party Problem*”, *Neural Computation*, vol. 17, pp. 1875-1902, 2005.
- [9] Cardoso, J.-F., Bose, S., and Friedlander, B., “*On optimal source separation based on second- and fourth-order cumulants*”, *Proc. IEEE SSAP Wkshp.*, Corfou, 1996.
- [10] Beluchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E., “*A Blind Source Separation Technique using seconde order statitics*”, *IEEE Transaction on Signal Processing*, Vol XX, No. Y, Feb. 1997.
- [11] Cardoso, J.-F., C.N.R.S. e E.N.S.T., “*Blind signal separation: statistical principles*,” *Proceedings of the IEEE*, VOL. 9, No. 10, PP. 2009-2025, Oct. 1998.
- [12] Hyvärinen, A., Karhunen, J., Oja E., “*Independent Component Analysis*,” John Wiley and Sons, New York, 2001.

- [13] Hyvärinen, A., “*Complexity Pursuit: Separating Interesting Components from Time Series*,” *Neural Computation*, 13, PP. 883-898, 2001.
- [14] Hyvärinen, A. and Oja, E., “*A fast fixed-point algorithm for independent component analysis*”, *Neural Computation*, 9(7), PP. 1483-1492, 1997.
- [15] Hyvärinen, and Oja, E., “*Independent Component Analysis: algorithms and applications*”, *Neural Networks*, 13, PP 411-430, Helsinki, 2000.
- [16] Stone, J. V., “*Independent Component Analysis A Tutorial Introduction*”, MIT Press, 2004.
- [17] Cover, T.M. and Thomas, J. A. “*Elements of Information Theory*” John Wiley and Sons, New York, 1991.
- [18] Friedman, J. H. and Tukey, J.W., “*A Projection Pursuit Algorithm for exploratory data analysis*”, *IEEE Transactions on Computers*, c-23, PP. 881-890, 1974.
- [19] Friedman, J. H., “*Exploratory Projection Pursuit*”, *Journal of American Statistical Association*, Vol. 82, No. 397, PP 249-266, 1987.
- [20] Huber, P., “*Projection Pursuit*”, *The Annals of Statistics*, 13(2), 435-475, 1985.
- [21] Jones, M.C. and Sibson, R., “*What is Projection Pursuit*”, *Journal of Royal Statistical Society*, Vol. 150, No.1, PP.1-37, 1987.
- [22] Papoulis, A., “*Probability, Random Variables, and Stochastic Processes*” McGraw-Hill, New York, 3rd Ed., 1991.
- [23] Moreto, F. A. L. e Ramirez, M. A., “*A Projection Pursuit Approach to Blind Source Separation*”, *International Workshop on Telecommunications*, pp. 227-231, Santa Rita do Sapucaí - MG – Brasil, Fev. 2007.
- [24] Haykin, S. S., Veen, B. V., “*Sinai e Sistemas*,” Bookman, 2001.
- [25] Lee, E. A., Messerschmitt D. G., “*Digital Communication*,” Kluwer Academic Publishers, Second Edition, 1994.
- [26] Roden, M. S., “*Analog and Digital Communication Systems*,” Prentice Hall, 4th Edition, 1996.
- [27] Stuart A., e Ord, K., “*Kendall’s advanced theory of statistics – Distribution Theory*”, Oxford University Press, 6th Edition, Vol. 1, New York, 2006.
- [28] Norman L. J., Kotz, S., Kemp, A. W., “*Univariate discrete distributions*”, Wiley-Interscience, 3rd Edition, 2005.
- [29] Nandi, A., editor, “*Blind Estimation Using Higher-Order Statistics*”, Kluwer, 1999.

- [30] Pierce, J. R., “*An Introduction to Information Theory – Symbols, Signals and Noise*”, 2nd Revised Edition, New York, 1980.
- [31] Delfosse, N., Loubaton, P., “*Adaptive blind separation of independent sources: a deflation approach*”, *Signal Processing*, 45:59–83, Paris, 1995.
- [32] Horn, R. A., Johnson, C. R., “*Matrix Analysis*”, Cambridge University Press, New York, 1985.
- [33] Oppenheim, A.V., Schafer, R.W, Buck, J.R., “*Discrete-Time Signal Processing*”, 2nd Edition, Prentice Hall, 1999.
- [34] Johnson, M. E., Lowe V. W. Jr., “*Bounds on the Sample Skewness and Kurtosis*”, *Technometrics*, Vol. 21, No. 3, pp. 377-378, Aug., 1979.
- [35] Friedman, J. H.; Stuetzle, W., “*Projection Pursuit Regression*”, *Journal of the American Statistical Association*, Vol. 76, No. 376, pp. 817-823, Dec., 1981.
- [36] Friedman, J. H.; Stuetzle, W., Schroeder, A., “*Projection Pursuit Density Estimation*” *Journal of the American Statistical Association*, Vol. 79, No. 387, pp. 599-608, Sep., 1984.
- [37] Duan, N., “*The Adjoint Projection Pursuit Regression*”, *Journal of the American Statistical Association*, Vol. 85, No. 412, pp. 1029-1038, Dec., 1990.
- [38] Pham, T. V. and Smeulders, A. W. M., “*Efficient projection pursuit density estimation for background subtraction*”, *Proceedings of the Sixth IEEE International Workshop on Visual Surveillance*, Graz, Austria, 2006.
- [39] Klinker, S.; Grassmann, J., “*Projection Pursuit Regression and Neural Networks*”, Humboldt - University of Berlin, *Quantification and Simulation of Economic Processes*, 1998.
- [40] Cichocki, A., Amari, S., Siwek, K., Tanaka, T., Phan, A. H., *et al.*, “*ICALAB Toolboxes*”, [OnLine], Available: <http://www.bsp.brain.riken.jp/ICALAB>, Mar. 2007.
- [41] Hyvärinen, A., “*The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis*”, *Neural Processing Letters*, 10: 1-5, 1999.
- [42] Cheriyyadat, A., Bruce, L. M., “*Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data*”, *Geoscience and Remote Sensing Symposium*, *Proceedings IEEE International*, pp. 3420 - 3422 vol.6, 2003.
- [43] Jolliffe, I.T., “*Principal Component Analysis*”, 2^a ed., Springer, New York, 2002.
- [44] Wallace, D. L., “*Asymptotic Approximations to Distributions*”, *The Annals of Mathematical Statistics*, Vol. 29, No. 3, pp. 635-654, Sep., 1958.

- [45] Tchebyshev, P. L., “*Sur Deux Théorèmes Relatifs aux Probabilités*”, Acta Math, Vol 14, pp 305-315, 1890.
- [46] Hyvärinen, A., “*New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit*”, Advances in Neural Information Processing System 10 (NIPS*97), pp. 273-279, MIT Press, 1998.
- [47] Shannon, C. E., “*A Mathematical Theory of Communication*”, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
- [48] Hyvärinen, A., “*Fast and Robust Fixed-Point Algorithms for Independent Component Analysis*”, IEEE Trans. on Neural Networks, 10(3):626-634, 1999.
- [49] Hyvärinen, A., “*Gaussian Moments for Noisy Independent Component Analysis*”, IEEE Signal Processing Letters, Vol. 6, No. 6, Jun. 1999.
- [50] Lay, D. C., “*Álgebra Linear e suas Aplicações*”, LTC, 1999.
- [51] Hyvärinen, A., “*Independent Component Analysis by General Nonlinear Hebbian-like Learning Rules*”, Signal Processing, vol. 64 no. 3, pp 301-313, 1998.
- [52] Gävert, H., Hurri, J., Särelä, J., e Hyvärinen, A., “*The FastICA package*”, [OnLine], Available: <http://www.cis.hut.fi/projects/ica/fastica/>, Oct. 2005.
- [53] Winston, W. L., “*Operations Research – Applications and Algorithms*”, Thomson Brooks/Cole, 2004.
- [54] Sun, W., Yuan, Y.-X., “*Optimization Theory and Methods - Nonlinear Programming*”, Springer, 2006.
- [55] Kuhn, H. W., Tucker, A. W., “*Nonlinear Programming*”, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp 481-492, Berkeley, California, 1951.
- [56] Karvanen, J., Eriksson, J., Koivunen, V., “*Pearson system based method for blind separation*”, ICA2000, Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, pp. 585–590, 2000.
- [57] Karvanen, J., Koivunen, V., “*Blind separation methods based on Pearson system and its extensions*”, Signal Processing 82, pp 663 – 673, 2002.
- [58] Ord, J. K., “*Families of frequency distributions*”, Griffin, London, 1972.
- [59] Cardoso, J.-F., e Laheld, B. H., “*Equivariant adaptive source separation*”, IEEE Trans. on Signal Processing, 44(12):3017–3030, 1996.

- [60] Tichavský, P., Koldovsky, Z., Oja, E., “*Performance Analysis of the FastICA Algorithm and Cramér–Rao Bounds for Linear Independent Component Analysis*”, IEEE Transactions on Signal Processing, Vol. 54, No. 4, Apr. 2006.
- [61] Koldovsky, Z., Tichavský P., Oja, E., “*Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound*”, IEEE Transactions on Neural Networks, Vol. 17, No. 5, Sep. 2006.
- [62] Cardoso J.-F., e Adalı, T., “*The Maximum Likelihood Approach to Complex ICA*”, Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings, Vol. 5, 2006.
- [63] Zarzoso, V., Comon, P., Kallel, M., “*How Fast is FastICA?*”, EUSIPCO-2006, Florence, Italia, Sep. 2006.
- [64] Nobre, J., Azevedo, C., “*O princípio da equivariância: conceitos e aplicações*”, Revista Colombiana de Estadística, Vol. 29 No. 2. pp. 195 a 220. Dec. 2006.
- [65] Comon, P., Jutten, C., e Héroult, J., “*Blind separation of sources, Part II: Problems Statement*”, Signal Processing, 24:11–20, 1991.
- [66] Sorouchyari, E., “*Blind separation of sources, Part III: Stability analysis*”, Signal Processing, 24:21–29, 1991.
- [67] Héroult, J., Jutten, C., and Ans, B., “*Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé*”, Actes du Xième colloque GRETSI, vol. 2, pp. 1017–1022, Nice, France, Mai 1985.
- [68] Amari S.-I., and Cardoso. J.-F., “*Blind source separation—semiparametric statistical approach*”. IEEE Trans. on Signal Processing, 45(11):2692–2700, 1997.
- [69] Cichocki A., and Unbehauen, R., “*Robust neural networks with on-line learning for blind identification and blind separation of sources*”, IEEE Trans. on Circuits and Systems, 43(11):894–906, 1996.
- [70] Cichocki, A., Unbehauen, R., and Rummert, E., “*Robust learning algorithm for blind separation of signals*”, Electronics Letters, 30(17):1386–1387, 1994.
- [71] Hüper, K., Shen, H., Seghouane, A.-K., “*Local Convergence Properties of FastICA and Some Generalizations*”, IEEE 31st International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp V-1009-V-1012, Toulouse, France, 2006.
- [72] Hyvärinen, A., “*A family of fixed-point algorithms for independent component Analysis*” in: Proc. ICASSP, pp. 3917–3920, Munich, Germany, Apr. 1997.

- [73] Chevalier, P., Albera, L., Comon, P., and Ferreol, A., “*Comparative performance analysis of eight blind source separation methods on radiocommunications signals*” in: Proc. Intl. Joint Conf. on Neural Networks, Budapest, Hungary, Jul. 2004.
- [74] Regalia, P. A., and Kofidis, E., “*Monotonic Convergence of Fixed-Point Algorithms for ICA*”, IEEE Transactions on Neural Networks, Vol. 14, No. 4, 2003.
- [75] Giannakopoulos, X., Karhunen, J., and Oja, E., “*Experimental comparison of neural algorithms for independent component analysis and blind separation*”, Int. J. Neural Syst., vol. 9, pp. 651–656, 1999.
- [76] Vincent, E., Gribonval R., and Févotte, C., “*Performance Measurement in Blind Audio Source Separation*”, IEEE Transaction On Audio and Language Processing, Vol. 14, No. 4, Jul. 2006.
- [77] Stuart A., and Ord, K., Arnold, S., “*Kendall’s Advanced Theory of Statistics – Classical Inference & the Linear Model*”, Oxford University Press, 6th Edition, Vol. 2A, New York, 2006.
- [78] Févotte, C., Gribonval, R., and Vincent, E., “*BSS_EVAL toolbox user guide*” IRISA, Rennes, France, Tech. Rep. 1706. [Online]. Available: http://www.irisa.fr/metiss/bss_eval, 2005.
- [80] Cardoso, J.-F., and Souloumiac, A., “*Blind beamforming for non-Gaussian signals*”, Radar and Signal Processing, IEE Proceedings F, Vol. 140, No. 6, pp. 362 -370, Dec. 1993.
- [81] Cardoso, J.-F., and Souloumiac, A., “*An efficient batch algorithm: JADE*”. [Online]. Available: <http://sig.enst.fr/~cardoso/guidesepsou.html>, 1993.
- [82] Meyer, C. D., “*Matrix Analysis and Applied Linear Algebra*”, SIAM, 2000.
- [83] Baker, E.S., DeGroat, R.D., “*Evaluating EVD and SVD errors in signal processing environments*”, Signals, Systems & Computers, Conference Record of the Thirty-Second Asilomar Conference on, Nov. 1998.
- [84] Maronna, R., Martin, D., Yohai, V., “*Robust Statistics – Theory and Methods*”, John Wiley & Sons, England, 2006.
- [85] James, C. J. and Gibson, O. J. (2002), “*Electromagnetic brain signal analysis using constrained ICA*”, Proceedings of 2nd European Medical and Biological Engineering Conference (EMBEC'02), Vienna, Austria, Part I, pp 426-427, Dec. 2002.
- [86] He T., Clifford G., Tarassenko, L.: “*Application of independent component analysis in removing artefacts from the electrocardiogram*,” Neural Comput. & Applic. 15(2): 105-116, 2006.

- [87] Wisbeck, J. O., Barros, A. K., and Ojeda, R. G. “*Application of ICA in the Separation of Breathing Artifacts in ECG Signals*,” Proceedings of ICONIP'98, Kyushu, Japan, Oct. 1998.
- [88] Ungureanu, M., Bigan, C., R. Strungaru, Lazarescum, V., “*Independent Component Analysis Applied in Biomedical Signal Processing*,” MEASUREMENT SCIENCE REVIEW, Bucharest, Romania, Volume 4, Section 2, 2004.
- [89] Martin, J., McKeown, Saab R., and Rafeef, A.-G., “*A Combined Independent Component Analysis (ICA)/ Empirical Mode Decomposition (EMD) Method to Infer Corticomuscular Coupling*,” Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering, Arlington, Virginia, Mar. 2005.
- [90] Vigário R., Jousmäki V., Hämäläinen, M., Hari, R., Oja, E., “*Independent component analysis for identification of artifacts in magneto encephalographic recordings*,” Proceedings of the 1997 conference on Advances in neural information processing systems 10, p.229-235, Denver, Colorado, United States, Jul. 1998.
- [91] Djuwari, D., Kumar, D. K., and Palaniswami, M., “*Limitations of ICA for Artefact Removal*,” Proceedings of the 2005 IEEE, Engineering in Medicine and Biology, 27th Annual Conference, Shanghai, China, Sep. 2005.
- [92] Feng, M.; Kammeyer, K.-D., “*Blind source separation for communication signals using antenna arrays*”, IEEE 1998 International Conference on Universal Personal Communications Page(s):665 - 669 vol.1, Oct. 1998.
- [93] Zarzoso, V., “*Exploiting independence for co-channel interference cancellation and symbol detection in multiuser digital communications*,” Seventh International Symposium on Proceedings of Signal Processing and Its Applications Page(s): 303 - 306 vol.2, Jul. 2003.
- [94] Gupta, M.; Santhanam, B., “*Prior ICA based blind multiuser detection in DS-CDMA systems*,” Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers Page(s): 2155 - 2159 Vol.2, Nov. 2004.
- [95] Cristescu, R., Joutsensalo, J., Karhunen, J., and Oja., E., “*A complexity minimization approach for estimating fading channels in CDMA communications*”, In Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000), pages 527–532, Helsinki, Finland, Jun. 2000.
- [96] Zarzoso, V., and Nandi, A. K., “*Improving MIMO Channel Equalization with Independent Component Analysis*”, Proceedings 6th IMA International Conference on Mathematics in Signal Processing, pp. 223-226, Cirencester, UK, Dec. 2004.

- [97] Zarzoso, V., and Nandi, A. K., “*Blind MIMO Equalization with Optimum Delay Using Independent Component Analysis*”, International Journal of Adaptive Control and Signal Processing, Vol. 18, No. 3, pp. 245-263, Apr. 2004.
- [98] Du, J., Lee, C.-H., Lee, H.-K., Suh, Y., “*BSS: a new approach for watermark attack*”, Proceedings Fourth International Symposium on Multimedia Software Engineering, Page(s): 182 – 187, 2002.
- [99] Lin, Q.-H., Yin, F.-L., Mei, T.-M., and Liang, H., “*A Blind Source Separation Based Method for Speech Encryption*”, IEEE Transactions on Circuits And Systems, Vol. 53, No. 6, Jun. 2006.
- [100] Calhoun, V. D., Adali, T., Hansen, L. K., Larsen, J., Pekar, J. J., “*ICA of Functional MRI Data: An Overview*”, 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, Apr. 2003.
- [101] Li, G., Zhang, J., “*Sphering and its Properties*”, Sankhya : The Indian Journal of Statistics, Volume 60, Series A, Pt. 1, pp. 119-133, 1998.
- [102] Vincent, E., Févotte, C., Gribonval, R., Röbel, A., Rodet, X., Carpentier, E. L., Benaroya, L., Bimbot, F., “*A Tentative Typology of Audio Source Separation Tasks*”, Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 715-720, Nara, Japan, 2003.
- [103] Mitianoudis N., and Davies, M. E., “*Audio source separation: solutions and problems*”, International Journal of Adaptive Control and Signal Processing, 18:299–314, 2004.
- [104] Asano, F.; Ikeda, S.; Ogawa, M.; Asoh, H.; Kitawaki, N., “*A combined approach of array processing independent component analysis for blind separation of acoustic signals*”, IEEE Transactions on Speech and Audio Processing, Volume 11, Issue 3 Page(s): 204 – 215, May 2003.
- [105] Murata, N., and Ikeda, S., “*An on-line algorithm for blind source separation on speech signals*”, In Proceedings of 1998 International Symposium on Nonlinear Theory and its Applications (NOLTA'98), pp.923-926, Crans-Montana, Switzerland, Sep. 1998.
- [106] Chien, J.-T., and Chen, B.-C., “*A New Independent Component Analysis for Speech Recognition and Separation*”, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, Jul. 2006.
- [107] Schobben, D.W.E., Torkkola, K., and Smaragdis, P., “*Evaluation of Blind Signal Separation Methods*”, First International Workshop on Independent Component Analysis and Blind Signal Separation, Aussois, France, Jan. 1999.

- [108] Virtanen T., “*Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria*”, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 3, Mar. 2007.
- [109] Liu, C., “*Enhanced independent component analysis and its application to content based face image retrieval*”, IEEE Transactions on Systems, Man and Cybernetics Volume 34, Issue 2, Page(s): 1117 – 1127, Apr. 2004.
- [110] Hoyer P.O., and Hyvärinen, A., “*Independent component analysis applied to feature extraction from color and stereo images*”, Network: Computation in Neural Systems, 11(3):191–210, 2000.
- [111] Malaroiu, S., Kiviluoto, K., and Oja, E., “*Time series prediction with independent component analysis*”, In Proc. Int. Conf. on Advanced Investment Technology, Gold Coast, Australia, 2000.
- [112] Cichocki, A., Amari, S., “*Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*”, Wiley, 2003.
- [113] Back A. D., and Weigend, A. S., “*A first application of independent component analysis to extracting structure from stock returns*”, Int. J. on Neural Systems, 8(4):473–484, 1997.
- [114] Parra, L. C., “*An introduction to independent component analysis and blind source separation*”, Sarnoff Corporation, Princeton, NJ, Apr. 1999.
- [115] Kiviluoto K., and Oja. E., “*Independent component analysis for parallel financial time series*”, In Proc. Int. Conf. on Neural Information Processing (ICONIP’98), volume 2, pages 895–898, Tokyo, Japan, 1998.
- [116] Rencher A. C., “*Methods of multivariate analysis*”, John Wiley & Sons, 2nd Ed., USA, 2002.
- [117] Tang, B., Shepherd, M., Heywood, M.I. and Luo, X., “*Comparing Dimension Reduction Techniques for Document Clustering*”, The Eighteenth Canadian Conference on Artificial Intelligence. Victoria, BC, Canada, May 2005.
- [118] Tang, B., Shepherd, M., Milios, E. and Heywood, M.I., “*Comparing and Combining Dimension Reduction Techniques for Efficient Text Clustering*”, International Workshop on Feature Selection for Data Mining - Interfacing Machine Learning and Statistics in conjunction with 2005 SIAM International Conference on Data Mining, Newport Beach, California, Apr. 2005.
- [119] Fodor, I.K., “*A survey of dimension reduction techniques*”, Technical Report, UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

- [120] Molgedey, L. and Schuster, H. G., “*Separation of a mixture of independent signals using time delayed correlations*”, Physical Review Letters, 72, No. 23, pp 3634–3636, 1994.
- [121] Bingham, E. and Hyvärinen, A., “*A fast fixed-point algorithm for independent component analysis of complex valued signals*”, International Journal of Neural Systems 10, No. 1, pp 1-8, World Scientific Publishing Company, 2000.
- [122] Oja, E., “*The nonlinear PCA learning rule and signal separation Mathematical analysis*”, Helsinki Univ. Technol., Rep. A26, Aug. 1995.
- [123] Luo, J., Hu, B., Ling, X.-T., and Liu, R.-W., “*Principal Independent Component Analysis*,” IEEE Transactions on Neural Networks, Vol. 10, No. 4, Jul 1999.
- [124] Haykin, S., “*Neural Networks: A Comprehensive Foundation*”, Prentice Hall; 2nd edition, 1999.
- [125] Cardoso, J.-F., “*Eigen-Structure of The Fourth-Order Cumulant Tensor With Application to The Blind Source Separation Problem*”, In Proc. ICASSP 90, pp 2655-2658, Albuquerque, NM, USA, 1990.
- [126] Härdle, W., Simar, L., “*Applied Multivariate Statistical Analysis*”, Springer, 2003.
- [127] Rao, R.C., “*Linear Statistical Inference and Its Application*”, John Wiley & Sons; 2nd edition, 2002.
- [128] Moreto, F. A. L., “*Análise de componentes independentes aplicada à separação de sinais de áudio por meio de busca de projeção*”, 1 CD-ROM Parte integrante da dissertação de mestrado Análise de componentes independentes aplicada à separação de sinais de áudio por meio de busca de projeção, EPUSP, 2008.
- [129] Schobben, D.W.E., “*Real-time Adaptive Concepts in Acoustics: Blind Signal Separation and Multichannel Echo Cancellation*”, Kluwer Academic Publishers, 2001.
- [130] Weinstein, E., Feder, M. e Oppenheim, A.V., “*Multi-channel signal separation by decorrelation*”, IEEE Trans. Speech Audio Processing, vol. 1, no. 4, pp. 405-413, 1993.
- [131] Fancourt, C., Parra, L., “*The Coherence Function in Blind Separation of Convolutional Mixtures of Nonstationary signals*”, Proc. IEEE Workshop on Neural Networks for Signal Processing, pp. 303-312, 2001.
- [132] Parra, L., and Spence, C., “*Convolutional blind separation of non-stationary sources*”, IEEE Trans. Speech Audio Proc., vol. 8, no. 3, pp. 320-327, 2000.
- [133] Schobben D.W.E. and Sommen, P.C.W., “*A new Blind Signal Separation Algorithm based on Second Order Statistics*”, Proc. of the IASTED International Conference Signal and Image Processing, Las Vegas, USA, Oct. 1998.

- [134] Bell, A.J. and Sejnowski, T.J., “*An information maximization approach to blind separation and blind deconvolution*”, Neural Computation 7, pp. 1129-1159, MIT Press, Cambridge MA, 1995.
- [135] Di Persia, L., Milone, D., Yanagida, M., Rufiner, H. L., “*Objective quality evaluation in blind source separation for speech recognition in a real room*”, Signal Processing, Vol. 87, Nº 8, pp. 1951-1965, Aug. 2007.
- [136] Mitianoudis, N., Davies, M., “*New fixed-point solutions for convolved mixtures*”, in 3rd International Conference on Independent Component Analysis and Blind Source Separation, San Diego, California, Dec. 2001.
- [137] Campani, C., Menezes, P., “*Introdução à Complexidade de Kolmogorov*”, Tutorial. [On Line]. URL: <http://minerva.ufpel.edu.br/~campani/grupo.htm>.
- [138] Lee, T.-W., Girolami, M, Bell, A. J., and Sejnowski, T. J., “*A Unifying Information-Theoretic Framework For Independent Component Analysis*,” Computers & Mathematics with Applications, Vol. 31 (11), pp. 1-21, Mar. 2000.
- [139] Jayant, N. S., Noll, P., “*Digital coding of waveforms*”, Englewood Cliffs: Prentice-Hall, 1984.