

ESTUDO DA MELHORIA DA TAXA  
DE APRENDIZAGEM DE UM  
ALGORITMO PARA SEPARAÇÃO  
CEGA DE FONTES UTILIZANDO  
TÉCNICAS VINDAS DA TEORIA DE  
REDES NEURAIS

FELIPE AUGUSTO PEREIRA DE  
FIGUEIREDO

MAIO/ 2011

Figueiredo, Felipe Augusto Pereira de  
F475e

Estudo da Melhoria da Taxa de Aprendizagem de um Algoritmo para Separação Cega de Fontes Utilizando Técnicas vindas da Teoria de Redes Neurais / Felipe Augusto Pereira de Figueiredo. – Santa Rita do Sapucaí, 2011.

135 p.

Orientador: PhD. Carlos Alberto Ynoguti  
Dissertação de Mestrado – Engenharia de Telecomunicações – Instituto Nacional de Telecomunicações – INATEL.  
Inclui anexo e bibliografia.

1. Separação Cega de Sinais 2. Redes Neurais 3. Filtragem Adaptativa 4. Análise de Componentes Independentes. I. Ynoguti, Carlos Alberto. II. Instituto Nacional de Telecomunicações – INATEL. III. Título.

CDU 621.39

# **Estudo da melhoria da taxa de aprendizagem de um algoritmo para Separação Cega de Fontes utilizando técnicas vindas da teoria de Redes Neurais**

FELIPE AUGUSTO PEREIRA DE FIGUEIREDO

Dissertação apresentada ao Instituto Nacional de Telecomunicações, como parte dos requisitos para obtenção do Título de Mestre em Telecomunicações.

Orientador: Prof. Dr. Carlos Alberto Ynoguti

**Santa Rita do Sapucaí  
2011**

Dissertação defendida e aprovada em 09/05/2011, pela comissão julgadora:

---

Prof. Dr. Carlos Alberto Ynoguti (Orientador) - Inatel

---

Prof. Dr. Luciano Leonel Mendes - Inatel

---

Prof. Dr. Marcelo de Oliveira Rosa - UTFPR

---

**Prof. Dr. Luciano Leonel Mendes**  
**Coordenador do Curso de Mestrado**

À minha mãe,  
minha primeira e grande professora.

## Agradecimentos

Primeiramente, agradeço a Deus por sempre estar ao meu lado, me dando saúde, força e perseverança para alcançar meus objetivos. Agradeço a Ele pela oportunidade de estar sempre trabalhando e aprendendo. Agradeço ao Espírito Santo por ter me inspirado e me dado os dons necessários para a realização deste trabalho.

A Aldair Ibrantina Pereira de Figueiredo, minha mãe amada, agradeço por sempre ser meu porto seguro, minha principal incentivadora e por forjar em mim as virtudes que tenho, sem sua ajuda eu não seria o que sou hoje. Agradeço a ela por todo o amor, carinho, compreensão e respeito que sempre recebi. Ao meu pai e ao meu irmão, que lá do céu tem me protegido e guiado nos caminhos da vida.

Ao meu orientador, Dr. Carlos Alberto Ynoguti, pela orientação, incentivo e preocupação despendidas com muita dedicação à realização deste trabalho. 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.

Aos amigos que compartilharam o árduo processo de desenvolvimento científico. Em especial ao casal Rodrigo Pimenta Carvalho e Lana Quele Cardoso Carvalho e aos amigos Breno Carvalho de Abreu e Suzana Barude Fernandes que me ajudaram com a gravação de uma base de dados para este trabalho. Faço um agradecimento especial aos amigos Antônio Tomaz de Carvalho e Robélia de Fátima Dias Pereira pelo inestimável apoio para que eu concluísse meu trabalho.

Meu sincero muito obrigado, a todos que contribuíram para realização desta etapa de minha vida, que foram e são muito importantes para mim. Ao INATEL que me ofereceu todos os recursos que necessitei.

À Deus toda glória e louvor.

# Índice

<b>Lista de Figuras.....</b>	<b>vii</b>
<b>Lista de Tabelas.....</b>	<b>ix</b>
<b>Lista de Abreviaturas e Siglas.....</b>	<b>x</b>
<b>Lista de Operadores e Símbolos.....</b>	<b>xii</b>
<b>Capítulo 1 - Introdução.....</b>	<b>16</b>
1.1 Histórico.....	16
1.2 Motivação do trabalho.....	20
1.3 Aplicações.....	23
1.4 Objetivos.....	24
1.5 Organização do trabalho.....	26
<b>Capítulo 2 - Principais Conceitos para a Separação Cega de Fontes.....</b>	<b>28</b>
2.1 Definição do problema.....	28
2.2 Modelos acústicos para a separação cega de fontes.....	31
2.2.1 Modelo para misturas instantâneas.....	31
2.2.2 Modelo para misturas convolutivas.....	33
2.3 Características das fontes e sua utilização na separação cega.....	34
2.3.1 Não-branca.....	36
2.3.2 Não-estacionariedade.....	37
2.3.3 Estrutura temporal das fontes.....	38
2.4 Ambigüidades na separação cega de fontes.....	39
2.5 Informação Mútua e a Divergência de Kullback-Leibler.....	40
2.6 Gradiente Descendente.....	43

2.7 Medida de desempenho.....	45
2.7.1 Medida da supressão das fontes de sinal interferente.....	46
<b>Capítulo 3 - Algoritmo para Separação Cega de Fontes Misturadas</b>	
<b>Convolutivamente baseado em Estatísticas de Segunda Ordem.....</b>	<b>48</b>
3.1 Introdução.....	48
3.2 Algoritmo para separação cega de sinais no domínio do tempo.....	52
3.2.1 Notação matricial para misturas convolutivas.....	53
3.2.2 Critério de otimização.....	58
3.2.3 Métodos para inicialização das matrizes de separação.....	63
3.2.4 Gradiente Natural.....	65
3.2.5 Métodos para estimação das matrizes de correlação.....	67
3.2.6 Normalização eficiente.....	70
<b>Capítulo 4 - Técnicas para adaptação do passo de aprendizagem.....</b>	<b>73</b>
4.1 Mínimos locais.....	73
4.2 Passo de aprendizagem fixo.....	74
4.3 Termo momentum.....	75
4.4 Bold driver.....	77
4.5 Decaimento exponencial.....	79
<b>Capítulo 5 - Resultados e experimentos.....</b>	<b>81</b>
5.1 Condições gerais dos experimentos.....	81
5.2 Avaliação das técnicas para adaptação do passo de aprendizagem.....	82
5.2.1 Misturas convolutivas sintéticas.....	82
5.2.2 Misturas convolutivas reais.....	91
5.3 Normalização eficiente.....	97
5.4 Estimação das matrizes de correlação através dos métodos da correlação e da covariância.....	101
5.5 Avaliação da restrição Sylvester e suas aproximações.....	103
5.6 Normalização dos sinais de saída do sistema de separação.....	105
5.7 Locutor em ambiente ruidoso e reverberante.....	107
<b>Capítulo 6 - Conclusões.....</b>	<b>110</b>
6.1 Sugestões para trabalhos futuros.....	115

<b>Anexo A.....</b>	<b>118</b>
C.1 Artigos publicados.....	118
C.2 Artigos submetidos.....	118
<b>Anexo B.....</b>	<b>119</b>
<b>Referências Bibliográficas.....</b>	<b>120</b>

## Lista de Figuras

1.1: Diagrama em blocos de um modelo genérico para SCF.....	22
2.1: Modelo para solução do problema da separação cega de sinais.....	29
2.2: Sistema MIMO convolutivo para SCF.....	34
2.3: Gráfico de contorno de uma função de custo com um mínimo local.....	44
3.1: Sistema MIMO linear para SCF.....	49
3.2: Divisão do sinal em segmentos sobrepostos.....	55
3.3: Ilustração de 3.20 para o caso 2 x 2.....	60
3.4: Arranjos para SCF que requerem (a) somente filtros FIR causais e (b) filtros FIR causais e não-causais.....	64
4.1: Diferentes tipos de superfícies formadas por funções de custo.....	74
4.2: Influência do termo momentum.....	77
5.1: Comparativo entre as técnicas para adaptação do passo de aprendizagem.....	84
5.2: Curva da SIR para a técnica do passo de aprendizagem fixo com 2000 épocas.....	85
5.3: Variação do valor da função de custo ao longo das iterações.....	88
5.4: Testes combinando o termo momentum com outras técnicas.....	88
5.5: Resposta em frequência e diagrama de recepção dos microfones.....	92
5.6: Arranjos para SCF que requerem (a) somente filtros FIR causais e (b) filtros FIR causais e não-causais.....	93
5.7: Comparativo entre as técnicas para adaptação do passo de aprendizagem.....	94

5.8: Curva da SIR para a técnica do passo de aprendizagem fixo com 1000 épocas.....	95
5.9: Comparação entre as formas de normalização da equação de atualização.....	98
5.10: Normalização eficiente com passo de aprendizagem igual a 0,009.....	100
5.11: Variação do passo de aprendizagem.....	100
5.12: Comparação entre os métodos de estimação das matrizes de correlação.....	101
5.13: Comparação entre a restrição Sylvester e suas aproximações eficientes.....	104
5.14: Curvas de SIR para a normalização e não normalização dos vetores de saída.....	106
5.15: Resultados de experimentos com ruídos.....	109
6.1: Variação da SIR em relação a variação do passo de aprendizagem.....	114

## Lista de Tabelas

5.1: Comparação da taxa de convergência e SIR máxima para cada técnica..	85
5.2: Tempo médio de processamento para cada iteração.....	86
5.3: Comparação da taxa de convergência e SIR máxima para as combinações de técnicas.....	90
5.4: Comparação entre as técnica de adaptação do passo de aprendizagem...	96
5.5: Comparação da SIR e tempo de simulação entre os métodos de normalização.....	99
5.6: Comparação da SIR e tempo de simulação entre os métodos de estimação.....	102
5.7: Tabela comparativa entre a restrição Sylvester e suas aproximações eficientes.....	105
5.8: Tabela comparativa entre a normalização e a não normalização dos vetores de saída y.....	107
5.9: SIR inicial e máxima para os experimentos com ruídos.....	109

## Lista de Abreviaturas e Siglas

<b>ASR</b>	<i>Automatic Speech Recognition</i> (Reconhecimento automático de fala).
<b>BPSK</b>	<i>Binary Phase Shift Keying</i> (Modulação por chaveamento de fase).
<b>BSS</b>	<i>Blind Source Separation</i> (Separação Cega de Fontes).
<b>EDICS</b>	<i>Editors Information Classification Scheme</i> .
<b>EEG</b>	<i>Electroencephalogram</i> (Eletroencefalograma).
<b>fdp</b>	<i>Função Densidade de Probabilidade</i> .
<b>FIR</b>	<i>Finite Impulse Response Filter</i> (Filtro com Resposta ao Impulso Finita).
<b>HOS</b>	<i>Higher Order Statistics</i> (Estatísticas de Ordem Superior).
<b>i.i.d.</b>	<i>Independent Identically Distributed</i> (Variáveis Aleatórias Independentes e Identicamente Distribuídas).
<b>ICA</b>	<i>Independent Component Analysis</i> (Análise de Componentes Independentes).
<b>IIR</b>	<i>Infinite Impulse Response</i> (Resposta Infinita ao Impulso).
<b>LTI</b>	<i>Linear Time-Invariant</i> (Sistema Linear invariante no tempo).
<b>MIMO</b>	<i>Multiple-inputs and Multiple-outputs</i> (Sistema com múltiplas entradas e múltiplas saídas).
<b>MLP</b>	<i>Multi Layer Perceptron</i> (Perceptron Multicamadas).
<b>PCA</b>	<i>Principal Component Analysis</i> (Análise de Componentes Principais).
<b>SCF</b>	<i>Separação Cega de Fontes</i> .
<b>SIR</b>	<i>Signal to Inteference Ratio</i> (Relação Sinal-Interferência).
<b>SISO</b>	<i>Single-Input and Single Output</i> (Sistema com entrada e saída únicas).
<b>SOM</b>	<i>Self Organized Maps</i> (Mapas Auto-Organizáveis).

<b>SOS</b>	<i>Second Order Statistics</i> (Estatística de segunda ordem).
<b>STFT</b>	<i>Short-Time Fourier Transform</i> (Transformada de Fourier de Tempo Curto).

## Lista de Operadores e Símbolos

### Operadores

$D(p  q)$	Divergência de Kullback-Leibler entre as variáveis aleatórias $p$ e $q$ .
$\text{diag} \{ A \}$	Operador que zera todos os valores não pertencentes a diagonal principal de uma matriz.
$\text{bdiag} \{ A \}$	Operador que zera todas as submatrizes não pertencentes a diagonal principal da matriz.
$\det \{ . \}$	Determinante de uma matriz quadrada.
$E \{ . \}$	Operador esperança.
$\hat{E} \{ . \}$	Estimativa da esperança através de média temporal.
$O(.)$	Ordem.
$\text{rank} \{ . \}$	Posto do argumento.
$SC \{ . \}$	Restrição <i>Sylvester</i> .
$SC_C \{ . \}$	Restrição <i>Sylvester</i> da coluna.
$SC_R \{ . \}$	Restrição <i>Sylvester</i> da linha.
$\text{tr} \{ . \}$	Traço de uma matriz, ou seja, soma dos elementos da diagonal principal.
$\nabla W$	Gradiente com relação a $W$ .
$\nabla_W^{NG}$	Gradiente natural com relação a $W$ .
$\nabla^2$	Operador Laplaciano.
$\partial/\partial x$	Derivada parcial com relação a $x$ .
$\delta_{i,j}$	Delta de Kronecker.

*	Operador de convolução.
$(\cdot)^T$	Transposto de $(\cdot)$ .
$(\cdot)^*$	Conjugado complexo de $(\cdot)$ .
$(\cdot)^H$	Hermitiano, ou seja, Conjugado transposto de $(\cdot)$ .
$(\cdot)^{-1}$	Operação inversa de $(\cdot)$ .
$ (\cdot) $	Valor absoluto de $(\cdot)$ , escalar.
$\ (\cdot)\ _F$	Norma de Frobenius (ou Norma Euclidiana) de $(\cdot)$ (vetor ou matriz).

## Símbolos

$C$	Número de amostras dependentes temporalmente da $q$ -ésima fonte.
$D$	Número de atrasos de tempo considerados para as correlações.
$f$	<i>Função estritamente convexa.</i>
$H$	Matriz que modelo o processo de mistura dos sinais originais.
$h_{qp,k}$	$k$ -ésimo coeficiente do filtro FIR que modela o caminho entre a $q$ -ésima fonte e o $p$ -ésimo sensor.
$J(m, W)$	Função de custo dependente do bloco atual de processamento e da matriz de parâmetros $W$ .
$K$	Matriz de covariância.
$L$	Número de coeficientes dos filtros de separação.
$N$	Comprimento dos segmentos de sinal de saída.
$P$	Número de sensores.
$p_{s,Q}(\cdot)$	Função densidade de probabilidade conjunta de dimensão $Q$ para todos os sinais originais.
$p_{s_q,C}(\cdot)$	fdp multidimensional com dimensão $C$ da $q$ -ésima fonte.
$p_{s_q,1}(\cdot)$	fdp unidimensional da $q$ -ésima fonte.
$p_{s,QC}(\cdot)$	fdp conjunta de todas as $Q$ fontes com dimensão $QC$ .
$Q$	Número de fontes de sinal original.
$R_{yy}$	Matriz de correlações.

$\mathbf{R}_{y_p y_p}$	Matriz de autocorrelação.
$\mathbf{R}_{xx}(m)$	Matriz de correlação de tempo curto dos sinais captados pelos sensores.
$\mathbf{R}_{yy}(m)$	Matriz de correlação de tempo curto dos sinais de saída do sistema de separação.
$\mathbf{R}_{y_p y_q}(i)$	Matriz de correlação cruzada de tempo curto para os canais de saída $p$ e $q$ .
$r_{y_p y_q}(i, u, v)$	Elemento da matriz $\mathbf{R}_{y_p y_q}(i)$ na $u$ -ésima linha e $v$ -ésima coluna.
$\tilde{\mathbf{R}}_{y_p y_q}(i)$	Matriz de correlação cruzada de tempo curto calculada através do método da correlação.
$\mathbf{s}(n)$	Vetor contendo os sinais originais.
$s_q(n)$	$q$ -ésima fonte de sinal original como uma função do tempo discreto $n$ .
$SIR_{y_q}$	Relação Sinal-Interferência para o $q$ -ésimo canal de saída.
$\mathbf{W}$	Matriz que contem os filtros de separação.
$w_{pq}(k)$	Coefficientes dos filtros de separação.
$\mathbf{W}(0)$	Ponto de partida do algoritmo para separação cega de sinais.
$\mathbf{W}_{pq}$	Matriz Toeplitz contendo os coeficientes do filtro que modelo o caminho entre o $p$ -ésimo sensor e a $q$ -ésimo saída.
$\mathbf{x}(n)$	Vetor contendo os sinais capturados pelos sensores.
$\mathbf{X}(m)$	Matriz contendo as amostras captadas pelos $P$ sensores.
$x_p(n)$	Sinal captado pelo $p$ -ésimo sensor como uma função do tempo discreto $n$ .
$\mathbf{X}_p(m)$	Matriz contendo $2L \times D$ amostras do sinal capturado pelo $p$ -ésimo sensor.
$\mathbf{y}(n)$	Vetor contendo os sinais de saída do sistema de separação.
$\mathbf{Y}(m)$	Matriz contendo as amostras de todos os $Q$ canais de saída do sistema de separação.
$y_{s,r,q}(n)$	Componente contendo a fonte de sinal original desejada $s_r(n)$ .
$y_{c,q}(n)$	Componente de diafonia presente no $q$ -ésimo canal de saída.

$\mathbf{Y}_q(m)$	Matriz contendo $N \times D$ amostras do $q$ -ésimo canal de saída do sistema de separação.
$\alpha$	Fator de sobreposição.
$\beta$	Parâmetro de influência de uma iteração passada sobre a atual.
$\delta_{y_q}$	Parâmetro para regularização dinâmica das matrizes de autocorrelação.
$\Lambda$	Número de épocas adotadas para a técnica do decaimento exponencial.
$\mu(n)$	Passo de aprendizagem como função do tempo discreto $n$ .
$\Psi(i)$	Termo momentum para $i$ -ésima iteração do algoritmo.

## Resumo

Este trabalho se dedica ao estudo da aplicação de algumas técnicas de adaptação dinâmica do passo de aprendizagem vindas da teoria de redes neurais ao problema da separação cega de sinais misturados convolutivamente. Estas técnicas são estudadas com o objetivo de analisar sua influência na convergência e estabilidade do algoritmo para separação cega adotado aqui. As técnicas para adaptação do passo de aprendizagem abordadas neste trabalho são utilizadas para a separação cega de misturas convolutivas reais e sintéticas.

Além do estudo da utilização das técnicas de adaptação dinâmica do passo de aprendizagem, avaliou-se também algumas modificações no algoritmo para separação cega com o intuito de diminuir sua complexidade computacional e melhorar sua convergência, ou seja, diminuir o tempo de convergência e obter o maior nível de separação possível. Por fim, apresenta-se os resultados de experimentos que mostram a utilidade do algoritmo adotado neste trabalho na de-reverberação e *denoising* de sinais de voz misturados convolutivamente com ruídos.

**Palavras-chave:** BSS, ICA, Redes Neurais.

## **Abstract**

The main purpose of this work is to study the adoption of some techniques originated in the neural network area to dynamically adapt the step size used during the separation process of convolved mixtures. These techniques are studied in order to analyze their influence on both the convergence and stability of the algorithm adopted in this essay. The techniques adopted on this work for the step size adaptation process are used to achieve the blind separation of real and synthetic convolved mixtures.

In addition to the study of the techniques to dynamically adapt the step size, some modifications on the adopted algorithm were evaluated with the purpose of diminishing the computational complexity and improving the convergence presented by it. Furthermore, simulations involving the separation of speech signals convolutively mixed with noises were also run and evaluated.

**Keywords:** BSS, ICA, Neural Networks.

# Capítulo 1

## Introdução

Este capítulo introdutório dá uma visão geral do problema da Separação Cega de Fontes e discute sua importância. Além disto, apresenta-se uma visão geral e os objetivos do trabalho.

### 1.1 Histórico

As técnicas para o processamento cego de sinais nasceram nos anos 80, quando os primeiros equalizadores adaptativos foram desenvolvidos para sistemas de comunicação digital [1][2][3][4]. O problema requeria a criação de um sistema que compensasse os efeitos causados por um canal linear estacionário com uma única entrada e uma única saída (*Single-Input Single-Output*, SISO) sem o conhecimento do sinal de entrada.

A comunidade científica cunhou o termo “cego” para nomear todos os métodos de inversão ou identificação que se baseiam somente em observações dos sinais de saída de um dado sistema. Na verdade, técnicas para processamento cego de sinais são empregadas quando o diagrama de olho está fechado, o que explica a terminologia “cega”. O diagrama de olho é uma ferramenta utilizada para avaliar o efeito combinado do ruído de canal e da interferência inter-simbólica no desempenho de um sistema de comunicação digital. Ele é definido como a superposição

sincronizada de todas as possíveis realizações do sinal de interesse (e.g., sinal recebido, saída do receptor) analisadas dentro de um intervalo de sinalização particular. O nome diagrama de olho vem do fato que ele se assemelha ao olho humano quando se analisa sinais binários. A região central do diagrama é chamada de abertura do olho. Tal diagrama é obtido quando se analisa a sincronicidade de sinais discretos, tais como sinais BPSK, através de um osciloscópio.

O termo “não-supervisionada” foi utilizado algumas vezes no início dos estudos, mas atualmente o termo “cega” é a terminologia adotada em todo o mundo, mesmo que não seja a mais correta, uma vez que só é compreensível no contexto das comunicações digitais.

O problema da Separação Cega de Fontes (SCF) difere do problema da equalização cega, estudado anteriormente por Sato [1], Godard [2] e Benveniste [3] pelo fato de que o sistema linear desconhecido consiste de várias entradas e saídas: tal sistema é conhecido como sistema com múltiplas entradas e múltiplas saídas (*Multiple-Inputs Multiple-Outputs*, MIMO). Inicialmente restrito a canais sem memória, o problema da SCF agora abrange misturas lineares e não-lineares, com ou sem memória.

O problema da SCF foi formulado pela primeira vez por volta de 1982 por Bernard Ans, Jeanny Hérault e Christian Jutten [5][6][7][8], no âmbito de redes neurais, para o entendimento da dinâmica dos movimentos musculares em animais vertebrados [9]. O problema da separação de fontes também foi esboçado de forma independente no âmbito das comunicações [10][11]. As primeiras contribuições em conferências de processamento de sinais [5] e redes neurais [7][8] apareceram por volta de 1985. Estes artigos atraíram imediatamente a atenção de pesquisadores da área de processamento de sinais, principalmente na França e mais tarde em toda a Europa.

Em 1985 e 1986, os primeiros estudos sobre separação de fontes foram apresentados em conferências sobre redes neurais, Cognitiva’85 (Paris, França) e Snowbird’86 (Utah, USA). Estes trabalhos atraíram a atenção de alguns pesquisadores, mas foram ofuscados por novos e interessantes estudos no campo das redes neurais. Por exemplo, na conferência Cognitiva’85, Le Cun [12] publicou um

trabalho sobre um novo algoritmo de aprendizagem para redes neurais multicamadas (*Multi Layer Perceptron*, MLP) o qual ficou conhecido como o algoritmo back-propagation. Durante a conferência Snowbird'86, muitos pesquisadores estavam bastante excitados com os modelos de Hopfield, Mapas Auto-organizáveis de Kohonen, MLP e algoritmos de back-propagation. O real interesse pela SCF surgiu somente bem mais tarde na comunidade de redes neurais, por volta de 1995.

Desde meados da década de 1990, o problema da SCF tem sido abordado por pesquisadores com experiência em vários campos: processamento de sinais, estatística e redes neurais. Numerosas sessões especiais sobre o assunto foram organizadas em conferências internacionais, tais como, GRETSI desde 1993 (França), NOLTA'95 (Las Vegas, USA), ISCAS (Atlanta, USA), EUSIPCO desde 1996, NIPS'97 (Denver, USA), ESANN'97 (Bruges, Bélgica), IWANN'99 (Alicante, Espanha), MaxEnt2006 (Paris, França).

O primeiro workshop internacional totalmente dedicado a este tópico, organizado em Aussois nos Alpes Franceses em Janeiro de 1999, atraiu 130 pesquisadores de todo o mundo. Após os primeiros artigos internacionais, publicados em 1991 na revista *Signal Processing* [13][14][15], várias revistas internacionais contribuíram para a divulgação da SCF: *Traitement du Signal* (em Francês), *Signal Processing*, *IEEE Transactions on Signal Processing*, *IEEE Transactions on Circuits and Systems*, *Neural Computation* e *Neural Networks*. Adicionalmente, um comitê técnico dedicado ao estudo das técnicas de separação cega foi criado em Julho de 2001 pela sociedade de circuitos e sistemas do IEEE (*IEEE Circuits and Systems Society*), e SCF é atualmente uma "EDICS" (*Editors Information Classification Scheme*) na *IEEE Transactions on Signal Processing* e em muitas conferências.

O problema da separação de fontes foi inicialmente investigado para o caso das misturas lineares instantâneas [5]. A generalização para o caso das misturas convolutivas foi considerado no início da década de 1990 [16]. Finalmente, o caso para misturas não-lineares, com exceção de alguns trabalhos, foi abordado no final da década de 1990 [17][18][19]. Além disso, a Análise de Componentes Independentes (*Independent Component Analysis*, ICA), que é uma técnica utilizada para estimar os sinais originais baseada na independência estatística destes sinais, foi introduzida em

1987 [20], e formalizada para misturas lineares por Comon em 1991 [21][22].

Vale ressaltar que na literatura o termo Separação Cega de Fontes, SCF ou BSS em inglês, é por vezes confundido ou mesmo usado como sinônimo para Análise de Componentes Independentes (ICA), embora se refiram a modelos iguais ou similares que podem ser solucionados com os mesmos algoritmos sob a suposição básica de que as fontes originais são estatisticamente independentes entre si. No entanto, para muitos casos envolvendo problemas reais, o objetivo da ICA e da SCF se torna um pouco diferente. De fato, o objetivo da SCF é o de estimar os sinais originais mesmo que eles não sejam completamente independentes entre si, enquanto que o objetivo da ICA é o de determinar uma transformação que garanta que os sinais de saída sejam tão independentes uns dos outros quanto possível. Deve ser salientado que em muitos casos os métodos para ICA utilizam estatísticas de ordem superior (*High Order Statistics*, HOS), enquanto que métodos para SCF estão aptos a utilizar somente estatísticas de segunda ordem (*Second Order Statistics*, SOS) [29]. Métodos de segunda ordem supõem que as fontes apresentam alguma estrutura temporal, ou seja, consideram que as fontes originais apresentam estatísticas de segunda ordem (correlações, variância e média) variantes ao longo do tempo, enquanto que métodos de ordem superior supõem somente sua independência mútua [27][28]. Outra diferença é a de que métodos utilizando estatísticas de ordem superior não podem ser aplicados a sinais que possuam distribuição de probabilidade gaussiana, enquanto que métodos de segunda ordem não apresentam tal restrição [27][28]. Portanto, métodos para SCF não substituem a Análise de Componentes Independentes (ICA) e vice versa, uma vez que cada uma das abordagens é baseada em suposições e frequentemente objetivos diferentes.

Claramente, a independência das fontes está ligada à área de probabilidade e estatística, e ferramentas trazidas deste campo de estudos têm ajudado imensamente na solução do problema da SCF. Por exemplo, a técnica da análise fatorial, que foi amplamente estudada nos anos 50, parece ser outra forma de formalizar ICA, especialmente para o caso do problema da separação. Embora muitos resultados estivessem disponíveis há muitos anos, os resultados de Darmois [23] foram trazidos à luz por Comon em 1991 [21], e mais recentemente, pesquisadores utilizaram alguns

teoremas apresentados no livro de estatística [24] publicado em 1973.

Além da separação de fontes, ICA também pode ser usada para a decomposição de dados complexos, tais como sinais e imagens em bases esparsas cujos componentes possuem a propriedade da independência mútua [25].

As palavras chave: separação de fontes e ICA não foram usadas desde o início. Por exemplo, os primeiros artigos [5][6] tinham títulos bastante longos e intrincados. Em 1986, a palavra discriminação de fontes foi usada, mas a palavra-chave separação de fontes se tornou aceita somente após 1987. Técnicas para separação cega também receberam uma denominação estranha por P. Duvaut 1990 [26], cuja tradução corresponde a “métodos de clarividência”. O termo ICA foi introduzido pela primeira vez em 1987 [20] mas o conceito só foi formalizado mais tarde, em 1991 por Comon [21][22].

O número de artigos publicados sobre os temas SCF e ICA é enorme: em Junho de 2009, 22 mil artigos científicos foram registrados pelo Google Scholar em engenharia, computação e matemática [27]. Entre os livros publicados sobre o tema da SCF se destacam [27][28][29].

## 1.2 Motivação do trabalho

Em várias situações práticas é necessário o uso de sensores para capturar informações. As informações a serem processadas são obtidas através de diferentes formas, variando de acordo com a aplicação. Por exemplo, os dados podem ser obtidos por meio de um conjunto de antenas, vários microfones espalhados em uma sala e sensores biomédicos. Cada um dos sensores capta diferentes conjuntos de dados, uma vez que cada sensor apresenta características distintas, tais como sua localização. Embora diferentes, as medições oriundas dos sensores podem ser relacionadas, visto que elas capturam o mesmo evento ou o mesmo conjunto de sinais originais.

Em muitos casos práticos, os sinais fornecidos pelos sensores são misturas dos diversos sinais (fontes) originais. Além disso, em geral, não é possível observar as fontes de sinais originais diretamente e tão pouco determinar a forma com que elas

foram misturadas. Esse problema é conhecido como o problema da Separação Cega de Fontes, SCF ou em inglês Blind Source Separation, BSS.

Portanto, a Separação Cega de Fontes consiste basicamente na recuperação das fontes de sinal originais a partir dos sinais misturados. Estes sinais misturados são tipicamente obtidos através de um conjunto de sensores, para os quais, cada saída representa uma combinação de diferentes fontes. O adjetivo “cego” é utilizado para salientar o fato de que pouco se sabe a respeito das características ou propriedades dos sinais originais e sobre o processo de mistura e/ou como eles influem nos sinais captados pelos sensores. Nestes casos, somente as informações presentes nas saídas dos sensores e algumas propriedades dos sinais originais são conhecidas. Devido a este fato, o desenvolvimento de métodos para SCF têm sido alvo de muitos trabalhos em diversas áreas de conhecimento, como por exemplo, bioengenharia, telecomunicações, economia, processamento de sinais e imagens, entre outras [27].

Na última década, o processamento cego de sinais vem se tornando uma área bastante promissora de pesquisa dentro do universo do processamento de sinais. Isto é devido ao seu sólido embasamento teórico e suas inúmeras aplicações em potencial. A principal diferença entre as técnicas adotadas para o processamento cego e as convencionais (e.g., equalizadores e filtros adaptativos) está no fato de que as técnicas para processamento cego não necessitam de dados ou seqüências de treinamento e desconsideram o conhecimento sobre os sinais originais e também sobre o sistema de mistura ao qual os sinais originais foram submetidos [21][28][29].

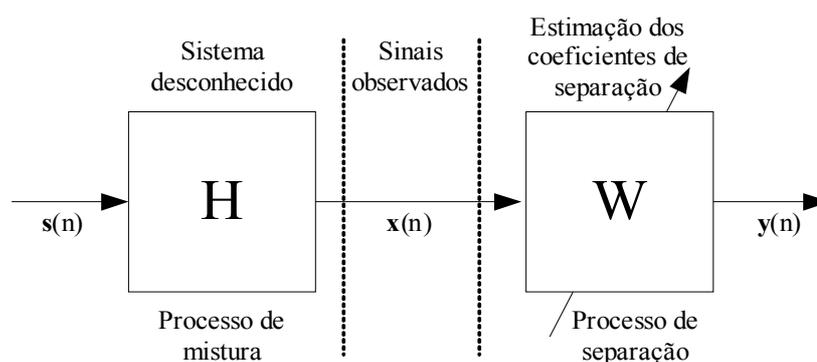
Com o intuito de solucionar o problema da separação cega de fontes é necessário suprir a falta de informações relacionadas aos sinais originais e ao sistema de mistura. Para suprir esta falta de informações, algumas considerações devem ser feitas. Estas considerações impõem restrições às propriedades estatísticas e/ou estruturais dos sinais originais e do sistema de mistura. Nesta dissertação, o sistema de mistura será considerado linear e invariante no tempo (*Linear Time-Invariant*, LTI). Outra consideração feita nesta dissertação é que o número de fontes originais é igual ao número de sensores.

Um problema bastante conhecido na área de processamento cego é o chamado de Cocktail Party. Neste problema, várias pessoas se encontram em uma sala

conversando ao mesmo tempo e além disso, existe uma música tocando ao fundo. Portanto, o objetivo aqui é o de aplicar um método de separação para encontrar o sinal de voz de cada um dos locutores e da música, a partir dos sinais captados pelos microfones.

O problema do Cocktail Party foi mencionado inicialmente em 1953 por Cherry [30], no contexto da capacidade humana de reconhecimento de mensagens de voz. Desde então, muitos pesquisadores têm buscado novas formas de representar esse problema [31][32][33].

O diagrama em blocos de um sistema genérico para SCF é mostrado na Figura 1.1. Como ilustrado na figura, assume-se que os sinais observados  $x(n)$ , onde  $x(n)$  é o vetor contendo os sinais captados pelos sensores, foram gerados através de uma mistura desconhecida dos sinais originais  $s(n)$ , onde  $s(n)$  é o vetor contendo os sinais originais. Portanto, o objetivo da SCF é o de encontrar um sistema de separação que consiga estimar os sinais originais. O vetor  $y(n)$ , entregue na saída do sistema mostrado na Figura 1.1, contém as estimativas produzidas durante o processo de separação.



**Figura 1.1:** Diagrama em blocos de um modelo genérico para SCF.

Existem duas abordagens básicas para se modelar o processo de mistura dos sinais originais. A primeira é o caso para misturas instantâneas, onde o processo de mistura é modelado através de uma matriz. Neste caso, cada sensor (microfone) capta uma combinação linear dos sinais originais.

A segunda abordagem para SCF, a qual é adotada neste trabalho, é para as

chamadas misturas convolutivas, onde o processo de mistura é modelado através de filtros. Neste caso, cada sensor capta versões atenuadas e atrasadas de todos os sinais originais, ou seja, cada sensor capta o resultado da convolução dos sinais originais com a resposta impulsiva do trajeto entre sua origem e o ponto de mistura, neste caso o sensor. As versões atrasadas dos sinais são devidas à reflexões dos sinais nas paredes da sala. No Capítulo 2 apresenta-se uma discussão mais detalhada das abordagens para se modelar o processo de mistura dos sinais originais.

### 1.3 Aplicações

O problema da SCF surge em muitos sistemas multi-sensores: arranjos de antenas, arranjos de sensores químicos e conjuntos de sensores para eletroencefalografia. Este amplo conjunto de possíveis aplicações é uma das razões pelas quais as técnicas para SCF têm obtido grande sucesso entre pesquisadores de diversas áreas. Tais técnicas têm sido empregadas intensivamente em três campos em particular:

- aplicações biomédicas tais como eletrocardiograma, eletroencefalograma, magneto-encefalografia, ressonância magnética, e.g. [34][35];
- separação de fontes de áudio, com aplicações especiais para música e fala, e.g. [36];
- aplicações em telecomunicações, e.g. [27][37][38].

Além disso, métodos para SCF também tem sido empregados no processamento de imagens, marcas d'água digital, pré-processamento para classificação de padrões e monitoramento de sistemas complexos. Uma vasta relação de aplicações nestas áreas se encontra disponível em [29].

Métodos para separação cega de fontes se baseiam essencialmente na estimação de parâmetros desconhecidos, o que normalmente requer um modelo para o sistema de separação, uma função de custo e um algoritmo de otimização. Desta

forma, para que uma dada aplicação atinja resultados relevantes é necessário que:

- as fontes satisfaçam umas das premissas básicas: independência estatística entre as fontes de sinal original (i.i.d i.e., independentes e identicamente distribuídos), não-estacionariedade e/ou esparcidade.
- o sistema de separação seja adequado ao modelo do sistema de mistura (modelo para misturas instantâneas ou convolutivas), o qual assume que o modelo físico que produz os sinais observados esteja correto.

Esparcidade é uma característica apresentada por alguns sinais como, por exemplo, os de voz, os quais possuem uma distribuição muito próxima a da distribuição de Laplace [121][123]. Estes sinais também são conhecidos como super-gaussianos [28], pois possuem valores muito próximos de zero durante grande parte do tempo, fazendo com que sua distribuição de probabilidade tenha um pico muito elevado e estreito ao redor de zero. Portanto, quando sinais são suficientemente esparsos, pode-se assumir que as fontes raramente se sobrepõem.

Caso estas restrições não sejam satisfeitas, ou seja, um modelo ou um critério errado for usado, o algoritmo de otimização irá fornecer estimativas dos sinais originais que são de fato ótimas com relação ao critério e ao modelo adotado, mas cuja relevância não é garantida.

## 1.4 Objetivos

Algoritmos para SCF, em sua grande maioria, baseiam-se em técnicas não-supervisionadas que empregam funções de custo as quais devem ser minimizadas com o intuito de se encontrar um sistema de separação que entregue em sua saída estimativas dos sinais originais. A minimização da função de custo adotada por um algoritmo para SCF é realizada através de métodos de otimização iterativos [28]. Estes algoritmos, em sua grande parte, se baseiam no gradiente da função de custo adotada.

Algoritmos para SCF apresentam frequentemente uma baixa taxa de convergência e uma alta dependência sobre o valor do passo de aprendizagem [28] [29]. Além disso, a forma da superfície formada pela função de custo, para a maioria das aplicações, apresenta irregularidades que são difíceis de se lidar através da técnica do gradiente descendente com um passo de aprendizagem fixo [39]. Dependendo da complexidade da superfície formada pela função de custo, poderão existir vários mínimos locais. Caso o algoritmo encontre uma região que contenha um mínimo local, ele certamente será direcionado para aquele mínimo e o processo de otimização ficará preso naquele ponto, gerando um resultado sub-ótimo, enquanto seria desejável que o algoritmo continuasse a procura até encontrar um mínimo global ou pelo menos um resultado aceitável (abaixo de um limiar predeterminado).

A separação cega de sinais que foram misturados convolutivamente, envolve geralmente, a estimação de um sistema de separação que pode conter centenas ou milhares de parâmetros (coeficientes dos filtros de separação). Este alto número de parâmetros faz com que a superfície formada pela função de custo seja bastante complexa, tornando o processo de busca pelo ponto de mínimo global mais difícil. Sabe-se da teoria de redes neurais que uma manipulação (adaptação) apropriada do passo de aprendizagem durante o processo de treinamento pode levar a uma melhoria da taxa de convergência do algoritmo além de evitar mínimos locais [28][29][40]. Um grande número de métodos para adaptação do passo de aprendizagem no contexto de redes neurais têm sido propostos ao longo dos anos [39].

O principal objetivo deste trabalho é estudar a utilização de algumas das técnicas de adaptação do passo de aprendizagem utilizadas no treinamento de redes neurais no contexto da separação cega de sinais misturados convolutivamente. Com o intuito de atingir tal objetivo, utiliza-se um algoritmo para SCF misturadas convolutivamente baseado em estatísticas de segunda ordem, também conhecidas como momentos de segunda ordem, primeiramente apresentado em [41]. Os resultados do estudo da adaptação do passo de aprendizagem são apresentados no Capítulo 5. Além destes resultados, estudou-se também algumas formas de se reduzir a complexidade computacional do algoritmo utilizado neste trabalho além de verificar sua utilização na separação de sinais misturados com ruídos da base de

dados Aurora [42]. O algoritmo para SCF foi desenvolvido com a ferramenta de simulações Matlab [43].

## 1.5 Organização do trabalho

Esta dissertação foi dividida em cinco partes: introdução ao problema da separação cega de fontes, estudo de um algoritmo para separação cega no domínio do tempo baseado em estatísticas de segunda ordem, descrição das técnicas utilizadas para adaptar o passo de aprendizagem, simulações computacionais e conclusões.

Este capítulo apresenta uma visão histórica e introdutória ao problema da separação cega de fontes, bem como a motivação e os objetivos deste trabalho. Além disso, são apresentadas algumas aplicações para técnicas de SCF.

O Capítulo 2 apresenta a definição do problema da SCF, além de fazer uma breve revisão dos conceitos mais importantes para a compreensão do processo de separação cega de sinais e do algoritmo apresentado no Capítulo 3, tais como, o gradiente descendente, a medida da informação mútua entre variáveis aleatórias e as propriedades dos sinais acústicos nas quais o algoritmo se baseia pra realizar a separação das fontes. No Capítulo 2 também são apresentadas as indeterminações inerentes ao processo de SCF devido às suposições feitas para a realização da separação. Por fim, são apresentados os modelos acústicos utilizados para modelar o processo de mistura dos sinais originais e a medida de desempenho utilizada para avaliar a qualidade da separação obtida pelo algoritmo.

O algoritmo para separação cega adotado neste trabalho é apresentado no Capítulo 3. Este algoritmo têm como objetivo a separação de sinais misturados convolutivamente, ou seja, sinais captados por sensores localizados em um ambiente reverberante. O algoritmo foi inicialmente introduzido em [41] e se baseia em estatísticas de segunda ordem para realizar a separação dos sinais. No Capítulo 4 são apresentadas as técnicas de adaptação do passo de aprendizagem vindas da teoria de redes neurais que serão adotadas neste trabalho para os experimentos mostrados no Capítulo 5. O Capítulo 5, além de apresentar os resultados da adoção das técnicas de adaptação do passo de aprendizagem mostradas no Capítulo 4 para misturas

convolutivas reais e sintéticas, apresenta também resultados de experimentos que visam a diminuição da complexidade do algoritmo utilizado no presente trabalho.

Finalmente, no Capítulo 6 são apresentadas as conclusões e as sugestões de trabalhos futuros com base nos resultados obtidos.

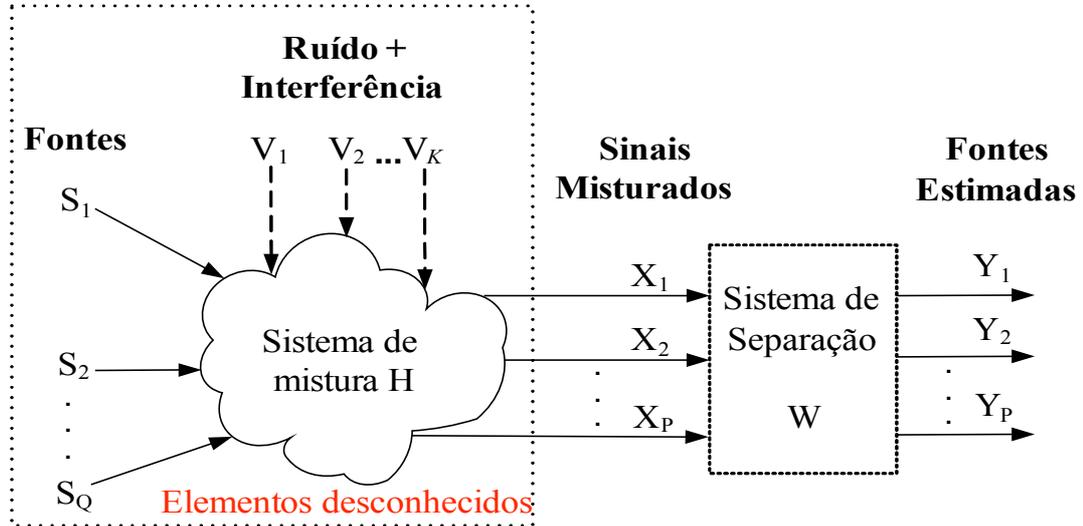
## Capítulo 2

# Principais Conceitos para a Separação Cega de Fontes

Este capítulo apresenta a definição do problema da Separação Cega de Fontes, além de fazer uma breve revisão de alguns dos conceitos mais importantes para compreensão da separação cega de fontes e do algoritmo apresentado no Capítulo 3. Ao final deste capítulo introduz-se a medida de desempenho adotada neste trabalho para avaliar a qualidade da separação obtida pelo algoritmo para SCF apresentado no Capítulo 3.

### 2.1 Definição do problema

O problema da separação cega de fontes considera a existência de um determinado número de sinais emitidos por fontes físicas. Estas fontes podem ser por exemplo, sinais elétricos sendo emitidos por diferentes áreas do cérebro, vários locutores falando simultaneamente em uma sala ou até mesmo ondas de rádio-frequência. Além disso, supõe-se a existência de sensores que estão localizados em posições diferentes, assim, cada sensor capta uma mistura diferente dos sinais originais. O objetivo da separação cega de fontes é recuperar os sinais originais a partir deste conjunto de sinais misturados captados pelos sensores. O problema da separação cega de sinais pode ser definido de forma geral da maneira que se segue. Seja o modelo definido na Figura 2.1.



**Figura 2.1:** Modelo para solução do problema da separação cega de sinais [29].

Neste modelo, considera-se que  $Q$  sinais  $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_Q(n)]^T$ , estão sendo emitidos, ou transmitidos, por alguma fonte ou fontes físicas através de um canal  $\mathbf{H}$ . A seguir, considera-se a existência de sensores em diferentes posições, os quais captam e armazenam as misturas das fontes originalmente transmitidas. No canal de transmissão, além da mistura, os sinais originais sofrem o efeito de ruídos e interferências, as quais serão representadas por  $\mathbf{v}(n)$ . Os sinais misturados captados pelos sensores são representados por  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_P(n)]^T$ .

O objetivo da separação cega de sinais é o de encontrar os sinais originais  $\mathbf{s}(n)$ . Uma forma de fazer isso é através de um sistema de separação  $\mathbf{W}$ , que forneça uma estimativa,  $\mathbf{y}(n) = [y_1(n), y_2(n), \dots, y_P(n)]^T$  destes sinais.

Para o caso geral de separação cega de sinais, considera-se  $P$  sinais misturados e normalmente tem-se que  $P \geq Q$ . Estes sinais, captados pelos sensores, podem ser o resultado da convolução e combinação linear entre os sinais originais  $\mathbf{s}(n)$  e o sistema de mistura  $\mathbf{H}$  ou somente da combinação linear entre as fontes e o sistema, dependendo do modelo de mistura adotado.

Neste trabalho assim como na literatura [27][28][29], a menos que dito o contrário, as fontes  $\mathbf{s}(n)$  são consideradas variantes no tempo, com média zero e

totalmente desconhecidas. Salvo considerações iniciais, o número  $Q$  de fontes é também desconhecido e pode variar ao longo do tempo. Algoritmos destinados à separação cega de sinais tem disponíveis apenas os  $P$  sinais captados pelos sensores  $\mathbf{x}(n)$ .

Portanto, para solucionar o problema descrito acima, é preciso adaptar, ou estimar, os pesos  $w_{ij}$  da matriz de separação  $\mathbf{W}$  para que as medidas provenientes dos sensores  $\mathbf{x}(n)$  possam gerar boas estimativas  $\mathbf{y}(n)$  dos sinais originais  $\mathbf{s}(n)$  ou seja, sinais  $\mathbf{y}(n)$  que se aproximem o máximo de propriedades conhecidas dos sinais originais  $\mathbf{s}(n)$ .

Os coeficientes ótimos para o sistema de separação são encontrados quando os sinais de saída  $\mathbf{y}(n)$  atendem a algum tipo de restrição ou consideração estabelecida à respeito das fontes que se deseja separar. Um tipo de restrição bastante adotado é o da independência estatística entre os sinais de saída do sistema de separação  $\mathbf{W}$ . Esta restrição é válida para uma variedade de sinais, tais como os sinais de voz de múltiplos locutores. Portanto, para que o problema da SCF seja solucionado, deve-se encontrar um sistema de separação cujas as saídas sejam estatisticamente independentes.

O problema descrito acima é frequentemente descrito como **Separação Cega de Fontes, SCF** [28][71][88]. O termo “cega” salienta o fato de que os sinais originais e o sistema de separação são considerados desconhecidos.

A primeira impressão para alguém que acaba de iniciar o estudo sobre SCF é que ela parece ser impossível de ser alcançada, uma vez que o termo “cego” insinua que se deve estimar as fontes originais sem o conhecimento dos parâmetros envolvidos no processo de mistura dos sinais originais, tais como a distribuição de probabilidade dos sinais e as características do canal. É difícil imaginar que se pode estimá-los de alguma forma. De fato, após estudo mais aprofundado, verifica-se que sem algum conhecimento *a priori*, não é possível encontrar-se estimativas únicas para os sinais originais. Entretanto, pode-se estimá-los até certas indeterminações [29]. Em termos matemáticos estas indeterminações ou ambigüidades, podem ser expressas como escalonamento arbitrário, permutação e filtragem dos sinais originais estimados.

Neste trabalho, o sistema de mistura  $\mathbf{H}$  é representado por filtros com resposta ao impulso finita (*Finite Impulse Response Filter*, FIR). Cada um dos filtros que compõem o sistema, modela a função de transferência entre as fontes de sinais e os sensores. Este modelo de sistema de separação é conhecido como **modelo para misturas convolutivas** pelo fato de que considera versões diretas e atrasadas das fontes. Adicionalmente, considera-se neste trabalho a ausência de ruído e que o número de fontes é igual ao número de sensores.

## 2.2 Modelos acústicos para a separação cega de fontes

Desde que Herault e Jutten apresentaram o problema da separação cega de fontes em [5] e [89], vários métodos para solucionar o problema tem sido propostos. Estes métodos, em geral, adotam uma das duas abordagens existentes para modelar o processo de mistura. A primeira abordagem é para o caso com misturas instantâneas, a qual modela o processo de mistura como sendo uma matriz de escalares. A segunda é para o caso que envolve misturas convolutivas. Esta abordagem modela o processo de mistura como um conjunto de filtros. Este trabalho focará na abordagem para misturas convolutivas a fim de modelar os ambientes acústicos nos quais sinais de vozes podem ser captados. A seguir, uma breve discussão dos modelos utilizados para misturas instantâneas e convolutivas é feita.

### 2.2.1 Modelo para misturas instantâneas

O caso mais simples de SCF lida com um modelo de misturas instantâneas, onde nenhuma versão atrasada dos sinais originais é considerada, ou seja, somente sinais diretos se encontram presentes nas misturas captadas pelos sensores. Isto pode ser descrito como um conjunto de observações  $x_p(n)$ ,  $p = 1, \dots, P$ , geradas como sendo uma mistura linear de componentes independentes  $s_q(n)$ ,  $q = 1, \dots, Q$ , pela equação

$$x_p(n) = \sum_{q=1}^P h_{qp} s_q(n) + v_p(n) \quad (2.1)$$

onde  $h_{qp}$  são os coeficientes de peso de cada fonte para cada sensor,  $v_p(n)$  é a possível contribuição de ruído em cada um dos sensores e  $n$  é o índice de tempo discreto.

O objetivo da SCF para o caso de misturas instantâneas é o de encontrar um sistema de separação com coeficientes  $w_{pq}$ , que recuperam estimativas  $y_q(n)$ ,  $q = 1, \dots, Q$  dos sinais originais  $s_q(n)$  através de

$$y_q(n) = \sum_{p=1}^P w_{pq} x_p(n) \quad (2.2)$$

Existem várias aplicações onde o modelo para misturas instantâneas é aplicável. Por exemplo, no estudo do cérebro, a SCF ajuda a identificar componentes fundamentais presentes nas atividades cerebrais captadas por um eletroencefalograma, EEG [29], ou em análises econômicas, onde a SCF é utilizada para encontrar fatores ocultos em dados financeiros [28]. Outra aplicação é no processamento de imagens, onde a extração de características independentes leva a uma melhoria na qualidade das imagens, por exemplo, de observações astronômicas [90]. Um tratamento detalhado do caso para misturas instantâneas e os respectivos algoritmos para sua solução podem ser encontrados em [28]. Neste trabalho considera-se o caso da SCF para ambientes acústicos, i.e., ambientes que irão gerar versões atrasadas dos sinais devido à reverberação, e sendo assim, o modelo para misturas instantâneas não é apropriado, pois este modelo não considera versões atrasadas dos sinais originais. Para a separação de misturas convolutivas faz-se necessário a adoção de um modelo que incorpore versões atrasadas dos sinais captados pelos sensores.

### 2.2.2 Modelo para misturas convolutivas

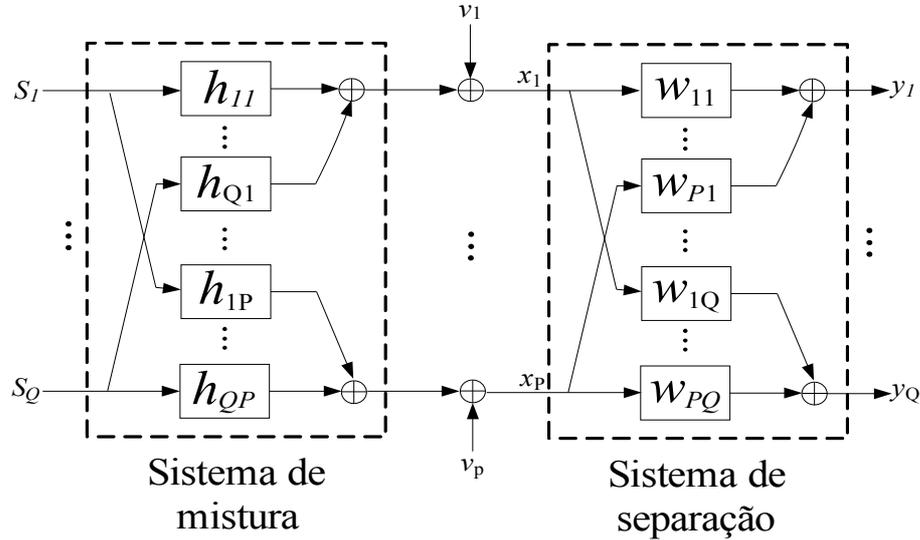
A extensão do modelo para misturas instantâneas através da adição de versões atrasadas dos sinais originais  $s_q(n)$  leva a um sistema de mistura que é formado por filtros com resposta ao impulso finita (FIR) ao invés de escalares. Um sistema de mistura com  $M$  coeficientes é descrito por

$$x_p(n) = \sum_{q=1}^Q \sum_{k=0}^{M-1} h_{qp,k} s_q(n-k) + v_p(n) \quad (2.3)$$

onde  $h_{qp,k}$ ,  $k = 0, \dots, M - 1$  são os coeficientes do filtro FIR que modela os caminhos percorridos pela  $q$ -ésima fonte até o  $p$ -ésimo sensor. Deve ser salientado que os sinais fonte devem ser considerados pontuais para que se possa modelar os vários caminhos percorridos por um sinal através de filtros FIR. Sinais de voz podem ser modelados como fontes pontuais, ou seja, fontes de sinal que emitem ondas sonoras esféricas cuja intensidade é função da distância entre a frente de onda e a fonte. Além dos sinais fonte, um ruído  $v_p(n)$  pode ser captado por cada sensor. Similarmente à SCF para misturas convolutivas, deseja-se encontrar um sistema de separação para o qual, os sinais de saída  $y_q(n)$  sejam representados por

$$y_q(n) = \sum_{p=1}^P \sum_{k=0}^{L-1} w_{pq,k} x_p(n-k) \quad (2.4)$$

O parâmetro  $L$  é o comprimento dos filtros FIR de separação,  $w_{pq,k}$ . Vale ressaltar que o número de taps dos filtros de separação  $L$  deve ser escolhido de tal forma que seja no mínimo igual a  $M$ , sendo que o valor ótimo de  $L$  é encontrado quando este é feito igual a  $M$  [53]. O modelo para misturas convolutivas juntamente com o sistema de separação é apresentado na Figura 2.2 através de um diagrama em blocos.



**Figura 2.2:** Sistema MIMO convolutivo para SCF.

Analisando-se a Figura 2.2 fica óbvio que a SCF pode ser classificada como uma técnica para sistemas com múltiplas entradas e múltiplas saídas (*Multiple Input Multiple Output*, MIMO). Os algoritmos para SCF são comumente desenvolvidos sob a suposição de que o número de fontes  $s_q(n)$ ,  $q = 1, \dots, Q$ , potencialmente ativas simultaneamente é igual ao número de sinais,  $x_p(n)$ ,  $p = 1, \dots, P$ , captados pelos sensores. Este modelo é denominado modelo-padrão para SCF. Além disso, no caso mais geral, onde têm-se um número arbitrário de fontes e sensores pode ser sempre reduzido para o modelo padrão de SCF através do uso de técnicas adequadas [28][45][47]. Ao longo deste trabalho adota-se o modelo-padrão para SCF. O foco principal deste trabalho é a SCF em ambientes acústicos, para os quais o modelo convolutivo se mostra mais apropriado. No Capítulo 3, um algoritmo para a solução do problema da SCF em ambientes acústicos é apresentado.

### 2.3 Características das fontes e sua utilização na separação cega

Nesta seção são discutidas as propriedades apresentadas por sinais acústicos

tais como fala e música, e sua utilização em algoritmos para separação cega. O pressuposto básico de algoritmos para a separação cega de fontes é que as fontes de sinais originais sejam estatisticamente independentes entre si. Estes algoritmos utilizam estatísticas de ordem superior (HOS) e são comumente denominados de algoritmos para ICA. Os primeiros algoritmos para SCF/ICA [5][7][8][21][22][71] foram derivados para o modelo de misturas instantâneas e não levavam em consideração dependências temporais exibidas pelos sinais originais. Quando os sinais originais são considerados estatisticamente independentes entre si e sem nenhuma estrutura temporal, estatísticas de ordem superior, HOS são essenciais para a solução do problema da SCF/ICA. Portanto, originalmente a independência mútua para sinais temporalmente brancos, i.e., sinais cujas amostras são independentes ao longo do eixo temporal, pode ser expressa por

$$p_{s,Q}([s_1(n), \dots, s_Q(n)]^T) = \prod_{q=1}^Q p_{s_q,1}(s_q(n)) \quad (2.5)$$

onde  $p_{s,Q}(\cdot)$  é a função densidade de probabilidade conjunta de dimensão  $Q$  para todos os sinais originais e  $p_{s_q,1}(\cdot)$  é a fdp unidimensional da  $q$ -ésima fonte.

Nos primórdios da pesquisa da separação cega de fontes, trabalhos da comunidade de processamento de sinais sugeriam que somente a decorrelação dos sinais captados pelos sensores, ou seja, a diagonalização de matrizes de correlação para múltiplos atrasos de tempo seria suficiente para se encontrar estimativas dos sinais originais [10][91]. Para misturas instantâneas de sinais não-brancos a afirmação anterior é verdadeira e comprovada em [92][93][94]. Entretanto, para o caso que lida com misturas convolutivas a decorrelação dos sinais captados não é suficiente para se obter a separação dos sinais [92]. Métodos para separação cega de misturas convolutivas que somente realizam a diagonalização de matrizes de correlação ao longo de múltiplos intervalos de tempo não apresentam solução única e na verdade, podem gerar estimativas dos sinais originais que são decorrelacionadas mas não estatisticamente independentes entre si [50]. Portanto, como claramente mostrado por Weisntein e colegas em [92], condições adicionais se fazem necessárias

à solução do problema da SCF para sinais misturados convolutivamente. Estas condições adicionais podem ser conseguidas através da exploração de uma das propriedades dos sinais apresentadas na seqüência.

### 2.3.1 Não-brancura

Sinais de áudio apresentam dependências temporais que podem ser exploradas por métodos para SCF. Para tal, assume-se que as amostras de cada um dos sinais originais não são independentes ao longo do eixo temporal, entretanto, as amostras de sinais pertencentes a diferentes fontes são consideradas mutuamente independentes. Isto leva a uma generalização da equação (2.5)

$$p_{s, QC}([\mathbf{s}_1(n), \dots, \mathbf{s}_Q(n)]^T) = \prod_{q=1}^Q p_{s_q, C}(s_q(n)) \quad (2.6)$$

onde  $\mathbf{s}_q(n)$  é o vetor que contém as  $C$  amostras dependentes temporalmente da  $q$ -ésima fonte,  $p_{s, QC}(\cdot)$  é a fdp conjunta, com dimensão  $QC$ , de todas as fontes e  $p_{s_q, C}(\cdot)$  é a fdp multidimensional com dimensão  $C$  da  $q$ -ésima fonte. Partindo do pressuposto da independência estatística mútua para sinais originais não-brancos, expressa pela equação (2.6), vários algoritmos podem ser encontrados na literatura [49][95][96][97].

A propriedade da não-brancura é explorada principalmente usando-se estatísticas de segunda ordem, através da diagonalização simultânea de matrizes de correlação de saída ao longo de vários atrasos de tempo. Como exemplo, pode-se citar [49][50][98][99] para misturas instantâneas e [93][100] para misturas convolutivas. Em [101] mostra-se que estatísticas de segunda ordem não são suficientes para realizar a separação de sinais brancos; esta somente é conseguida através do uso de estatísticas ordem mais elevada. Isto se deve ao fato de que sinais brancos não geram restrições suficientes para se solucionar o sistema linear que modela o problema da separação cega de fontes. Sinais brancos são sinais que possuem função de correlação idêntica a apresentada pelo processo estocástico

conhecido como ruído branco gaussiano [121], ou seja, são sinais que não apresentam nenhuma correlação temporal (temporalmente independentes) [27][28].

Deve-se notar que algoritmos para separação cega de misturas convolutivas que se baseiam na independência mútua estatística expressa pela equação (2.5) ao invés de (2.6), têm por objetivo a remoção de dependências temporais dos sinais estimados, ou seja, sua descorrelação temporal (branqueamento). Este branqueamento temporal causará alguma distorção aos sinais separados, caso os sinais originais sejam temporalmente não-brancos [28].

### 2.3.2 Não-estacionariedade

As correlações ou dependências de tempo curto para sinais de áudio são geralmente consideradas como sendo variantes no tempo. Portanto, em muitas das aplicações da SCF para ambientes acústicos a não-estacionariedade dos sinais originais pode ser explorada através da diagonalização simultânea de matrizes de correlação de tempo curto de saída em instantes de tempo diferentes, como exemplo, pode-se citar [62][92] para misturas instantâneas e [102][103][104] para misturas convolutivas. A estacionariedade das fontes tem um papel decisivo na separação cega de misturas convolutivas. Se as fontes possuem variâncias não-estacionárias, estatísticas espaço-temporais de segunda ordem são suficientes para se obter a separação, conforme discutido em [92][102]. Sinais não-estacionários geram restrições suficientes para solucionar o sistema linear através do qual o problema da SCF é modelado. Para o cálculo das matrizes de correlação de saída de tempo curto é necessário a separação dos sinais misturados em segmentos, dentro dos quais se considera os sinais como sendo estacionários no sentido amplo. Na seção 3.2.4 apresentam-se os métodos de estimação baseados em segmentos utilizados para o cálculo das matrizes de correlação de tempo curto.

### 2.3.3 Estrutura temporal das fontes

Enquanto que métodos utilizados para Análise de Componentes Independentes (ICA) consideram a mistura dos sinais originais como sendo a mistura linear de variáveis aleatórias independentes, em muitas aplicações entretanto, o que se tem não são variáveis aleatórias mas sim sinais temporais ou séries temporais; ou seja, sinais cujas características variam ao longo do tempo e cujas amostras possuem uma seqüência (ordem), a qual não pode ser alterada. Isto contrasta com o modelo básico para ICA onde as amostras dos sinais captados pelos sensores não possuem ordem. Os sinais poderiam ser embaralhados de qualquer forma e isto não teria nenhum efeito na validade do modelo e tampouco nos métodos de estimação empregados para ICA. Se os sinais originais são sinais temporais, a situação se torna bastante diferente.

De fato, se os sinais captados são séries temporais, eles podem conter muito mais estrutura (informação estatística) do que simples variáveis aleatórias. Por exemplo, as autocorrelações e correlações cruzadas de séries temporais ao longo de vários atrasos de tempo são estatísticas adicionais bem definidas que podem ser utilizadas para melhorar a estimação do modelo de separação. Além disto, esta informação adicional pode tornar possível a estimação do modelo para casos onde os métodos básicos de ICA não podem fazê-lo (por exemplo, para o caso onde os sinais originais são gaussianos mas correlacionados ao longo do tempo).

Modelos padrão para ICA fazem uso de estatísticas espaciais das misturas para estimar cegamente um modelo de separação espacial. Como dito anteriormente, em geral, estatísticas espaciais de ordem superior (*Higher Order Statistics*, HOS) são necessárias para se atingir este objetivo [28]. Entretanto, se os sinais originais são correlacionados temporalmente, estatísticas espaço-temporais de segunda ordem são suficientes para a separação cega sob algumas condições, conforme mostrado em [50]. Em contraste, a separação cega para misturas convolutivas deve utilizar estatísticas espaço-temporais das misturas para estimar um sistema de separação espaço-temporal.

## 2.4 Ambigüidades na separação cega de fontes

Existem algumas ambigüidades inerentes ao processo de SCF dado que o conceito de separação se baseia puramente na suposição da independência mútua entre os sinais originais. O modelo de SCF para misturas instantâneas apresenta as seguintes indeterminações [28]:

- **Fator de escala:** os sinais estimados (separados) podem somente ser estimados até um fator de escala, ou seja, não se pode determinar as variâncias (energias) dos sinais originais. Como consequência disto, não se pode fixar a amplitude dos sinais estimados.
- **Permutação:** a ordem dos sinais estimados não pode ser determinada, ou seja, não se sabe em qual saída do sistema de separação um dado sinal original que foi estimado irá estar presente. É impossível determinar a ordem dos sinais originais estimados sem nenhuma informação *a priori* [29].

Embora não seja muitas vezes citado na literatura como uma ambigüidade, é importante observar que em alguns casos, algoritmos para SCF causam também uma alteração na fase, ou seja, mudança de sinal, no sinal recuperado. Felizmente, esta indeterminação é em muitos casos insignificante [67].

Como ambos, os sinais originais e o sistema de separação são desconhecidos, um possível escalonamento e permutação dos sinais originais pode sempre ser desfeito por um sistema de separação diferente. Devido a impossibilidade de se distinguir se o escalonamento e a permutação ocorreram nos sinais originais ou no sistema de mistura, estas ambigüidades não podem ser resolvidas sem o uso de informações *a priori*, caso somente os sinais captados pelos sensores sejam conhecidos. Desta forma, os sinais originais podem ser recuperados somente até um escalonamento e uma permutação dos sinais.

No caso da SCF para misturas convolutivas as indeterminações são:

- **Filtragem:** os sinais originais podem somente ser estimados até uma operação de filtragem arbitrária.
- **Permutação:** a ordem dos sinais estimados (recuperados) continua ainda arbitrária.

Como no caso para misturas instantâneas, as indeterminações da permutação e da filtragem não podem ser resolvidas sem a incorporação de informações *a priori*. Por exemplo, caso se conheça as posições dos sensores, então a posição de cada fonte pode ser determinada através do sistema de separação [105][106], o que para algumas aplicações é suficiente para resolver o problema da permutação. A ambigüidade da filtragem pode ser reduzida à ambigüidade do escalonamento caso o valor ótimo para tamanho dos filtros de separação seja escolhido [52].

Embora estas indeterminações pareçam ser limitações bastante severas, para um grande número de aplicações elas não são importantes para a solução do problema da separação das fontes originais, uma vez que grande parte da informação relevante sobre elas está contida em suas formas de onda e não em suas amplitudes ou na ordem em que elas são disponibilizadas na saída do sistema de separação [28] [50]. Para alguns modelos dinâmicos, entretanto, não se pode garantir que os sinais estimados terão exatamente as mesmas formas de onda dos sinais originais, e então, os requisitos devem se tornar, em alguns casos, mais flexíveis, na medida em que as formas de onda dos sinais estimados serão versões distorcidas (filtradas) das fontes originais [107][108].

## 2.5 Informação Mútua e a Divergência de Kullback-Leibler

A informação mútua é uma medida da quantidade de informação que uma variável aleatória contém a respeito de outra, ou seja, ela mede o grau de dependência

mútua entre variáveis. Esta medida pode ser interpretada como uma medida de distância empregando-se o que se chama de divergência de Kullback-Leibler. A função de custo adotada neste trabalho é baseada na medida de informação mútua entre variáveis aleatórias fornecida pela divergência de Kullback-Leibler. Uma das principais utilidades desta medida é a de possibilitar a exploração simultânea de várias propriedades dos sinais originais. Esta vantagem é empregada na formulação do algoritmo para separação cega de fontes apresentado no Capítulo 3.

A divergência de Kullback-Leibler é definida entre duas funções densidade de probabilidade  $n$ -dimensionais  $p$  e  $q$  como

$$D(p||q) = \int p(\mathbf{x}) \log \left[ \frac{p(\mathbf{x})}{q(\mathbf{x})} \right] d\mathbf{x} \quad (2.7)$$

A Divergência de Kullback-Leibler pode ser considerada como um tipo de medida de distância entre duas funções densidade de probabilidade, uma vez que ela é sempre não-negativa, e zero se e somente se as duas distribuições forem iguais. Esta é uma consequência direta da convexidade apresentada pelo logaritmo negativo e da aplicação da desigualdade de Jensen [55]. A desigualdade de Jensen define que para qualquer função estritamente convexa  $f$  e qualquer variável aleatória  $y$ , têm-se

$$E \{ f(y) \} \geq f(E \{ y \}) \quad (2.8)$$

Fazendo-se  $f(y) = -\log(y)$  e assumindo-se que  $y = q(x)/p(x)$  onde  $x$  possui a mesma distribuição dada por  $p$ . Então têm-se

$$\begin{aligned} D(p||q) &= E \{ f(y) \} = E \left\{ -\log \left[ \frac{q(\mathbf{x})}{p(\mathbf{x})} \right] \right\} = \int p(\mathbf{x}) \left\{ -\log \left[ \frac{q(\mathbf{x})}{p(\mathbf{x})} \right] \right\} d\mathbf{x} \\ &\geq f(E \{ y \}) = -\log \left\{ \int p(\mathbf{x}) \left[ \frac{q(\mathbf{x})}{p(\mathbf{x})} \right] d\mathbf{x} \right\} \\ &= -\log \left\{ \int q(\mathbf{x}) d\mathbf{x} \right\} = 0 \end{aligned} \quad (2.9)$$

Além disso, a desigualdade em (2.8) somente se torna uma igualdade se e somente se  $y$  for constante. No caso abordado aqui,  $y$  é constante se e somente se as duas distribuições forem idênticas e portanto prova-se a propriedade da Divergência de Kullback-Leibler citada anteriormente.

A Divergência de Kullback-Leibler não é realmente uma medida de distância, pois ela não é simétrica, i.e.,  $D(p||q) \neq D(q||p)$ .

Para aplicar a divergência de Kullback-Leibler como medida do grau de independência entre variáveis aleatórias, considera-se que as variáveis aleatórias  $x_i$  sejam independentes. Desta forma, sua densidade de probabilidade conjunta pode ser fatorada de acordo com a definição de independência, ou seja, elas podem ser fatoradas como o produto das fdp's das  $n$  variáveis aleatórias. Desta forma, poderia-se medir a independência das variáveis  $x_i$  como sendo a divergência de Kullback-Leibler entre a densidade real  $p_{\mathbf{x}}(\mathbf{x})$  e a densidade fatorizada  $p_{x_1}(x_1)p_{x_2}(x_2)\dots p_{x_n}(x_n)$ , onde  $p_{x_i}(x_i)$  são as densidades marginais das variáveis  $x_i$ . Através de manipulações algébricas, mostra-se que esta quantidade se iguala à medida de informação mútua [28][55]. Portanto, como exemplo, pode-se mostrar que o grau de independência, ou seja, a informação mútua entre variáveis aleatórias  $x_i$  com distribuição normal é dada por

$$D(p||q) = \frac{1}{2} \{ \log(\|\text{bdiag } \mathbf{K}\|) - \log(\|\mathbf{K}\|) \} \quad (2.10)$$

onde  $\mathbf{K}$  é a matriz de covariância e  $\text{bdiag}$  é uma operação aplicada à matrizes formadas por várias submatrizes e faz com que todas as submatrizes não pertencentes à diagonal principal da matriz sejam igualadas a zero.

A interpretação da Informação Mútua como sendo a Divergência de Kullback-Leibler implica na seguinte propriedade: *a informação mútua é sempre não negativa e, é igual à zero se e somente se as variáveis aleatórias são independentes.* Esta é uma consequência direta das propriedades da divergência de Kullback-Leibler.

## 2.6 Gradiente Descendente

Muitos dos critérios adotados para SCF almejam a minimização de uma função de custo  $J(m, \mathbf{W})$  em relação a uma matriz de parâmetros  $\mathbf{W}$  a cada iteração  $m$ .

Para a solução do problema da minimização de uma função n-dimensional, a abordagem mais utilizada é a do gradiente descendente que é também conhecida como Máxima Declividade (*Steepest Descent*). Nesta abordagem, a função de custo  $J(m, \mathbf{W})$  é minimizada iterativamente iniciando-se de um ponto  $\mathbf{W}(0)$ . Em seguida calcula-se o valor do gradiente de  $J(m, \mathbf{W})$  naquele ponto e então atualiza-se (move-se)  $\mathbf{W}$  na direção do gradiente negativo através de um pequeno passo. Uma vez lá, repete-se o mesmo procedimento a partir do ponto atual, e assim por diante. Através do processo de otimização iterativo descrito acima, a matriz  $\mathbf{W}$ , para a iteração atual pode ser obtida através da seguinte equação de atualização recursiva

$$\mathbf{W}(m) = \mathbf{W}(m-1) - \mu(m) \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \quad (2.11)$$

com o gradiente calculado no ponto  $\mathbf{W}(m-1)$ . O parâmetro  $\mu(m)$  é o tamanho do passo em direção ao gradiente negativo. Tal parâmetro é conhecido como taxa de aprendizado ou tamanho do passo [56]. A iteração (2.11) prossegue até a convergência de  $\mathbf{W}$ , o que na prática acontece quando a distância Euclidiana entre duas soluções consecutivas,  $\|\mathbf{W}(m) - \mathbf{W}(m-1)\|$ , cai abaixo de um valor predefinido. A diferença entre o valor atual e o anterior de  $\mathbf{W}$  é dada por

$$\mathbf{W}(m) - \mathbf{W}(m-1) = \Delta \mathbf{W} \quad (2.12)$$

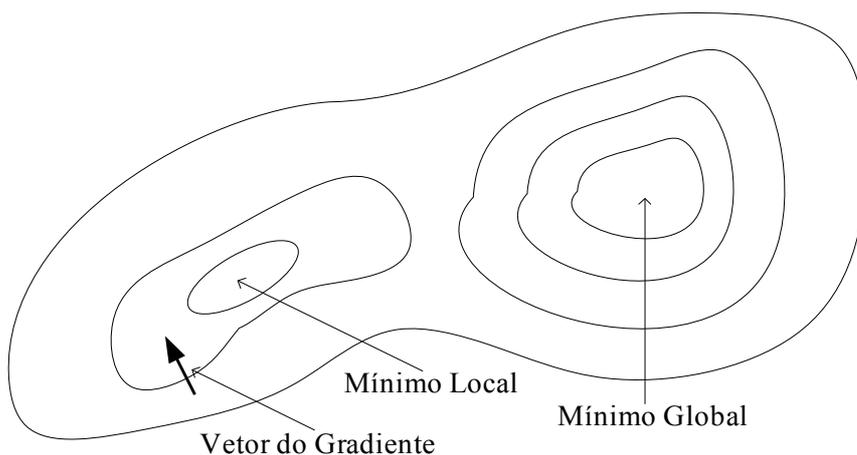
Pode-se escrever a regra (2.11) como

$$\Delta \mathbf{W} = -\mu(m) \frac{\partial J(m, \mathbf{W})}{\partial \mathbf{W}} \quad (2.13)$$

Em muitos casos, a taxa de aprendizado  $\mu(m)$  precisa e deve ser dependente do tempo [27][28].

Geometricamente, a abordagem do gradiente descendente pode ser imaginada como a descida de um morro em direção a um vale. Cada iteração do algoritmo faz com que se aproxime mais da base do vale. Esta interpretação também revela a desvantagem da utilização do gradiente descendente: a menos que a função de custo  $J(m, \mathbf{W})$  seja bem simples e suave, o gradiente descendente irá convergir em direção ao mínimo local mais próximo ao invés do mínimo global. Desta forma, o método do gradiente descendente não oferece nenhuma forma de se evitar mínimos locais. Funções de custo não-quadráticas podem apresentar muitos máximos e mínimos locais e portanto, bons valores iniciais são muito importantes para a inicialização do algoritmo [28][56].

Como exemplo, considera-se o caso mostrado na Figura 2.3. A função  $J(m, \mathbf{W})$  é mostrada como um gráfico de contorno. Na região mostrada na figura existe um mínimo local e um mínimo global. Do ponto inicial mostrado na figura, onde o vetor do gradiente foi plotado, é muito provável que o algoritmo convirja para o mínimo local.



**Figura 2.3:** Gráfico de contorno de uma função de custo com um mínimo local.

Geralmente, a velocidade de convergência pode ser bem baixa nas proximidades de um ponto de mínimo. Isto se deve ao fato de que o gradiente

aproxima-se de zero nestas regiões [28]. A velocidade de convergência do algoritmo depende grandemente do passo de aprendizado  $\mu(m)$  e portanto a escolha de um valor apropriado para este parâmetro é essencial à convergência do método: valores pequenos resultarão em convergência lenta e valores muito grandes causarão instabilidade podendo fazer com que o ponto de solução do algoritmo fique “zigzagueando” em torno do ponto de mínimo.

Uma extensão simples do método do gradiente descendente, popular entre as regras de aprendizagem utilizadas em redes neurais, é o uso de iterações com dois passos ao invés de somente um, como o utilizado em (2.11), o que leva ao método conhecido como momentum [79]. Encontra-se na literatura de redes neurais um grande número de heurísticas, as quais têm como objetivo melhorar a aprendizagem de métodos que utilizam o gradiente descendente [56][117][122].

## 2.7 Medida de desempenho

A qualidade da separação de um algoritmo para SCF é algo complexo de ser avaliado e que depende da área de aplicação. Em alguns casos deseja-se preservar alguma característica de interesse ao invés de simplesmente recuperar-se um sinal que seja o mais próximo possível do original. Por exemplo, em aplicações de reconhecimento de locutor (*Automatic Speaker Recognition*, ASR), é mais importante se preservar algumas características acústicas do que a inteligibilidade do sinal, enquanto que em aplicações para ajuda auditiva, a reconstrução perfeita da fonte original não é tão importante quanto a inteligibilidade [109].

O desempenho de algoritmos para SCF é, em geral, medido utilizando-se métodos derivados de outras áreas do processamento de sinais. Tais métodos podem ser classificados em duas áreas: avaliação subjetiva, na qual é verificada alguma qualidade subjetiva percebida a partir dos sinais estimados [110] ou diferenças subjetivas entre as formas das fontes estimadas e as fontes originais [110][111]; e avaliação objetiva, na qual uma quantidade numérica é associada à qualidade da separação [109][112][113]. No presente trabalho se utilizará uma forma de avaliação objetiva a qual é apresentada a seguir.

### 2.7.1 Medida da supressão das fontes de sinal interferente

Apresenta-se nesta seção a medida de desempenho utilizada neste trabalho para avaliar a qualidade da separação realizada pelo algoritmo adotado. Em geral, algoritmos para SCF focam na supressão de interferências de fontes pontuais e portanto, para ilustrar este aspecto e definir uma medida objetiva apropriada decompõe-se os sinais de saída  $y_q(n)$ ,  $q = 1, \dots, P$  do sistema de separação como

$$y_q(n) = y_{s,r,q}(n) + y_{c,q}(n) \quad (2.14)$$

onde  $y_{s,r,q}(n)$  é a componente contendo a fonte de sinal original desejada  $s_r(n)$ . Em geral, a fonte de sinal original estimada presente no  $q$ -ésimo canal de saída pode ser qualquer uma das fontes originais devido à ambigüidade da permutação. Portanto, nos experimentos apresentados neste trabalho, os canais de saída do sistema de separação foram ordenados manualmente a fim de evitar qualquer permutação. Desta forma, nenhuma permutação dos canais de saída ocorre e então pode se afirmar que  $r = q$ . A componente  $y_{c,q}(n)$  é a componente de diafonia (*crosstalk*) presente no  $q$ -ésimo canal de saída decorrente das fontes remanescentes que não puderam ser suprimidas pelo algoritmo de SCF. Diafonia é a interferência causada por uma fonte não desejada ao canal de saída de uma dada fonte desejada, ou seja, é a presença de um certo nível de sinal de uma fonte não desejada no canal de saída onde se esperava a presença de somente uma fonte de sinal. Portanto, pode-se dizer que a Relação Sinal-Interferência (SIR), é uma medida quantitativa da diafonia presente num dado canal de saída de um sistema de separação cega, pois ela mede o quanto o sinal interferente foi suprimido pelo sistema em relação ao sinal desejado.

A decomposição do sinal de saída em componentes desejada e interferente permite avaliar a supressão do sinal interferente através do cálculo da Relação-Sinal-Interferência (*Signal-to-Interference Ratio*, SIR) para cada canal de saída  $q = 1, \dots, P$ . A relação sinal-interferência é definida como a relação entre a energia do sinal

separado e a energia do sinal interferente, sendo dada por

$$\begin{aligned} SIR_{y_q} &= 10 \log \frac{E \{ y_{s,r,q}^2(n) \}}{E \{ y_{c,q}^2(n) \}} \\ &\approx 10 \log \frac{\sum_n y_{s,r,q}^2(n)}{\sum_n y_{c,q}^2(n)} \end{aligned} \quad (2.15)$$

Como mostrado em (2.15) o operador esperança  $E \{.\}$  deve ser substituído na prática por uma média temporal. Conforme dito em [86], a SIR é a definição de qualidade mais adequada aos métodos de SCF, uma vez que SCF se trata da separação de sinais e não da identificação de sistemas. A SIR é utilizada para avaliar o desempenho do algoritmo apresentado no Capítulo 3 e é dada em decibel (dB).

O sinal  $y_q(n)$  é separado nas componentes  $y_{s,r,q}(n)$  e  $y_{c,q}(n)$  gravando-se cada um dos locutores separadamente. Portanto, as misturas (sinais captados pelos sensores) são obtidas somando-se as contribuições independentes de cada locutor. No caso de misturas reais, a contribuição independente de cada locutor já se encontra convoluída com a resposta ao impulso do ambiente. No caso de misturas sintéticas, o sinal do locutor precisa ainda ser convoluído com a resposta ao impulso do ambiente que esta sendo simulado.

A medida da qualidade da separação apresentada pela SIR é utilizada em [118] como função-objetivo de um algoritmo genético para separação cega de sinais misturados convolutivamente.

## Capítulo 3

# Algoritmo para Separação Cega de Fontes Misturadas Convolutivamente baseado em Estatísticas de Segunda Ordem

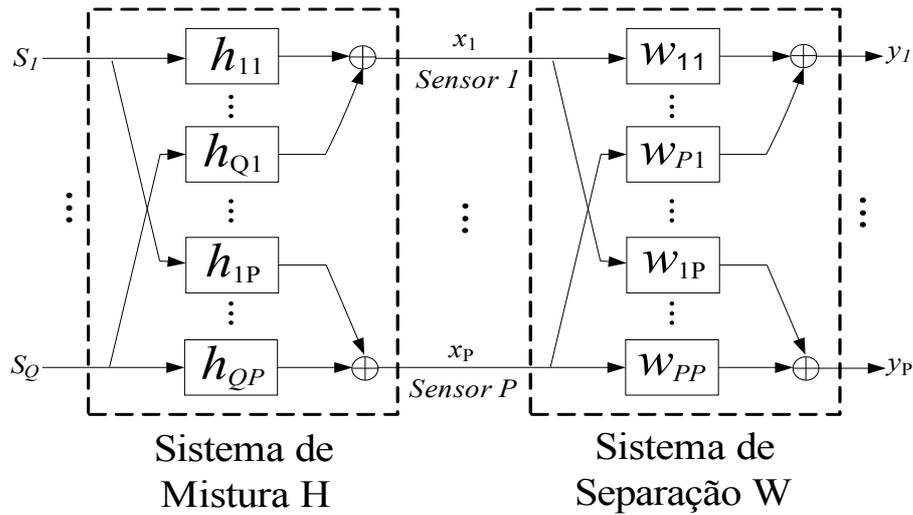
Este capítulo apresenta um algoritmo para separação cega de sinais misturados convolutivamente que se baseia em estatísticas de segunda ordem e explora simultaneamente as propriedades da não-brancura e não-estacionariedade dos sinais originais.

### 3.1 Introdução

A tarefa da separação cega de misturas convolutivas oriundas de séries temporais desconhecidas corresponde à idéia de extrair sinais de informação de sinais que são o resultado da convolução entre os sinais de informação originais e as respostas ao impulso dos diversos caminhos entre a origem dos sinais e os sensores. Este problema emerge em áreas tão distintas quanto a análise de portfólios [44] e a engenharia biomédica [119][120].

Um exemplo proeminente é o chamado problema do Coquetel (*Cocktail Party Problem*) [30], onde se deseja recuperar os sinais de voz de múltiplos locutores que se encontram falando simultaneamente em uma sala. Esta sala poderá geralmente apresentar certo nível de reverberação devido às reflexões nas paredes, isto é, os

sinais originais  $s_q(n)$ ,  $q = 1, \dots, Q$  podem sofrer um processo de filtragem imposto por um sistema com múltiplas entradas e múltiplas saídas (*Multiple Input Multiple Output*, MIMO) antes de serem captados pelos sensores. A Figura 3.1 apresenta o modelo adotado neste trabalho para abordar o problema da separação cega de sinais misturados convolutivamente.



**Figura 3.1:** Sistema MIMO linear para SCF.

Onde  $s_q(n)$ ,  $q = 1, \dots, Q$  são as  $Q$  fontes de sinal original,  $h_{qp}$ ,  $q = 1, \dots, Q$ ,  $p = 1, \dots, P$  são os filtros que modelam o ambiente acústico onde se encontram as fontes de sinais originais,  $x_p(n)$ ,  $p = 1, \dots, P$  são as  $P$  combinações dos sinais originais captados pelos sensores,  $w_{pq}$ ,  $p = 1, \dots, P$ ,  $q = 1, \dots, Q$  são os filtros utilizados para separar os sinais captados pelos sensores e  $y_q(n)$ ,  $q = 1, \dots, P$  são os sinais de saída do sistema de separação cega.

A maioria dos algoritmos utilizados para a separação cega de fontes é construída assumindo-se que o número  $Q$  de fontes de sinais  $s_q(n)$  é igual ao número de  $P$  de sinais captados pelos sensores  $x_p(n)$ . O caso mais geral, entretanto, com um número arbitrário de fontes e sensores pode sempre ser reduzido para o modelo mais conhecido e simples de Separação Cega de Fontes (SCF), ou seja, o caso onde  $P = Q$ .

O caso onde o número de fontes é menor do que o número de sensores é conhecido como SCF sobre-determinado ( $P > Q$ ). A principal abordagem utilizada para simplificar o problema da separação neste caso é aplicar Análise de Componentes Principais (*Principal Component Analysis*, PCA) [28] para retirar as primeiras  $P$  componentes e então usar um dos algoritmos padrão para SCF.

O caso mais difícil onde  $P < Q$  é conhecido como SCF sub-determinado ou SCF com bases incompletas. Para este caso, a esparcidade das fontes pode ser utilizada para se supor que em um dado intervalo de tempo somente uma das fontes de sinal original se encontra ativa [45][46], conforme mencionado na seção 1.3 do Capítulo 1. Esta suposição é utilizada em [47], onde mostra-se que a esparcidade pode ser explorada com o intuito de eliminar  $Q - P$  fontes, e assim, pode-se aplicar novamente uma das técnicas padrão de SCF ao problema. A Transformada de Fourier de Tempo Curto (*Short-Time Fourier Transform*, STFT) é frequentemente adotada para se explorar a esparcidade das fontes [46][48].

Portanto, neste trabalho considera-se o modelo padrão de SCF, onde o número  $Q$  de fontes  $s_q(n)$  é igual ao número de sinais  $x_p(n)$ ,  $p = 1, \dots, P$  (Figura 3.1) capturados pelos sensores. Um sistema de mistura, como o mostrado na Figura 3.1, com  $M$  coeficientes pode ser descrito por

$$x_p(n) = \sum_{q=1}^P \sum_{k=0}^{M-1} h_{qp}(k) s_q(n-k) \quad (3.1)$$

onde  $h_{qp}(k)$ ,  $k = 0, \dots, M - 1$  são os coeficientes do filtro que modela o canal formado pelos diversos caminhos percorridos pelo sinal da  $q$ -ésima fonte até o  $p$ -ésimo sensor.

A SCF busca encontrar um sistema de separação como o mostrado na Figura 3.1, onde as saídas  $y_q(n)$ ,  $q = 1, \dots, P$  são descritas por

$$y_q(n) = \sum_{p=1}^P \sum_{k=0}^{L-1} w_{pq}(k) x_p(n-k) \quad (3.2)$$

onde  $w_{pq}(k)$ ,  $k = 0, \dots, L - 1$  são os coeficientes dos filtros que compõem o sistema MIMO de separação do  $p$ -ésimo sensor para o  $q$ -ésimo canal de saída e  $L$  é o número de coeficientes dos filtros. O parâmetro  $L$  deve ser escolhido de tal forma que seja no mínimo igual à  $M$ , possibilitando assim o modelamento de todas as reflexões. Neste trabalho, o melhor valor para o parâmetro  $L$  é encontrado de forma empírica, i.e., através de experimentos onde seu valor é variado até que o algoritmo apresente a melhor separação possível. Várias simulações foram necessárias para se encontrar tal valor. O valor de  $L$  deve ser encontrado de forma empírica pois não se conhece as características do ambiente acústico. Os coeficientes  $w_{pq}(k)$  do sistema de separação podem reconstruir os sinais originais exceto pelo fato de que eventualmente eles podem estar permutados e apresentar uma energia diferente da original [28], como já mostrado no Capítulo 2.

Para se estimar os  $P^2 L$  coeficientes  $w_{pq}(k)$  do sistema MIMO considera-se neste trabalho uma abordagem que utiliza estatísticas de segunda ordem. É sabido que para sinais reais que apresentam alguma estrutura temporal, estatísticas de segunda ordem geram informações (restrições) suficientes para se solucionar o problema da SCF, em princípio, através do uso de uma das seguintes propriedades destes sinais [28]:

- **Não-brancura:** explorada através da diagonalização simultânea de matrizes de correlação calculadas para os sinais de saída através de múltiplos atrasos de tempo, e.g., [49][50].
- **Não-estacionariedade:** explorada através da diagonalização simultânea de matrizes contendo correlações de tempo curto (*short-time correlation*) dos sinais de saída em diferentes instantes de tempo.

Enquanto existem vários algoritmos que lidam com o problema da separação de sinais misturados convolutivamente utilizando-se unicamente uma das propriedades acima, somente alguns poucos exploram ambas simultaneamente [51] [52]. No trabalho apresentado em [53] é mostrado que a abordagem adotada aqui, de

fato, implica na necessidade de se considerar simultaneamente ambas as propriedades. Isto é provado em [54], onde mostra-se que tal abordagem de segunda ordem é a abordagem ótima para o problema que lida com misturas convolutivas, no sentido em que minimiza a informação mútua [55] entre as fontes de sinais originais. Em [53], alguns resultados são apresentados, demonstrando-se que o algoritmo possibilita a separação eficiente de sinais de voz em ambientes reverberantes e sem ruído.

Um problema que pode ocorrer durante o processo de estimação dos sinais é a possível ocorrência de overfitting. Overfitting ou treinamento excessivo ocorre quando o sistema MIMO de separação memoriza os dados sendo processados (dados de treinamento) ao invés de extrair as características gerais que permitirão a estimação de um sistema de separação. Ele ocorre após certo número de iterações, quando o número de coeficientes  $L$  adotados para os filtros de separação é muito maior do que o número de coeficientes  $M$  dos filtros que modelam o sistema de mistura, acarretando em perda da capacidade de generalização do sistema de separação. O overfitting faz com que a qualidade da separação alcançada pelo sistema de separação seja degradada [56], e portanto, a escolha adequada de  $L$  é essencial para a qualidade da separação alcançada pelo algoritmo apresentado a seguir. Uma alternativa bastante utilizada para se evitar o overfitting é a de parar o algoritmo quando o valor da função de custo começar a aumentar excessivamente [40]. Outro problema que pode aparecer durante a fase de estimação é o underfitting. Ele ocorre quando o sistema de separação possui menos parâmetros do que o necessário, ou seja, quando  $L$  é muito menor do que  $M$ , fazendo com que o sistema MIMO de separação não seja capaz de generalizar o que faz com que o algoritmo não convirja [56].

### **3.2 Algoritmo para separação cega de sinais no domínio do tempo**

Nesta seção apresenta-se a formulação matricial necessária à derivação do

algoritmo para separação cega de fontes no domínio do tempo adotado neste trabalho. Este algoritmo foi inicialmente proposto em [41] e baseia-se em estatísticas de segunda ordem para extrair os sinais originais das misturas captadas pelos sensores.

A parte principal do algoritmo é uma função de custo que leva em consideração as propriedades da não-brancura e não-estacionariedade dos sinais originais. Em seguida considera-se a chamada propriedade da equivariância para o caso convolutivo e a partir daí encontra-se um termo de atualização baseado no gradiente natural.

### 3.2.1 Notação matricial para misturas convolutivas

Nota-se através da Figura 3.1 que em um dado instante de tempo  $n$ , os sinais de saída  $y_q(n)$ ,  $q = 1, \dots, P$  do sistema de separação podem ser descritos pela equação abaixo.

$$y_q(n) = \sum_{p=1}^P \mathbf{x}_p^T(n) \mathbf{w}_{pq} \quad (3.3)$$

onde

$$\mathbf{x}_p(n) = [x_p(n), x_p(n-1), \dots, x_p(n-L+1)]^T \quad (3.4)$$

é um vetor que contém as últimas  $L$  amostras do sinal captado pelo sensor  $x_p$ , e são referentes ao  $p$ -ésimo canal e onde

$$\mathbf{w}_{pq} = [w_{pq,0}, w_{pq,1}, \dots, w_{pq,L-1}]^T \quad (3.5)$$

é o vetor que contém os  $L$  coeficientes do filtro formado pelo caminho entre o  $p$ -ésimo sensor e o  $q$ -ésimo canal de saída.

Além do comprimento  $L$  dos filtros de separação e do número de canais  $P$ , é necessária a introdução de dois novos parâmetros:

- O número de atrasos de tempo  $D$  considerados para as correlações presentes na função de custo ( $1 \leq D \leq L$ ). Estes  $D$  atrasos de tempo são incorporados à função de custo para se explorar a propriedade da não-brancura das fontes originais [41];
- O comprimento  $N$  dos segmentos de sinal de saída que são utilizados como base para a estimativa das correlações de tempo curto.

Para assegurar que a matriz de correlações  $\mathbf{R}_{yy}$ , a qual será apresentada adiante, seja não-singular e, conseqüentemente, tenha determinante diferente de zero, é necessário que se defina um vetor de sinais de saída  $\mathbf{y}_q$  com comprimento  $N$ , onde  $N \geq PD$ , isto é necessário para se obter uma matriz de correlação  $\mathbf{R}_{yy}$  com posto máximo em (3.19) [41][54]. Cada um destes vetores forma um segmento com  $N$  amostras dos sinais de saída  $\mathbf{y}_q$  do sistema de separação e pode ser expresso como mostrado abaixo.

$$\mathbf{y}_q(m) = \sum_{p=1}^P \mathbf{U}_p^T(m) \mathbf{w}_{pq} \quad (3.6)$$

com  $m$  sendo o índice de tempo do bloco/segmento, e

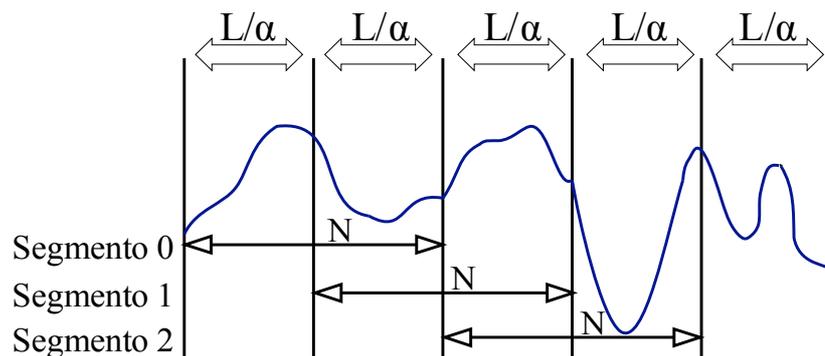
$$\mathbf{y}_q(m) = [y_q(mL), \dots, y_q(mL+N-1)]^T \quad (3.7)$$

$$\mathbf{U}_p(m) = [\mathbf{x}_p(mL), \dots, \mathbf{x}_p(mL+N-1)] \quad (3.8)$$

$$\mathbf{U}_p(m) = \begin{bmatrix} x_p(mL) & \cdots & x_p(mL+N-1) \\ x_p(mL-1) & \cdots & x_p(mL+N-2) \\ \vdots & \ddots & \vdots \\ x_p(mL-L+1) & \cdots & x_p(mL+N-L) \end{bmatrix} \quad (3.9)$$

Conforme pode ser observado,  $\mathbf{U}_p(m)$ ,  $p = 1, \dots, P$  é claramente uma matriz que apresenta estrutura Toeplitz e que possui dimensões  $L \times N$ . Uma matriz Toeplitz possui o valor de cada diagonal descendente da esquerda para a direita constante, i.e., são matrizes que obedecem a igualdade:  $\mathbf{A}_{i,j} = \mathbf{A}_{i-1,j-1}$ . Estas matrizes possuem propriedades bastante úteis na solução de equações lineares.

Analogamente à filtragem adaptativa supervisionada baseada em blocos [57] [58], a abordagem adotada aqui pode também ser executada com segmentos de dados sobrepostos, aumentando-se assim a taxa de convergência. A sobreposição é introduzida trocando-se o índice de tempo  $mL$  nas equações anteriores por  $m(L/\alpha)$ , onde o fator de sobreposição deve variar entre  $1 \leq \alpha \leq L$ . A Figura 3.2 mostra um fator de sobreposição de 50%. O fator de sobreposição será omitido para clareza da formulação matricial.



**Figura 3.2:** Divisão do sinal em segmentos sobrepostos.

Com o intuito de se incorporar  $D$  atrasos de tempo à função de custo adotada neste trabalho, faz-se necessária a introdução da seguinte matriz

$$\mathbf{Y}_q(m) = \begin{bmatrix} y_q(mL) & \dots & y_q(mL-D+1) \\ y_q(mL+1) & \ddots & y_q(mL-D+2) \\ \vdots & \ddots & \vdots \\ y_q(mL+N-1) & \dots & y_q(mL-D+N) \end{bmatrix} \quad (3.10)$$

a qual possui dimensões  $N \times D$ . Esta matriz incorpora  $D$  vetores  $y_q(m)$  os quais são definido em (3.7) e contém  $N$  amostras subseqüentes do sinal de saída do sistema de separação.

Usando-se a definição acima, a equação (3.6) pode ser estendida para

$$\mathbf{Y}_q(m) = \sum_{p=1}^P \mathbf{X}_p(m) \mathbf{W}_{pq} \quad (3.11)$$

onde  $\mathbf{X}_p(m)$  são matrizes obtidas dobrando-se a dimensão das matrizes Toeplitz  $\mathbf{U}_p(m)$ , conforme mostrado abaixo.

$$\mathbf{X}_p(m) = [\mathbf{U}_p^T(m), \mathbf{U}_p^T(m-1)] \quad (3.12)$$

e  $\mathbf{W}_{pq}$  são matrizes com dimensões  $2L \times D$  definidas como

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq,0} & 0 & \dots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & w_{pq,L-1} \\ 0 & \dots & 0 & 0 \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix} \quad (3.13)$$

Com o intuito de se garantir convoluções lineares para todos os elementos de  $\mathbf{Y}_q(m)$  até o número máximo de atrasos de tempo  $D = L$ , as dimensões de  $\mathbf{X}_p(m)$  e  $\mathbf{W}_{pq}$  devem ser, respectivamente  $N \times 2L$  e  $2L \times D$ , conforme mostrado em [41].

As matrizes  $\mathbf{U}_p^T(m-1)$ ,  $p = 1, \dots, P$ , também apresentam estrutura Toeplitz de modo que a primeira linha de  $\mathbf{X}_p(m)$  contém  $2L$  amostras capturadas pelo  $p$ -ésimo sensor e cada linha subsequente é deslocada de uma amostra para a direita e portanto, contém uma nova amostra.

Matrizes que exibem a estrutura Toeplitz especial como a mostrada em (3.13) são conhecidas como matrizes *Sylvester*, de convolução ou de transmissão [59][60][61] e possibilitam expressar a operação de convolução como o produto entre uma matriz e um vetor. No restante deste trabalho irá se utilizar o termo matrizes *Sylvester* para se referir a estas estruturas.

Analisando-se a matriz definida em (3.13), nota-se que para o caso geral  $1 \leq D \leq L$ , ou seja, o caso onde o parâmetro  $D$  pode variar entre 1 e  $L$ , as últimas  $L-D+1$  linhas devem ser preenchidas com zeros para assegurar a compatibilidade com  $\mathbf{X}_p(m)$ . Vale ressaltar que quando  $D = 1$ , somente um atraso de tempo é considerado e portanto, o algoritmo resultante poderá somente explorar a propriedade da não-estacionariedade dos sinais originais. Desta forma, percebe-se que a propriedade da não-brancura só é explorada caso  $D > 1$ .

Para possibilitar uma notação conveniente do algoritmo, todos os canais são combinados gerando-se a seguinte equação compacta:

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W} \quad (3.14)$$

com as matrizes

$$\mathbf{Y}(m) = [\mathbf{Y}_1(m), \dots, \mathbf{Y}_p(m)] \quad (3.15)$$

$$\mathbf{X}(m) = [\mathbf{X}_1(m), \dots, \mathbf{X}_p(m)] \quad (3.16)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix} \quad (3.17)$$

### 3.2.2 Critério de otimização

Após ter-se definido a formulação matricial compacta (3.14), utilizada para modelar a filtragem MIMO sofrida pelos sinais captados pelos sensores ao serem processados pelo sistema de separação, apresenta-se nesta seção o critério de otimização primeiramente derivado em [41] como uma generalização de [62] que, simultaneamente, explora as propriedades da não-branca e da não-estacionariedade dos sinais originais.

Com o intuito de se encontrar um critério de otimização que simultaneamente explore as propriedades da não-branca e não-estacionariedade das fontes, deve-se definir primeiramente as seguintes matrizes de correlação de tempo curto

$$\mathbf{R}_{xx}(m) = \mathbf{X}^H(m) \mathbf{X}(m) \quad (3.18)$$

$$\mathbf{R}_{yy}(m) = \mathbf{Y}^H(m) \mathbf{Y}(m) \quad (3.19)$$

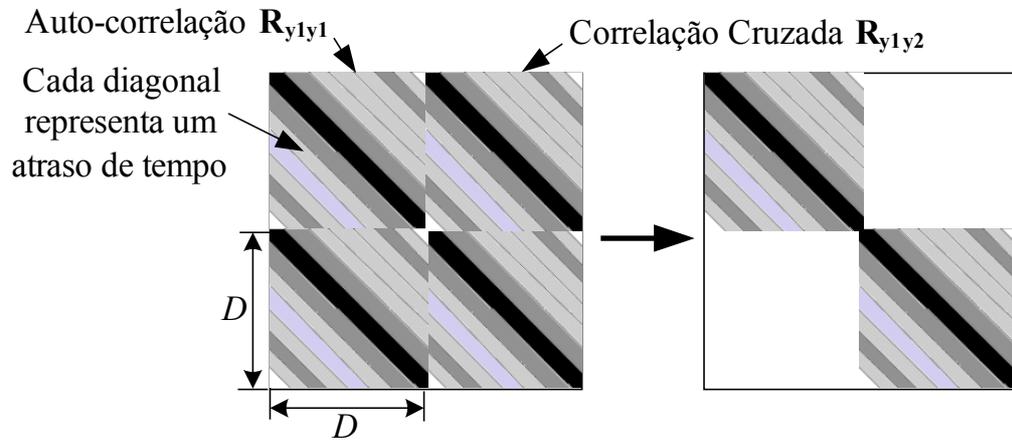
onde  $^H$  indica o transposto conjugado da matriz. No presente trabalho, os sinais acústicos são considerados sinais com valores reais e desta forma, o transposto conjugado tem o mesmo efeito do transposto e portanto,  $^H$  pode ser substituído pelo sobrescrito  $^T$ . A função de custo adotada neste trabalho é apresentada na equação abaixo.

$$\begin{aligned} \mathbf{J}(m, \mathbf{W}) &= \frac{1}{M} \sum_{i=0}^{M-1} \{ \log |\text{bdiag } \mathbf{Y}^T(i) \mathbf{Y}(i)| - \log |\mathbf{Y}^T(i) \mathbf{Y}(i)| \} \\ &= \frac{1}{M} \sum_{i=0}^{M-1} \{ \log |\text{bdiag } \mathbf{R}_{yy}(i)| - \log |\mathbf{R}_{yy}(i)| \} \end{aligned} \quad (3.20)$$

Esta função de custo é baseada na Informação Mútua de Shannon [55][63] e é utilizada com o intuito de minimizar a informação mútua entre os canais de saída. Analisando-se a equação (3.20) percebe-se que os sinais de saída  $y_q, q = 1, \dots, P$  do sistema de separação são divididos em  $M$  segmentos para que se possa realizar a medida da informação mútua entre os sinais de saída para cada um dos segmentos. Ao final, o somatório das informações mútuas ao longo dos  $M$  segmentos é normalizado pelo número de segmentos em que o sinal foi dividido. Como os sinais de saída são considerados não-estacionários, cada segmento desses sinais pode apresentar valores diferentes para as  $PD \times PD$  correlações incorporadas à matriz  $\mathbf{R}_{yy}(m)$ , gerando, assim, informações (restrições) adicionais para a separação dos sinais [53][54].

As matrizes (3.18) e (3.19), com dimensões  $2PL \times 2PL$  e  $PD \times PD$  respectivamente, devem ser avaliadas em vários instantes de tempo e incorporam vários atrasos de tempo com o objetivo de explorar ambas propriedades [116]. Existem dois métodos para se estimar as matrizes de correlação para sinais de saída não-estacionários: o método da correlação e o método da covariância. Ambos são utilizados nos problemas de predição linear [64] e serão apresentados na seção 3.2.6. Enquanto o método da correlação leva à uma complexidade computacional menor devido à estrutura Toeplitz das matrizes  $\mathbf{R}_{yy}$ , o método da covariância apresenta maior precisão. Todos os experimentos apresentados neste trabalho consideram o método da covariância e somente onde mencionado utiliza-se o método da correlação.

A função de custo (3.20) torna-se zero se e somente se todos os elementos das submatrizes não pertencentes à diagonal principal de  $\mathbf{Y}^T \mathbf{Y}$ , isto é, as correlações cruzadas entre os vários sinais de saída do sistema de separação, forem iguais a zero para todos os atrasos de tempo considerados (veja a Figura 3.3). Portanto, em adição à não-estacionariedade, a equação (3.20) explicitamente também explora a propriedade da não-brancura dos sinais de saída.



**Figura 3.3:** Ilustração de 3.20 para o caso 2 x 2.

Como foi adotada a formulação matricial (3.14) para se efetuar os cálculos das matrizes de correlação de tempo curto  $\mathbf{Y}^T(m)\mathbf{Y}(m)$ , a função de custo inclui todos os  $D$  atrasos de tempo de todas as auto-correlações e correlações cruzadas dos sinais de saída do sistema de separação cega. A desigualdade de Oppenheim  $\sum_q \log |\mathbf{Y}_q^T \mathbf{Y}_q| \geq \log |\mathbf{Y}^T \mathbf{Y}|$  [65] assegura que o primeiro termo dentro dos colchetes na equação (3.20) é sempre maior ou igual ao segundo termo, sendo que a igualdade só ocorre se todos os elementos não pertencentes à diagonal principal de  $\mathbf{Y}^T \mathbf{Y}$ , isto é, as correlações cruzadas de saída se tornarem iguais a zero.

O algoritmo utilizado neste trabalho é baseado no gradiente de primeira ordem. Para expressar as equações de atualização dos coeficientes dos filtros de separação exclusivamente através de matrizes *Sylvester*  $\mathbf{W}$ , toma-se o gradiente em relação à  $\mathbf{W}$ .

$$\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W}) = \frac{\partial \mathbf{J}(m, \mathbf{W})}{\partial \mathbf{W}} \quad (3.21)$$

E como resultado tem-se

$$\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W}) = \frac{2}{M} \sum_{i=0}^{M-1} \mathbf{R}_{xy}(i) \mathbf{R}_{yy}^{-1}(i) \{ \mathbf{R}_{yy}(i) - \text{bdiag} \mathbf{R}_{yy}(i) \} \text{bdiag}^{-1} \mathbf{R}_{yy}(i) \quad (3.22)$$

Com um procedimento de otimização iterativo, a matriz de separação corrente pode ser obtida através da equação de atualização recursiva

$$\mathbf{W}(m) = \mathbf{W}(m-1) - \mu(m) \Delta \mathbf{W}(m) \quad (3.23)$$

onde  $\mu$  é o tamanho do passo em direção ao mínimo da função de custo.

Este parâmetro é conhecido da teoria de redes neurais como taxa de aprendizagem e  $\Delta \mathbf{W}(m)$  é a matriz de atualização, a qual é feita igual a  $\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W})$  para a adaptação do gradiente descendente. Um fato importante a ser notado é que devido ao processo de adaptação do gradiente descendente, a matriz de separação se torna variante no tempo. O problema da inversão das matrizes de correlação que normalizam a equação (3.22) é discutido na seção 3.2.6.

Conforme demonstrado em [66][67], a matriz  $\Delta \mathbf{W}(m)$ , a qual contém os coeficientes de atualização dos filtros de separação, não apresenta estrutura *Sylvester*. Como pode ser visto em (3.13), a matriz  $\mathbf{W}_{pq}$  possui estrutura *Sylvester* e sendo assim, a matriz  $\Delta \mathbf{W}(m)$  deve apresentar também a mesma estrutura [67]. Desta forma, deve-se aplicar algum tipo de restrição à matriz de atualização  $\Delta \mathbf{W}(m)$  que faça com que ela passe a apresentar estrutura *Sylvester*. Esta restrição será a partir de agora denominada de restrição *Sylvester*, e será simbolizada por **SC** (*Sylvester Constraint*).

A restrição *Sylvester* é imposta tomando-se a média aritmética dos elementos em cada uma das diagonais das matrizes  $\Delta \mathbf{W}_{pq}$  para  $p, q = 1, \dots, P$  [67][115]. Esta forma de restrição apresenta uma alta carga computacional, pois para se calcular a média aritmética, todos os  $P^2 LD$  elementos do gradiente descendente  $\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W})$  devem ser calculados. Entretanto, existem duas formas de

implementação que possibilitam aproximar a restrição *Sylvester* calculando-se apenas alguns elementos do gradiente  $\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W})$  [115]. A primeira aproximação calcula somente os primeiros  $L$  elementos da primeira coluna de cada uma das submatrizes de  $\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W})$  e repete estes elementos formando-se assim matrizes com estrutura *Sylvester*, esta aproximação é chamada de restrição da coluna e é denotada por  $SC_C$ . Em [41][67] mostra-se que o parâmetro  $D$  deve ser feito igual a  $L$  para que todos os elementos de  $\mathbf{W}$  possam ser ajustados independentemente. A segunda forma eficiente de aproximação da restrição *Sylvester* calcula somente os elementos da  $L$ -ésima linha de cada uma das submatrizes de  $\nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W})$  e também repete os elementos, formando-se assim matrizes *Sylvester*, esta aproximação é chamada de restrição da linha e é denotada por  $SC_R$ . Para a segunda aproximação supõe-se que o número de atrasos de tempo  $D$  levados em consideração para o cálculo das matrizes de correlação de saída seja igual à  $L$ , desta forma, a  $L$ -ésima linha das submatrizes conterá os  $L$  coeficientes de atualização  $w_{pq,k}$ ,  $k = 0, \dots, L-1$  dos filtros. A eficiência computacional da restrição *Sylvester* e suas aproximações eficientes são comparadas no Capítulo 5. A restrição *Sylvester* da linha  $SC_R$  é a aproximação que apresenta menor complexidade computacional e resultados muito próximos aos obtidos quando se utiliza a restrição *Sylvester*  $SC$  conforme é mostrado no Capítulo 5.

Dependendo da restrição *Sylvester* escolhida, diferentes seqüências de valores  $y_q(n)$  serão utilizadas para se calcular as correlações cruzadas usadas para se obter a matriz de atualização  $\Delta \mathbf{W}(m)$ , ou seja, cada coluna da matriz de correlações cruzadas tem seus valores calculados com base em uma referência de tempo diferente. Quando se utiliza a restrição da coluna ( $SC_C$ ) os coeficientes dos filtros de separação são multiplicados (convolidos) somente com os valores da primeira coluna da matriz de correlações cruzadas, enquanto que quando se utiliza a restrição da linha ( $SC_R$ ) multiplica-se os coeficientes dos filtros de separação com todas as seqüências de correlação presentes na matriz de correlações cruzadas, conforme mostrado em [41][66].

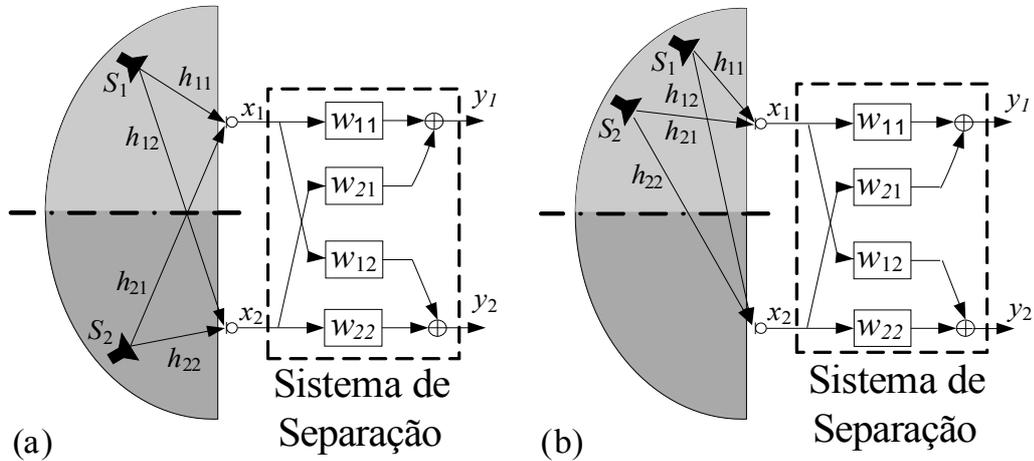
Examinando-se a multiplicação entre a matriz  $W_{pq}$  e a segunda parte da equação (3.26), observa-se que devido à estrutura *Sylvester*, a submatriz superior esquerda com dimensões  $L \times L$  de  $\Delta W_{pq}$  é completamente determinada pela submatriz superior esquerda  $L \times L$  de  $W_{pq}$  e a segunda parte de (3.26). Além disso, analisando-se a matriz definida em (3.13), percebe-se que sua primeira coluna e sua  $L$ -ésima linha contêm todos os  $L$  coeficientes do filtro de separação, validando assim os procedimentos apresentados acima para se obter aproximações para a restrição *Sylvester*. Portanto, conclui-se que o valor ótimo para o parâmetro  $D$ , ou seja, o número de atrasos de tempo que devem ser levados em consideração para o cálculo das matrizes de correlação, é obtido quando este é feito igual a  $L$ . Esta conclusão é confirmada em [41] e [53] através de várias simulações sob várias condições.

As formas eficientes de restrição eliminam o processo de média dos elementos mas ainda possibilitam a obtenção dos  $P^2L$  valores, os quais são realmente relevantes à atualização dos filtros. Adicionalmente pode ser notado que para ambas aproximações a multiplicação matriz/matriz em (3.22) se torna uma multiplicação entre uma matriz e um vetor [68]. Como mostrado em [68], a aproximação da restrição *Sylvester* ( $SC$ ) através de ( $SC_C$ ) ou de ( $SC_R$ ) não leva a uma degradação significativa da separação. Resultados apresentados em [67] e no Capítulo 5 deste trabalho, mostram que a restrição *Sylvester* da linha ( $SC_R$ ) é a aproximação mais adequada para a restrição *Sylvester* ( $SC$ ).

### 3.2.3 Métodos para inicialização das matrizes de separação

Quando se aplica o algoritmo para SCF apresentado neste trabalho a ambientes reais com  $P = 2$ , ou seja, duas fontes e dois sensores (microfones), pode-se distinguir dois tipos de cenários acústicos, conforme mostrado na Figura 3.4. De acordo com a localização das fontes de sinal original, métodos diferentes para a inicialização da matriz  $W$  são necessários. No arranjo bidimensional mostrado na Figura 3.4a cada uma das fontes está localizada em uma metade do plano formado

pelos sensores, e portanto apenas filtros FIR causais são necessários para se obter a separação dos sinais; isto é comprovado em [66][67]. Desta forma, os filtros de separação  $W_{pp}$  ( $p = 1, \dots, P$ ) podem ser inicializados com um impulso unitário no primeiro coeficiente, ou seja,  $w_{pp,0} = 1$  enquanto que todos os outros coeficientes dos filtros  $w_{pq}$ ,  $p \neq q$  são feitos iguais a zero. Por outro lado, se as fontes se encontram localizadas no mesmo semiplano, como mostrado na Figura 3.4b, um filtro FIR  $W_{21}$  não-causal se faz necessário [67]. Na prática, o problema da não-causalidade pode ser solucionado através da inicialização dos filtros FIR  $W_{pp}$  ( $p = 1, \dots, P$ ) com impulsos unitários deslocados, similarmente ao adotado por algoritmos para processamento adaptativo de sinais espaciais (*Adaptive Beamforming*) [69]. O deslocamento apropriado é determinado pela distância entre os sensores e pelo atraso máximo de chegada dos sinais entre os sensores. Em [66] o deslocamento do impulso unitário é feito igual a  $L/2$ , ou seja,  $w_{pp,L/2} = 1$ .



**Figura 3.4:** Arranjos para SCF que requerem (a) somente filtros FIR causais e (b) filtros FIR causais e não-causais.

A escolha do método de inicialização das matrizes de separação determina a aproximação da restrição *Sylvester* ( $SC$ ) mais adequada a ser adotada. Para o caso onde se tem misturas causais (Figura 3.4a), isto é, inicialização com  $w_{pp,0} = 1$  ambas as aproximações ( $SC_C$ ) e ( $SC_R$ ) podem ser empregadas. Como descrito

acima, para misturas não-causais é necessário que se inicialize as matrizes de separação com um impulso unitário deslocado, como por exemplo  $w_{pp,L/2} = 1$ . Em [54][67] mostra-se que para o caso com misturas convolutivas, a matriz resultante da aplicação da restrição *Sylvester* da coluna ( $SC_C$ ) sempre terá os coeficientes  $\Delta w_{pq,k}$  para  $0 \leq k \leq L/2 - 1$  iguais a zero; isto é, estes coeficientes não podem ser estimados. Portanto, para a inicialização com um impulso unitário deslocado, somente a restrição *Sylvester* da linha ( $SC_R$ ) pode ser utilizada.

A restrição *Sylvester* da linha ( $SC_R$ ) juntamente com a inicialização utilizando o impulso unitário deslocado devem ser adotadas quando não se possui nenhuma informação a priori sobre a localização das fontes. Isto se deve ao fato da alta generalidade obtida por esta combinação [67].

### 3.2.4 Gradiente Natural

É sabido que o método do gradiente descendente, isto é,  $\Delta W(m) = \nabla_w J(m, W)$ , apresenta uma convergência lenta para vários problemas práticos devido à sua dependência dos dados sendo processados [41].

O gradiente descendente de uma função  $J(m, W)$  aponta em direção ao menor valor daquela função em um sistema de coordenadas Euclidianas Ortogonais. Entretanto, o espaço formado pelos parâmetros da função de custo  $J(m, W)$  nem sempre é Euclidiano, podendo apresentar uma estrutura métrica Riemanniana, como foi mostrado por Amari em [70]. Em tais casos, a direção da máxima declividade é dada pelo método chamado de Gradiente Natural [70] ou Gradiente Relativo [71]. Com o intuito de usar-se o gradiente natural como fator de atualização  $\nabla_w J(m, W)$ , a seguinte modificação deve ser aplicada ao gradiente descendente (3.22) [28][29] [41]

$$\nabla_w^{NG} J(m, W) = \nabla_w J(m, W) W^T W \quad (3.24)$$

Para a abordagem baseada em (3.14), deve-se reorganizar o lado direito da equação

acima, ou seja

$$\nabla_{\mathbf{W}}^{NG} \mathbf{J}(m, \mathbf{W}) = \mathbf{W}\mathbf{W}^T \nabla_{\mathbf{W}} \mathbf{J}(m, \mathbf{W}) = \mathbf{W}\mathbf{W}^T \frac{\partial \mathbf{J}(m, \mathbf{W})}{\partial \mathbf{W}} \quad (3.25)$$

o que leva à seguinte expressão:

$$\nabla_{\mathbf{W}}^{NG} \mathbf{J}(m, \mathbf{W}) = \frac{2}{M} \sum_{i=0}^{M-1} \mathbf{W}(m) \{ \mathbf{R}_{yy}(i) - \text{bdiag} \mathbf{R}_{yy}(i) \} \text{bdiag}^{-1} \mathbf{R}_{yy}(i) \quad (3.26)$$

a qual é utilizada como atualização  $\Delta \mathbf{W}(m)$  em (3.23). O operador  $\text{bdiag}$  é uma operação aplicada à matrizes formadas por várias submatrizes e faz com que todas as submatrizes não pertencentes à diagonal principal da matriz sejam igualadas a zero.

A modificação introduzida acima tem como objetivo remover a dependência apresentada pelo método do gradiente descendente nos dados sendo processados. Esta dependência pode ser notada em (3.22), onde percebe-se que o fator de atualização depende do sinal de entrada do sistema de separação  $\mathbf{X}(m)$ .

A equação (3.26) apresenta desempenho uniforme ou seja, seu desempenho não depende do valor atual da matriz de misturas. Isto significa que o resultado da separação dos sinais misturados através da otimização daquela equação é sempre o mesmo para qualquer que seja a matriz de misturas [28][71]. O desempenho uniforme apresentado pela equação (3.26) é devido à propriedade da equivariância, introduzida na equação de atualização dos coeficientes dos filtros de separação multiplicando-se a equação (3.22) pela matriz  $\mathbf{W}\mathbf{W}^T$ , [61][72]. Outra vantagem da utilização do gradiente natural é a redução da complexidade computacional, uma vez que não se torna mais necessária a inversão da matriz  $\mathbf{R}_{yy}$  de dimensão  $PD \times PD$  em (3.22). Neste caso, somente submatrizes com dimensão  $D \times D$  precisam ser invertidas [67].

### 3.2.5 Métodos para estimação das matrizes de correlação

A função de custo apresentada na Equação (3.26) envolve o cálculo de matrizes de autocorrelação e de correlação cruzada para se obter a matriz  $\mathbf{R}_{yy}$ . Esta seção apresenta métodos para se calcular tais matrizes.

Dos problemas de predição linear é sabido que, em princípio, existem duas técnicas para a estimação de matrizes de autocorrelação e correlação cruzada de tempo curto, os chamados método da correlação e da covariância [73][74][75]. Ambos baseiam-se em estatísticas de segunda ordem e são utilizados com sinais não-estacionários. Enquanto o método da correlação apresenta uma complexidade computacional baixa devido à estrutura Toeplitz das matrizes  $\mathbf{R}_{yy}$ , o método da covariância apresenta uma maior precisão para os cálculos das correlações necessárias em (3.26) [64]. Ambos são apresentados na seqüência.

#### Método da covariância

A matriz de correlação cruzada baseada em estatísticas de segunda ordem é definida para cada canal de saída como sendo:

$$\mathbf{R}_{y_p y_q}(i) = \frac{1}{N} \mathbf{Y}_p^T(i) \mathbf{Y}_q(i) \quad (3.27)$$

com a matriz  $\mathbf{Y}_p(i)$  sendo definida abaixo.

$$\begin{aligned} \mathbf{Y}_p(i) &= [\mathbf{y}_p(iL), \dots, \mathbf{y}_p(iL - D + 1)] \\ &= \begin{bmatrix} y_p(iL) & \dots & y_p(iL - D + 1) \\ y_p(iL + 1) & \ddots & y_p(iL - D + 2) \\ \vdots & \ddots & \vdots \\ y_p(iL + N - 1) & \dots & y_p(iL - D + N) \end{bmatrix} \end{aligned} \quad (3.28)$$

A definição da matriz de correlação cruzada em (3.27) descreve o método de estimação da matriz de correlação real  $E\{\mathbf{y}_q^T \mathbf{y}_p\}$ , o qual é comumente conhecido como método da covariância [64]. Neste caso, o elemento de  $\mathbf{R}_{y_p y_q}(i)$  na  $u$ -ésima linha e  $v$ -ésima coluna ( $u, v \in \{0, \dots, D-1\}$ ) é dado por:

$$r_{y_p y_q}(i, u, v) = \frac{1}{N} \sum_{n=iL}^{iL+N-1} y_p(n-u) y_q^*(n-v) \quad (3.29)$$

onde o superscrito  $*$  é o operador conjugado. Isto leva em geral a uma matriz de correlação cruzada que não exhibe nenhuma estrutura especial, uma vez que cada elemento da matriz é computado avaliando-se os sinais em instantes de tempo diferentes. Portanto, os elementos da matriz dependem dos deslocamentos de tempo  $u$  e  $v$  dos respectivos canais. A combinação de todos os canais leva a

$$\begin{aligned} \mathbf{R}_{yy}(i) &= \frac{1}{N} \sum_{j=0}^{N-1} \mathbf{y}^T(iL+j) \mathbf{y}(iL+j) \\ &= \frac{1}{N} \mathbf{Y}^T(i) \mathbf{Y}(i), \end{aligned} \quad (3.30)$$

com

$$\mathbf{Y}(i) = [\mathbf{Y}_1(i), \dots, \mathbf{Y}_P(i)] \quad (3.31)$$

Este método de estimação da matriz de correlação é normalmente chamado de método da covariância pois possui as propriedades da matriz de covariâncias [76].

## Método da correlação

Quando a matriz de correlação é estimada usando-se o método da correlação, seus elementos não dependem mais dos deslocamentos de tempo  $u$  e  $v$  mas somente do atraso de tempo relativo  $\tilde{v} = v - u$  ( $\tilde{v} \in \{-D+1, \dots, D-1\}$ ) dos sinais

$y_p(n-u)$  e  $y_p(n-v)$ . Isto leva à seguinte definição

$$\tilde{r}_{y_p y_q}(i, \tilde{v}) = \begin{cases} \frac{1}{N} \sum_{n=iL}^{iL+N-\tilde{v}-1} y_p(n+\tilde{v}) y_q^*(n) & \text{para } \tilde{v} \geq 0 \\ \frac{1}{N} \sum_{n=iL-\tilde{v}}^{iL+N-1} y_p(n+\tilde{v}) y_q^*(n) & \text{para } \tilde{v} < 0 \end{cases} \quad (3.32)$$

onde pode ser visto que o número de elementos do somatório decresce de acordo com o aumento do atraso de tempo  $\tilde{v}$ . A adoção do método da correlação leva a uma matriz de correlação  $D \times D$  com estrutura Toeplitz. Para indicar a estimação através do método da correlação, o símbolo “til ( $\sim$ )” é usado e, portanto, a matriz de correlação com estrutura Toeplitz é definida como mostrado abaixo

$$\tilde{\mathbf{R}}_{y_p y_q}(i) = \begin{bmatrix} \tilde{r}_{y_p y_q}(i, 0) & \dots & \tilde{r}_{y_p y_q}(i, D-1) \\ \tilde{r}_{y_p y_q}(i, -1) & \ddots & \tilde{r}_{y_p y_q}(i, D-2) \\ \vdots & \ddots & \vdots \\ \tilde{r}_{y_p y_q}(i, -D+1) & \dots & \tilde{r}_{y_p y_q}(i, 0) \end{bmatrix} \quad (3.33)$$

Analogamente ao método da covariância, pode-se expressar a matriz de correlação cruzada (3.33) para o método da correlação também como o produto de matrizes.

$$\tilde{\mathbf{R}}_{y_p y_q}(i) = \frac{1}{N} \tilde{\mathbf{Y}}_p^T(i) \tilde{\mathbf{Y}}_q(i) \quad (3.34)$$

com a matriz  $N+D-1 \times D$

$$\tilde{\mathbf{Y}}_p(i) = \begin{bmatrix} y_p(iL) & 0 & \dots & 0 \\ y_p(iL+1) & y_p(iL) & \ddots & 0 \\ \vdots & y_p(iL+1) & \dots & \vdots \\ y_p(iL+N-1) & \vdots & \ddots & 0 \\ 0 & y_p(iL+N-1) & \dots & y_p(iL) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_p(iL+N-1) \end{bmatrix} \quad (3.35)$$

A combinação de todos os canais leva a

$$\tilde{\mathbf{R}}_{yy}(i) = \frac{1}{N} \tilde{\mathbf{Y}}^T(i) \tilde{\mathbf{Y}}(i) \quad (3.36)$$

com

$$\tilde{\mathbf{Y}}(i) = [\tilde{\mathbf{Y}}_1(i), \dots, \tilde{\mathbf{Y}}_p(i)] \quad (3.37)$$

Através da definição da matriz  $\tilde{\mathbf{Y}}_p(i)$  nota-se que o método da correlação requer somente  $N$  amostras do sinal para a estimação da matriz de correlação cruzada, enquanto que o método da covariância necessita de  $N+D-1$  amostras. Devido a estrutura Toeplitz das matrizes de correlação cruzada, o método da correlação leva a uma menor complexidade computacional quando implementado [67].

### 3.2.6 Normalização eficiente

Analisando-se a equação (3.26), percebe-se que ela apresenta uma normalização através das matrizes de auto-correlação  $\mathbf{R}_{y_p y_p}$  que é inerente à função de custo adotada. Esta normalização é conhecida da teoria de filtragem adaptativa supervisionada [73]. A normalização através das matrizes de auto-correlação garante

uma boa convergência dos filtros adaptativos mesmo para aqueles com um grande número de coeficientes e onde os sinais de entrada são correlacionados. Por outro lado, esta normalização traz como consequência uma grande complexidade computacional devido à inversão de  $P$  matrizes com dimensão  $L \times L$ . A complexidade é de  $O(L^3)$  quando se usa o método da covariância e para o método da correlação é de  $O(L^2)$  devido à estrutura Toeplitz das matrizes envolvidas no cálculo das auto-correlações. Para ambientes reais, o valor do parâmetro  $L$  pode estar entre 100 e 1000 coeficientes, fazendo com que o algoritmo se torne muito lento. Portanto, aproximações que reduzam a complexidade computacional do algoritmo com um mínimo de degradação do desempenho da separação são desejáveis.

Uma possível solução para a redução da complexidade computacional é a de se aproximar as matrizes de auto-correlação  $\mathbf{R}_{y_p y_p}(i)$  através de uma matriz diagonal, ou seja, através das potências dos sinais de saída  $y_p$ ,  $p = 1, \dots, P$ . conforme explicado a seguir.

$$\mathbf{R}_{y_q y_q}(i) = \frac{1}{N} \text{bdiag} \{ \mathbf{Y}_q^T(i) \mathbf{Y}_q(i) \} \quad (3.38)$$

Este caso corresponde a uma simplificação do método da covariância, onde os valores das diagonais não são idênticos. Para o método da correlação assume-se estacionariedade dentro de cada bloco  $i$  e portanto a matriz de correlação  $\tilde{\mathbf{R}}_{y_q y_q}$  pode ser simplificada como

$$\begin{aligned} \tilde{\mathbf{R}}_{y_p y_q}(i) &= \frac{1}{N} \sum_{n=iL}^{iL+N-1} |y_q(n)|^2 \mathbf{I} \\ &= \sigma_{y_q}^2(i) \mathbf{I} \end{aligned} \quad (3.39)$$

para  $q = 1, \dots, P$ . O operador  $\text{bdiag} \{ A \}$  faz com que todos os elementos da matriz não pertencentes à diagonal principal sejam feitos iguais à zero.

Para blocos contendo pausas na locução e baixo ruído de fundo, a

normalização através das matrizes de auto-correlação  $\mathbf{R}_{y_p, y_p}$  leva à inversão de matrizes mal condicionadas. No caso das aproximações (3.38) e (3.39), a inversão pode levar à divisão por valores de  $\sigma_{y_q}^2$  muito pequenos ou mesmo à divisão por zero, o que torna a estimação dos coeficientes dos filtros bastante sensível. Para uma adaptação robusta,  $\mathbf{R}_{y_p, y_p}$  é substituído por uma versão regularizada  $\mathbf{R}_{y_p, y_p} + \delta_{y_q} \mathbf{I}$ . A principal característica da regularização é o compromisso entre a fidelidade aos dados e a fidelidade à informação a priori sobre a solução [77]. Neste trabalho utiliza-se uma regularização dinâmica similar às utilizadas para filtragem adaptativa supervisionada [58][78]

$$\delta_{y_q} = \delta_{max} e^{-\sigma_{y_q}^2 / \sigma_0^2} \quad (3.40)$$

com dois parâmetros  $\delta_{max}$  e  $\sigma_0^2$ . Este método proporciona uma transição suave entre a regularização para potências de saída  $\sigma_{y_q}^2$  com valores pequenos e fidelidade aos dados sempre que a potência de saída for alta o suficiente.

## Capítulo 4

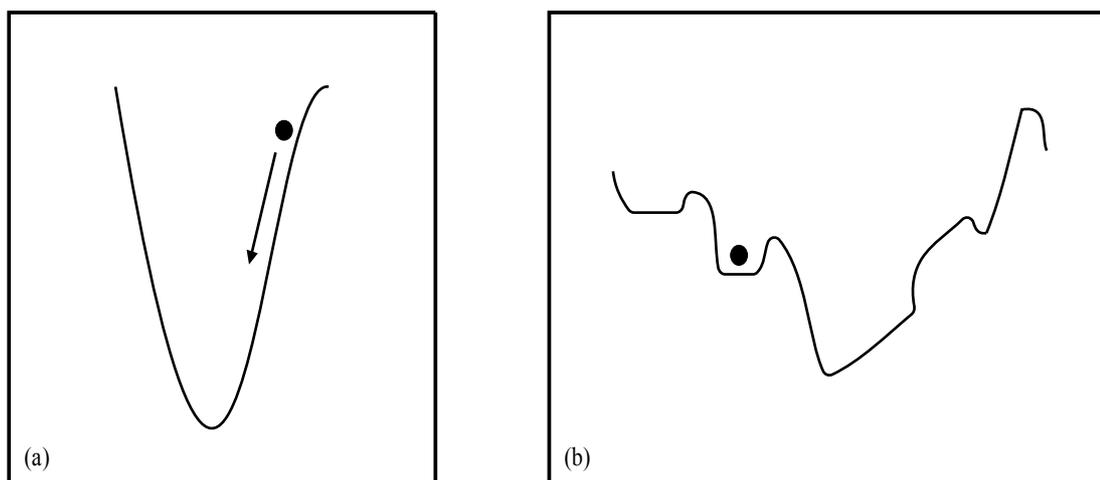
# Técnicas para adaptação do passo de aprendizagem

Este capítulo consiste na contribuição original deste trabalho. Ele apresenta técnicas trazidas da teoria de redes neurais que são incorporadas ao algoritmo para Separação Cega de Fontes (SCF) apresentado no Capítulo 3. A adição destas técnicas ao algoritmo tem como intuito melhorar sua velocidade de convergência e sua Relação Sinal-Interferência (SIR) [117].

### 4.1 Mínimos locais

O método do gradiente descendente é inicializado em algum ponto da superfície da função de custo  $J(m, W)$ , a qual é definida pelos coeficientes dos filtros de separação. Em seguida, o método tenta se mover em direção ao mínimo local mais próximo da função de custo. A Figura 4.1a apresenta a superfície de uma função de custo para a qual o processo de busca do mínimo global se mostra bastante simplificado. Entretanto, para problemas reais, a superfície se mostra bem mais complexa, muitas vezes se assemelhando à situação mostrada na Figura 4.1b. Como pode ser percebido, o algoritmo pode ficar aprisionado em um dos vários mínimos locais mostrados na figura. Mínimos locais são pontos na superfície formada pela função de custo que apresentam uma solução estável, mas não ótima. A solução ótima é somente atingida quando um mínimo global é encontrado.

Caso o algoritmo entre em uma região com um mínimo local, a minimização será direcionada para ele e irá parar quando encontrá-lo, enquanto que o desejado seria que a busca continuasse em direção ao mínimo global.



**Figura 4.1:** Diferentes tipos de superfícies formadas por funções de custo.

## 4.2 Passo de aprendizagem fixo

O algoritmo inicialmente apresentado em [41] adota um passo de aprendizagem fixo, o que por muitas vezes faz com que o algoritmo leve muitas épocas (ou iterações) para atingir uma separação razoável dos sinais originais. Em alguns casos, a superfície formada pela função de custo é bastante complexa, causando assim, lentidão na aprendizagem do algoritmo. Além da taxa de convergência lenta, o algoritmo apresenta uma alta dependência sobre o valor do passo de aprendizagem.

A escolha de um valor apropriado para o passo de aprendizagem  $\mu$  é essencial para a convergência do algoritmo: valores pequenos resultarão em convergência lenta e valores muito grandes causarão instabilidade, podendo fazer com que o algoritmo fique ziguezagueando em torno de um ponto de mínimo [56]. Por exemplo, se em algum momento o algoritmo estiver em uma ravina (tipo de vale estreito e íngreme), é desejável que se tenha um passo de aprendizagem pequeno, pois caso contrário, o algoritmo poderia oscilar entre seus dois lados, movendo-se

lentamente em direção ao gradiente negativo da função [76]. Se, pelo contrário, o algoritmo se encontrar no meio de um platô (área plana da superfície formada pela função de custo situada numa ligeira elevação) e o ponto de mínimo estiver distante, então um aumento do passo de aprendizagem possibilitará que o algoritmo se mova mais rápido. A seguir, apresenta-se algumas técnicas para se adaptar o passo de aprendizagem com o intuito de evitar mínimos locais e melhorar a convergência do algoritmo.

### 4.3 Termo momentum

Entre as técnicas da teoria de redes neurais utilizadas para acelerar o processo de treinamento e evitar mínimos locais, a do termo momentum é uma das mais frequentemente adotadas [40]. Ela foi primeiramente utilizada em redes neurais no trabalho de Rumelhart, Hinton e Williams [79], onde foi usada para melhorar o desempenho do procedimento de treinamento para redes neurais conhecido como Backpropagation [56]. A idéia do termo momentum consiste em incorporar à atualização atual dos coeficientes do sistema de separação informação de iterações anteriores. Sua grande utilização é influenciada pelo fato de ser uma técnica simples e efetiva. O termo momentum é representado pela equação (4.1).

$$\Psi(i) = \beta (W(i-1) - W(i-2)) \quad (4.1)$$

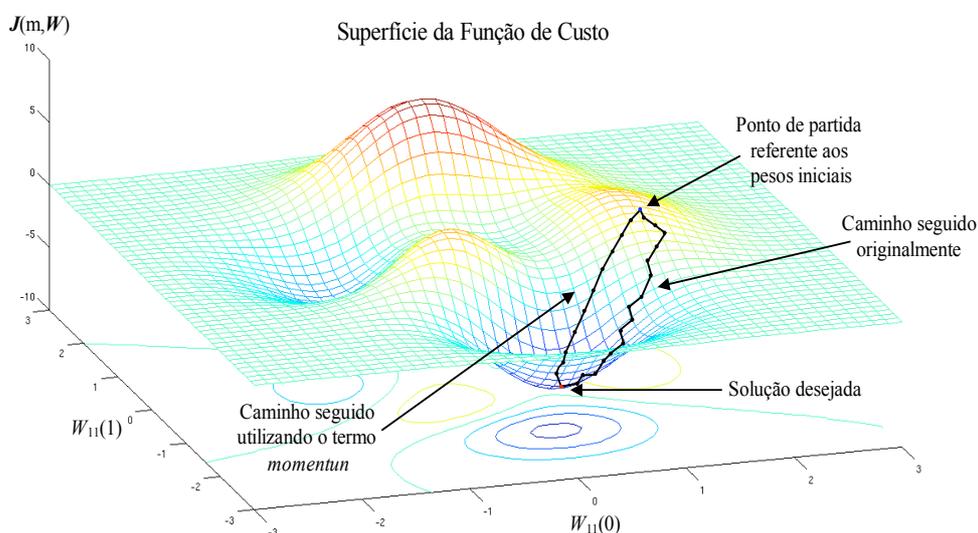
onde  $\beta$  é chamado de parâmetro momentum e determina o quanto uma iteração passada influenciará a atual. O parâmetro  $i$  corresponde à atual iteração do algoritmo. Deve-se ressaltar que para a técnica do termo momentum o parâmetro  $\mu$  não é variável no tempo.

A fórmula completa de ajuste dos coeficientes dos filtros que modelam o sistema de separação utilizando o termo momentum passa então a ser:

$$W(i) = W(i-1) - \mu \Delta W(i-1) + \Psi(i) \quad (4.2)$$

A adoção do termo momentum introduz um efeito de “amortecimento” no procedimento de procura do mínimo da função de custo. Isto evita oscilações dos valores dos coeficientes dos filtros de separação em áreas irregulares, tais como vales e ravinas, através do cálculo da média dos componentes do gradiente com sinais opostos [80]. Isto é, quando o gradiente permanece mudando de direção durante várias iterações, o termo momentum amortece (suaviza) estas variações. Em muitos pontos sobre a superfície formada pela função de custo o gradiente não aponta em direção ao mínimo e sucessivos passos do gradiente descendente podem oscilar de um lado para o outro, progredindo bem lentamente em direção ao ponto de mínimo. Além disto, a introdução do termo momentum acelera a convergência do algoritmo em longas áreas planas, e.g., platôs [79]. Isto é, quando o gradiente permanece apontando na mesma direção, o termo momentum aumenta o tamanho dos passos tomados em direção ao mínimo. Deve haver um compromisso entre  $\mu$  e  $\beta$  pois caso se utilize valores grandes para ambos parâmetros corre-se o risco de se fazer o algoritmo sair da região com um mínimo;

Em algumas situações, o termo momentum consegue evitar que o algoritmo fique parado em um mínimo local, ajudando-o a saltar por sobre estas regiões [80]. Ele pode ser considerado como uma aproximação de um método de segunda ordem, uma vez que utiliza informações oriundas de iterações anteriores [56]. Portanto, a inclusão do termo momentum na fórmula de ajustes dos coeficientes dos filtros de separação aumenta a velocidade de aprendizado (aceleração), reduzindo o perigo de instabilidade [39]. A Figura 4.2 mostra o efeito da utilização do termo momentum no caminho seguido pelo algoritmo durante o processo de separação.



**Figura 4.2:** Influência do termo momentum na procura pelo mínimo.

#### 4.4 *Bold driver*

A técnica heurística conhecida como *Bold Driver* foi introduzida por Battiti em 1989 [81]. Ela consiste num método para processamento em lotes (*batch* ou ciclo) que é utilizado para adaptar o passo de aprendizagem  $\mu$ . Sua operação é bastante simples: inicia-se o passo de aprendizagem  $\mu$  com um valor pequeno e a cada iteração (ciclo) do algoritmo compara-se os valores atual e passado da função de custo  $J(m, W)$  definida na equação (3.20). Caso o valor atual da função de custo seja menor do que o valor apresentado para a iteração anterior (indicando convergência), o passo de aprendizagem é normalmente incrementado por um fator de 1% a 10% de seu valor presente. No caso em que o valor atual da função de custo é maior ou igual ao valor anterior acrescido de um valor pré-determinado (indicando divergência), o passo de aprendizagem deve ser bruscamente decrementado, tipicamente por um fator de 50%. Isto indica que o valor do passo de aprendizagem adotado para a iteração atual foi demasiadamente grande, fazendo com que o algoritmo passasse através do ponto de mínimo e indo parar do outro lado de um

vale. O processo de redução do valor do passo de aprendizagem é repetido a cada iteração até que o valor atual da função de custo se torne menor do que o valor anterior; isto irá ocorrer inevitavelmente, uma vez que a direção de procura do algoritmo é aquela apontada pelo gradiente negativo. Caso o valor atual da função de custo seja menor do que o valor anterior acrescido de um valor pré-determinado, nada é feito, ou seja, o valor do passo de aprendizagem é mantido igual ao da iteração anterior.

A técnica do *bold driver* fará com que o passo de aprendizagem  $\mu$  cresça lentamente até que o algoritmo se encontre dando um passo que claramente o levará ao outro lado de um vale da superfície da função de custo. Uma vez que isto significa que o algoritmo alcançou uma área complicada (irregular) da superfície, faz sentido reduzir-se drasticamente o passo de aprendizagem nestes pontos. A técnica do *bold driver* pode ser resumida através da equação (4.3).

$$\mu(i+1) = \begin{cases} a \cdot \mu(i) & \text{se } J(i, \mathbf{W}) < J(i-1, \mathbf{W}), a > 1, \\ b \cdot \mu(i) & \text{se } J(i, \mathbf{W}) \geq c \cdot J(i-1, \mathbf{W}), b < 1, c > 1, \\ \mu(i) & \text{caso contrário} \end{cases} \quad (4.3)$$

onde  $a$  é o valor escolhido para se incrementar o passo de aprendizagem caso o valor atual da função de custo seja menor do que o da iteração anterior. Normalmente um valor entre 1,01 e 1,1 é adotado, conforme mencionado anteriormente. Os valores 0,5 e 1,3 são adotados para os parâmetros  $b$  e  $c$  respectivamente. O parâmetro  $i$  corresponde à iteração atual do algoritmo.

Na equação (4.3), a função de custo definida pela equação (3.20) deve ser calculada para todos os  $M$  segmentos em que os sinais captados pelos sensores foram separados. A equação (3.20) envolve o cálculo do determinante da matriz  $\mathbf{Y}^T \mathbf{Y}$  com dimensões  $PL \times PL$ . Com o intuito de se evitar a adição deste dispendioso cálculo e, conseqüentemente, não aumentar a carga computacional apresentada pelo algoritmo, utiliza-se as matrizes de correlação cruzada já obtidas anteriormente para o cálculo da

equação (3.26). O cálculo da norma L2, também conhecida como norma Euclidiana, das matrizes de correlação cruzada tem como resultado um valor escalar, o qual pode ser utilizado em substituição ao cálculo da equação (3.20). Analisando-se a equação (3.20) percebe-se que esta se torna igual a zero caso as matrizes de correlação cruzada sejam nulas para todos os  $M$  segmentos. Esta observação valida o uso da norma L2 como uma forma indireta de cálculo da equação (3.20). Testes realizados mostraram que a adoção da norma L2 diminui o tempo de simulação em aproximadamente 46% quando comparado com o algoritmo que calcula (3.20) a cada iteração. Os experimentos apresentados na seção 5.2 do Capítulo 5 utilizam a norma L2 como aproximação à equação (3.20).

## 4.5 Decaimento exponencial

Outra técnica heurística bastante efetiva que pode ser utilizada para acelerar o processo de estimação dos coeficientes dos filtros de separação é a técnica do decaimento exponencial, que é uma técnica não-adaptativa, ou seja, não se baseia nos valores de saída do algoritmo. Foi introduzida por Kohonen em 1982 [82] e é utilizada em Mapas Auto-Organizáveis (*Self Organized Maps*, SOM) durante a fase de organização ou ordenação. A seguinte equação, a qual foi empiricamente obtida, é adotada para atualizar o passo de aprendizagem:

$$\mu(i) = \mu_0 e^{-\frac{i^2}{\Lambda * \eta}} \quad (4.4)$$

onde  $\mu_0$  é o valor inicial do passo de aprendizagem,  $\eta$  é o de fator de decaimento da exponencial,  $\Lambda$  é o número de épocas adotado e  $i$  é o valor da iteração atual do algoritmo.

Como pode ser percebido, a equação (4.4) varia com as épocas, i.e., é variante no tempo e decresce gradualmente a cada nova iteração do algoritmo. A equação (4.4) faz com que a aprendizagem para as primeiras iterações do algoritmo seja rápida, uma vez que o valor de  $\mu$  para estas iterações é grande. Isto faz com que o

algoritmo encontre rapidamente o caminho para o mínimo da função de custo. Após várias iterações, o valor de  $\mu$  se torna muito pequeno, quase não afetando mais o processo de estimação dos coeficientes dos filtros de separação. Isto se torna mais claro se a seguinte analogia é feita, quando alguém inicia o aprendizado de uma nova língua, todas as informações novas são rapidamente assimiladas (alta taxa de aprendizagem), mas após um certo tempo o número de novas informações diminui e, conseqüentemente, a taxa de aprendizagem decresce [82][83].

A seguir serão apresentados os resultados da adição destas técnicas ao algoritmo apresentado no Capítulo 3 deste trabalho, além de implementações eficientes de tal algoritmo.

## Capítulo 5

### Resultados e experimentos

Nesta seção apresenta-se os resultados referentes aos experimentos realizados com o algoritmo definido no Capítulo 3.

#### 5.1 Condições gerais dos experimentos

Os experimentos apresentados neste capítulo foram realizados com sinais convoluídos com respostas ao impulso sintéticas (também chamadas de misturas convolutivas sintéticas) que simulam o comportamento acústico de uma sala real e também com sinais reais, ou seja, sinais captados por microfones em um ambiente acústico real (misturas convolutivas reais). O comprimento dos filtros de separação depende do ambiente acústico simulado ou real, e será citado para cada um dos experimentos apresentados a seguir. O fator de sobreposição, conforme apresentado no Capítulo 3, é o valor utilizado para determinar o quanto o segmento de voz atual sobrepõe o segmento anterior, é feito igual a 1.

Para todos os experimentos realizados neste trabalho, faz-se  $P = Q = 2$ , ou seja, o número de fontes é igual ao número de sensores (microfones). A qualidade da separação alcançada pelo algoritmo é avaliada através da medida da Relação Sinal-Interferência (SIR) apresentada pelos canais de saída  $y_p$ ,  $p = 1, 2$ . O cálculo da SIR foi apresentado no Capítulo 3. As curvas de SIR apresentadas nos experimentos a seguir são o resultado da média entre a SIR obtida para ambos os canais de saída do

sistema de separação.

A localização das fontes de sinal originais tanto para ambientes acústicos reais quanto para simulados foi escolhida de tal forma que cada uma das fontes se localiza em uma das metades do plano, conforme mostrado na Figura 3.4 do Capítulo 3. Desta forma, somente filtros FIR causais são necessários e portanto, pode-se iniciar os filtros de separação com impulsos unitários, conforme discutido na seção 3.2.2 do Capítulo 3.

Em todos os experimentos apresentados a seguir, exceto quando expressamente dito o contrário, utilizam-se a aproximação da restrição *Sylvester* da linha ( $SC_R$ ), o método da correlação para estimação das matrizes de correlação cruzada e a técnica da normalização eficiente da equação de atualização apresentada na seção 3.2.5 do Capítulo 3.

Para os experimentos apresentados na seqüência, cada um dos locutores profere uma frase distinta da do outro.

O algoritmo para Separação Cega de Fontes adotado neste trabalho foi totalmente desenvolvido e simulado com a ferramenta de simulações Matlab [43].

## **5.2 Avaliação das técnicas para adaptação do passo de aprendizagem**

Nesta seção compara-se a SIR apresentada pelo algoritmo quando se utiliza um fator de aprendizagem fixo com a SIR de cada uma das técnicas para aceleração da aprendizagem descritas no Capítulo 4. São apresentados resultados para misturas convolutivas sintéticas e reais.

### **5.2.1 Misturas convolutivas sintéticas**

Os sinais de áudio utilizados nesta seção foram gravados especificamente para este trabalho. Estes sinais possuem aproximadamente 6 segundos de duração e correspondem às vozes de um homem e de uma mulher. As gravações foram

realizadas em um ambiente acústico com baixo ruído de fundo, com frequência de amostragem igual a 11025 Hz e com 16 bits de resolução.

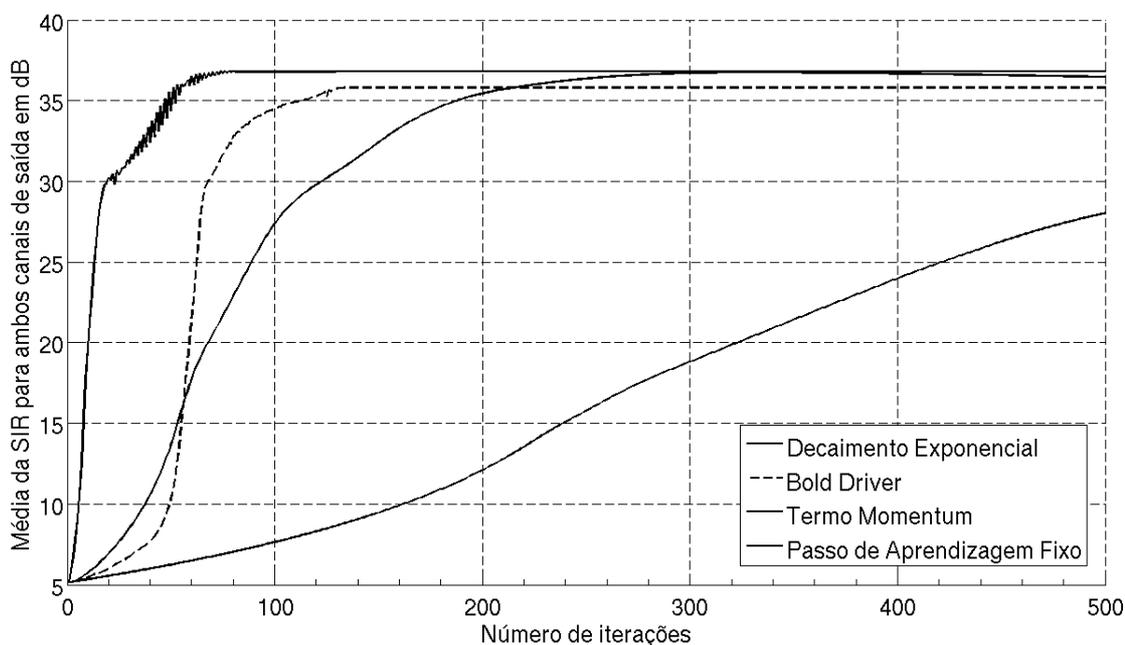
As características de reverberação e absorção de ondas sonoras em um ambiente acústico (sala) podem ser simuladas através da convolução da resposta ao impulso do ambiente e os sinais de voz originais. A resposta ao impulso de um ambiente acústico pode ser simulada precisamente e eficientemente pelo método da imagem [84][85]. Os experimentos apresentados nesta seção foram conduzidos utilizando-se sinais de voz convoluídos com respostas ao impulso sintéticas de salas obtidas através do método da imagem. O número  $L$  de coeficientes dos filtros de separação é feito igual a 447, que é o número de coeficientes dos filtros de mistura gerados pelo método da imagem.

O resultado da utilização das técnicas para adaptação do passo de aprendizagem, apresentadas no capítulo anterior é mostrado na Figura 5.1. A figura possibilita uma comparação da velocidade de aprendizagem e SIR entre as técnicas.

O valor do passo de aprendizagem é feito igual a 0,002 para todas as técnicas exceto para a técnica do decaimento exponencial, onde  $\mu_0$  é feito igual a 0,07. Como mencionado no Capítulo 4, o valor inicial para o passo de aprendizagem  $\mu_0$  deve ser alto para proporcionar uma rápida aprendizagem.

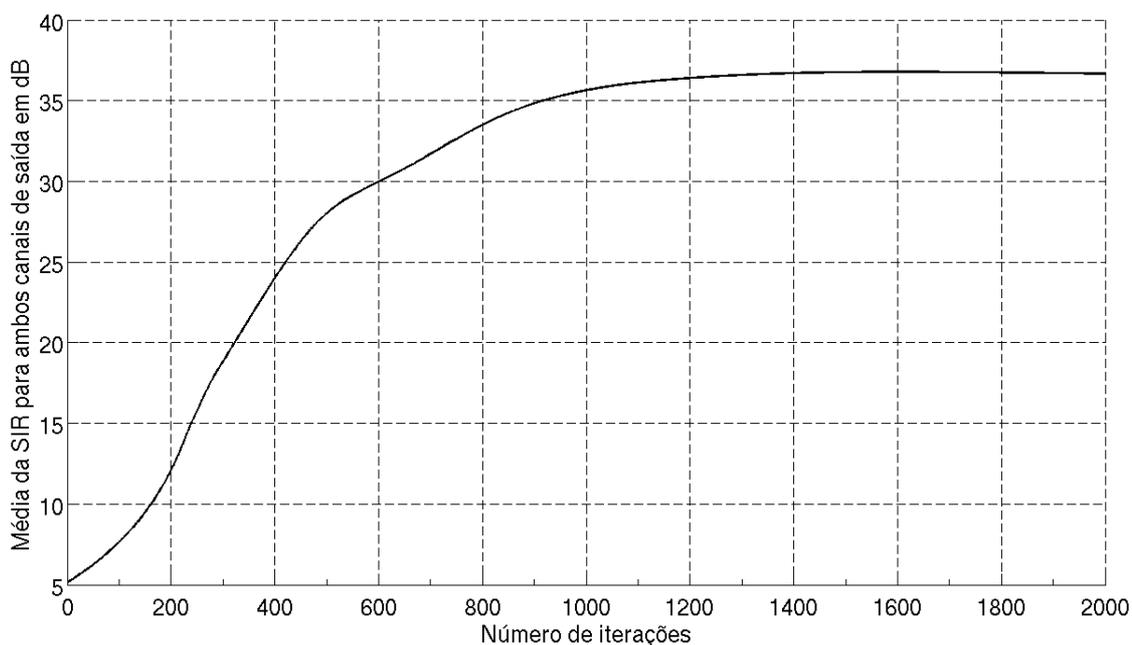
Alguns dos parâmetros das técnicas utilizadas neste trabalho devem ser encontrados empiricamente. Várias simulações foram executadas com o intuito de se encontrar os melhores valores para tais parâmetros. Após algumas simulações, o melhor valor para o parâmetro momentum  $\beta$  foi encontrado como sendo 0,8, o fator de decaimento  $\eta$  é feito igual a 10. Para a técnica do *bold driver*, os parâmetros  $a$ ,  $b$  e  $c$  são feitos iguais a 1,1, 0,5 e 1,3, respectivamente.

Os resultados apresentados nesta seção foram publicados no artigo [117].



**Figura 5.1:** Comparativo entre as técnicas para adaptação do passo de aprendizagem.

Como pode ser notado através da Figura 5.1, a curva para o algoritmo utilizando um passo de aprendizagem fixo não converge para um valor estável de SIR durante as 500 épocas adotadas para a comparação entre as técnicas de aprendizagem. Pode ser observado que as técnicas do decaimento exponencial e do termo momentum tendem para um mesmo valor fixo (constante) de SIR, entretanto, a técnica do *bold driver* atinge um valor menor de SIR. A conclusão para esta diferença entre os valores finais é que o algoritmo com a técnica do *bold driver* encontra um ponto de mínimo local pior (menos profundo) do que o encontrado pelo algoritmo quando utiliza as técnicas do decaimento exponencial ou do termo momentum. A existência de vários mínimos locais se deve à complexidade da superfície formada pela função de custo adotada neste trabalho. A Figura 5.2 apresenta a curva da SIR para a técnica do passo de aprendizagem fixo para uma simulação com 2000 épocas.



**Figura 5.2:** Curva da SIR para a técnica do passo de aprendizagem fixo com 2000 épocas.

Analisando-se as Figuras 5.1 e 5.2 e a Tabela 5.1, pode-se notar que os resultados para todas as técnicas apresentadas neste trabalho são melhores do que o apresentado pela técnica do passo de aprendizagem fixo. Nota-se também que a técnica do decaimento exponencial apresenta o melhor resultado: ela faz com que o algoritmo convirja mais rapidamente e apresente o maior valor de SIR entre todas as técnicas apresentadas neste trabalho.

**Tabela 5.1:** Comparação da taxa de convergência e SIR máxima para cada técnica.

<b>Técnica</b>	<b>Época da convergência</b>	<b>SIR máxima (dB)</b>
Decaimento exponencial	80	36,9158
<i>Bold driver</i>	129	35,8116
Termo momentum	270	36,7553
Passo de aprendizagem fixo	1596	36.7790

As simulações mostram que o tempo de convergência do algoritmo adotado neste trabalho pode ser reduzido por um fator de aproximadamente 20 vezes quando se utiliza a técnica do decaimento exponencial.

A Tabela 5.2 apresenta o tempo médio para cada ciclo (iteração) de processamento realizado pelo algoritmo. Estes valores de tempo foram obtidos em um computador com processador Intel Core 2 Duo e 2 Gbytes de memória RAM e sistema operacional Windows. Como pode ser percebido, nenhuma das técnicas faz com que o tempo médio para cada iteração aumente consideravelmente. A técnica do *bold driver* apresenta um tempo médio para cada iteração 5,78 % maior do que o tempo médio apresentado pela técnica do passo de aprendizagem fixo. Isto pode ser explicado pelo fato de que a cada iteração do algoritmo utilizando a técnica do *bold driver*, a norma L2 das matrizes de correlação cruzada precisa ser calculada, causando assim, este pequeno aumento no tempo médio de cada iteração apresentado por aquela técnica.

A melhoria da taxa de convergência e o pequeno aumento do tempo médio para cada iteração, justifica o uso das técnicas de adaptação dinâmica do passo de aprendizagem descritas no Capítulo 4.

**Tabela 5.2:** Tempo médio de processamento para cada iteração.

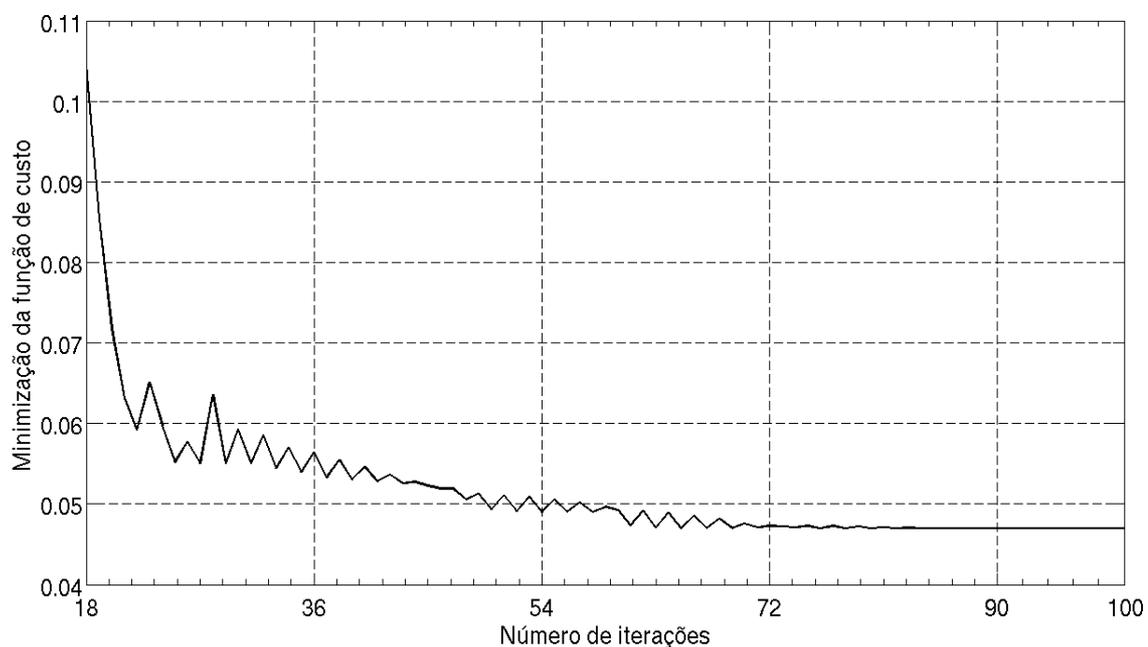
<b>Técnica</b>	<b>Tempo médio para cada iteração [s]</b>
Passo de aprendizagem fixo	14,71
Decaimento exponencial	15,08
Termo momentum	15,10
<i>Bold driver</i>	15.56

O comportamento oscilatório apresentado pela curva da SIR para a técnica do decaimento exponencial é devido à mudanças do sinal do gradiente, isto é, mudanças na direção do gradiente descendente a cada iteração do algoritmo. Como visto na Figura 5.1, a taxa de convergência para a técnica do decaimento exponencial é bastante elevada devido ao alto valor de  $\mu$  para as primeiras épocas do algoritmo.

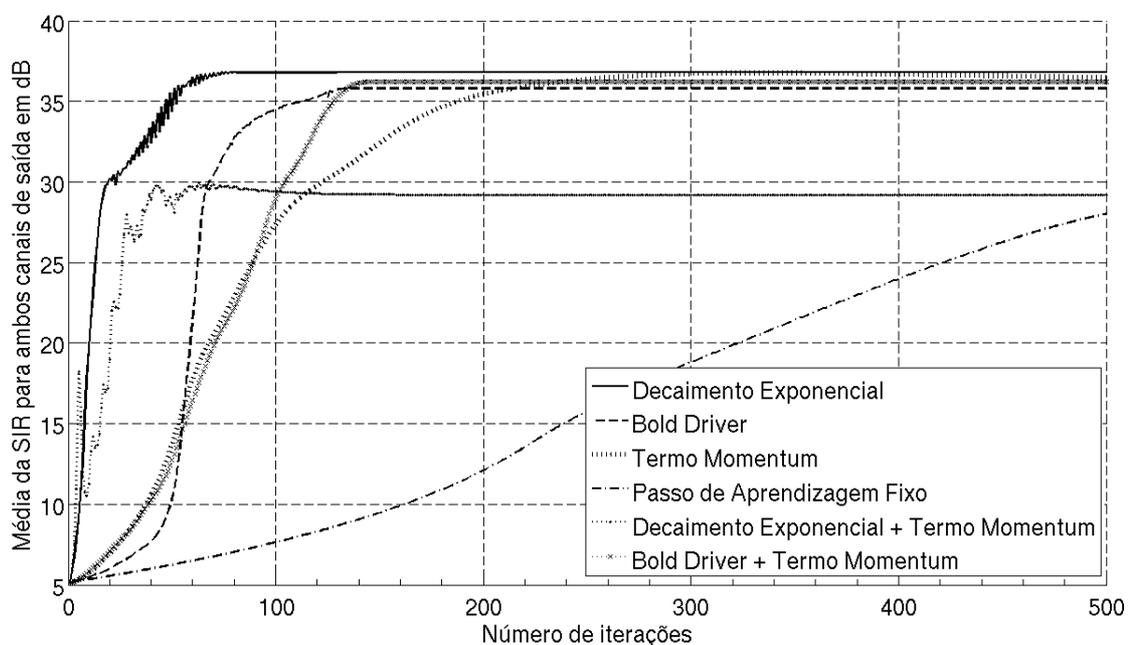
Isto faz com que o algoritmo encontre o ponto de mínimo rapidamente. Em alguns casos, o ponto de mínimo se encontra em um vale profundo e estreito da superfície formada pela função de custo. Este é o caso mostrado na Figura 5.1, pois quando o algoritmo utilizando a técnica do decaimento exponencial encontra tal vale, o valor do passo de aprendizagem  $\mu$  ainda é suficientemente grande a ponto de fazer com que o algoritmo oscile entre ambos os lados do vale. Conforme o valor de  $\mu$  é decrementado, o movimento oscilatório em direção à base do vale diminui, se estabilizando a medida que  $\mu$  tende a zero. Portanto, devido ao fato de não ter seus parâmetros ajustados dinamicamente, a técnica do decaimento exponencial pode fazer com que o algoritmo apresente um comportamento oscilatório como o mostrado na Figura 5.1. Por terem seus parâmetros ajustados dinamicamente, as técnicas do *bold driver* e do termo momentum podem diminuir ou até mesmo eliminar um possível comportamento oscilatório do algoritmo. A Figura 5.3 apresenta o gráfico da variação do valor da função de custo calculada através da norma L2. Conforme pode ser notado, o valor da função apresenta oscilação na mesma faixa de iterações do que a apresentada pela curva de SIR na Figura 5.1. A figura mostra que na média a função de custo está sendo minimizada, entretanto existem algumas oscilações devidas à variações na direção apontada pelo gradiente. Pode-se perceber que a partir da iteração de número 84 as oscilações cessam e o valor da função permanece estável.

Testes adicionais foram executados combinando-se a técnica do termo momentum com as técnicas do *bold driver* e do decaimento exponencial. Os testes foram executados com o mesmo conjunto de parâmetros definidos acima. A Figura 5.4 mostra o resultado de tais experimentos.

Observado-se a Figura 5.4, nota-se que a combinação da técnica do termo momentum com a do *bold driver* ou com a do decaimento exponencial não resulta em nenhum dos dois casos em uma nova técnica que apresente velocidade de convergência ou mesmo SIR maior do que as apresentadas pelas técnicas discutidas aqui quando utilizadas de forma separada.



**Figura 5.3:** Variação do valor da função de custo ao longo das iterações.



**Figura 5.4:** Testes combinando o termo momentum com outras técnicas.

A combinação da técnica do termo momentum com a do *bold driver* tem

como resultado uma curva de SIR que apresenta velocidade de convergência maior do que a apresentada pela técnica do termo momentum. Entretanto, a curva da SIR para a combinação de técnicas apresenta convergência mais lenta quando comparada com a curva para a técnica do *bold driver*. A técnica do *bold driver* faz com que o algoritmo convirja na época número 129, já a combinação de técnicas faz com que este convirja na época número 141, conforme mostrado na Tabela 5.3.

Comparando-se os valores de SIR máxima fornecidos pelas Tabelas 5.1 e 5.3, percebe-se que a combinação das técnicas do termo momentum e do *bold driver* apresenta um valor de SIR máxima maior do que a apresentada pela técnica do *bold driver* mas menor do que a SIR máxima para a técnica do termo momentum.

A segunda combinação de técnicas para acelerar a convergência do algoritmo adotado neste trabalho une as técnicas do decaimento exponencial e do termo momentum. Como mencionado no Capítulo 4, a técnica do decaimento exponencial faz com que o passo de aprendizagem  $\mu$  diminua a cada iteração do algoritmo, tendendo a zero após algumas iterações. Isto faz com que o termo de atualização dos coeficientes dos filtros de separação  $\nabla W(m)$  seja zerado. O experimento que combina as técnicas do decaimento exponencial e do termo momentum encontra um ponto de mínimo totalmente diferente do encontrado pelas outras técnicas. Isto explica o pior resultado apresentado por tal combinação.

O termo *momentum* é considerado como uma aproximação de um método de segunda ordem, uma vez que ele utiliza informações de iterações anteriores. Tais métodos utilizam a informação contida na derivada de segunda ordem da função de custo. A adoção do termo momentum tende a suavizar a taxa de convergência do algoritmo, evitando-se assim comportamentos oscilatórios indesejados [40]. Desta forma, quando o valor do passo de aprendizagem  $\mu$  tende a zero, devido à técnica do decaimento exponencial, a curva da SIR para esta combinação de técnicas se torna plana. Isto faz com que valores consecutivos dos filtros de separação  $W(m)$  se tornem idênticos, fazendo com que o termo momentum se torne igual a zero.

**Tabela 5.3:** Comparação da taxa de convergência e SIR máxima para as combinações de técnicas.

<b>Técnica</b>	<b>Época da convergência</b>	<b>SIR máxima (dB)</b>
Termo Momentum + <i>Bold driver</i>	141	36,2502
Termo Momentum + Decaimento exponencial	120	29.9176

As oscilações mostradas na Figura 5.4 durante as épocas iniciais da combinação entre as técnicas do decaimento exponencial e do termo momentum são devidas a combinações entre os valores de  $\mu$  e do termo momentum que fazem com que o algoritmo fique oscilando entre os lados de um vale. Em alguns casos, o algoritmo se distancia bastante do vale, tendo que subir uma colina (indicado pelo decréscimo do valor da SIR) até que encontre novamente o caminho em direção ao mínimo. Estas oscilações podem ser diminuídas encontrando-se empiricamente valores para  $\mu_0$ ,  $\eta$  e  $\beta$ .

Nesta seção mostrou-se a utilização de algumas técnicas vindas da teoria de redes neurais com o intuito de melhorar a taxa de aprendizagem e qualidade da separação do algoritmo proposto por H. Buchner e colegas [54] para a separação de misturas convolutivas sintéticas. Os resultados apresentados aqui provam que todas as três técnicas utilizadas neste trabalho, decaimento exponencial, termo momentum e *bold driver*, reduzem o tempo de convergência do algoritmo além de aumentarem a qualidade da separação. Dentre as três técnicas adotadas neste trabalho, a do decaimento exponencial é a que apresenta melhor resultado, ela faz com que o algoritmo convirja mais rapidamente e alcance uma SIR maior.

## 5.2.2 Misturas convolutivas reais

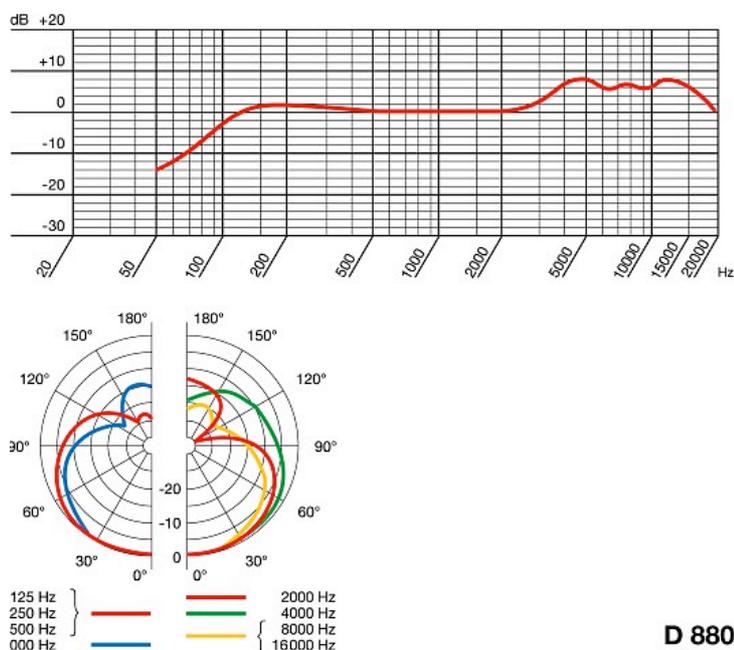
Nesta seção são apresentados os resultados obtidos com a separação de sinais gravados em um ambiente acústico real. Os sinais de áudio usados para os experimentos desta seção foram gravados em um estúdio de uma rádio com 2,49 m de altura, 4,30 m de comprimento e 3,44 m de largura. Estes sinais possuem aproximadamente 4 segundos de duração e correspondem às vozes de um homem e de uma mulher. As gravações foram realizadas em um ambiente acústico com baixo ruído de fundo. Os sinais foram amostrados a uma frequência de 8000 Hz e possuem resolução de 16 bits.

Quando gravações reais são utilizadas em experimentos para separação cega de fontes a única informação disponível são os sinais captados pelos microfones (sensores). A qualidade da separação obtida pelo sistema de separação não pode ser determinada diretamente através dos sinais captados pelos microfones.

A abordagem adotada para avaliar a qualidade da separação obtida pelo algoritmo proposto neste trabalho foi a de realizar gravações em uma sala quando somente um dos locutores estava ativo (falando). Para estas gravações ambos os microfones se encontravam em aberto, ou seja, os dois microfones captaram versões das locuções. Estas gravações são denotadas por  $x_{p,s_q}$ , isto é, são as contribuições da  $q$ -ésima fonte de sinal original para o  $p$ -ésimo microfone. Os sinais misturados são obtidos somando-se estas contribuições independentes, ou seja,  $x_p = \sum_q x_{p,s_q}$ . Portanto  $x_p$  representa os sinais que seriam captados pelos microfones caso todas as fontes (locutores) estivessem ativos (falando) simultaneamente. Esta soma aritmética dos sinais independentes é justificada para aplicações acústicas dado que ondas sonoras são aditivas também [86]. Deve-se salientar que todos os locutores estavam presentes durante as gravações individuais uma vez que a acústica do ambiente é influenciada por suas presenças.

Foram utilizados 2 microfones da marca AKG, modelo D880, espaçados entre si de 20 cm. Estes microfones são do tipo cardióide, também chamados de unidirecionais. A curva de resposta em frequência e o diagrama de recepção destes

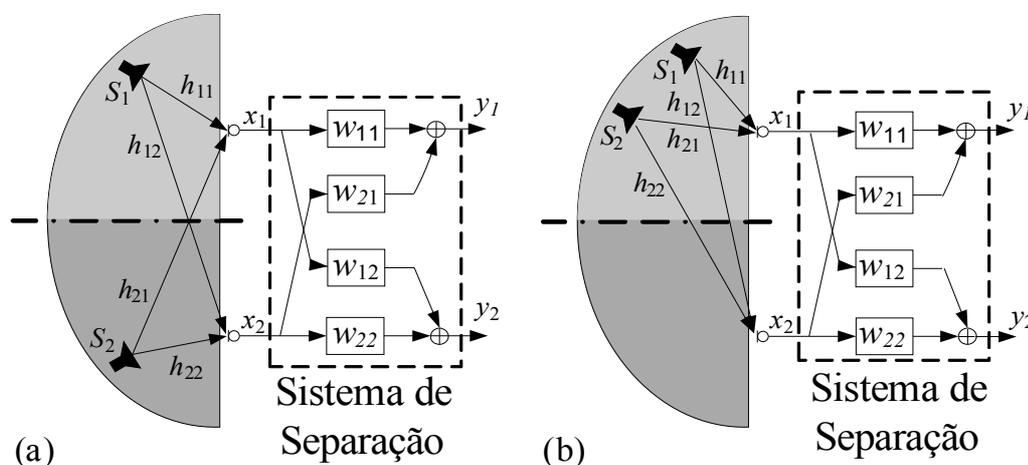
microfones são mostrados na Figura 5.5.



**Figura 5.5:** Resposta em frequência e diagrama de recepção dos microfones.

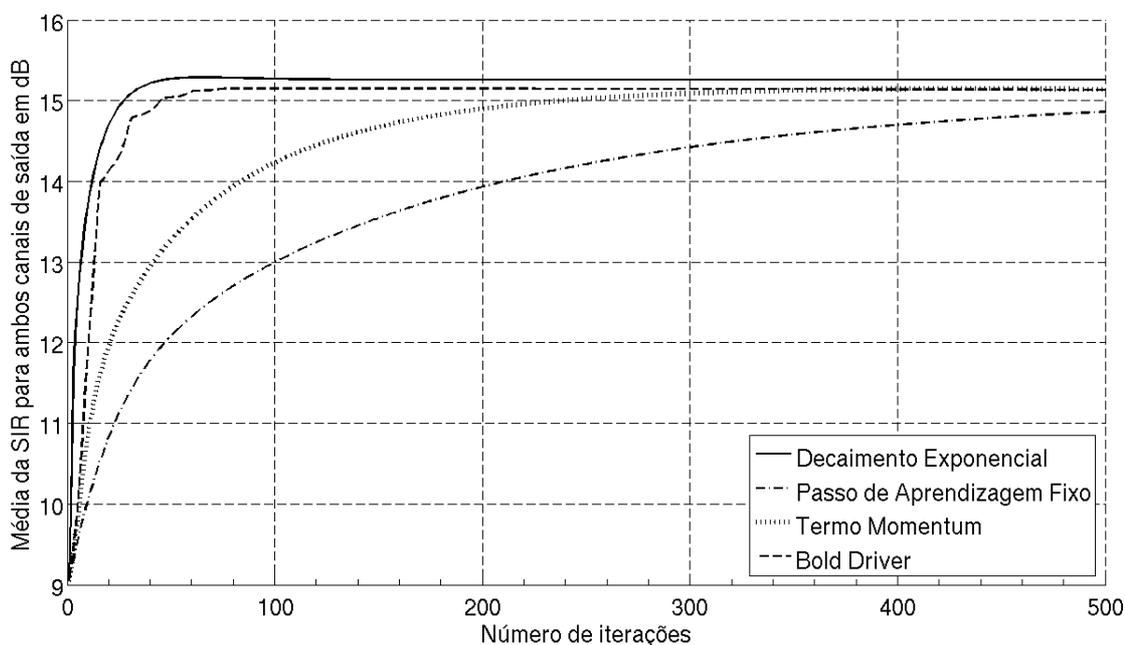
A distância entre os locutores e os microfones foi escolhida como 75 cm. Os locutores foram dispostos de tal forma que ficassem posicionados respectivamente a  $45^\circ$  e  $-45^\circ$  em relação ao eixo formado pelo pelo centro dos microfones, conforme pode ser visto na Figura 3.4 do Capítulo 3, a qual é reproduzida abaixo na Figura 5.6. Isto faz com que apenas filtros FIR causais sejam necessários para a separação dos sinais.

Foram necessários uma série de experimentos com o intuito de se encontrar o melhor valor para o número de coeficientes  $L$  dos filtros de separação, uma vez que não se possui informações sobre as características acústicas da sala onde os sinais originais foram gravados. Além dos experimentos citados acima, outros experimentos também foram executados a fim de se encontrar os melhores valores para os parâmetros necessários às técnicas de adaptação do passo de aprendizagem avaliadas aqui.



**Figura 5.6:** Arranjos para SCF que requerem (a) somente filtros FIR causais e (b) filtros FIR causais e não-causais.

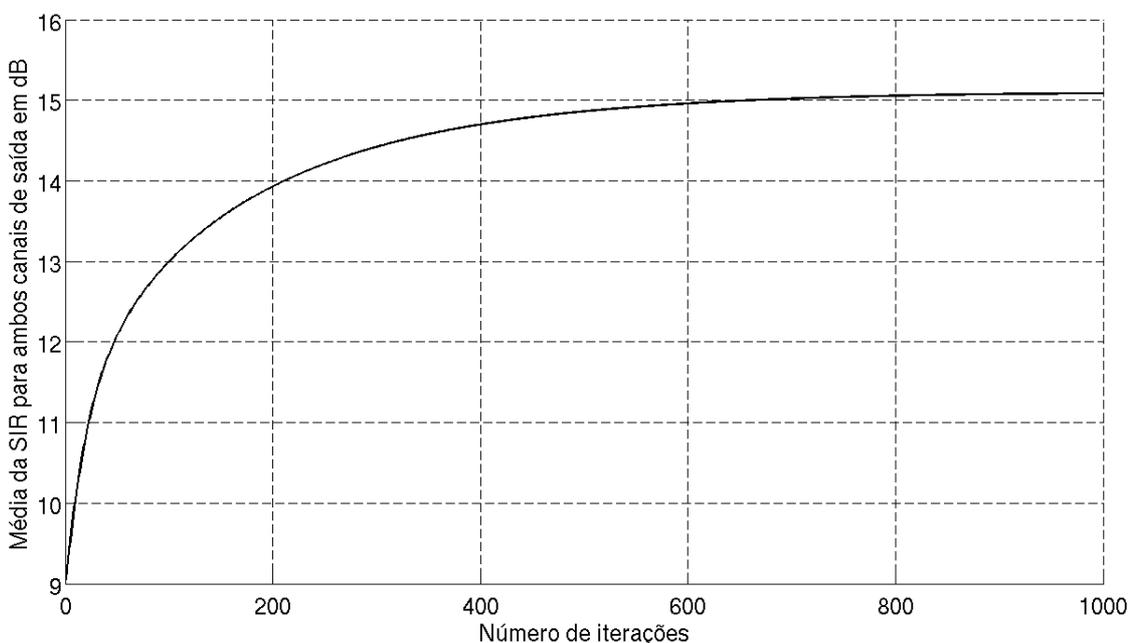
Durante os experimentos realizados para esta seção do trabalho, o número  $L$  de coeficientes dos filtros de separação foi variado de 64 à 1512 em passos de 64 coeficientes, sendo que o melhor valor (mais alto) de SIR foi encontrado quando  $L$  é feito igual a 1024. O resultado da utilização das técnicas para adaptação do passo de aprendizagem é mostrado na Figura 5.7. A figura possibilita a comparação da velocidade de aprendizagem e SIR entre as técnicas. O melhor valor para o passo de aprendizagem  $\mu$  foi encontrado empiricamente como sendo igual a 0,006. Este valor foi utilizado em todas as técnicas exceto para a técnica do decaimento exponencial, onde  $\mu_0$  foi encontrado como sendo igual a 0,09. Após algumas simulações, o melhor valor para o parâmetro momentum,  $\beta$ , foi encontrado como sendo 0,6 e o fator de decaimento  $\eta$  é feito igual a 10. Para a técnica do *bold driver*, os parâmetros  $a$ ,  $b$  e  $c$  são feitos iguais a 1,1, 0,5 e 1,3, respectivamente.



**Figura 5.7:** Comparativo entre as técnicas para adaptação do passo de aprendizagem.

A curva da Relação Sinal-Interferência para o algoritmo utilizando um passo de aprendizagem fixo não converge para um valor fixo (constante) de SIR durante as 500 épocas adotadas para a comparação entre as técnicas de aprendizagem, conforme pode ser visto na Figura 5.7. Pode ser observado que as técnicas do *bold driver* e do termo momentum tendem para um mesmo valor constante de SIR, entretanto, a técnica do decaimento exponencial atinge um valor maior de SIR. A conclusão para esta diferença entre os valores finais alcançados é que o algoritmo que adota a técnica do decaimento exponencial encontra um ponto de mínimo local melhor (mais profundo) do que o encontrado pelo algoritmo quando adota as técnicas do *bold driver* ou do termo momentum. A existência de vários mínimos locais se deve como dito anteriormente, à complexidade da superfície formada pela função de custo adotada neste trabalho. A Figura 5.8 apresenta a curva da SIR para a técnica do passo de aprendizagem fixo para uma simulação com 1000 épocas. Conforme mostrado na figura, o algoritmo utilizando tal técnica de adaptação converge somente após um grande número de iterações (ciclos), justificando o estudo de técnicas que acelerem o

processo de separação realizado pelo algoritmo.



**Figura 5.8:** Curva da SIR para a técnica do passo de aprendizagem fixo com 1000 épocas.

Analisando-se as Figuras 5.7 e 5.8 e a Tabela 5.4 conclui-se que os resultados apresentados pelas técnicas introduzidas no Capítulo 4 são melhores do que o resultado apresentado pela técnica do passo de aprendizagem fixo. Além de acelerar a convergência e atingir valores razoáveis de SIR, nenhuma das técnicas apresentadas no Capítulo 4 aumenta consideravelmente o tempo de simulação apresentado pelo algoritmo, ver Tabela 5.4. Nota-se que a técnica do decaimento exponencial apresenta novamente o melhor resultado, ela faz com que o algoritmo convirja mais rapidamente, atinja o maior valor de SIR entre todas as técnicas de adaptação do passo de aprendizagem e introduz o menor aumento do tempo médio de iteração do algoritmo.

A Tabela 5.4 apresenta o tempo médio para cada ciclo (iteração) de processamento realizado pelo algoritmo. Estes valores de tempo foram obtidos em um computador com processador Intel Core 2 Duo e 2 Gbytes de memória RAM e

sistema operacional Windows.

**Tabela 5.4:** Comparação entre as técnicas de adaptação do passo de aprendizagem.

Técnica	SIR Inicial (dB)	Época da convergência	SIR máxima (dB)	Tempo médio para cada iteração [s]
Decaimento exponencial	9.2433	66	15,2876	33,22
<i>Bold driver</i>		121	15,1478	34,88
Termo momentum		382	15,1490	33,24
Passo de aprendizagem fixo		882	15,0858	33,13

As simulações (ver Tabela 5.4) mostram que o tempo de convergência do algoritmo adotado neste trabalho pode ser reduzido por um fator de aproximadamente 13,36 vezes em relação ao tempo de convergência da técnica do passo de aprendizagem fixo quando se utiliza a técnica do decaimento exponencial. A técnica do *bold driver* apresenta um tempo médio para cada iteração 5,28 % maior do que o tempo médio apresentado pela técnica do passo de aprendizagem fixo. Conforme dito na seção 5.2.1, este comportamento é devido ao fato de que a cada iteração do algoritmo utilizando a técnica do *bold driver*, a norma L2 das matrizes de correlação cruzada precisa ser calculada.

Conforme pode ser visto através dos valores de SIR máxima mostrados na Tabela 5.4, a separação das misturas convolutivas reais é mais difícil do que a separação das misturas convolutivas sintéticas apresentada na seção 5.2.1. Isto é explicado pelo fato de que quanto maior o número de coeficientes  $L$ , mais complexa é a superfície  $L$ -dimensional formada pela função de custo, fazendo com que a busca pelo mínimo local se torne mais complicada.

A aceleração da taxa de convergência, os bons valores de SIR e o pequeno impacto no tempo médio de cada iteração do algoritmo justificam mais uma vez a utilização das técnicas de adaptação do passo de aprendizagem discutidas no Capítulo 4.

### 5.3 Normalização eficiente

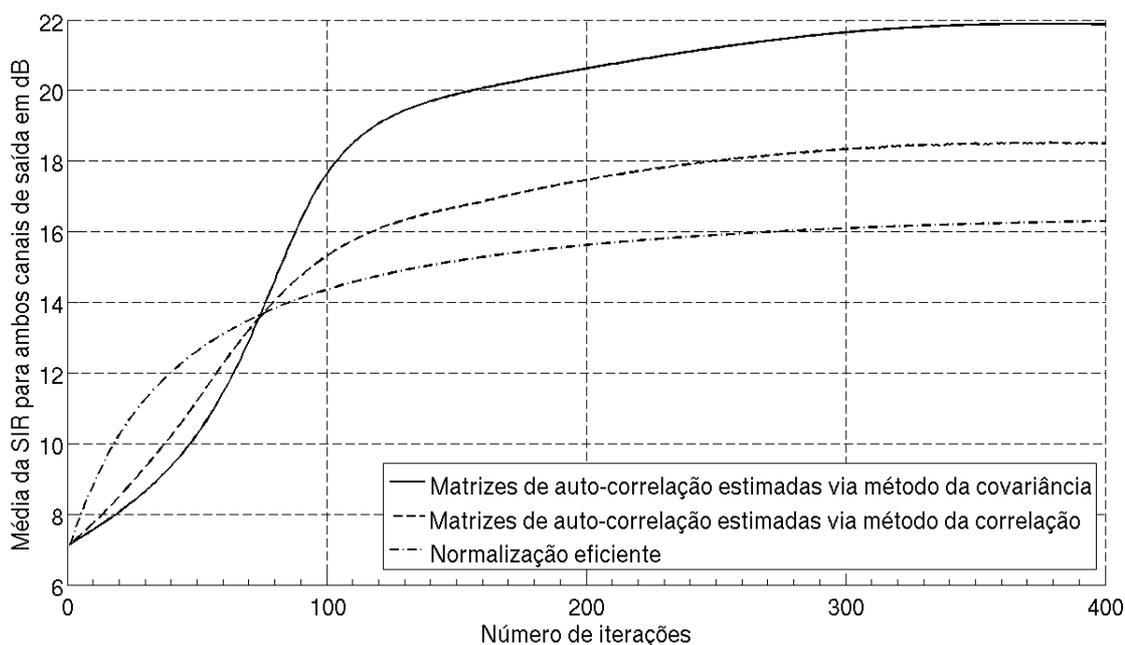
A equação de atualização dos coeficientes dos filtros de separação (3.26) definida no Capítulo 3, apresenta uma normalização das matrizes de correlação cruzada por meio de matrizes de auto-correlação  $\mathbf{R}_{y_p y_p}$ . Esta normalização introduz uma grande complexidade computacional ao algoritmo devido à inversão de  $P$  matrizes com dimensão  $L \times L$ . Na seção 3.2.5 daquele capítulo apresenta-se uma técnica de normalização da equação de atualização que reduz a complexidade do algoritmo. Esta técnica utiliza matrizes diagonais formadas pelas potências dos sinais de saída ao invés de matrizes de auto-correlação, conforme descrito na seção 3.2.5.

Esta seção apresenta a comparação entre o método da normalização eficiente e a normalização através das matrizes de auto-correlação. As matrizes de auto-correlação serão estimadas através dos métodos da correlação e da covariância.

Os sinais de áudio utilizados nestes experimentos possuem aproximadamente 5 segundos de duração e correspondem às vozes de um homem e de uma mulher. As gravações foram realizadas em um ambiente acústico com baixo ruído de fundo, com frequência de amostragem igual a 11025 Hz e com 16 bits de resolução. Os sinais de voz foram convoluídos com filtros obtidos através do método da imagem. O número  $L$  de coeficientes dos filtros de separação é feito igual a 1012, que é o número de coeficientes dos filtros de mistura gerados pelo método da imagem.

A técnica do passo de aprendizagem fixo foi utilizada nos experimentos desta seção e o valor do passo de aprendizagem  $\mu$  foi fixado em 0,002. A Figura 5.9 compara a normalização eficiente com a normalização realizada com matrizes de auto-correlação  $\mathbf{R}_{y_p y_p}$  estimadas através do método da covariância e da correlação.

A Tabela 5.5 apresenta a comparação entre os valores de SIR e tempo de simulação entre os métodos de normalização avaliados nesta seção. Os valores mostrados na tabela foram obtidos em um computador com processador Intel Core 2 Duo e 2 Gbytes de memória RAM e sistema operacional Windows.



**Figura 5.9:** Comparação entre as formas de normalização da equação de atualização.

Como pode ser percebido analisando-se os valores apresentados na tabela, a normalização eficiente da equação de atualização (3.26) reduz o tempo médio de cada iteração em 56,09% quando comparada com a normalização através de matrizes de auto-correlação estimadas via método da covariância e em 46,96% quando comparada com a normalização através de matrizes de auto-correlação estimadas via método da correlação.

Com o passo de aprendizagem feito igual a 0,002 o máximo valor para a Relação Sinal-Interferência obtido com a normalização eficiente é 5,5485 dB menor do que a máxima SIR obtida com a normalização através das matrizes de auto-correlação estimadas via método da covariância. Experimentos adicionais com a técnica da normalização eficiente mostraram que se o valor do passo de aprendizagem for aumentado, obtém-se um valor de SIR bem próximo ao obtido pela normalização através de matrizes de correlação estimadas via método da covariância.

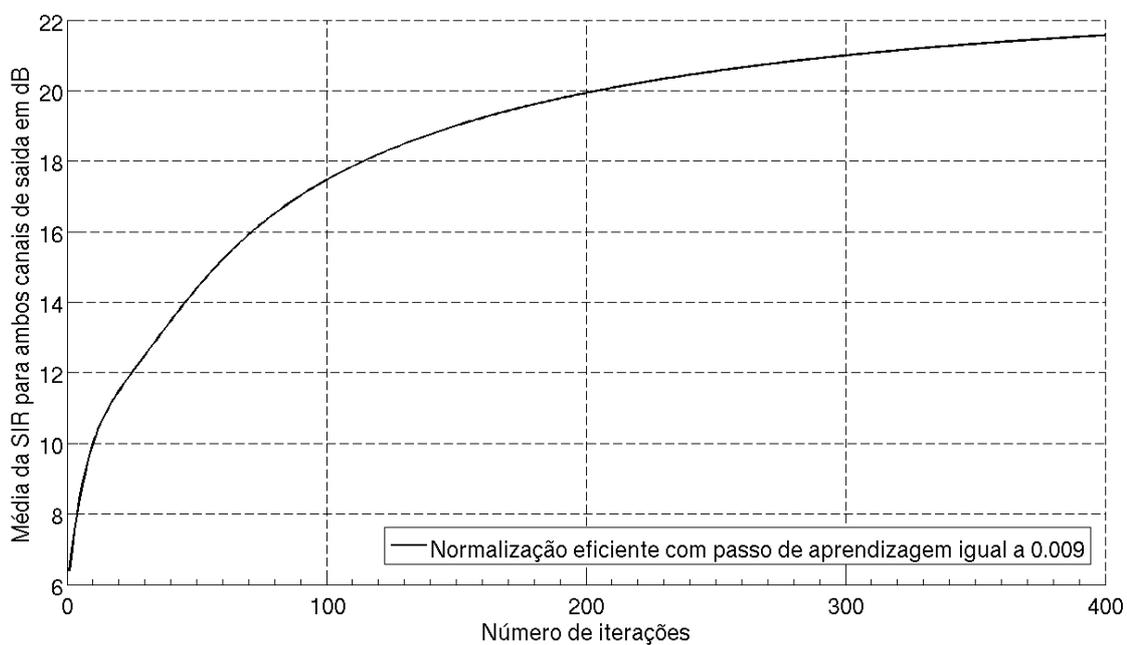
**Tabela 5.5:** Comparação da SIR e tempo de simulação entre os métodos de normalização.

Normalização	SIR Inicial (dB)	SIR Final (dB)	Tempo médio para cada iteração [s]	Tempo total [h]
$R_{y_p, y_p}$ (Covariância)	7,1430	21,8734	67,44	7,49
$R_{y_p, y_p}$ (Correlação)		18.5001	55.83	6,20
Eficiente		16.3249	29.61	3.29

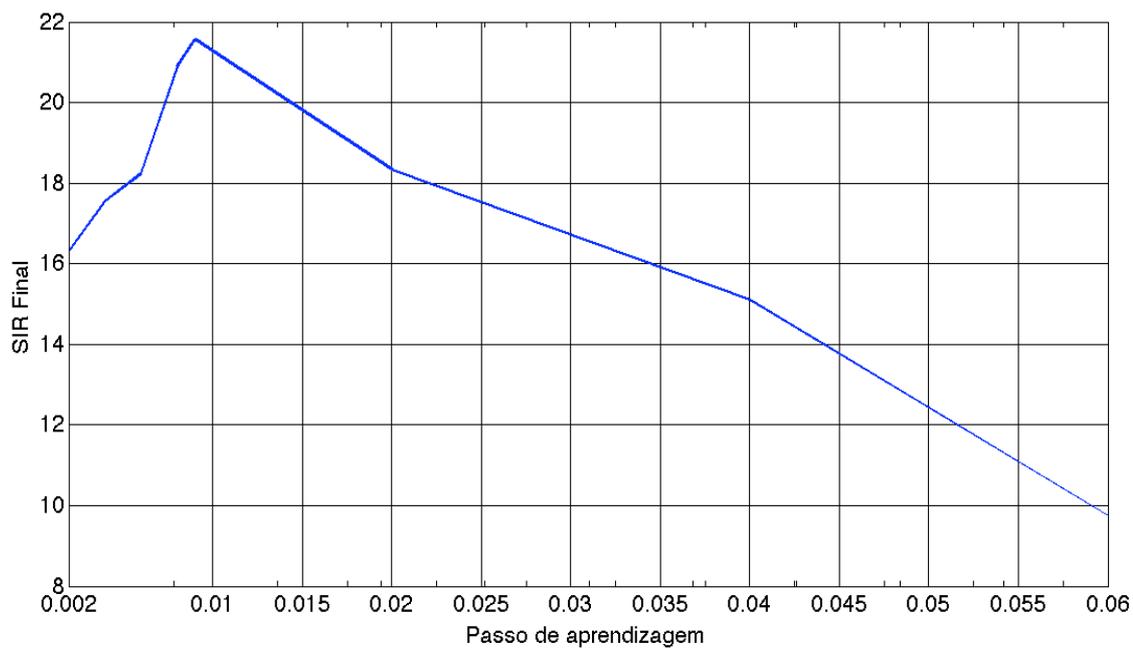
A Figura 5.10 mostra o resultado de um experimento onde utilizou-se os mesmos filtros e os mesmos sinais de áudio dos experimentos mostrados na Figura 5.9, mas com o valor do passo de aprendizagem igual a 0,009. O valor de SIR final para este experimento é de 21,5680 dB, provando o que foi dito no parágrafo anterior.

A Figura 5.11 mostra o gráfico de SIR para a variação do passo de aprendizagem. Percebe-se que o valor de SIR aumenta até que alcance o valor máximo de 21,57 dB e depois começa a decair devido a problemas de instabilidade provocados pelo alto valor do passo de aprendizagem  $\mu$ . Após análise da figura, fica evidente a dependência do valor de SIR alcançado pelo algoritmo e o valor do passo de aprendizagem  $\mu$ .

Os resultados apresentados nesta seção mostram que a técnica da normalização eficiente converge para valores de SIR muito próximos aos obtidos com a normalização através das matrizes de auto-correlação estimadas via método da covariância quando o valor do passo de aprendizagem é corretamente escolhido. Mostrou-se também que a adoção da técnica da normalização eficiente reduz a duração das simulações, justificando portanto seu uso.



**Figura 5.10:** Normalização eficiente com passo de aprendizagem igual a 0,009.

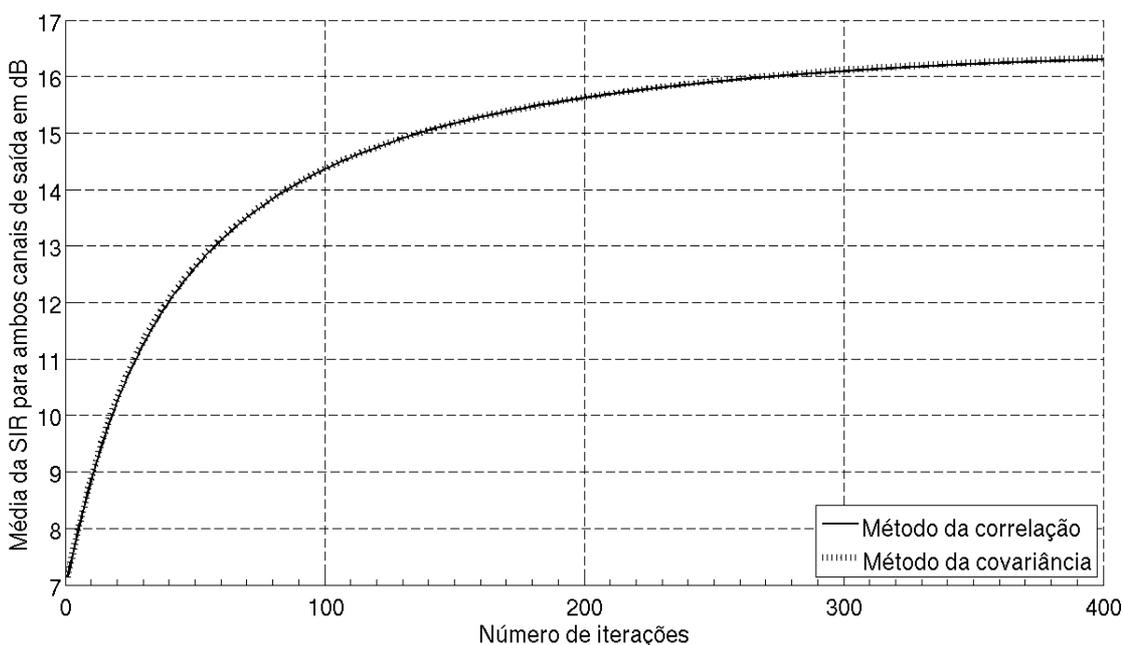


**Figura 5.11:** Variação do passo de aprendizagem.

## 5.4 Estimação das matrizes de correlação através dos métodos da correlação e da covariância

As matrizes de correlação que aparecem na equação de atualização dos coeficientes dos filtros de separação podem ser estimadas através de dois métodos, o da correlação ou o da covariância, conforme discutido na seção 3.2.4 do Capítulo 3. A diferença entre a adoção de ambos os métodos de estimação será avaliada utilizando-se os filtros e sinais de áudio utilizados na seção 5.3. A técnica do passo de aprendizagem fixo foi adotada e o passo de aprendizagem feito igual a 0,002.

O resultado apresentado na Figura 5.12 mostra que ambos os métodos de estimação possuem uma curva de SIR muito semelhante e que convergem para valores muito próximos, conforme mostrado na Tabela 5.6.



**Figura 5.12:** Comparação entre os métodos de estimação das matrizes de correlação.

Em geral, o método da covariância exibe desempenho um pouco melhor do que o apresentado pelo método da correlação. Isto se deve ao fato de que o método da covariância não assume estacionariedade dentro dos segmentos em que os sinais

captados pelos microfones são divididos. Entretanto, a estimação das matrizes de correlação através do método da covariância é mais dispendiosa do que a estimação através do método da correlação, conforme pode ser notado através dos valores de tempo apresentados na Tabela 5.6.

**Tabela 5.6:** Comparação da SIR e tempo de simulação entre os métodos de estimação.

<b>Método</b>	<b>SIR Inicial (dB)</b>	<b>SIR Final (dB)</b>	<b>Tempo médio para cada iteração [s]</b>	<b>Tempo total [h]</b>
Covariância	7.1430	16,3839	40,43	4,49
Correlação		16,3249	29,61	3.29

A estimação das matrizes de correlação através do método da correlação faz com que haja um decréscimo de 26,76% no tempo de simulação quando comparado à duração da simulação que utiliza o método da covariância. Além disto, a utilização do método da correlação não causa grandes impactos na separação alcançada, ou seja, no valor da Relação Sinal-Interferência final.

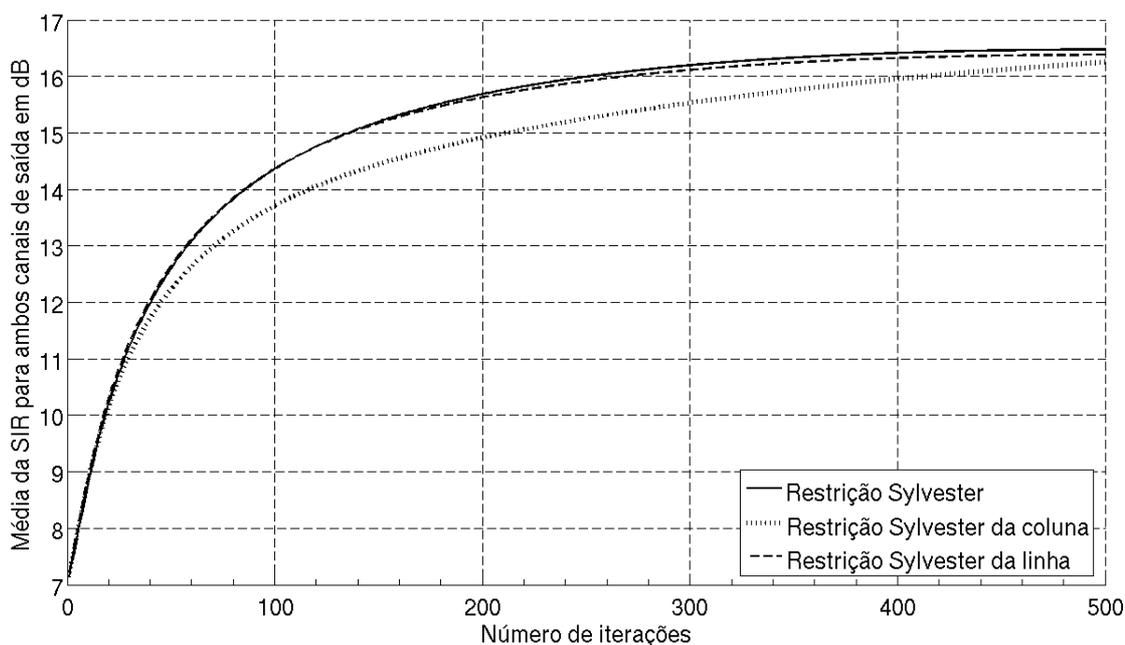
Através dos resultados apresentados na Figura 5.12 e Tabela 5.6 conclui-se que para sinais de voz, o método da correlação é a abordagem mais adequada para a estimação das matrizes de correlação. A utilização do método da correlação tem como vantagem a redução da complexidade computacional do algoritmo devido à estrutura Toeplitz das matrizes de correlação e conseqüentemente apresenta menor tempo médio para iteração. Resultados similares aos apresentados nesta seção foram observados em estudos de predição linear onde o método da correlação é a abordagem adotada devido a sua baixa complexidade computacional e boa estabilidade [64][67][87].

## 5.5 Avaliação da restrição *Sylvester* e suas aproximações

Na seção 3.2.2 do Capítulo 3 mostrou-se que um tipo de restrição conhecida como *Sylvester* é necessária para garantir a estrutura *Sylvester* das matrizes de atualização dos coeficientes dos filtros de separação  $\Delta W(m)$ . Conforme discutido naquele capítulo, a restrição *Sylvester* (SC) é imposta tomando-se a média aritmética dos elementos (valores para atualização dos coeficientes dos filtros de separação) presentes em cada uma das diagonais das matrizes  $\Delta W_{pq}$  para  $p, q = 1, \dots, P$ . Adicionalmente, foram propostas duas formas eficientes de aproximação da restrição *Sylvester* (SC) com o intuito de melhorar o desempenho apresentado pelo algoritmo. Ao invés de se realizar o cálculo da média das diagonais das matrizes de atualização, foi proposto na seção 3.2.2 que se utilizasse somente a primeira coluna ou a  $L$ -ésima linha da matriz  $\Delta W_{pq}$  como forma de aproximação da restrição *Sylvester* (SC). Estas duas formas eficientes de restrição são chamadas de restrição *Sylvester* da coluna e da linha e são simbolizadas por  $(SC_C)$  e  $(SC_R)$ , respectivamente.

Nesta seção são comparadas as diferentes formas de implementação da restrição *Sylvester*. Os resultados apresentados aqui foram obtidos convoluindo-se os sinais de voz utilizados na seção 5.4 com os filtros gerados pelo método da imagem também utilizados naquela seção. A técnica do passo de aprendizagem fixo foi adotada e o passo de aprendizagem  $\mu$  feito igual a 0,002.

A Figura 5.13 apresenta a comparação entre as três formas de implementação da restrição *Sylvester*. Nota-se que a restrição *Sylvester* original (SC) alcança a maior separação entre as formas implementações da restrição *Sylvester*. A aproximação da restrição *Sylvester* através da restrição *Sylvester* da coluna  $(SC_C)$  ou da linha  $(SC_R)$  não leva a uma degradação significativa do desempenho da separação apresentada pelo algoritmo.



**Figura 5.13:** Comparação entre a restrição *Sylvester* e suas aproximações eficientes.

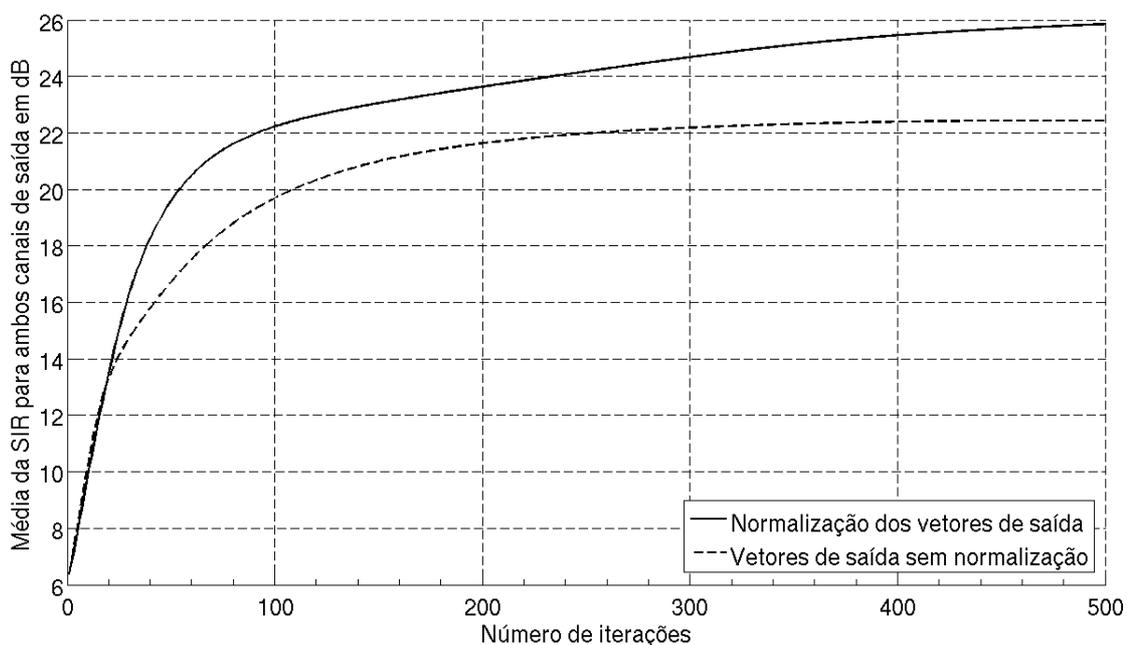
O resultado deste experimento mostra que a restrição *Sylvester* da linha ( $SC_R$ ) é uma aproximação adequada para a complexa restrição *Sylvester* ( $SC$ ). A Tabela 5.7 mostra que o tempo de simulação para a restrição *Sylvester* da linha ( $SC_R$ ) é 34,98 % menor do que o tempo de simulação para a restrição *Sylvester* original, ( $SC$ ). A SIR final alcançada pelo algoritmo utilizando a restrição *Sylvester* da linha ( $SC_R$ ) é apenas 0,0791 dB menor do que a SIR alcançada quando se utiliza a restrição original ( $SC$ ).

**Tabela 5.7:** Tabela comparativa entre a restrição *Sylvester* e suas aproximações eficientes.

Restrição	SIR Inicial (dB)	SIR Final (dB)	Tempo médio para cada iteração [s]	Tempo total [h]
( $SC$ )	7,1430	16,4767	44,20	6,14
( $SC_C$ )		16,2550	32,05	4,45
( $SC_R$ )		16,3976	28,74	3,99

## 5.6 Normalização dos sinais de saída do sistema de separação

Nesta seção apresenta-se o resultado referente à normalização dos vetores de saída do sistema de separação. Os vetores de saída  $y_q(n)$ ,  $q = 1, \dots, P$  do sistema de separação passam por um processo de normalização, onde cada amostra de saída do sistema tem o valor da média temporal subtraído e dividido pelo desvio padrão. Utilizou-se nesta seção o sinal de ruído “*babble*” (reunião) disponibilizado pela base de dados ETSI/Aurora [42]. O ruído foi misturado convolutivamente com o áudio de um locutor do sexo masculino, o qual foi gravado especificamente para este trabalho. Os sinais foram amostrados a uma frequência de 8000 Hz, possuem resolução de 16 bits e duração de 4 segundos. Utilizou-se o filtro gerado pelo método da imagem adotado nos experimentos realizados na seção 5.2.1. A técnica do passo de aprendizagem fixo foi utilizada para o experimento conduzido aqui e o melhor valor para o parâmetro  $\mu$  foi encontrado como sendo 0,06 após vários testes. A Figura 5.14 apresenta as curvas de Relação Sinal-Interferência (SIR) obtidas pelo algoritmo quando se normaliza e não se normaliza os vetores de saída do sistema de separação.



**Figura 5.14:** Curvas de SIR para a normalização e não normalização dos vetores de saída.

A Tabela 5.8 apresenta uma comparação entre os valores obtidos com a normalização e a não normalização dos vetores de saída. Os valores apresentados na tabela mostram que houve um aumento de 3,4196 dB na separação conseguida pelo algoritmo quando se normaliza os vetores de saída  $y_q(n)$ ,  $q = 1, \dots, P$  do sistema de separação. Nota-se também que há um aumento de 97,84 % no tempo levado pelo algoritmo para realizar as 500 iterações. Este aumento no tempo de simulação é devido ao processo de normalização dos vetores de saída para cada uma das iterações do algoritmo.

**Tabela 5.8:** Tabela comparativa entre a normalização e a não normalização dos vetores de saída  $y$ .

Vetor de saída $y$	SIR Inicial (dB)	SIR Final (dB)	Tempo médio para cada iteração [s]	Tempo total [h]
Não normalizado	6,3846	22,44	16,72	2,32
Normalizado		25,86	33,06	4,59

O objetivo da normalização é o de fazer com que a variância (que aqui é igual à energia do sinal pois a média é igual a zero) e a média dos sinais utilizados pelo algoritmo, durante o processo de separação, sejam respectivamente iguais a um e a zero. A normalização tende a fazer com que o processo de separação dos sinais misturados se torne mais comportado através da melhoria das condições numéricas do processo de otimização. O algoritmo se comporta melhor e produz resultados melhores. A principal vantagem da normalização dos vetores de saída é o de evitar que conjuntos de amostras com um grande intervalo de variação dominem conjuntos com um pequeno intervalo de variação.

Os resultados apresentados nesta seção mostram que a normalização dos vetores de saída tem como resultado uma melhoria do desempenho do algoritmo mas com um aumento do tempo de simulação.

## 5.7 Locutor em ambiente ruidoso e reverberante

O realce de sinais de voz em ambientes reverberantes e ruidosos é um problema desafiador que possui muitas aplicações e é bastante útil nas áreas de reconhecimento robusto de voz e de telecomunicações.

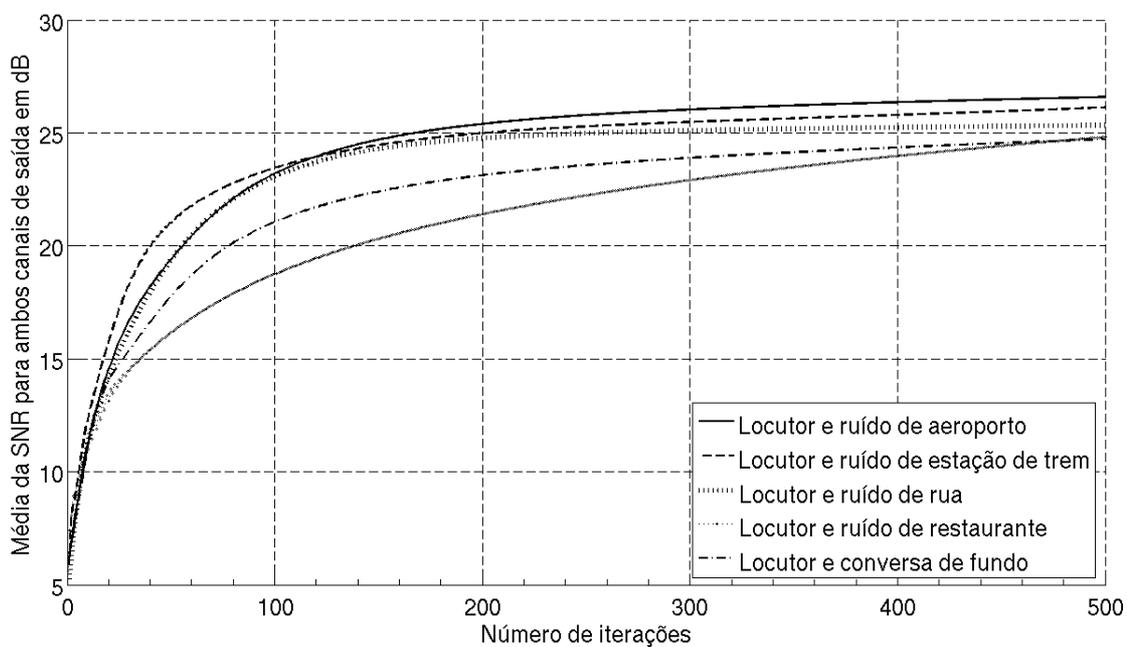
A dificuldade em realçar sinais de voz depende fortemente das condições do ambiente em que o locutor se encontra. Caso ele se encontre próximo a um

microfone, os efeitos de reverberação são mínimos e métodos tradicionais podem lidar com níveis moderados de ruído. Entretanto, caso o locutor se encontre distante do microfone, existirão distorções mais severas, que incluem grandes quantidades de ruído e perceptível reverberação. Denoising e de-reverberação de sinais de voz nestas condições tem se provado um problema bastante difícil.

Esta seção apresenta resultados de experimentos que mostram a utilidade do algoritmo apresentado neste trabalho na de-reverberação e denoising de sinais de voz misturados convolutivamente com ruídos disponibilizados pela base de dados ETSI/Aurora [42]. A base de dados Aurora contem a gravação de 5 ruídos captados nos seguintes ambientes: aeroporto, reunião, restaurante, rua e estação de trem. Estes ruídos foram misturados convolutivamente com o áudio de um locutor gravado especificamente para este trabalho. Os sinais foram amostrados a uma frequência de 8000 Hz, possuem resolução de 16 bits e duração de 4 segundos. Utilizou-se o filtro gerado pelo método da imagem adotado nos experimentos realizados na seção 5.2.1. A técnica do passo de aprendizagem fixo foi utilizada nestes experimentos e o melhor valor para o parâmetro  $\mu$  foi encontrado como sendo 0,04 após vários testes. Nesta seção, a medida de supressão do sinal interferente (SIR) pode ser interpretada como sendo a medida da SNR, pois como visto, o sinal interferente aqui é um ruído. Portanto, se utilizará tal termo daqui por diante. A Figura 5.15 mostra a curva da SNR obtida pelo algoritmo para cada uma das misturas entre o locutor e os ruídos da base de dados Aurora.

Como pode ser observado, o algoritmo obtém um alto grau de separação entre o sinal do locutor e os ruídos em todos os casos. A Tabela 5.9 apresenta os valores de SNR inicial e final para os experimentos realizados.

O ganho é superior a 17 dB para o pior caso e de mais de 20 dB para o melhor. Este ganho proporcionado pelo algoritmo poderia ser utilizado para aumentar o desempenho de sistemas de reconhecimento de fala em ambientes ruidosos.



**Figura 5.15:** Resultados de experimentos com ruídos.

**Tabela 5.9:** SNR inicial e máxima para os experimentos com ruídos.

Ruído	SNR Inicial (dB)	SNR Final (dB)	Ganho (dB)
Aeroporto	6,1320	26,5912	20,4592
Reunião	6,3846	24,7258	18,3412
Restaurante	7,4158	24,8305	17,4147
Rua	5,2653	25,3379	20,0726
Estação de trem	5,7279	26,1338	20,4059

## Capítulo 6

### Conclusões

Neste capítulo são apresentadas as principais contribuições da pesquisa realizada no presente trabalho, bem como sugestões para pesquisas futuras, com base nos resultados obtidos.

Nos últimos anos, muitas pesquisas têm sido dedicadas ao estudo do problema da Separação Cega de Fontes misturadas convolutivamente. Existem várias razões que explicam porque há um esforço muito grande por parte dos pesquisadores para se aplicar técnicas para separação cega à interfaces homem-máquina, onde atualmente esquemas de seletividade espacial (*beamforming*) adaptativo ainda predominam. Uma razão, é que abordagens para SCF somente se baseiam na suposição da independência mútua das fontes de sinal original e não necessitam de mais nenhuma informação a priori, tais como disposição dos sensores ou a posição das fontes de sinal desejado e interferente. Outra razão se deve ao fato de que as características espectrais (resposta em frequência) dos microfones utilizados não afetam o desempenho de tais algoritmos.

O problema da SCF consiste em recuperar ou identificar uma ou mais fontes originais quando somente suas misturas estão disponíveis para observação. Nos últimos 15 anos, tem crescido o número de pesquisas relacionadas a este assunto devido, principalmente, à variedade de aplicações que envolvem múltiplos sensores e múltiplas fontes. Como consequência, cresce o sentimento de que as técnicas para separação cega constituem uma ferramenta universal capaz de encontrar aplicações

nas mais diferentes áreas.

O objetivo principal desta dissertação foi o de estudar formas de se acelerar a convergência de algoritmos para Separação Cega de Fontes através de algumas técnicas trazidas da teoria de Redes Neurais e realizar simulações comparativas entre estas técnicas. Além disso, estudou-se também formas de se diminuir a complexidade computacional apresentada pelo algoritmo (normalização eficiente, estimação das matrizes de correlação através do método da correlação e uso da restrição *Sylvester* da linha) adotado neste trabalho e sua utilização na separação de sinais de voz misturados com ruídos.

O resultado da adoção de três técnicas vindas da teoria de Redes Neurais (Termo Momentum, Decaimento Exponencial e *Bold Driver*), utilizadas para acelerar a convergência e aumentar a qualidade da separação apresentada pelo algoritmo discutido no Capítulo 3 foram estudadas. Simulações envolvendo misturas convolutivas sintéticas e reais comprovam que todas as três técnicas fazem com que o algoritmo apresente uma taxa de convergência mais rápida do que a apresentada pelo algoritmo quando utilizando um passo de aprendizagem fixo. Dentre todas as técnicas estudadas, a do decaimento exponencial é a que apresenta um melhor resultado, devendo ser escolhida sempre que possível. A única desvantagem desta técnica, é que seus parâmetros necessitam ser ajustados empiricamente.

Adicionalmente, foram estudadas algumas modificações no algoritmo adotado neste trabalho com o objetivo de diminuir sua complexidade computacional. A redução da complexidade computacional viabiliza a utilização de técnicas mais custosas, tais como a do *bold driver*, que conforme mostrado no Capítulo 5 apresenta o maior valor de tempo médio entre as técnicas estudadas. Isto é devido a necessidade do cálculo da norma L2 a cada iteração do algoritmo. Vale ressaltar que a redução da complexidade computacional e conseqüente redução do tempo de simulação que é obtida através das modificações discutidas no Capítulo 5 pode ser utilizada para diminuir o tempo de ajuste dos parâmetros de técnicas como a do decaimento exponencial, que possuem alguns parâmetros a serem ajustados. Estas modificações levaram a implementações eficientes do algoritmo utilizado aqui e mantiveram uma boa qualidade de separação dos sinais misturados. As modificações

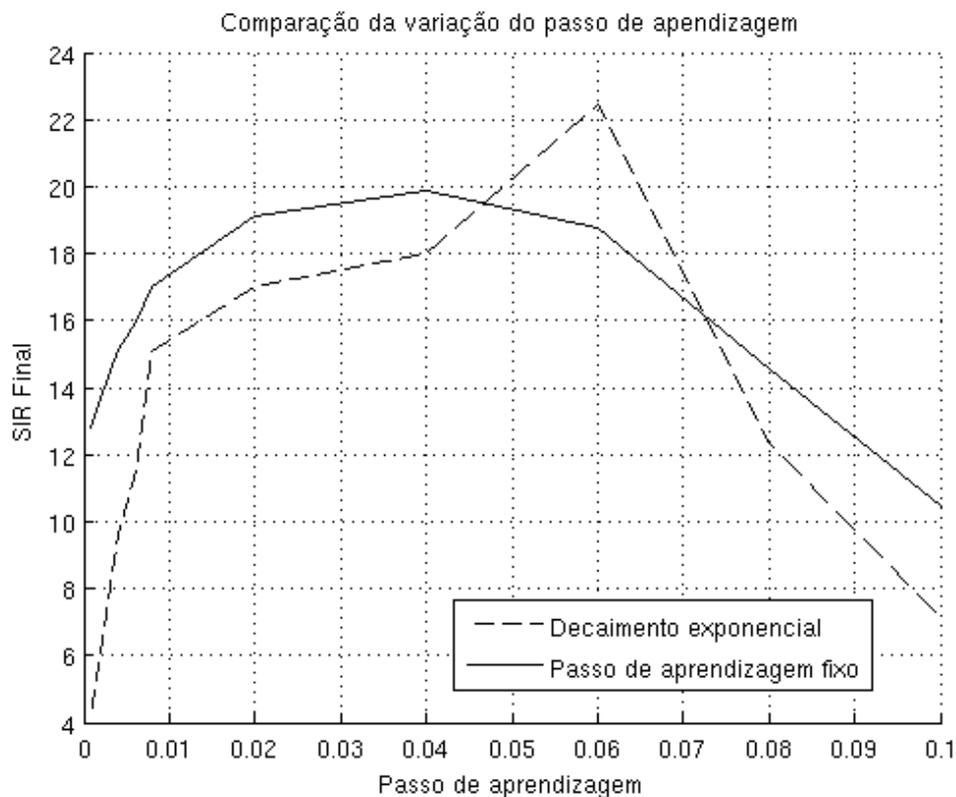
apresentadas no Capítulo 5 podem ser utilizadas de forma separada ou em conjunto, pois cada uma modifica uma parte distinta do algoritmo. A combinação de todas as modificações discutidas no Capítulo 5 leva a uma grande redução do tempo de simulação apresentado pelo algoritmo. Entretanto tem-se como consequência a diminuição da qualidade da separação. Quando se combina as três modificações discutidas no Capítulo 5, normalização eficiente, estimação das matrizes de correlação cruzada através do método da correlação e restrição *Sylvester* da linha ( $SC_R$ ) obtém-se um decréscimo de 67,26 % no tempo de processamento quando comparado ao tempo de processamento do algoritmo sem nenhuma das modificações discutidas anteriormente. Entretanto, a qualidade de separação do algoritmo sofre uma degradação: durante os testes realizados verificou-se que esta degradação é de aproximadamente 1,4 % em relação ao algoritmo sem nenhuma modificação para melhoria do desempenho computacional. Esta degradação é insignificante e pode ser eliminada quando se utiliza umas das técnicas para adaptação do passo de aprendizagem discutidas neste trabalho. A considerável diminuição do tempo de simulação aliada a insignificante degradação causada pelas modificações discutidas no Capítulo 5 fazem com que elas sejam de grande valia para algoritmos de separação cega de fontes que adotam abordagens de segunda ordem (SOS) no domínio do tempo.

Constatou-se também através de simulações que o algoritmo apresenta excelentes resultados quando este tem como entrada sinais de voz misturados com vários tipos de ruídos de fundo (*denoising*) disponibilizados pela base de dados ETSI/Aurora [42]. O algoritmo adotado aqui pode ser utilizado como uma etapa de pré-processamento em sistemas de reconhecimento de fala em ambientes ruidosos.

Durante as simulações utilizando-se a abordagem do passo de aprendizagem fixo, verificou-se que a escolha de seu valor inicial afeta o desempenho do algoritmo. Valores muito grandes fazem com que o algoritmo fique instável e valores muito pequenos fazem com que o processo de separação se torne lento. O melhor valor para o passo de aprendizagem para todos os conjuntos de simulações foi encontrado através de um processo empírico. Dentre todas as técnicas de adaptação do passo de aprendizagem apresentadas neste trabalho, a do decaimento exponencial é a mais

sensível a escolha do valor inicial do passo de aprendizagem. Isto se deve ao fato desta técnica não ter o valor do passo de aprendizagem ajustado dinamicamente a cada iteração do algoritmo. O valor do fator de decaimento da exponencial utilizada por esta técnica também deve ser encontrado empiricamente.

A Figura 6.1 apresenta um gráfico onde compara-se as variações da SIR obtida pelo sistema de separação quando se varia o passo de aprendizagem. A comparação é feita entre as técnicas do passo de aprendizagem fixo e do decaimento exponencial. Estes experimentos foram realizados utilizando-se as mesmas condições dos experimentos apresentados na seção 5.7 do Capítulo 5. Conforme pode ser visto, a técnica do decaimento exponencial é realmente sensível à escolha do passo de aprendizagem inicial: valores muito pequenos fazem com que o algoritmo não encontre o ponto de mínimo e conseqüentemente não atinja bons valores de SIR quando compara-se com o mesmo valor de  $\mu$  para a técnica do passo de aprendizagem fixo. Vale notar também que para valores pequenos de  $\mu$ , a técnica do passo de aprendizagem fixo faz com que o algoritmo atinja valores razoáveis de SIR. Isto se deve ao fato de que o valor do passo de aprendizagem é constante ao longo das iterações do algoritmo, o que não ocorre com a técnica do decaimento exponencial, onde  $\mu$  decai, tendendo a zero ao longo das iterações. Desta forma, o algoritmo não consegue chegar até a área com o ponto de mínimo, pois o tamanho dos passos tomados em direção à aquele ponto são diminuídos ao longo das iterações. Conforme aumenta-se o valor de  $\mu$ , a técnica do decaimento exponencial faz com que o algoritmo encontre rapidamente o ponto de mínimo e conseqüentemente atinja valores de SIR maiores do que os atingidos quando se adota a técnica do passo de aprendizagem fixo com o mesmo valor de  $\mu$ . O aumento deliberado do passo de aprendizagem faz com que o algoritmo fique instável, i.e., quando o ponto de mínimo é encontrado, o alto valor de  $\mu$  faz com que o algoritmo passe o ponto e se distancia bastante do ponto de mínimo. O resultado desta instabilidade são os valores baixos de SIR quando  $\mu$  é muito alto.



**Figura 6.1:** Variação da SIR em relação a variação do passo de aprendizagem.

As técnicas do termo momentum e do *bold driver* não são drasticamente afetadas pela escolha do passo de aprendizagem, pois são técnicas que tem o passo de aprendizagem ajustado a cada iteração do algoritmo.

Após análise dos resultados obtidos durante as simulações conclui-se que a técnica do decaimento exponencial é a que apresenta melhor resultado. Ela aumenta a taxa de aprendizagem e conseqüentemente diminui o tempo necessário para a que o algoritmo convirja. Adicionalmente, a técnica do decaimento exponencial melhora a qualidade da separação obtida pelo algoritmo, i.e., aumenta o valor da Relação Sinal-Interferência (SIR) de saída do sistema de separação. Além disto, dentre todas as outras técnicas de adaptação do passo de aprendizagem estudadas aqui, a do decaimento exponencial é a que apresenta menor tempo médio para cada iteração do algoritmo. O tempo médio de processamento para cada iteração do algoritmo é muito próximo do tempo levado quando se utiliza a abordagem do passo de aprendizagem

fixo. Entretanto, como citado anteriormente, seus parâmetros necessitam ser ajustados empiricamente.

## 6.1 Sugestões para trabalhos futuros

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

- Uma das desvantagens de muitos algoritmos para SCF é que o número de fontes deve ser conhecido à priori para que o sistema de separação seja estimado de forma correta. Assim uma técnica que descubra o número de fontes à priori, ou seja, antes do início do algoritmo se faz necessária.
- Todos os experimentos realizados aqui foram feitos de forma "*offline*", ou seja, todo o sinal a ser processado se encontrava armazenado e o algoritmo sempre o processava completamente em todas as iterações. A habilidade de um algoritmo em se adaptar em tempo real é bastante útil em varias aplicações. Teleconferências e dispositivos para auxílio na audição são aplicações onde o processamento em tempo real é necessário. Portanto, extensões e comparações do algoritmo adotado aqui para processamento em tempo real necessitam ser exploradas.
- A escolha de valores para parâmetros tais como passo de aprendizagem, termo momentum, número de instantes de tempo considerados para o cálculo das matrizes de correlação e fator de decaimento exponencial das técnicas estudadas neste trabalho foram encontrados através de um processo empírico. Este processo é bastante custoso e, conseqüentemente, formas de se obter estimativas ótimas destes parâmetros seriam um tópico de estudo bastante útil.
- O estudo de estratégias para se obter estimativas iniciais dos coeficientes dos

filtros de separação, fariam com que o algoritmo convergisse mais rapidamente, diminuindo o tempo gasto para se atingir um valor predeterminado de SIR.

- O desempenho do algoritmo quando submetido a sinais de entrada com diferentes frequências de amostragem deve produzir resultados melhores, uma vez que quanto maior a frequência de amostragem maior a quantidade de informação e, portanto, talvez maior seja o grau de separação dos sinais misturados.
- Modelos de Misturas Gaussianas (*Gaussian Mixture Models*, GMM) pertencem à classe de técnicas para reconhecimento de padrões. Eles modelam a função densidade de probabilidade de observações de variáveis utilizando-se misturas de densidades de probabilidade gaussianas multivariáveis. Dado uma série de entradas, os pesos de cada uma das distribuições é refinado através de algoritmos de maximização da esperança (*Expectation-Maximization*, EM). GMMs são utilizados em sistemas de reconhecimento de locutor. Quando uma locução de teste é fornecida ao sistema é possível verificar qual dos modelos de locutor mais correlacionados com as propriedades das locução inserida. Este locutor é então eleito pelo sistema de identificação como o mais correlacionado com as propriedades da locução de teste previamente disponibilizada. O estudo da existência de uma relação entre a probabilidade de ocorrência de uma locução dado um modelo de locutor,  $P(O|M)$ , e a Relação Sinal-Interferência (SIR) entregue na saída do sistema de separação seria de grande valia para a área de processamento cego de sinais. Caso exista uma relação, poderia-se utilizar o valor de  $P(O|M)$  entregue por um GMM para prever o grau de separação que pode ser alcançado por tais algoritmos.
- Estudo da utilização de filtros passa-faixa, passa-baixa ou passa-alta ao invés de filtros passa-tudo na inicialização das matrizes que modelam o sistema de

separação.

## **Anexo A**

### **Artigos referentes à dissertação**

#### **C.1 Artigos publicados**

- Figueiredo, F. A. P. e Ynoguti, C. A. “*On the improvement of the learning rate in Blind Source Separation using techniques from Artificial Neural Networks theory,*” Proceedings of the International Workshop on Telecommunications – IWT/09.
- Figueiredo, F. A. P. e Ynoguti, C. A. “*Blind Source Separation in Reverberant Environment Using Genetic Algorithms,*” Proceedings of the International Workshop on Telecommunications – IWT/11.

#### **C.2 Artigos submetidos**

- Figueiredo, F. A. P. e Ynoguti, C. A. “*Noise Reduction in Reverberant Environments with a Blind Source Separation Algorithm,*” artigo submetido ao XXIX Simpósio Brasileiro de Telecomunicações – SBrT 2011.

## **Anexo B**

### **Códigos da implementação**

Os códigos criados para implementar o algoritmo de separação cega de fontes se encontram num DVD fornecido juntamente com este trabalho.

## Referências Bibliográficas

- [1] Sato, Y. "A method of self recovering equalization for multilevel amplitude-modulation systems," IEEE Trans. Com. 23, pages 679–682, 1975.
- [2] Godard, D. "Self recovering equalization and carrier tracking in two dimensional data communication systems," IEEE Trans. Com. 28, pages 1867–1875, 1980.
- [3] Benveniste, A., Goursat, M. e Ruget, G. "Robust identification of a non-minimum phase system," IEEE Trans. Auto. Contr. 25, pages 385–399, 1980.
- [4] Donoho, D. "On minimum entropy deconvolution," in: Applied Time Series Analysis II, Tulsa, 1980, pages 565–608.
- [5] Héroult, J., Jutten, C. e Ans, B. "Detection de grandeurs primitives dans un message composite par une architecture de calcul neuromimetique em apprentissage non supervise," Actes du Xeme colloque GRETSI, 1985.
- [6] Héroult, J. e Ans, B. "Circuits neuronaux à synapses modifiables: décodage de messages composites par apprentissage non supervisé," C. -R. de l'Acad mie des Sciences, 299(III-13):525-528, 1984.
- [7] Ans, B., Héroult, J. e Jutten, C. "Adaptive neural architectures: detection of primitives," In Proc. of COGNITIVA'85, pages 593-597, Paris, France, 1985.
- [8] Héroult, J. e Jutten, C. "Space or time adaptive signal processing by neural networks models," in: Intern. Conf. on Neural Networks for Computing, Snowbird (Utah, USA), 1986, pp. 206–211.
- [9] Roll, J.-P. "Contribution à la proprioception musculaire, à la perception et au contrôle du mouvement chez l'homme," PhD thesis, Université d'Aix-Marseille 1, 1981.
- [10] Bar-Ness, Y., Carlin, J. e Steinberger, M. "Bootstrapping adaptive cross-pol

*canceller for satellite communications,*” in Proc. IEEE Int. Conf. Communications, 1982, pp. 4F.5.1-4F.5.5.

[11] Bar-Ness, Y., Carlin, J. e Steinberger, M. “*Bootstrapping adaptive interference cancelers: some practical limitations,*” in: Proc. The Globecom. Conference, 1982, pp. 1251–1255.

[12] Le Cun, Y. “*A learning scheme for asymmetric threshold network,*” in: Proceedings of COGNIT- ICA’85, Paris, France, 1985, pp. 599–604.

[13] Jutten, C. e Hérault, J. “*Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture,*” Signal Process. 24 (1991) 1–10.

[14] Comon, P., Jutten, C. e Hérault, J. “*Blind separation of sources, Part II: Problem statement,*” Signal Process. 24 (1991) 11–20.

[15] Sorouchyari, E. “*Blind separation of sources, Part III: Stability analysis,*” Signal Process. 24 (1991) 21–29.

[16] Comon, P. “*Analyse en Composantes Indépendantes et identification aveugle,*” Traitement du Signal 7 (1990) 435–450. Special issue.

[17] Burel, G. “*Blind separation of sources: A nonlinear neural algorithm,*” Neural Netw. 5 (1992) 937–947.

[18] Hyvärinen, A. e Pajunen, P. “*Nonlinear independent component analysis: Existence and uniqueness results,*” Neural Netw. 12 (1999) 429–439.

[19] Taleb, A. e Jutten, C. “*Source separation in post nonlinear mixtures,*” IEEE Trans. SP 47 (1999) 2807–2820.

[20] Jutten, C. “*Calcul neuromimétique et traitement du signal: analyse en composantes indépendantes,*” Thèse d’état ès sciences physiques, UJF-INP Grenoble, 1987.

[21] Comon, P. “*Independent Component Analysis,*” Republished in Higher-Order Statistics, J. L. Lacoume (Ed.), Elsevier, 1992, pp. 29–38.

[22] Comon, P. “*Independent component analysis, a new concept?,*” Signal Process. 36 (1994) 287–314.

[23] Darmois, G. “*Analyse générale des liaisons stochastiques,*” Rev. Inst. Intern. Stat. 21 (1953) 2–8.

[24] Kagan, A., Linnik, Y. e Rao, C. “*Characterization Problems in Mathematics*

*Statistics*,” John Wiley & Sons, 1973.

[25] Donoho, D. “*Nature vs. math: Interpreting independent component analysis in light of computational harmonic analysis*,” in: Independent Component Analysis and Blind Source Separation, ICA2000, Helsinki, June 2000, pp. 459–470.

[26] Duvaut, P. “*Traitement du Signal, Special issue on nonlinear and non-Gaussian techniques*,” December 1990.

[27] Comon, P. e Jutten, C. “*Handbook of Blind Source Separation, Independent Component Analysis and Applications*,” Elsevier, 2010.

[28] Hyvärinen, A., Karhunen, J. e Oja, E. “*Independent Component Analysis*,” Wiley & Sons, Inc., New York, 2001.

[29] Cichocki, A. e Amari, S. “*Adaptive Blind Signal and Image Processing*,” Wiley & Sons, Ltd., Chichester, UK, 2002.

[30] Cherry, E. C. “*Some Experiments on the Recognition of Speech, with One and with Two Ears*,” *Journal of the Acoustical Society of America*, 24, pp. 975-979, 1953.

[31] Haykin, S. e Chen, Z. “*The Cocktail Party Problem*”, *Neural Computation*, vol. 17, pp. 1875-1902, 2005.

[32] 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.

[33] Lee, T.-W., Ziehe, A., Orglmeister, R. e Sejnowski, 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.

[34] Cohen, D. “*Magnetoencephalography: Evidence of magnetic fields produced by alpha rhythm currents*,” *Science* 161 (1972) 664–666.

[35] Makeig, S., Bell, A.J., Jung, T.-P. e Sejnowski, T. J. “*Independent component analysis of electroencephalographic data*,” *Advances in Neural Information Processing Systems* 8 (1996) 145–151.

[36] Makino, S., Lee, T.-W. e Sawada, H. “*Blind Speech Separation*,” Springer, Dordrecht, the Netherlands, 2007.

[37] Veen, A.V.D., Talwar, S. e Paulraj, A. “*A subspace approach to blind space-time*

*signal processing for wireless communication systems,*” IEEE Trans. Signal Process. 45 (1997) 173–190.

[38] Gorokhov, A. e Loubaton, P. “*Subspace-based techniques for blind separation of mixtures with temporally correlated sources,*” IEEE Trans. Circuit. Syst. 44 (1997) 813–820.

[39] Moreira, M. e Fiesler, E. “*Neural networks with adaptive learning rate and momentum terms,*” IDIAP technical report , 1995.

[40] Braga, A. P., Carvalho, A. P. L. F. e Ludermir, T. B. “*Redes Neurais Artificiais: Teoria e Aplicações,*” LTC Livros Técnicos, 2000.

[41] Buchner, H., Aichner, R. e Kellermann, W. “*A generalization of a class of blind source separation algorithms for convolutive mixtures,*” Proc. IEEE Int. Symposium on Independent Component Analysis and Blind Signal Separation (ICA), Nara, Japan, Apr. 2003, pp. 945-950.

[42] “*Aurora project database,*” <http://aurora.hsr.de/aurora-5.html>.

[43] Disponível em: <http://www.mathworks.com/products/matlab/>

[44] Back, A.D. e Weigend, A. S. “*A first application of independent component analysis to extracting structure from stocks returns,*” International Journal of Neural Systems, 8(4): 473-484, 1997.

[45] Zibulevsky, M. e Pearlmutter, B.A. “*Blind source separation by sparse decomposition in a signal dictionary,*” Neural Computation, vol. 13, pp. 863-882, 2001.

[46] Araki, S., Makino, S., Sawada, H. e Mukay, R. “*Underdetermined Blind Separation of Convolutive Mixtures of Speech Using Time-Frequency Mask and Mixing Matrix Estimation,*” in IEICE Transactions on Fundamentals od Electronics, Communications and Computer Science, Vol. E88-A Issue 7, July 2005.

[47] Araki, S., Makino, S., Blin, A., Mukai, R. e Sawada, H. “*Blind Separation of More Speech than Sensors with Less Distortion by Combining Sparseness and ICA,*” in Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Kyoto, Japan, Sep. 2003, pp. 271-274.

[48] Rodrigues, G. F., de Souza, A. C. S., de Araújo, L. C. e Yehia, H. C., “*Blind Separation of Speech Convolutive Mixtures via Time-Frequency Masking,*” in Proceedings of the international Workshop on Telecommunications – IWT/07, pp 204-

208, 2007.

[49] Molgedey, L. e Schuster, H.G. “*Separation of a mixture of independent signals using time delayed correlations,*” Physical Review Letters, vol. 72, pp. 3634-3636, 1994.

[50] Tong, L., Liu, R.-W., Soon, V.C. e Huang, Y.-F. “*Indeterminacy and identifiability of blind identification,*” IEEE Trans. on Circuits and Systems, vol. 38, pp. 499-509, 1991.

[51] Nishikawa, T., Saruwatari, H., e Shikano, K. “*Comparison of time-domain ICA, frequency-domain ICA and multistage ICA for blind source separation,*” in Proc. European Signal Processing Conference (EUSIPCO), Sep. 2002, vol. 2, pp. 15-18.

[52] Aichner, R., Araki, S., Makino, S., Nishikawa, T. e Saruwatari, H. “*Time-domain blind source separation of non-stationary convolved signals with utilization of geometric beamforming,*” in Proc. Int. Workshop on Neural Networks for Signal Processing (NNSP), Martigny, Switzerland, 2002, pp. 445-454.

[53] Buchner, H., Aichner, R. e Kellermann, W. “*A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics,*” IEEE Transactions on Speech and Audio Processing, Jan 2005.

[54] Buchner, H., Aichner, R., e Kellermann, W. “*Blind Source Separation for Convolutive Mixtures Exploiting Nongaussianity, Nonwhiteness, and Nonstationarity,*” Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Kyoto, Japan, September 2003, pp. 275-278.

[55] Cover, T.M. e Thomas, J.A. “*Elements of Information Theory,*” Wiley & Sons, New York, 1991.

[56] Haykin, S. “*Neural Networks: A Comprehensive Foundation,*” Prentice Hall, 1998.

[57] Moulines, E., Amrane, O. A. e Grenier, Y. “*The generalized multidelay adaptive filter: structure and convergence analysis,*” IEEE Trans. Signal Processing, vol. 43, pp. 14-28, Jan. 1995.

[58] Buchner, H., Benesty, J. e Kellermann, W. “*Multichannel Frequency-Domain Adaptive Algorithms with Application to Acoustic Echo Cancellation,*” in J.Benesty and Y.Huang (eds.), Adaptive signal processing: Application to real-world problems,

Springer-Verlag, Berlin/Heidelberg, Jan. 2003.

[59] Ekstrom, M.P. “*A spectral characterization of the ill-conditioning in numerical deconvolution,*” IEEE Trans. Audio Electroacoust., AU-21(4):344– 348, August 1973.

[60] Miyoshi, M. e Kaneda, Y. “*Inverse filtering of room acoustics,*” IEEE Trans. Acoust., Speech, Signal Processing, 36(2):145–152, February 1988.

[61] Milinazzo, F., Zala, C. e Barrodale, I. “*On the rate of growth of condition numbers for convolution matrices,*” IEEE Trans. Acoust., Speech, Signal Processing, ASSP-35(4):471–475, April 1987.

[62] Matsuoka, K., Ohya, M. e Kawamoto, M. “*A neural net for blind separation of nonstationary signals,*” Neural Networks, vol. 8, no. 3, pp. 411-419, 1995.

[63] Shannon, C.E. e Weaver, W. “*The mathematical theory of communication,*” University of Illinois Press, Urbana, Illinois 1949.

[64] Markel, J.D. e Gray, A. H. “*Linear Prediction of Speech,*” Springer-Verlag, Berlin, 1976.

[65] Oppenheim, A. “*Inequalities connected with definite hermitian forms,*” J. London Math. Soc., vol. 5, pp. 114-119, 1930.

[66] Aichner, R., Buchner, H. e Kellermann, W. “*On the causality problem in time-domain blind source separation and deconvolution algorithms,*” In Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), volume 5, pages 181–184, Philadelphia, PA, USA, March 2005.

[67] Aichner, R. “*Acoustic Blind Source Separation in Reverberant and Noisy Environments,*” PhD Thesis, University Erlangen-Nuremberg, Germany, Oct. 2007.

[68] Aichner, R., Buchner, H., Yan, F. e Kellermann, W. “*A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments,*” Signal Processing, 86(6):1260–1277, June 2006.

[69] Van Veen, B. e Buckley, K. “*Beamforming: A versatile approach to spatial filtering,*” IEEE Trans. Acoust., Speech, Signal Processing, pages 4–24, April 1988.

[70] Amari, S. “*Natural gradient works efficiently in learning,*” Neural Computation, vol. 10, pp. 251-276, 1998.

[71] Cardoso, J. F. “*Blind signal separation: Statistical principles,*” Proc. IEEE, vol.

86, pp. 2009-2025, Oct. 1998.

[72] Schervish, M. J. “*Theory of Statistics*,” Springer, 1995.

[73] Haykin, S. “*Adaptive Filter Theory*,” 3rd ed., Prentice Hall., Englewood Cliffs, NJ, 1996.

[74] Deller, J. R., Hansen, J. H. L., e Proakis, J.G. “*Discrete-Time Processing of Speech Signals*,” IEEE Press, New York, 2000.

[75] Markel, J. D. e Gray, A. H. “*Linear Prediction of Speech*,” Springer, Berlin, 1976.

[76] Makhoul, J., e Wolf, J. “*Linear Prediction and the Spectral Analysis of Speech*,” BBN Report No. 2304, August 1972.

[77] Cichocki, A. e Unbehauen, R. “*Neural Networks for Optimization and Signal Processing*,” John Wiley & Sons, Chichester, 1994.

[78] Buchner, H., Benesty, J., e Kellermann, W. “*Generalized multichannel frequency-domain adaptive filtering: Efficient realization and application to hands-free speech communication*,” *Signal Processing*, 85:549–570, 2005.

[79] Rumelhart, D. E., Hinton, G. E., e Williams, R. J. “*Learning internal representations by error propagation*,” in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, volume I: Foundations, chapter 8, pages 318-362. MIT Press, Cambridge, Massachusetts, 1986.

[80] Qian, N. “*On the momentum term in gradient descent learning algorithms*,” *Neural Networks Volume 12*, pp. 145–151, 1999.

[81] Battiti, R. “*Accelerated Backpropagation Learning: Two Optimization Methods*,” *Complex Systems*, 3:331-342, 1989.

[82] Kohonen, T. “*Self-Organized Formation of Topologically Correct Feature Maps*,” *Biological Cybernetics*, Volume 43, pp. 59-69, 1982.

[83] Kohonen, T. “*The Hypermap Architecture*,” In T. Kohonen, K. Makisara, O. Simula, and J. Kangas (Eds), *Artificial Neural Networks*, pp. 1357-1360, 1991.

[84] Rindel, J. H. “*The use of computer modeling in room acoustics*,” *Journal of Vobroengineering*, vol. 3, no. 4, pp. 41–72, 2000.

[85] Savioja, L. “*Modeling techniques for virtual acoustics*,” Ph.D. dissertation, Helsinki University of Technology, august 2000.

- [86] Schobben, D., Torkkola, K., Smaragdis, P. “*Evaluation of Blind Signal Separation Methods*,” in Proceedings Int. Workshop Independent Component Analysis and Blind Signal Separation, J.F. Cardoso, Ch. Jutten, and Ph. Loubaton, Eds., Aussois, France, pp. 261-266, January 1999.
- [87] Barnwell III, T.P., Nayebi, K., e Richardson, C. H. “*Speech Coding: A Computer Laboratory Textbook*,” Digital Signal Processing Laboratory Series. Georgia Tech, 1996.
- [88] Amari, S., Cichocki, A., e Yang, H. H. “*A new learning algorithm for blind signal separation*,” in Advances in neural information processing systems, 8, Cambridge, MA, MIT Press, 1996, pp. 757-763.
- [89] Herault, J. e Jutten, C. “*Space and time adaptive signal processing by neural network models*,” Neural Network for Computing, AIP Conference Proceedings 151, pp. 207-211, 1986.
- [90] Cardoso, J. F. “*Independent component analysis of the cosmic microwave background*,” In Proc. Int. Symp. Independent Component Analysis and Blind Signal Separation (ICA), pages 1111–1116, Nara, Japan, April 2003.
- [91] Bradwood, D. “*Cross-coupled cancellation systems for improving cross-polarisation discrimination*,” in Proc. IEEE Int. Conf. Antennas and Propagation, Nov. 1978, vol. I, pp. 41-45.
- [92] Weinstein, E., Feder, M., e Oppenheim, A. “*Multi-channel signal separation by decorrelation*,” IEEE Trans. on Speech and Audio Processing, vol 1, no. 4, pp. 405-413, Oct. 1993.
- [93] Gerven, S. V. e Compernelle, D. V. “*Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness*,” in IEEE Trans. Signal Processing, vol. 43, no. 7, pp. 1602-1612, 1995.
- [94] Gerven, S. V. e Compernelle, D. V. “*Signal separation in a symmetric adaptive noise canceler by output decorrelation*,” in Proc. ICASSP 92, 1992, vol. IV, pp. 221-224.
- [95] Tong, L., Soon, V.C., Inouye, Y., Huang, Y., e Liu, R. “*Waveform-preserving blind estimation of multiple sources*,” In Proc. IEEE Conference on Decision and Control, pages 2388–2393 vol.3, Brighton, UK, 1991.

- [96] Tong, L., Inouye, Y., e Liu, R. “*A finite-step global algorithm for the parameter estimation of multichannel MA processes,*” IEEE Trans. Signal Proc., 40(10):2547–2558, 1992.
- [97] Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., e Moulines, E. “*Second-order blind separation of correlated sources,*” In Proc. Int. Conference on Digital Sig. Processing, pages 346–351, Cyprus, 1993.
- [98] Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., e Moulines, E. “*A blind source separation technique using second-order statistics,*” IEEE Trans. Signal Processing, 45(2):434–444, February 1997.
- [99] Wu, H. C. e Principe, J. C. “*A unifying criterion for blind source separation and decorrelation: simultaneous diagonalization of correlation matrices,*” In Proc. IEEE Int. Workshop Neural Networks for Signal Processing (NNSP), pages 496–505, Amelie Island, Fl, USA, 1997.
- [100] Krongold, B. S. e Jones, D. L. “*Blind source separation of nonstationary convolutively mixed signals,*” In Proc. IEEE Workshop on Statistical Signal and Array Processing (SSAP), pages 53–57, Pocono Manor, PA, USA, August 2000.
- [101] Liu, R., Inouye, Y., e Luo, H. “*A system-theoretic foundation for blind signal separation of MIMO- FIR convolutive mixtures - a review,*” in ICA’00, 2000, pp. 205–210.
- [102] Parra, L. e Spence, C. “*Convolutive blind source separation based on multiple decorrelation,*” in Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP’97), Cambridge, UK, 1998.
- [103] Parra, L. e Spence, C. “*Convolutive blind source separation of non-stationary sources,*” IEEE Trans. Speech and Audio Processing, pp. 320-327, May 2000.
- [104] Kawamoto, M., Matsuoka, K., e Ohnishi, N. “*A method of blind separation for convolved non-stationary signals,*” Neurocomputing, vol. 22, pp. 157-171, 1998.
- [105] Buchner, H., Aichner, R., Stenglein, J., Teutsch, H., e Kellermann, W. “*Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering,*” In Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), volume 3, pages 97–100, Philadelphia, PA, USA, March 2005.
- [106] Aichner, R., Buchner, H., Wehr, S., e Kellermann, W. “*Robustness of acoustic*

*multiple-source localization in adverse environments,*” In Proc. Of 7th ITG-Fachtagung Sprachkommunikation, Kiel, Germany, April 2006.

[107] Chan, D. B., Rayner, P. J. W., e Godsill, S. J. “*Multi-channel signal separation,*” In Proc. ICASSP, pages 649–652, May 1996.

[108] Weinstein, E., Feder, M., e Oppenheim, A. V. “*Multi-channel signal separation by decorrelation,*” IEEE Trans. on Speech and Audio Processing, 1(4):405–413, Oct. 1993.

[109] Di Persia, L., Milone, D. H., 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.

[110] Mitianoudis, N., Davies, M. “*New fixed-point solutions for convolved mixtures,*” in 3<sup>rd</sup> International Conference on Independent Component Analysis and Blind Source Separation, San Diego, California, Dec. 2001.

[111] 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.

[112] Vincent, E., Gribonval, R., e Févotte, C. “*Performance Measurement in Blind Audio Source Separation,*” IEEE Transaction On Audio and Language Processing, Vol. 14, No. 4, Jul. 2006.

[113] Fancourt, C. L. e Parra, L. “*The coherence function in blind source separation of convolutive mixtures of non-stationary signals,*” in Proc. Int. Workshop on Neural Networks for Signal Processing (NNSP), 2001.

[114] Cardoso, J. F. “*Entropic contrasts for source separation: geometry and stability,*” In Unsupervised adaptive filtering, Volume 1, pages 139-189. S. Haykin editor, Wiley 2000.

[115] Aichner, R., Buchner, H., e Kellermann, W. “*The TRINICON framework for adaptive MIMO signal processing with focus on the generic Sylvester constraint,*” Proc. ITG Conference on Speech Communication, Aachen, Germany, Oct. 2008.

[116] Buchner, H. e Kellermann, W. “*TRINICON for Dereverberation of Speech and Audio Signals,*” In P.A. Naylor and N.D. Gaubitch (eds.), Speech Dereverberation, Springer-Verlag, London, pp. 311-385, Jul. 2010.

- [117] Figueiredo, F. A. P. e Ynoguti, C. A. “*On the improvement of the learning rate in Blind Source Separation using techniques from Artificial Neural Networks theory.*” Proceedings of the International Workshop on Telecommunications – IWT/09.
- [118] Figueiredo, F. A. P. e Ynoguti, C. A. “*Blind Source Separation in Reverberant Environment Using Genetic Algorithms,*” paper submitted to the International Workshop on Telecommunications – IWT/11.
- [119] Nam, H., Yim, T. G., Han, S., Oh, J. B. e Lee, S. “*Independent component analysis of ictal EEG in medial temporal lobe epilepsy,*” (SLOHSVLD 43:160–164, 2002.
- [120] James, C. J. e Wang, S., “*Blind source separation in single-channel EEG analysis: an application to BCI,*” In, Proceedings 28th Annual International Conference IEEE Engineering in Medicine and Biology Society (EMBS), 2006.
- [121] Papoulis, A. “*Probability, Random Variables, and Stochastic Processes,*” 3rd ed. WCB/McGraw-Hill, 1991.
- [122] Gallant, S. I. “*Neural network learning and expert systems,*” The MIT Press, 1995.
- [123] Rabiner, L. e Juang, B. H. “*Fundamentals of Speech Recognition,*” Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.