

PROPOSTAS PARA SEPARAÇÃO CEGA E SUPERVISIONADA DE FONTES

Diego Barreto Haddad

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO
DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. José Gabriel Rodríguez Carneiro Gomes, Ph.D.

Prof. Lisandro Lovisoló, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2008

HADDAD, DIEGO BARRETO

Propostas para Separação Cega e
Supervisionada de Fontes [Rio de
Janeiro] 2008

XIV, 155 p 29,7 cm (COPPE/UFRJ,
M.Sc., Engenharia Elétrica, 2008)

Dissertação - Universidade Federal do
Rio de Janeiro, COPPE

- 1.Separação de Fontes
- 2.Análise de Componentes Independentes
- 3.Banco de filtros
- 4.Análise de Componentes Esparsos

I.COPPE/UFRJ II.Título (série)

AGRADECIMENTOS

A gratidão está *demodé*. Tal fato não me faz moossa, já que não me sacrifico ao Moloch da modernidade. Repito o apócrifo verso do grego Anacreonte: “*Se tenho ou não a fronte encalvecida, não sei*” e, inspirando-me em Nelson Rodrigues, acrescentaria: *mas sou um septuagenário nato*. Minha única diferença é que não nasci de fraque como, a crermos no escritor brasileiro, vieram ao mundo Rui Barbosa e outros idosos de nascença do século XIX.

Sendo assim, cumpre retribuir, por mínima que seja a paga, o apoio de todos os meus familiares e amigos, o qual foi de excepcional valia nesta minha saga de mestrando. Foram todos de inexcedível compreensão. Apoio maior seria impossível.

Agradeço à minha orientadora, por seus incentivos, por sua compreensão ilimitada, por suas sugestões sempre arrazoadas, por seu gabarito tecnológico indiscutível.

Agradeço ao meu co-orientador (informal) Paulo B. Batalheiro, pela sua inestimável ajuda. Embora muito atarefado, sempre esteve presente.

Por fim, sou profundamente grato à minha companheira, Angélica, a qual nunca reclamou de ver-me comprometido com o mestrado, mesmo julgando-me, aqui e ali, algo exagerado na dedicação.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PROPOSTAS PARA SEPARAÇÃO CEGA E SUPERVISIONADA DE FONTES

Diego Barreto Haddad

Junho/2008

Orientador: Mariane Rembold Petraglia

Programa de Engenharia Elétrica

As técnicas que contemplam o problema de separação de fontes encontram-se em grande desenvolvimento. Neste trabalho, foram propostas modificações no algoritmo Infomax de Análise de Componentes Independentes, as quais recorrem a algum conhecimento estatístico acerca dos coeficientes da matriz de misturas. Este procedimento foi generalizado para um algoritmo de deconvolução cega em sinais de módulo constante. Foram testadas modificações em algoritmos de separação convolutiva de fontes, as quais introduzem, entre outras, taxas de aprendizagem variantes no tempo, estimativas recursivas de matrizes de correlações e estimativas de direção de chegadas das fontes. Foi proposta uma alteração em uma função custo no domínio da frequência, e verificado se outra função custo no domínio do tempo permite a escolha de filtros IIR na sua formulação direta. Em análises de componentes esparsos, propõe-se um novo algoritmo de estimativa da matriz de misturas e métodos que auxiliam na reconstrução das fontes, entre eles o recurso a banco de filtros não criticamente decimados. No caso monosensor de separação de fontes, foi proposta uma técnica de separação de fontes supervisionada que baseia-se em redes neurais e na estrutura harmônica de sinais musicais, com ganhos significativos em relação a uma técnica concorrente.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PROPOSALS FOR BLIND AND SUPERVISED SOURCE SEPARATION

Diego Barreto Haddad

June/2008

Advisor: Mariane Rembold Petraglia

Department: Electrical Engineering

The techniques addressing the source separation problem are in growing. In this work, proposed modifications in the Infomax algorithm of Independent Component Analysis, that use some statistical knowledge about the mixture matrix coefficients, were proposed. This procedure was generalized for a blind deconvolution algorithm designed for constant modulus signals. Some modifications in convolutive source separation algorithms were tested, introducing, among other features, time variant learning factors, recursive estimates of correlation matrices and direction of arrival estimates. A modification of a frequency-domain cost function was proposed, and it was analysed whether other time domain cost function permit us to choose IIR filters in direct formulation. In sparse component analysis, a new algorithm for mixture matrix estimation was proposed, as well as new methods for improving the source reconstruction step, among them the use of non-critically decimated filter banks. In the monosensor case, a new technique for supervised source separation was proposed, that uses neural networks and the harmonic structure of musical signals, with significant gains compared to another technique.

Sumário

1	Introdução	1
1.1	Configurações de Mistura	2
2	Separação de Fontes	5
2.1	Mistura Linear e Instantânea (MLI)	6
2.2	Separação de fontes no caso MLI	7
2.2.1	Separabilidade	8
2.2.2	Princípios de Separação	11
2.2.2.1	O Método Infomax sob um Paradigma Bayesiano . .	12
2.2.2.2	Empregando Conhecimento acerca da Matriz de Misturas	17
2.2.2.3	Conhecimento Imperfeito acerca das Estatísticas da Matriz de Misturas	18
2.2.2.4	Aplicação em Comunicações Digitais	20
2.2.3	Avaliação	21
2.2.4	Simulações	22
2.2.4.1	Experimento I: Infomax	22
2.2.4.2	Experimento II: Infomax com Informação Imperfeita	23
2.2.4.3	Experimento III: ext-MIMO-CM	24
2.3	Considerações Finais	26
3	Misturas Convolutivas	27
3.1	Evolução dos Métodos de Separação em Misturas Convolutivas	28
3.2	Generalização de Algoritmos de Separação Cega de Fontes para Misturas Convolutivas Utilizando Estatísticas de Segunda Ordem (GASFC)	31

3.2.1	Formulação Matricial para Misturas Convolutivas com Ex- tensão para Múltiplos Intervalos Temporais	32
3.2.2	Função Custo	35
3.2.3	Minimização da função custo	37
3.2.4	Gradiente Natural	40
3.2.5	Implementações <i>On-Line</i> , <i>Block On-Line</i> e <i>Off-Line</i>	41
3.2.6	Avaliação do Método GASFC	41
3.3	Método II: Separação Cega de Fontes Explorando as Dependências de Alta Ordem entre Raias	42
3.3.1	Avaliação do Método EDAO	45
3.4	Estimativa do Ângulo de Chegada	45
3.5	Propostas	48
3.5.1	Adoção de uma Taxa de Aprendizagem Distinta para Cada Parâmetro	48
3.5.2	Atualização Recursiva das Matrizes	54
3.5.3	Aumento do Comprimento dos Filtros de Separação	56
3.5.4	Emprego de Filtros IIR	58
3.5.5	Avaliação de Eventuais Permutações por meio da Direção de Chegada	68
3.5.6	Mudança no Método EDAO por Modificações na Função Custo	70
3.6	Considerações Finais	72
4	Misturas Indeterminadas	74
4.1	Introdução	74
4.2	Problema de Estimativa da Matriz de Misturas	76
4.2.1	Varredura de Zibulevsky	77
4.2.2	Modelos de Misturas de Exponenciais	81
4.2.3	Clusterização <i>K-means</i>	84
4.2.4	Clusterização <i>Fuzzy-C</i>	85
4.2.5	Clusterização Hierárquica	86
4.2.6	Método DUET	87
4.2.7	Extensão dos Métodos DUET e TIFROM	92
4.2.8	Método Proposto: Modelos de Misturas de Gaussianas	95

4.2.9	Simulações	97
4.3	Problema da reconstrução das fontes	101
4.3.1	Proposta I: Decomposição das fontes por meio de bancos de filtros não criticamente decimados	102
4.3.2	Proposta II: Minimização de norma l_p , $p \neq 1$	103
4.3.3	Proposta III: Introdução de um critério de continuidade	104
4.3.4	Simulações	105
4.3.4.1	Avaliação da dificuldade da etapa de reconstrução	105
4.3.4.2	Banco de filtros não criticamente decimados	109
4.3.4.3	Variação do número de fontes nulas a cada instante	115
4.3.4.4	Minimização da norma l_p ($p \neq 1$)	119
4.3.4.5	Adoção de um Critério de Continuidade	121
4.4	Considerações Finais	121
5	Separação de Fontes no Caso de Uma Única Mistura	123
5.1	O método FGMM	125
5.2	Método Proposto: FNH	128
5.2.1	Análise DFT para sinais harmônicos	129
5.2.2	Formação dos padrões	129
5.2.3	Análise com Redes Neurais	130
5.2.4	Mascaramento	130
5.3	Simulações	132
5.4	Considerações Finais	134
6	Conclusão	139

Lista de Figuras

1.1	Diferentes configurações de misturas lineares.	4
2.1	Problemas da Separação Cega de Fontes.	9
2.2	Estrutura de um algoritmo de separação cega de fontes.	12
2.3	$\log p(\mathbf{H} \mathbf{X}, \mathbf{I})$, sob hipótese de distribuição uniforme dos coeficientes da matriz de mistura.	19
2.4	$\log p(\mathbf{H} \mathbf{X}, \mathbf{I})$, sob hipótese de distribuição exponencial dos coeficientes da matriz de mistura.	20
2.5	Trajetória dos coeficientes $w_{12} \times w_{11}$ das matrizes de separação (colunas normalizadas), com 2000 amostras das misturas disponíveis. A trajetória contínua é do método Infomax padrão e a tracejada do método modificado.	24
2.6	Melhora no SINR (em dB) em função do número de amostras dos métodos ext-MIMO-CM e ext-MIMO-CM modificado. Resultado com 1500 amostras oriundo de média Monte Carlo com 500 simulações.	25
2.7	Melhora no SINR (em dB) em função do número de amostras (sem ruído) dos métodos ext-MIMO-CM e ext-MIMO-CM modificado. Resultado oriundo de média Monte Carlo com 1000 simulações.	25
3.1	Ilustração das posições do sensor j e da fonte i , com a origem no ponto \mathbf{O}	46
3.2	Ilustração da suposição de paralelismo do modelo de campo distante.	47
3.3	SIR médio (em dB) dos métodos GASFC (original) e GASF-VM-I (proposto), para $\alpha = 1,01$, $L = 6$ e $\mu = 3 \times 10^{-4}$	50
3.4	SIR médio (em dB) dos métodos GASF (original) e GASF-VM-I (proposto), para $\alpha = 1,005$, $L = 6$ e $\mu = 7 \times 10^{-4}$	50

3.5	Evolução dos parâmetros $\mu_{pq}(k)$, em função do tempo t (em segundos). (a) $\mu_{11}(0)$; (b) $\mu_{12}(1)$; (c) $\mu_{21}(2)$ e (d) $\mu_{22}(3)$	51
3.6	SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-II (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $L = 6$ e $C = 8$	52
3.7	SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-II (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $L = 6$ e $C = 2$	53
3.8	SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-III (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $C_0 = 10$, $L = 6$ e $\lambda = 0,01$	54
3.9	SIR médio (em dB) dos métodos GASFC (original) e GASFC _{rec} (proposto), para $\lambda = 0,95$, $\mu = 3 \times 10^{-4}$ e $L = 5$	55
3.10	SIR médio (em dB) dos métodos GASFC (original) e GASFC _{rec} (proposto), para $\lambda = 0,99$, $\mu = 3 \times 10^{-4}$ e $L = 5$	56
3.11	SIR máximo obtido para diferentes comprimentos L dos filtros de separação.	57
3.12	Evolução do SIR para diferentes comprimentos L dos filtros de separação.	58
3.13	Configuração de mistura e os filtros de separação obtidos pelo algoritmo GASFC.	59
3.14	Evolução do SIR utilizando algoritmo GASFC no formato <i>block online</i> , $\mu = 10^{-4}$, $L = 2$ e número de blocos igual a 50.	60
3.15	Realização na forma direta da estrutura IIR utilizada, utilizando entrada $x(n)$ com saída $y(n)$	60
3.16	SIR em função dos parâmetros a_{11} e b_{11}	63
3.17	Função custo em função dos parâmetros a_{11} e b_{11} , com $N = 4$	64
3.18	Função custo em função dos parâmetros a_{11} e b_{11} , com $N = 14$	64
3.19	SIR em função dos parâmetros a_{12} e b_{12}	65
3.20	Função custo em função dos parâmetros a_{12} e b_{12} , com $N = 6$	65
3.21	Função custo em função dos parâmetros a_{12} e b_{12} , com $N = 16$	66
3.22	SIR em função dos parâmetros a_{21} e b_{21}	66
3.23	Função custo em função dos parâmetros a_{21} e b_{21} , com $N = 12$	67
3.24	SIR em função dos parâmetros a_{22} e b_{22}	67
3.25	Função custo em função dos parâmetros a_{22} e b_{22} , com $N = 10$	68

3.26	Estimativa da direção de chegada de cada fonte em função da frequência.	69
3.27	Estimativa da direção de chegada de cada fonte em função da frequência.	70
3.28	Razão sinal interferência em função do parâmetro D .	72
4.1	Exemplo de fontes maximamente esparsas: (a) fonte 1; (b) mistura 1; (c) fonte 2; (d) mistura 2; (e) fonte 3; (f) gráfico de $x_2(n) \times x_1(n)$.	78
4.2	Gráfico de $x_2(n) \times x_1(n)$; à esquerda, coeficientes numa transformada <i>wavelet packet</i> com 7 níveis de decomposição e à direita, amostras no domínio do tempo.	79
4.3	Gráfico de $x_3(n) \times x_2(n) \times x_1(n)$, amostras no domínio do tempo.	80
4.4	Gráfico de $\tilde{x}_3(n) \times \tilde{x}_2(n) \times \tilde{x}_1(n)$; coeficientes numa transformada <i>wavelet packet</i> com 7 níveis de decomposição.	80
4.5	Função potencial típica, com 3 fontes e $\lambda = 5$.	82
4.6	Direções geradas pelas matrizes \mathbf{H}_i nos três casos estudados.	98
4.7	Medida logerr em função do número de misturas para $N = 7, 8, 9$ e 10 .	108
4.8	Medida logerr em função do número de misturas para $N = 11, 12, 13$ e 14 .	108
4.9	Medida logerr em função do número de misturas para o caso $M \times M + 1$.	109
4.10	Comparação dos erros de reconstrução dos métodos caminho mais curto e <i>melhor caso</i> .	118
4.11	Distribuição de $\hat{s}_{act}(n)$ via <i>melhor caso</i> com diferentes números de fontes. (a) $N = 7$, (b) $N = 8$, (c) $N = 9$ e (d) $N = 10$.	118
4.12	Distribuição de $\hat{s}_{act}(n)$ via <i>melhor caso</i> com diferentes números de fontes. (a) $N = 11$, (b) $N = 12$, (c) $N = 13$ e (d) $N = 14$.	119
4.13	Erro de reconstrução em função da minimização da norma p .	120
4.14	Erro de reconstrução em função do parâmetro α .	120
5.1	Exemplo de mascaramento via redes neurais e sua interpolação.	131
5.2	Fontes originais. (a) bateria, (b) piano.	134
5.3	Fontes estimadas pelo método FGMM, com 4 gaussianas e janela de comprimento 512. (a) bateria, (b) piano.	135
5.4	Fontes estimadas pelo método FGMM, com 16 gaussianas e janela de comprimento 512. (a) bateria, (b) piano.	135

5.5	Fontes estimadas pelo método FGMM, com 4 gaussianas e janela de comprimento 1024. (a) bateria, (b) piano.	136
5.6	Fontes estimadas pelo método FGMM, com 8 gaussianas e janela de comprimento 1024. (a) bateria, (b) piano.	136
5.7	Fontes estimadas pelo método FNH, com 3 neurônios na camada escondida e janela de comprimento 1024. (a) bateria, (b) piano.	137
5.8	Padrão do mascaramento no espaço tempo \times frequência em um pequeno trecho de sinal.	137

Lista de Tabelas

2.1	Denominação dos termos que constam no teorema de Bayes acima formulado.	13
2.2	SIR final - SIR inicial (em dB) para o método Infomax, Infomax modificado (com informação estatística acerca da matriz de mistura) e a diferença entre os resultados do método modificado e do padrão, para diferentes comprimentos das misturas.	23
2.3	SIR final - SIR inicial para o Infomax modificado, variando o parâmetro μ' da distribuição <i>a priori</i> aproximada e utilizando 2000 amostras. O parâmetro D indica a divergência de Kullback Leibler entre as distribuições real e aproximada.	26
3.1	Propriedades exploradas e suas respectivas influências na definição da função custo.	31
3.2	Propriedades exploradas e suas respectivas influências na definição da função custo.	34
3.3	Matrizes utilizadas e suas respectivas dimensões.	35
3.4	SIR obtido com diferentes parâmetros N e para filtros de separação distintos, admitindo o menor valor da função custo do método GASFC.	62
4.1	Valores reais e estimativas oriundas de métodos distintos para os três casos abordados.	99
4.2	Erros quadráticos médios oriundos de métodos distintos para os três casos abordados.	100
4.3	Erros quadráticos médios globais dos quatro métodos que foram bem-sucedidos em todos os casos.	101
4.4	A medida logerr para diferentes números de fontes e de misturas.	107

4.5	Resultados finais para sinais de voz com wavelet bior6.8. L é o número de níveis de decomposição, e D é o fator de decimação.	111
4.6	Resultados finais para sinais de voz com wavelet coif5. L é o número de níveis de decomposição, e D é o fator de decimação.	112
4.7	Resultados finais para sinais de voz com wavelet db16. L é o número de níveis de decomposição, e D é o fator de decimação.	113
4.8	Resultados finais para sinais de voz com wavelet sym9. L é o número de níveis de decomposição, e D é o fator de decimação.	114
4.9	Melhora obtida nas medidas SAR e SDR quando não se recorre à decimação. Foi utilizada uma decomposição de 7 níveis.	115
4.10	Melhora via <i>melhor caso</i> no erro quadrático médio de reconstrução para o caso $N - 1 \times N$	117
4.11	Distribuição de $\hat{s}_{act}(n)$ para diferentes números de fontes.	117
5.1	Medidas SIR e SAR obtidas pelos métodos FGMM e FNH.	133

Capítulo 1

Introdução

Os diversos tipos de sensores, ao capturarem sinais, deles nos fornecem uma versão em geral corrompida por distorções, ruídos, atenuações e interferências. A presença de interferências implica termos acesso a misturas de diferentes fontes. Cabe ressaltar que estas misturas podem em alguns casos ser intencionais como, por exemplo, em gravações de músicas.

Em análise de sinais, cumpre por vezes recuperar cada uma das fontes a partir das misturas supracitadas. Por exemplo, no caso de ondas cerebrais [1], se desejaria ter acesso a cada uma das fontes de impulsos elétricos. O mesmo ocorreria no caso de ondas sonoras, em aplicações como transcrição musical ou reconhecimento de fala [2].

Este problema é denominado “separação de fontes”, e na literatura a ele se associou a célebre expressão *cocktail party*, a qual faz referência à habilidade humana de discernir, em uma festa, os sons desejados (como os de uma conversa) do barulho ambiente. A facilidade com que nos desincumbimos desta tarefa faz-nos freqüentemente subestimar sua complexidade.

As técnicas de separação de fontes são importantes na medida em que estas apresentam inúmeras aplicações, além das já citadas, entre as quais: áudio [3], sensoriamento remoto [4], análise de sinais sísmicos [5], processamento de imagens [6] e em comunicações digitais [7]. Estudadas intensivamente desde meados da década de 90, as técnicas de Análise de Componentes Independentes têm sido largamente empregadas em BSS (do inglês *Blind Source Separation* - Separação Cega de Fontes; o adjetivo “cega” será explicado mais à frente), de tal forma que alguns chegam a

usar os termos ICA e BSS de forma intercambiável. Entretanto, a abordagem via ICA não contempla os casos, na prática comuns, onde número de misturas é superado pelo número de fontes, além de requerer adaptações em configurações convolutivas. Para tratar destas outras configurações, surgem como ferramentas primordiais (entre outras) a Análise de Componentes Esparsos e as Técnicas de Separação no Domínio da Frequência.

1.1 Configurações de Mistura

Em geral, podemos classificar os tipos de misturas em duas grandes classes: lineares e não lineares. O grupo das misturas lineares é o contemplado nesta dissertação, podendo ser subdividido em dois casos: misturas instantâneas e misturas convolutivas.

Seja N o número de fontes, M o número de sensores e $s_i(n)$ a n -ésima amostra da i -ésima fonte. O vetor $\mathbf{s}(n)$ é formado através da concatenação de todas as fontes no instante n , de forma que $\mathbf{s}(n) = [s_1(n) \ s_2(n) \ \dots \ s_N(n)]^T$, com $(\cdot)^T$ significando transposição. Considerando $x_j(n)$ a n -ésima amostra da j -ésima mistura, podemos definir o vetor $\mathbf{x}(n)$ de forma semelhante: $\mathbf{x}(n) = [x_1(n) \ x_2(n) \ \dots \ x_M(n)]^T$. A relação entre os vetores $\mathbf{x}(n)$ e $\mathbf{s}(n)$ é dada por:

$$\mathbf{x}(n) = \mathbf{H} * \mathbf{s}(n), \quad (1.1)$$

onde \mathbf{H} (oriunda do sistema de mistura) é denominada *matriz de mistura* de dimensões $M \times N$ e $*$ significa convolução. Considerando-se sistemas causais, os elementos de \mathbf{H} podem ser expressos na forma $\sum_{q=0}^{\infty} h_{ij}(q)z^{-q}$ (no domínio da transformada- z); na eventualidade de o sistema de mistura apresentar filtros com resposta ao impulso finita (FIR, do inglês *Finite Impulse Response*) e de comprimento máximo Q , os elementos de \mathbf{H} podem ser expressos por $\sum_{q=0}^{Q-1} h_{ij}(q)z^{-q}$, caso conhecido como convolutivo. Esta formulação possui dois casos particulares muito importantes: o caso $Q = 1$ (ou caso *instantâneo*, onde os filtros se reduzem a meras constantes) e o caso de cada elemento assumir a forma $h_{ij}(q_{ij})z^{-q_{ij}}$, com pelo menos um q_{ij} positivo (caso denominado *anecóico*).

Em sinais de comunicações digitais, os filtros h_{ij} correspondem aos múltiplos caminhos percorridos pelo sinal até sua chegada aos sensores. No caso de sinais

de áudio, estes filtros refletem a reverberação do ambiente, consistindo, no caso de câmaras anecóicas, em versões diferentemente atrasadas e atenuadas dos sinais em cada sensor (daí o nome do caso anecóico).

O grau de dificuldade (quanto à separação) das misturas lineares é oriundo de quatro fatores. A independência deles motiva-nos a considerá-los quais eixos de dificuldade. O primeiro eixo de dificuldade reflete o número de coeficientes do sistema de mistura. Este eixo contempla, por ordem crescente de dificuldade, as configurações de misturas instantâneas, as ditas anecóicas e, por fim, as misturas convolutivas (caso mais geral).

O segundo eixo de dificuldade é o da relação entre o número de misturas e o número de fontes. O aumento do número de misturas implica uma maior facilidade para as técnicas de separação, já que resulta em aumento da informação disponível. Como cada sensor encontra-se numa localização distinta da dos demais, as contribuições de cada fonte diferem nos diversos sensores; daí dizermos que um maior número de misturas aumenta a *diversidade espacial* a que temos acesso. É comum a divisão das misturas quanto a este eixo em três tipos, a saber: misturas superdeterminadas, determinadas e indeterminadas, onde o número M de misturas, respectivamente, supera, iguala e é superado pelo número N de fontes. Optamos por destacar das misturas indeterminadas o caso no qual temos acesso a apenas uma única mistura (caso mono-sensor), devido ao fato de este caso exigir, em geral, técnicas bastante distintas das aplicáveis às misturas indeterminadas.

O terceiro fator a influenciar a dificuldade das técnicas de separação reside no conhecimento que temos a respeito das fontes e da matriz de misturas. As técnicas de separação de fontes que não requerem nenhum (ou quase nenhum) conhecimento deste tipo são classificadas como *cegas*.

O quarto eixo que contribui para o grau de dificuldade do sistema de separação consiste no número de amostras das misturas a que temos acesso. O aumento deste número implica um aumento de informação, facilitando a estimativa dos parâmetros do sistema de separação.

As diferentes configurações para o caso de misturas lineares são mostradas na Figura 1.1, onde são evidenciados os dois primeiros eixos de dificuldade citados.

Nesta dissertação, abordamos problemas de separação de fontes em con-

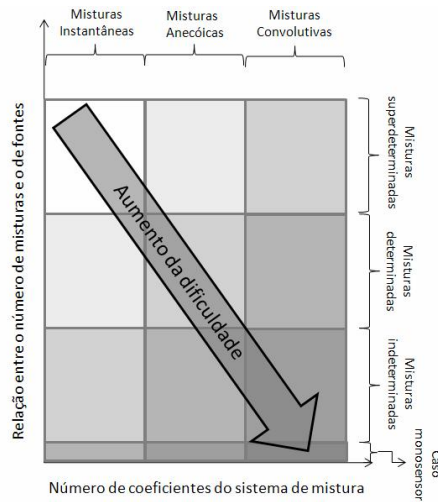


Figura 1.1: Diferentes configurações de misturas lineares.

figurações instantâneas e convolutivas.

No Capítulo 2, contemplamos o caso instantâneo e determinado, verificando se algum conhecimento estatístico, ainda que inexato, acerca da matriz de misturas pode nos auxiliar na separação.

No Capítulo 3, o caso tratado é o convolutivo e determinado. Foram testadas técnicas de separação tanto no domínio do tempo quanto no da frequência. Propostas de aceleração de convergência, de alteração da função custo e de utilização de filtros com resposta ao impulso infinita são discutidas.

No Capítulo 4, as configurações instantâneas e indeterminadas (com pelo menos duas misturas) são contempladas. São feitas diversas propostas para estas estirpes de mistura, entre as quais: um método para estimativa da matriz de misturas, a influência do *aliasing* oriundo de bancos de filtros decimados e a inclusão de critérios de continuidade e de minimização de norma l_p , com $p \neq 1$.

No Capítulo 5 tratamos de configurações de misturas onde temos acesso a apenas um único sensor. Propomos neste capítulo uma abordagem fundamentada em redes neurais, a qual atenta para a estrutura harmônica de sinais musicais. Este capítulo precede o último, o qual trata das conclusões da dissertação.

Capítulo 2

Separação de Fontes

O paradigma da “separação de fontes” possui outras aplicações, além das já citadas na Introdução. Por exemplo, ele é também solicitado como etapa de pré-processamento de obtenção de dados não contaminados para análise posterior [8]. Como vimos na Introdução, quando não possuímos nenhum¹ conhecimento *a priori* acerca das fontes, tampouco acerca do sistema de mistura, classificamos como “cega” a separação. A popularidade destas técnicas deve-se não somente ao seu vasto campo de aplicabilidade, mas também porque estas técnicas abordam problemas interessantes em estatísticas de mais alta ordem e sistemas não-lineares [9].

As formas de misturas podem, obviamente, variar de infinitas formas; embora inicialmente restritos às misturas lineares e instantâneas (configuração anteriormente considerada “aparentemente impossível” [10]), os modelos atuais já contemplam as configurações (mesmo as com duração de poucos segundos) convolutivas em ambientes reais, onde as misturas são efetuadas por filtros FIR dotados de um número de coeficientes da ordem de milhares (no caso de ondas acústicas) [11],[12]. Em geral, podemos classificar os tipos de misturas em duas grandes classes: lineares e não lineares, sendo aquelas costumeiramente modeladas de duas formas: misturas lineares instantâneas e misturas lineares convolutivas.

Esta dissertação explora métodos de separação de misturas lineares, cuja configuração mais simples é a denominada *mistura instantânea*, a qual introduz uma

¹Na realidade, nenhum método de separação de fontes é completamente “cego”, pois sempre se parte de algumas hipóteses suficientemente gerais, como as de independência, esparsidade, diferenças morfológicas, distribuição dos coeficientes da matriz de misturas, entre outros.

série de conceitos muito úteis para os casos mais gerais.

Neste capítulo há, primeiramente, uma descrição matemática do problema, seguida do tópico que trata da separabilidade. Então apresentamos, dentro de um paradigma bayesiano, um método clássico (Infomax) dedicado à separação cega de fontes. O paradigma utilizado permitirá avaliar a melhora de desempenho do método na eventualidade de possuímos algum conhecimento estatístico, mesmo que inexato, acerca da matriz de misturas. Estas modificações nos inspiraram a modificar um método de separação dedicado à configuração de fontes de módulo constante.

2.1 Mistura Linear e Instantânea (MLI)

Sejam N fontes² $s_i(n)$ ($i = 1, \dots, N$). O vetor $\mathbf{s}(n)$ é definido pelas fontes s_i no instante n :

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_N(n) \end{bmatrix}. \quad (2.1)$$

Seja \mathbf{H} uma matriz (de dimensões $M \times N$) denominada matriz de mistura. Considerando a mistura linear, instantânea e não ruidosa, podemos expressar o vetor $\mathbf{x}(n)$ que contém as n -ésimas amostras das M misturas pela equação:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n). \quad (2.2)$$

Concatenando os vetores $\mathbf{s}(n)$ e $\mathbf{x}(n)$, obtemos, respectivamente, a matriz das fontes \mathbf{S} e a matriz das misturas \mathbf{X} (não confundir com a matriz *de mistura* \mathbf{H}), da forma a seguir:

$$\mathbf{S} = \begin{bmatrix} s_1(1) & s_1(2) & \dots & s_1(K) \\ s_2(1) & s_2(2) & \dots & s_2(K) \\ \vdots & \vdots & \vdots & \vdots \\ s_N(1) & s_N(2) & \dots & s_N(K) \end{bmatrix}, \quad (2.3)$$

²Algumas destas definições já constam na Introdução; a redundância deve-se a razões de clareza, já que na Introdução foi apresentada uma visão mais genérica do problema de separação de fontes.

$$\mathbf{X} = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(K) \\ x_2(1) & x_2(2) & \dots & x_2(K) \\ \vdots & \vdots & \vdots & \vdots \\ x_M(1) & x_M(2) & \dots & x_M(K) \end{bmatrix}, \quad (2.4)$$

onde K é o número de amostras de cada fonte. Cada mistura (linha da matriz das misturas \mathbf{X}) é uma combinação linear das fontes (linhas da matriz das fontes \mathbf{S}).

2.2 Separação de fontes no caso MLI

A abordagem mais comum para separação de fontes no caso MLI consiste em aplicar uma matriz \mathbf{W} de separação à matriz das misturas, efetuando o produto \mathbf{WH} . Esta é a forma de separação mais tradicional, e a analisada neste capítulo. No Capítulo 4, contemplaremos outros métodos de separação, mais recentes, baseados especialmente em esparsidade, particularmente úteis na possibilidade de haver menos sensores do que fontes (sob a restrição $2 \leq M < N$).

Efetuando o produto matricial da matriz \mathbf{W} com a matriz das misturas, obtemos a estimativa \mathbf{Y} das fontes, da forma a seguir:

$$\mathbf{Y} = \mathbf{WX}, \quad (2.5)$$

ou, reescrevendo-a por recurso à Equação (2.2):

$$\mathbf{Y} = \mathbf{WHS}. \quad (2.6)$$

A equação acima já permite uma conclusão acerca da *separabilidade* (ou seja, a existência de uma matriz \mathbf{W} a qual permite separar os sinais das fontes): a matriz de misturas \mathbf{H} deve possuir posto igual a N . Quando tal condição é satisfeita, dizemos que o conjunto das misturas é *separável*³ (pois sempre é possível encontrar uma matriz \mathbf{W} tal que o produto \mathbf{WH} seja a identidade, tornando ao menos possível, por meio de uma escolha adequada, obter em \mathbf{Y} a matriz \mathbf{S}); uma condição necessária

³É possível, a depender da matriz de mistura, recuperar *qualquer uma* das fontes, em separado, mesmo no caso de M ser menor do que N . Porém, esta abordagem exige matrizes de mistura muito específicas e mais do que uma única matriz de separação; ou seja, as fontes não são obtidas de forma simultânea [10].

para tal é termos $M \geq N$. Daí decorre que a separabilidade independe da matriz das fontes \mathbf{S} .

Mesmo sendo separáveis as misturas, encontrar a matriz \mathbf{W} não é algo trivial, dado que \mathbf{H} não é conhecida de forma explícita. A tarefa de encontrar uma estimativa $\widehat{\mathbf{W}}$ da matriz \mathbf{W} (ou mesmo $\widehat{\mathbf{H}}$, fazendo posteriormente $\widehat{\mathbf{W}} = \widehat{\mathbf{H}}^{-1}$) implica o recurso a certos princípios, denominados *princípios de separação*. Seguindo [10], analisaremos em separado as questões da separabilidade e dos princípios de separação.

2.2.1 Separabilidade

Como vimos, a separabilidade das fontes é uma propriedade intrínseca das combinações lineares nos sinais medidos (misturas) [10], ou seja, é uma propriedade inerente à matriz \mathbf{H} .

Seja \mathbf{C} a matriz dada pelo produto \mathbf{WH} . Não tendo controle total de \mathbf{C} (pois só temos controle parcial sobre \mathbf{H} ⁴), basta-nos, para propósitos de separabilidade, que \mathbf{H} tenha posto N . Idealmente, neste caso, deveríamos ter:

$$\mathbf{C} = \mathbf{I}, \tag{2.7}$$

onde \mathbf{I} é a matriz identidade. Porém, mesmo quando o conjunto de misturas é separável, não é sensato exigir a igualdade expressa na Equação (2.7); a falta de informação tanto sobre as fontes quanto sobre a matriz de misturas se revela muito restritiva para este propósito, como veremos a seguir.

Primeiramente, é fácil observar que multiplicar a coluna i da matriz \mathbf{H} por uma constante α possui o mesmo efeito para a matriz das misturas \mathbf{X} do que multiplicar a i -ésima fonte por essa mesma constante. Acrescentando a isso nossa ausência de informações, nota-se uma *ambigüidade de escalamento* impossível de ser contornada. Este fato impõe-nos ser menos exigentes; em outras palavras, consideraremos a separação bem-sucedida não somente quando \mathbf{C} for a identidade, mas uma matriz diagonal.

Em segundo lugar, a falta de qualquer informação *a priori* sobre as fontes não nos permite determinar a ordenação destas; ou seja, tendo acesso à matriz \mathbf{S} com

⁴Alterar a localização dos medidores implica modificar a matriz \mathbf{H} .

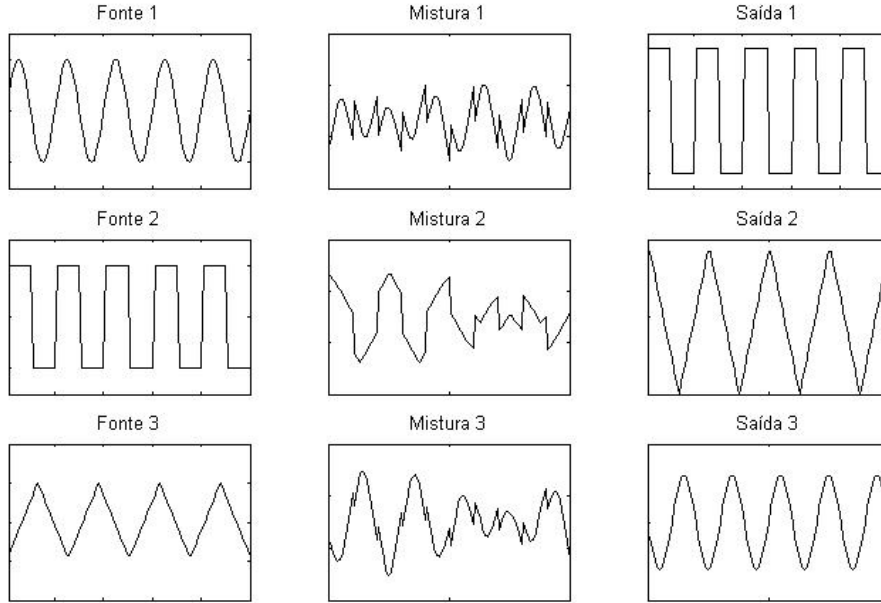


Figura 2.1: Problemas da Separação Cega de Fontes.

as linhas numa permutação aleatória, não poderíamos desfazer a permutação. Isto significa que não podemos esperar que sempre a i -ésima linha de \mathbf{Y} corresponderá à i -ésima linha de \mathbf{S} (ou seja a i -ésima fonte) escalada. Eis qualificado o problema da *permutação*.

A Figura 2.1 ilustra os problemas de ambigüidade de escalamento e de permutação inerentes à separação cega de fontes. Uma interpretação concisa destes problemas, utilizando o argumento (levemente alterado) de [13], pode ser efetuada da forma como se segue. Seja $\mathbf{\Lambda}$ uma matriz diagonal não-degenerada e \mathbf{P} uma matriz de permutação. Assim,

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) = \mathbf{H}\mathbf{P}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{s}(n) = \mathbf{H}\mathbf{P}\mathbf{\Lambda}\tilde{\mathbf{s}}(n), \quad (2.8)$$

onde $\tilde{\mathbf{s}}(n) = \mathbf{\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{s}(n)$ possui as mesmas propriedades estatísticas de $\mathbf{s}(n)$ (a menos da permutação e do escalamento), não podendo ser distingüido de $\mathbf{s}(n)$ sem mais informações *a priori*.

Para o caso $N = M = 3$, portanto, a solução

$$\mathbf{C} \approx \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & b \\ c & 0 & 0 \end{bmatrix}, \quad (2.9)$$

com $abc \neq 0$ é considerada tão boa quanto $\mathbf{C} \approx \mathbf{I}$ (aqui não empregamos o sinal de igualdade que na prática não costuma ocorrer). Importa notar que o problema da permutação pode tornar a matriz \mathbf{C} não-diagonal, enquanto que o problema do escalamento torna as constantes a, b e c distintas da unidade. Podemos caracterizar de forma rigorosa a estrutura comum a todas as matrizes \mathbf{C} que separam as fontes ⁵, a menos dos problemas da permutação e do escalamento, baseados em uma análise estreitamente semelhante à efetuada em [10].

Seja a partição do conjunto de N números inteiros $A = \{1, 2, \dots, N\}$ em M (com $M \leq N$) subconjuntos disjuntos $A_i, i = 1, 2, \dots, M$ (ou seja, $A_i \cap A_j = \emptyset, i \neq j$ e $\bigcup_{i=1}^M A_i = A$). Uma matriz \mathbf{P} de dimensões $M \times N$ é chamada uma *matriz de partição* se seu elemento p_{ij} obedece à seguinte regra:

$$p_{ij} = \begin{cases} 1, & j \in A_i \\ 0, & j \notin A_i \end{cases} \quad (2.10)$$

Dados M e N , o número de matrizes de partição é dado por:

$$\sum_{n_1 + \dots + n_M = N} \frac{n!}{n_1! n_2! \dots n_M!}, \quad (2.11)$$

o que iguala o número de escolhas para associar n_i inteiros entre $1, 2, \dots, n$ para o conjunto $A_i, i = 1, 2, \dots, m$. A matriz \mathbf{P} possui posto completo quando cada coluna possui um elemento não-nulo e quando cada linha possui ao menos um elemento não-nulo. Neste caso, podemos denominá-la *matriz de permutação*.

Uma matriz de dimensões $M \times N$ é denominada *matriz de partição generalizada* quando pode ser escrita qual um produto entre uma matriz de partição de dimensões $M \times N$ por uma matriz diagonal não singular Λ , de dimensões $N \times N$.

Esta formulação nos permite então afirmar que uma matriz \mathbf{C} , dada por \mathbf{WH} , separa as fontes caso **(1)** tenha posto completo e **(2)** seja uma matriz de partição generalizada. Assim, podemos reescrever a Equação (2.9) como um produto da forma a seguir:

⁵Na realidade, quando há menos misturas do que fontes, efetuar o produto de uma matriz \mathbf{W} pela matriz das misturas nunca permitirá a separação de todas as fontes *simultaneamente*, havendo ao menos a aglutinação de duas fontes em uma das saídas.

$$\mathbf{C} \approx \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & b \\ c & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} c & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}. \quad (2.12)$$

Esta caracterização é, obviamente, geral, não se restringindo ao caso 3×3 . Em resumo, se \mathbf{P} é uma matriz de partição generalizada, a solução \mathbf{PAS} é tão aceitável quanto \mathbf{S} , para os propósitos de separação.

2.2.2 Princípios de Separação

Supondo que exista uma matriz de separação \mathbf{W} que separe as fontes, cumpre recorrer a alguns princípios para determiná-la. Estes princípios são denominados “Princípios de Separação”.

Uma técnica amplamente disseminada para separação de fontes é a Análise de Componentes Independentes (ICA, do inglês *Independent Component Analysis*). Sua justificação deriva de uma afirmação, adaptada de [14], após um teorema de Darmois [10]:

“Para um vetor \mathbf{s} de entradas independentes com no máximo uma entrada gaussiana e para qualquer matriz inversível \mathbf{C} , se as componentes de $\mathbf{y} = \mathbf{C}\mathbf{s}$ são independentes, então \mathbf{y} é uma cópia de \mathbf{s} ” (traduzido de [15]).

As técnicas de separação de fontes baseadas em ICA comumente atuam por meio da maximização (ou minimização) de uma função custo, a qual é uma medida escalar de alguma propriedade da distribuição da saída \mathbf{Y} . Esta função pode ser a informação mútua [17] ou a *kurtosis* [18] (a qual emprega momentos de ordem superior), para citar dois exemplos comuns. A otimização costuma ser iterativa, sendo a atualização dos coeficientes obtida geralmente mediante métodos do tipo gradiente (descendente ou ascendente). A Figura 2.2 mostra a estrutura genérica de um método de separação cega de fontes.

Este capítulo aborda o método Infomax [17] (o mais tradicional), porém de uma forma algo diferente da mais disseminada. Seguindo de forma bastante semelhante o desenvolvimento em [19], inserimos este método num paradigma bayesiano. Esta inserção permite-nos generalizá-lo para situações onde há informações acerca

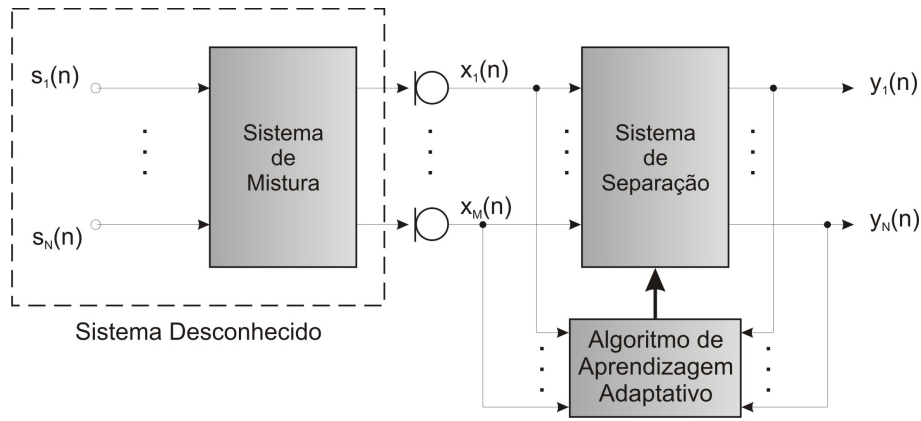


Figura 2.2: Estrutura de um algoritmo de separação cega de fontes.

dos coeficientes da matriz de mistura.

2.2.2.1 O Método Infomax sob um Paradigma Bayesiano

É comum afirmar-se que um método de separação “cega” é desprovido de quaisquer conhecimentos acerca das fontes, bem como de quaisquer informações a respeito da matriz de misturas. Eis uma interpretação algo equivocada, na medida em que todos estes métodos utilizam hipóteses mínimas que, exatamente por serem fracas, aplicam-se a uma grande variedade de configurações. Entre estas hipóteses, destacamos duas muito comuns: distribuição não-gaussiana (freqüentemente super-gaussiana⁶) das fontes e distribuição uniforme dos coeficientes da matriz de mistura.

A importância da primeira hipótese já foi devidamente ressaltada, sendo um elemento-chave na formulação dos métodos de separação (por exemplo, na maximização da *kurtosis* na família de métodos *FastICA* [18]). Já a segunda hipótese (distribuição uniforme da matriz de mistura) é bastante útil na maioria dos casos, pois em geral não possuímos um conhecimento estatístico acerca da matriz de mistura.

Enquanto a primeira hipótese é de há muito conhecida (sendo claramente explicitada no método Infomax [17], por exemplo), a segunda só recebeu tratamento adequado em [19]. Porém, mesmo em [19] (e no subsequente tutorial do mesmo autor [20]), não há simulações que verifiquem quantitativamente a contribuição oriunda de um conhecimento estatístico acerca da matriz de mistura. Neste capítulo,

⁶Uma definição qualitativa de “distribuição super-gaussiana” é a de que esta é uma distribuição mais “concentrada” do que uma distribuição gaussiana.

avaliamos quantitativamente esta melhora no caso específico de coeficientes da matriz de misturas amostrados segundo uma distribuição exponencial.

Quando possuímos um conhecimento específico *a priori* acerca da configuração da mistura, uma pergunta natural é se este conhecimento pode ser útil na melhoria de desempenho de um método de separação cega de fontes. A partir do momento em que este conhecimento é empregado, podemos intitular a separação como “informada” ou “semi-cega” (em contraposição à separação “cega”).

Central na metodologia bayesiana, podemos expressar o teorema de Bayes na seguinte forma:

$$p(\text{modelo}|\text{dados},I) = p(\text{modelo}|I)\frac{p(\text{dados}|\text{modelo},I)}{p(\text{dados}|I)}. \quad (2.13)$$

A Tabela 2.1 apresenta as denominações dos termos que constam na Equação (2.13).

A probabilidade *a posteriori* é a probabilidade de um modelo específico

Tabela 2.1: Denominação dos termos que constam no teorema de Bayes acima formulado.

Termo	Denominação
I	informação <i>a priori</i>
$p(\text{modelo} \text{dados},I)$	probabilidade <i>a posteriori</i>
$p(\text{modelo} I)$	probabilidade <i>a priori</i>
$p(\text{dados} \text{modelo},I)$	verossimilhança
$p(\text{dados} I)$	evidência

descrever de forma acurada o problema, a partir dos dados (misturas) e de nossa informação *a priori*. A probabilidade *a priori* reflete o grau em que acreditamos que um modelo específico descreve corretamente o problema, antes de termos acesso a qualquer tipo de dados - codificando portanto nosso conhecimento acerca dos prováveis valores dos parâmetros do modelo (admitindo um modelo parametrizado). A verossimilhança exhibe o grau em que acreditamos que um modelo produziu os dados observados (medida que envolve tanto o processo de efetuar predições com nosso modelo hipotético quanto o processo de comparar estas predições com os dados a que temos acesso). Para o nosso paradigma, a evidência, dado que é fixa num determinado problema, atua apenas como um fator de normalização e, portanto,

não é objeto de atenção especial.

O teorema de Bayes transforma o problema de separação de fontes num problema de busca, onde procuramos encontrar o modelo “mais provável” para explicar os dados que possuímos. Apliquemos agora a metodologia bayesiana a um modelo de mistura instantânea de fontes.

Admitindo um modelo de propagação instantânea, N fontes e N sensores, podemos descrever a i -ésima mistura (sinal obtido por meio do i -ésimo sensor) como:

$$x_i(n) = \sum_{j=1}^N h_{ij}s_j(n), \quad (2.14)$$

onde $s_j(n)$ é o valor da j -ésima fonte no instante n e h_{ij} são os elementos que formam uma matriz de mistura \mathbf{H} de dimensões $N \times N$, podendo ser interpretados como os coeficientes de combinações lineares envolvendo as fontes.

Aplicando o teorema de Bayes para obter a probabilidade do nosso modelo (expresso nas matrizes \mathbf{H} e \mathbf{S}), a partir do momento em que possuímos os dados das misturas (matriz \mathbf{X}), temos:

$$p(\mathbf{H}, \mathbf{S}|\mathbf{X}, I) = p(\mathbf{H}, \mathbf{S}|I) \frac{p(\mathbf{X}|\mathbf{H}, \mathbf{S}, I)}{p(\mathbf{X}|I)}. \quad (2.15)$$

Assumindo que os sinais das fontes independem da propagação (cujas perdas refletem-se na matriz \mathbf{H}), podemos fatorar $p(\mathbf{H}, \mathbf{S}|I)$, obtendo:

$$p(\mathbf{H}, \mathbf{S}|\mathbf{X}, I) = p(\mathbf{H}|I)p(\mathbf{S}|I) \frac{p(\mathbf{X}|\mathbf{H}, \mathbf{S}, I)}{p(\mathbf{X}|I)}. \quad (2.16)$$

Já que a evidência independe dos parâmetros de nosso modelo, podemos simplificar a busca dos valores de \mathbf{H} e de \mathbf{S} admitindo uma proporcionalidade:

$$p(\mathbf{H}, \mathbf{S}|\mathbf{X}, I) \propto p(\mathbf{H}|I)p(\mathbf{S}|I)p(\mathbf{X}|\mathbf{H}, \mathbf{S}, I). \quad (2.17)$$

Assumindo que o modelo de mistura é exato e que não há ruído, temos:

$$p(x_i(n)|\mathbf{H}, \mathbf{s}(n), I) = \delta \left(x_i(n) - \sum_{j=1}^N h_{ij}s_j(n) \right), \quad (2.18)$$

de onde obtemos:

$$p(\mathbf{X}|\mathbf{H}, \mathbf{S}, I) = \prod_{i=1}^N \prod_{n=1}^K \delta \left(x_i(n) - \sum_{j=1}^N h_{ij}s_j(n) \right), \quad (2.19)$$

onde K é o número de amostras disponíveis.

Admitindo fontes super-gaussianas (ou leptokurtóticas [20]), associamos uma probabilidade *a priori* correspondente para as amplitudes dos sinais emitidos pelas fontes, podendo escrever:

$$p(s_j(n)|\mathbf{I}) = q_j(s_j(n)), \quad (2.20)$$

onde $q_j(s_j(n))$ é a probabilidade de a j -ésima fonte possuir uma dada amplitude em qualquer instante n (a função q pode ser interpretada como um parâmetro do tipo *nuisance*, significando que não estamos primordialmente interessados nela, embora conhecê-la aproximadamente ou estimá-la seja necessário para obtermos as estimativas de interesse [15]). Cabe notar que, embora haja a admissão de independência entre amostras (o que não corresponde à realidade), na prática o método não falha quando na presença de uma dependência entre amostras; apenas não utiliza informações oriundas desta dependência [16]. Podemos, desta forma, escrever $p(\mathbf{S}|\mathbf{I})$ como:

$$p(\mathbf{S}|\mathbf{I}) = \prod_{j=1}^N \prod_{n=1}^K q_j(s_j(n)). \quad (2.21)$$

Assumindo a ausência de conhecimento acerca da matriz de mistura, podemos supor que sua distribuição *a priori* seja a uniforme. O paradigma bayesiano permite que explicitemos esta hipótese, implícita no método Infomax. Importa observar que, devido à ambigüidade de escalamento inerente aos métodos de separação sem informação sobre as fontes, os coeficientes da matriz de mistura atuam como parâmetros de escala - tornando mais apropriada que a distribuição uniforme uma distribuição não informativa de Jeffrey [19] (no entanto, em [18], conclui-se que a distribuição de Jeffrey não possui efeito nenhum na minimização, não podendo reduzir o problema de supertreinamento). Dito isto, expressamos $p(h_{ij}|\mathbf{I})$, a partir de um intervalo razoável $[h_{min}, h_{max}]$, por:

$$p(h_{ij}|\mathbf{I}) = \begin{cases} c & \text{se } h_{min} \leq h_{ij} \leq h_{max} \\ 0 & \text{se } h_{ij} < h_{min}, h_{max} < h_{ij} \end{cases} \quad (2.22)$$

onde $c = \frac{1}{h_{max} - h_{min}}$. Para toda a matriz \mathbf{H} , a distribuição conjunta *a priori* e uniforme é dada por:

$$p(\mathbf{H}|\mathbf{I}) = \prod_{i=1}^N \prod_{j=1}^N p(h_{ij}|\mathbf{I})$$

$$= \begin{cases} C & \text{se } \forall h_{ij}, h_{min} \leq h_{ij} \leq h_{max} \\ 0 & \text{se } \exists h_{ij} | h_{ij} < h_{min} \text{ ou } h_{max} > h_{ij} \end{cases} \quad (2.23)$$

com $C = cN^2$ (e não cN^2 , como consta em [19]).

A busca para determinar \mathbf{S} e \mathbf{H} pode ser muito simplificada caso nos restringamos a obter \mathbf{H} (ou $\mathbf{W} = \mathbf{H}^{-1}$), pois, a partir do fato de que $\mathbf{X} = \mathbf{HS}$, podemos obter \mathbf{S} por meio de $\mathbf{S} = \mathbf{H}^{-1}\mathbf{X}$. Para nos restringirmos à busca de \mathbf{H} , cabe marginalizar a Equação (2.17) sobre todos os valores possíveis dos sinais das fontes:

$$\begin{aligned} p(\mathbf{H}|\mathbf{X}, \mathbf{I}) &= \int d\mathbf{S} p(\mathbf{H}, \mathbf{S}|\mathbf{X}, \mathbf{I}) \\ &\propto \int d\mathbf{S} p(\mathbf{H}|\mathbf{I}) p(\mathbf{S}|\mathbf{I}) p(\mathbf{X}|\mathbf{H}, \mathbf{S}, \mathbf{I}) \\ &\propto p(\mathbf{H}|\mathbf{I}) \int d\mathbf{S} p(\mathbf{S}|\mathbf{I}) p(\mathbf{X}|\mathbf{H}, \mathbf{S}, \mathbf{I}). \end{aligned} \quad (2.24)$$

Substituindo $p(\mathbf{H}|\mathbf{I})$, $p(\mathbf{S}|\mathbf{I})$ e $p(\mathbf{X}|\mathbf{H}, \mathbf{S}, \mathbf{I})$, obtém-se (vide [19] para maiores detalhes):

$$p(\mathbf{H}|\mathbf{X}, \mathbf{I}) \propto p(\mathbf{H}|\mathbf{I}) \frac{1}{\det \mathbf{H}} \prod_{j=1}^N \prod_{n=1}^K q_j \left(\sum_{i=1}^N w_{ij} x_i(n) \right). \quad (2.25)$$

Aplicar o logaritmo em ambos os lados resulta em:

$$\begin{aligned} \log p(\mathbf{H}|\mathbf{X}, \mathbf{I}) &= M + \log p(\mathbf{H}|\mathbf{I}) \\ &\quad - \log \det \mathbf{H} + \sum_{j=1}^N \sum_{n=1}^K \log q_j \left(\sum_{i=1}^N w_{ij} x_i(n) \right), \end{aligned} \quad (2.26)$$

com M sendo o logaritmo da constante de proporcionalidade da Equação (2.25). Derivar a expressão acima em relação à inversa de \mathbf{H} motiva o uso de uma busca da matriz de separação via gradiente ascendente. Uma modificação no gradiente que o torna mais rápido e robusto ao mau condicionamento da matriz de misturas, denominado *gradiente relativo* [15] ou *gradiente natural* [21], consiste em pós-multiplicá-lo por $\mathbf{W}^T \mathbf{W}$ (onde $\mathbf{W} = \mathbf{H}^{-1}$).

Admitindo uma distribuição *a priori* uniforme para a matriz de misturas, o termo $p(\mathbf{H}|\mathbf{I})$ não influencia o gradiente. Assim, podemos expressar a equação de atualização (utilizando o gradiente natural) por [22]:

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta \{ \mathbf{I}_N - \mathbf{E}[\Phi(\mathbf{Y})\mathbf{Y}^T] \} \mathbf{W}_i, \quad (2.27)$$

onde $\Phi(y_i)$, também chamada de *função score*, é definida por $\Phi(y_i) = -\frac{\partial}{\partial y_i} \log q(y_i)$, $\mathbf{Y} = \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{H}\mathbf{S}$, $\mathbb{E}[\cdot]$ é o operador de média estatística e η é o fator de aprendizagem. Nesta dissertação, assumiremos $\Phi(y_i) = \tanh(y_i)$, procedimento muito comum, quando os sinais envolvidos seguem uma distribuição super-gaussiana.

2.2.2.2 Empregando Conhecimento acerca da Matriz de Misturas

Como vimos, o algoritmo Infomax utiliza, de forma implícita, a hipótese de distribuição uniforme da matriz de misturas. Por esta razão, o termo $p(\mathbf{H}|\mathbf{I})$ não influencia o gradiente e, por conseguinte, a equação de atualização. Caso tenhamos algum conhecimento *a priori* acerca da matriz de mistura, a abordagem bayesiana nos informa que basta adicionar ao gradiente natural o termo:

$$\frac{\partial \log p(\mathbf{H}|\mathbf{I})}{\partial \mathbf{H}^{-1}} (\mathbf{H}^{-1})^T \mathbf{H}^{-1}. \quad (2.28)$$

Fazendo $\mathbf{W} = \mathbf{H}^{-1}$, temos a nova equação de atualização:

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta \left[(\mathbf{I}_n - \mathbb{E}[\Phi(\mathbf{Y})\mathbf{Y}^T]) + \frac{\partial \log p(\mathbf{H}|\mathbf{I})}{\partial \mathbf{W}_i} \mathbf{W}_i^T \right] \mathbf{W}_i. \quad (2.29)$$

Esta modificação no gradiente incorpora nosso conhecimento acerca da matriz de mistura, incluindo na função objetivo a maximização da probabilidade *a priori* da matriz de mistura. Este termo pode revelar-se valioso quando a quantidade de amostras que possuímos é relativamente pequena (situação comum e crítica [3]), orientando a busca quando a informação oriunda da mistura não é muito confiável.

Admitiremos, para efeitos de avaliação, que a matriz de mistura possui coeficientes que sejam amostrados segundo uma distribuição exponencial, dada por:

$$p(h_{ij}) = \frac{1}{2\mu} e^{-\frac{|h_{ij}|}{\mu}} \quad (2.30)$$

Supondo um conhecimento exato desta distribuição, podemos escrever:

$$p(\mathbf{H}|\mathbf{I}) = \prod_{i,j} p(h_{ij}) = \frac{1}{(2\mu)^{N^2}} e^{-\sum_{i,j} \frac{|h_{ij}|}{\mu}} \quad (2.31)$$

Aplicando o logaritmo em ambos os lados:

$$\log p(\mathbf{H}|\mathbf{I}) = \log \frac{1}{(2\mu)^{N^2}} - \sum_{i,j} \frac{|h_{ij}|}{\mu}, \quad (2.32)$$

onde o termo $\log \frac{1}{(2\mu)^{N^2}}$, sendo uma constante, não interfere no gradiente. O termo $\frac{\partial \log p(\mathbf{H}|\mathbf{I})}{\partial \mathbf{W}_i} \mathbf{W}_i^T \mathbf{W}_i$ a adicionar no gradiente natural é calculado mediante a utilização de matemática simbólica de programas como Maple ou MATLAB. Como o resultado possui uma extensão razoável, não será exposto aqui. Cabe apenas ressaltar que devemos expressar \mathbf{H} em função de sua inversa \mathbf{W} para obter o gradiente desejado.

Em [19], a distribuição *a priori* não era uma exponencial, mas uma distribuição proporcional ao quadrado da distância entre a fonte e o sensor correspondente.

A Figura 2.3 apresenta $\log p(\mathbf{H}|\mathbf{X}, \mathbf{I})$ sob a hipótese de distribuição uniforme da matriz de misturas. Já na Figura 2.4, $\log p(\mathbf{H}|\mathbf{X}, \mathbf{I})$ é calculada sob a hipótese de distribuição exponencial (com $\mu = 0, 1$). Em ambas as figuras, o resultado foi normalizado; foram utilizadas 100 amostras das misturas, cada qual seguindo a distribuição:

$$p(x) = p(s) = 1/(2\sqrt{2\pi}\sigma)e^{-\frac{(x-1)^2}{2\sigma^2}} + 1/(2\sqrt{2\pi}\sigma)e^{-\frac{(x+1)^2}{2\sigma^2}}, \quad (2.33)$$

com $q = p$. Impondo $\mathbf{X} = \mathbf{S}$, a “separação” é perfeita quando $\mathbf{W} = \mathbf{I}$. O desvio em relação à identidade é efetuado por meio de dois parâmetros, v e u , da forma a seguir [15]:

$$\mathbf{H}(u, v) = \begin{bmatrix} \cosh(u) & \sinh(u) \\ \sinh(u) & \cosh(u) \end{bmatrix} \begin{bmatrix} \cos(v) & -\sin(v) \\ \sin(v) & \cos(v) \end{bmatrix}, \quad (2.34)$$

onde $\mathbf{H}(0, 0) = \mathbf{I}$. A conveniência desta exploração da vizinhança da matriz identidade é melhor justificada quando observamos que, para u e v pequenos, temos:

$$\mathbf{H}(u, v) \approx \mathbf{I} + u \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + v \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (2.35)$$

Em ambas as figuras, vemos a tendência de o máximo da função $\log p(\mathbf{H}|\mathbf{X}, \mathbf{I})$ se localizar próximo da identidade (como era de se esperar) e a comparação revela que as superfícies possuem grosseiramente a mesma aparência, com ligeiras diferenças.

2.2.2.3 Conhecimento Imperfeito acerca das Estatísticas da Matriz de Misturas

Admitir que possuímos um conhecimento perfeito acerca da distribuição da matriz de misturas é uma hipótese forte. Assim, cabe verificar até onde um conhe-

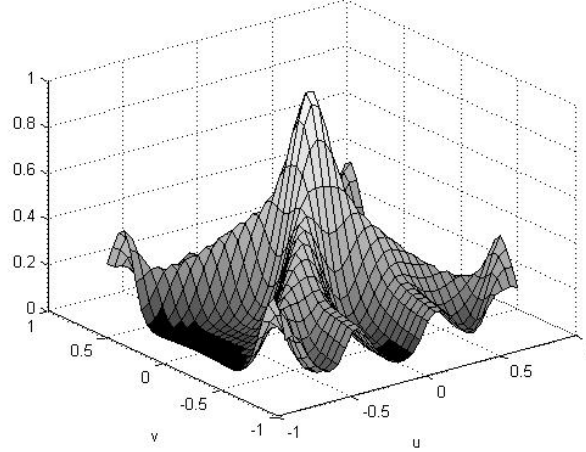


Figura 2.3: $\log p(\mathbf{H}|\mathbf{X}, \mathbf{I})$, sob hipótese de distribuição uniforme dos coeficientes da matriz de mistura.

imento imperfeito pode contribuir para a separação. Uma forma simples seria obter coeficientes da matriz de mistura com distribuição exponencial dotada de um determinado parâmetro μ , e utilizar como distribuição *a priori* uma exponencial com parâmetro μ' . Para avaliar a discrepância entre a distribuição exata e a aproximada, poderíamos utilizar a distância $|\mu - \mu'|$. Porém, optamos por utilizar uma medida mais significativa da discrepância entre distribuições: a divergência de Kullback Leibler, oriunda da Teoria da Informação [23], dada por:

$$D(p||q) = \int_{-\infty}^{+\infty} p(h) \log \frac{p(h)}{q(h)} dh, \quad (2.36)$$

onde p é a distribuição real dos coeficientes e q a distribuição aproximada. D é uma medida não-negativa, tornando-se nula se, e somente se, $p = q$. A partir de que valor de D não é mais interessante recorrer à aproximação dada por q é uma questão significativa, que será contemplada na seção de resultados deste capítulo.

Um comentário interessante acerca desta divergência é que esta constitui um caso particular de uma família de medidas de divergência, denominada α -divergência [24], com $\alpha = \pm 1$. Para $\alpha \neq \pm 1$, a α -divergência é dada por:

$$D^{(\alpha)}(p||q) = \frac{4}{1 - \alpha^2} \left\{ 1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} \right\}, \quad (2.37)$$

cabendo notar que para $\alpha = 0$, temos a *distância de Hellinger* [25], e a medida de divergência torna-se simétrica (mesmo na divergência de Kullback-Leibler, $D(p||q) \neq$

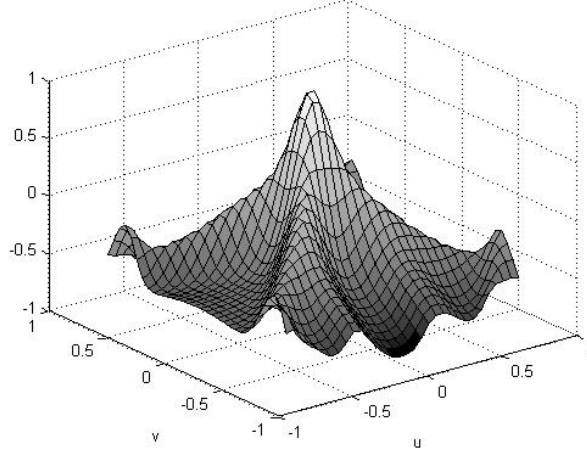


Figura 2.4: $\log p(\mathbf{H}|\mathbf{X}, \mathbf{I})$, sob hipótese de distribuição exponencial dos coeficientes da matriz de mistura.

$D(q||p)$, em geral). A α -divergência é estreitamente relacionada com a α -entropia de Rényi [25].

2.2.2.4 Aplicação em Comunicações Digitais

Em [26], é proposto um método (aqui denominado ext-MIMO-CM) para separação de sinais de comunicações digitais que apresentam módulo constante. Rigorosamente falando, este método não pode ser considerado cego, à medida em que admite a informação de módulo constante dos sinais a separar. Porém, podemos considerá-lo “cego” para nossos propósitos, já que não é utilizada nenhuma informação acerca da matriz de mistura. A função custo a minimizar envolve estatísticas de ordem superior à segunda e é dada por:

$$J = \sum_{i=1}^N \mathbb{E} \left[\left| |y_i|^2 - 1 \right|^2 \right] + \frac{\alpha}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \left| \mathbb{E}[y_i y_j^*] \right|^2, \quad (2.38)$$

onde o primeiro termo enfatiza a restrição de módulo unitário e o segundo contempla a correlação cruzada entre diferentes saídas (evitando a possibilidade de a mesma fonte ser extraída em saídas distintas).

Para a i -ésima coluna de \mathbf{W} (vetor \mathbf{w}_i), o gradiente é dado por:

$$\nabla_{\mathbf{w}_i} J = 2\mathbb{E} \left[(|y_i|^2 - 1) y_i^* \mathbf{x} \right] + \alpha \sum_{j=1, j \neq i}^N \mathbb{E}[y_i^* y_j] \mathbb{E}[y_j^* \mathbf{x}]. \quad (2.39)$$

Utilizando, para estimativa do valor esperado, duas amostras dos dados, temos um método estocástico de gradiente descendente, cuja equação de atualização pode ser expressa por:

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) - \eta \left\{ 2 \sum_{k=n-1}^n [|y_i(k)|^2 - 1] y_i^*(k) \mathbf{x}(k) + \alpha \cdot \sum_{j=1, j \neq i}^N \left[\sum_{k=n-1}^n y_i^*(k) y_j(k) \right] \left[\sum_{k=n-1}^n y_j^*(k) \mathbf{x}(k) \right] \right\} \quad (2.40)$$

A atualização de cada coluna da matriz \mathbf{W} pode ser disposta numa matriz, de forma a atualizar toda a matriz \mathbf{W} . Este algoritmo pode ser facilmente posto num paradigma bayesiano, como feito anteriormente com o método Infomax. Neste paradigma, um conhecimento *a priori* acerca da matriz de misturas pode ser incorporado ao gradiente mediante um termo adicional dado por $\frac{\partial \log p(\mathbf{H}|\mathbf{I})}{\partial \mathbf{W}_i}$ (havendo a opção de não utilizar-se o logaritmo). A principal diferença em relação à abordagem anterior reside no gradiente, que agora não está mais na sua formulação “natural”. Neste capítulo só contemplaremos o caso de misturas reais, portanto, a operação conjugado complexo pode ser ignorada.

2.2.3 Avaliação

Existem diversas formas de avaliar-se o desempenho de um método de separação de fontes [27],[28]. Neste capítulo, utilizamos duas medidas: SIR médio (razão sinal-interferência, em dB, para o método Infomax) e SINR médio (razão sinal-ruído mais interferência, em dB, para o método ext-MIMO-CM).

A matriz \mathbf{C} , cujos elementos são utilizados a seguir, é a resultante dos sistemas de mistura e de separação, sendo dada por $\mathbf{C} = \mathbf{W}\mathbf{H}$.

O SIR da i -ésima fonte na k -ésima saída é dado por:

$$\text{SIR}_i[\mathbf{w}_k] = \frac{\text{E} [c_{ki} s_i(n)^2]}{\sum_{j=1, j \neq i}^N \text{E} [c_{kj} s_j(n)^2]} \quad (2.41)$$

O SINR da i -ésima fonte na k -ésima saída, para fontes com módulo constante, é dado por (vide [26]):

$$\text{SINR}_i[\mathbf{w}_k] = \frac{|c_{ki}|^2}{\sum_{l=1, l \neq i}^K |c_{kl}|^2 + \mathbf{w}_k^T \mathbf{R}_{nn} \mathbf{w}_k}, \quad (2.42)$$

onde \mathbf{R}_{nn} é a matriz de autocorrelação do ruído.

2.2.4 Simulações

Em todas as simulações abaixo, foram utilizadas duas misturas (instantâneas) e duas fontes. Estes resultados foram apresentados em [29].

Para o método Infomax, foram utilizados sinais de voz (uma masculina, a outra feminina), amostrados em 16 kHz, sem adição de ruído; a avaliação de desempenho foi executada mediante a diferença entre a razão sinal-interferência final e a inicial (ambas em dB). As matrizes foram amostradas segundo uma distribuição exponencial com parâmetro $\mu = 0, 1$.

Já para o método ext-MIMO-CM foram utilizadas fontes que apresentam valores +1 ou -1. A avaliação foi efetuada pela diferença $\text{SINR}_{fin} - \text{SINR}_{in}$ (ou seja, a diferença entre o SIR final e o inicial, em dB).

2.2.4.1 Experimento I: Infomax

A partir de trechos de sinais de voz implementamos misturas instantâneas não-ruídosas. Para verificar quando é benéfica a informação *a priori*, foram realizados testes com trechos que continham 2000, 3000, 4000, 6000 e 7000 amostras, sendo efetuadas as médias do tipo Monte Carlo oriundas de 1000 simulações distintas, utilizando os dois métodos (Infomax e Infomax semi-cego), ambos com parâmetro $\eta = 0, 1$.

A partir dos resultados que constam na Tabela 2.2, verificamos que o método proposto apresenta melhoras quando a quantidade de amostras disponíveis é aproximadamente inferior ou igual a 6000. Quando o número de amostras aumenta, a verossimilhança (implícita na função custo) torna-se mais confiável e a inserção da informação estatística acerca da matriz de mistura acaba resultando em perdas de desempenho do algoritmo. A Figura 2.5 apresenta a trajetória dos coeficientes da primeira linha da matriz de separação em um caso específico (ambos os algoritmos foram inicializados no mesmo ponto). As retas que aparecem nesta figura correspondem ao lugar geométrico dos pontos que implicam uma separação perfeita. Nesta figura, verificamos que a trajetória dos coeficientes do algoritmo modificado é bastante diferente da trajetória do algoritmo original, e que no exemplo dado a separação obtida pelo método proposto apresentou resultado bastante superior.

Tabela 2.2: SIR final - SIR inicial (em dB) para o método Infomax, Infomax modificado (com informação estatística acerca da matriz de mistura) e a diferença entre os resultados do método modificado e do padrão, para diferentes comprimentos das misturas.

Comprimento	Infomax	Infomax _m	Diferença
2000	14,4359	16,2515	1,8186
3000	14,6114	16,9181	2,3067
4000	14,3246	16,4855	2,1609
5000	12,536	14,6377	2,1017
6000	16,3986	17,3028	0,9042
7000	22,2478	17,1682	-5,0796
8000	20,176	16,8931	-3,2829
9000	22,0291	18,0836	-3,9455
50000	35,68	33,158	-2,522

2.2.4.2 Experimento II: Infomax com Informação Imperfeita

Supondo que a distribuição real dos coeficientes da matriz de misturas seja uma exponencial com parâmetro $\mu = 0, 1$ e que a distribuição *a priori* (modelagem utilizada no gradiente natural) seja inexata, esta seção verifica experimentalmente até quando uma informação aproximada é benéfica. A medida de distância utilizada entre a distribuição real e a aproximada é a divergência de Kullback Leibler. Utilizando um número de amostras igual a 2000, alteramos o parâmetro μ' da distribuição aproximada e avaliamos o desempenho na separação. Os resultados, obtidos a partir de 300 médias do tipo Monte Carlo (exceto o caso $\mu' = \mu$, obtido da seção anterior com 1000 médias e ressaltado em negrito), encontram-se na Tabela 2.3. A diferença média entre o SIR final e o inicial no método Infomax padrão, com 2000 amostras, foi de 14, 4359 dB (ver Tabela 2.2). O método modificado apresentou resultado melhor sempre que a distância de Kullback Leibler foi inferior a um valor em torno de 0,81. Este resultado, embora não possa ser generalizado (para qualquer número de amostras, distribuições distintas da exponencial ou até mesmo para outros parâmetros da exponencial exata diferentes de $\mu = 0, 1$), significa que mesmo

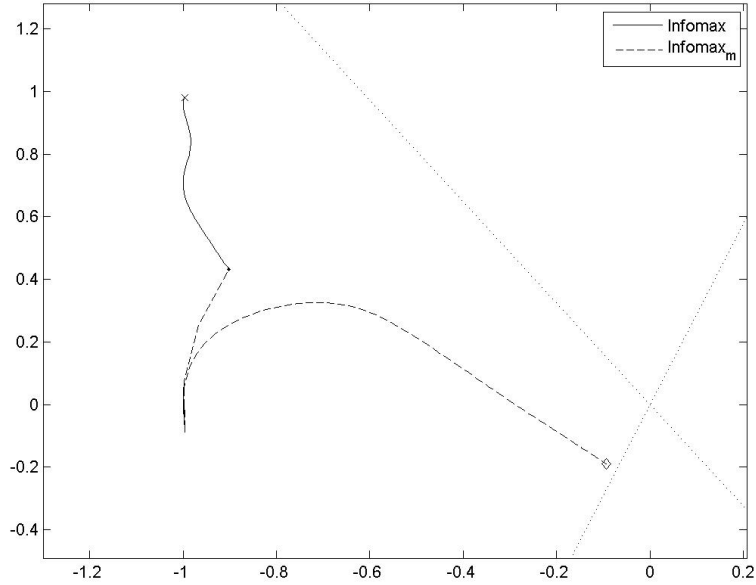


Figura 2.5: Trajetória dos coeficientes $w_{12} \times w_{11}$ das matrizes de separação (colunas normalizadas), com 2000 amostras das misturas disponíveis. A trajetória contínua é do método Infomax padrão e a tracejada do método modificado.

uma informação bastante inexata (por exemplo, arbitrar $\mu' = 0,4$ quando $\mu = 0,1$) da distribuição real dos coeficientes da matriz de mistura pode contribuir para a separação (em configurações onde o número de amostras disponíveis é pequeno).

2.2.4.3 Experimento III: ext-MIMO-CM

A comparação entre os métodos ext-MIMO-CM e ext-MIMO-CM modificado é mostrada nas Figuras 2.6 e 2.7, tratando esta do caso sem ruído e aquela do caso ruidoso, numa configuração bastante difícil para o algoritmo padrão ($\mu = 0,2$). Na Figura 2.6, arbitramos $\alpha = 1$, $\eta = 5.10^{-2}$, $\mu = \mu' = 0,2$ e variância do ruído branco gaussiano aditivo igual a 10^{-4} e na Figura 2.7 temos $\alpha = 0,1$, $\eta = 0,9.10^{-2}$ e $\mu = \mu' = 0,9$. Verifica-se que a informação *a priori* oferece uma contribuição significativa para o desempenho do algoritmo. A contribuição do método proposto pode se revelar particularmente importante em sistemas de comunicações digitais, devido ao fato de freqüentemente ser necessário efetuar a separação possuindo relativamente poucas amostras.

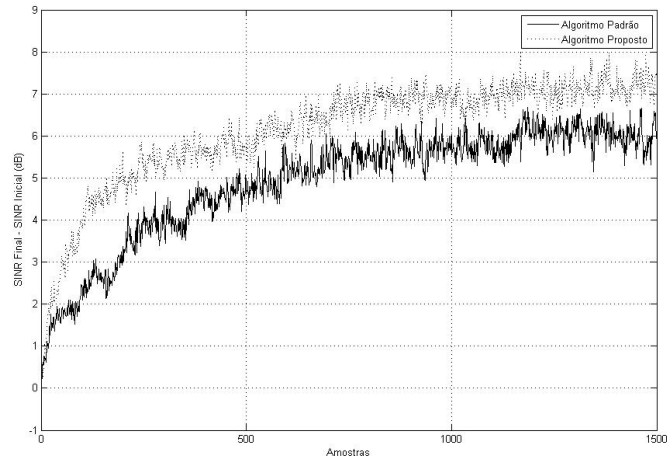


Figura 2.6: Melhora no SINR (em dB) em função do número de amostras dos métodos ext-MIMO-CM e ext-MIMO-CM modificado. Resultado com 1500 amostras oriundo de média Monte Carlo com 500 simulações.

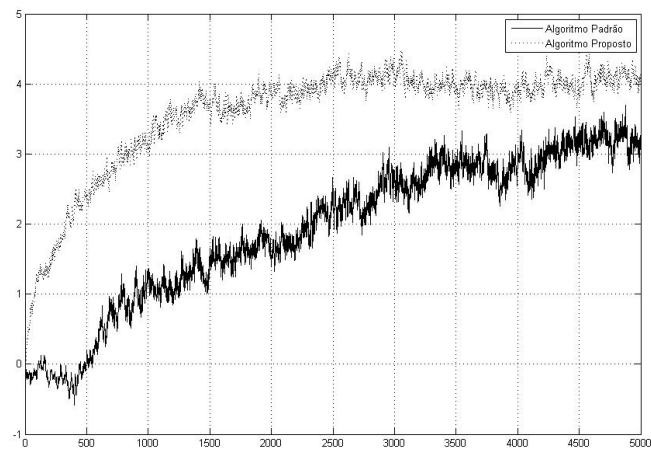


Figura 2.7: Melhora no SINR (em dB) em função do número de amostras (sem ruído) dos métodos ext-MIMO-CM e ext-MIMO-CM modificado. Resultado oriundo de média Monte Carlo com 1000 simulações.

Tabela 2.3: SIR final - SIR inicial para o Infomax modificado, variando o parâmetro μ' da distribuição *a priori* aproximada e utilizando 2000 amostras. O parâmetro D indica a divergência de Kullback Leibler entre as distribuições real e aproximada.

μ'	D	$SIR_{fin}-SIR_{in}$
0,005	16,0043	11,2441
0,01	6,6974	14,2325
0,1	0	16,2515
0,15	0,0721	16,3689
0,2	0,1931	15,9174
0,3	0,4319	15,4142
0,4	0,6363	15,3601
0,5	0,8094	14,4773
0,6	0,9584	14,1948

2.3 Considerações Finais

Neste capítulo, introduzimos o problema de separação de fontes na sua formulação mais simples (misturas instantâneas e lineares). Inserimos um método de separação clássico (Infomax) num paradigma bayesiano, o que permitiu estendê-lo para um caso em que tenhamos informações acerca da matriz de misturas. Foram efetuadas avaliações quantitativas as quais demonstram que mesmo um conhecimento estatístico inexato sobre os coeficientes da matriz de misturas pode auxiliar a solução do problema, quando é reduzida a informação disponível das misturas. Por fim, utilizamos o mesmo paradigma num método bem conhecido de separação de fontes voltado para sinais de comunicações digitais (sob a condição de módulo constante) e também verificamos uma melhora de seu desempenho (particularmente quando o número de amostras disponíveis é reduzido).

Capítulo 3

Misturas Convolutivas

O modelo de misturas lineares e instantâneas é demasiado simples. Não modela o atraso entre sensores, tampouco a reverberação (no caso de sinais de áudio) ou o fenômeno de múltiplos percursos (no caso de sinais de comunicações digitais). Um método de separação de fontes, para poder tratar de casos mais realistas, deve estender sua aplicabilidade para o caso de misturas convolutivas (doravante denominado apenas “caso convolutivo”). Neste modelo, a n -ésima amostra da k -ésima mistura (oriunda do k -ésimo sensor) das P fontes¹ pode ser expressa como:

$$x_k(n) = \sum_{p=1}^P \sum_{q=0}^{Q-1} h_{pk}(q) s_p(n-q), \quad (3.1)$$

onde Q é o comprimento dos filtros de mistura (h_{pk}). Para efetuar a separação, a solução mais comum é a baseada em filtros FIR (w_{pq}) de separação, cada qual de comprimento L . Assim, a k -ésima saída do método de separação pode ser descrita como:

$$y_k(n) = \sum_{p=1}^P \sum_{l=0}^{L-1} w_{pk}(l) x_p(n-l). \quad (3.2)$$

No caso convolutivo, torna-se necessária uma distinção entre duas soluções para a tarefa de separação:

(I) Deconvolução cega, onde o objetivo é obter:

$$y_k(n) = d_{k\Delta_j} s_j(n - \Delta_j), \quad \text{para } k = 1, \dots, P, \quad (3.3)$$

onde $d_{k\Delta_j}$ e Δ_j são, respectivamente, fatores de ganho e de atraso.

¹por motivos de facilidade de nomenclatura, neste capítulo P significará o número de fontes, o qual será suposto igual ao número de sensores.

(II) Separação cega de fontes, a qual é considerada bem-sucedida quando a seguinte condição é satisfeita:

$$y_k(n) = \sum_{i=0}^{\infty} d_{ki}s_j(n-i), \quad (3.4)$$

onde os coeficientes d_{ki} são coeficientes oriundos da função de transferência resultante do sistema de mistura e de separação.

Embora a deconvolução cega seja um caso particular da separação cega de fontes, a distinção é importante porque, em alguns casos, é imperioso preservar a estrutura temporal das fontes (o que somente ocorre na deconvolução cega). Isto é particularmente importante, por exemplo, em sinais de comunicações digitais.

3.1 Evolução dos Métodos de Separação em Misturas Convolutivas

Desde meados da década de 90, surgiram muitos artigos destinados à separação cega de fontes no caso de misturas convolutivas (a partir de agora abreviado por CBSS - *convolutive blind source separation*). Porém, a estrutura dos primeiros algoritmos propostos [30]-[34] atentava para a estimativa das funções de transferência dos canais e, a partir destas estimativas, recuperava as fontes originais. Isto nos permite classificá-los como métodos de identificação. Estes métodos, infelizmente, exigem um processamento inadequado às aplicações em tempo real. Por esta razão, a pesquisa recente concentra-se nos métodos ditos de “busca do inverso do sistema”, os quais estimam diretamente os filtros de separação. Este último tipo de algoritmo, que utiliza filtros FIR para separação, constitui o foco deste capítulo. Entre os primeiros artigos que recorreram a este paradigma, destacam-se [35] e [36], sendo este último dedicado à deconvolução cega.

Os métodos de CBSS totalmente implementados no domínio do tempo costumam apresentar grande custo computacional. Devido à equivalência entre convolução no domínio do tempo e produto no domínio da frequência, muitos métodos utilizam os sinais no domínio da frequência e empregam métodos baseados em ICA nas raias da DFT. Cada raia é tratada de forma diferente das demais, o que motivou a classificação destes métodos como de “banda estreita” (*narrowband*, em inglês) [11].

Estes métodos de banda estreita apresentam uma grande vantagem: custo computacional reduzido.

No entanto, eles introduzem uma série de dificuldades antes inexistentes. Eis as principais: (a) a ambigüidade de escalamento torna-se não trivial; (b) a equivalência entre a convolução no domínio do tempo e o produto das DFTs só ocorre quando a convolução é circular, diferentemente do que ocorre na prática, onde não raro a convolução é linear; e (c) o problema da permutação passa a ser extremamente crítico.

A dificuldade (a) implica distorções intoleráveis na estimativa das fontes, já que uma freqüência pode ser amplificada por uma constante de escalamento de magnitude muito distinta de outra freqüência, mesmo que próxima. Esta dificuldade quase sempre é contornada mediante o princípio da distorção mínima [37], o qual atua sobre a matriz de separação $\mathbf{W}(f)$ após a convergência do método, da forma a seguir:

$$\mathbf{W}(f) \leftarrow \text{diag} [\mathbf{W}^{-1}(f)] \mathbf{W}(f), \quad (3.5)$$

onde o operador $\text{diag}(\cdot)$ anula todos os coeficientes de uma matriz que não estejam em sua diagonal principal.

Este procedimento é justificado pelo seguinte raciocínio: supondo que a separação seja razoável e (sem perda de generalidade) que não há permutação, podemos aproximar a matriz $\mathbf{W}(f)$ por:

$$\mathbf{W}(f) \approx \mathbf{D}(f)\mathbf{H}^{-1}(f), \quad (3.6)$$

onde $\mathbf{D}(f)$ é uma matriz diagonal que contém os coeficientes de escalamento. Esta aproximação nos permite verificar que $\text{diag} [\mathbf{W}^{-1}(f)] \mathbf{W}(f) \approx \text{diag} [\mathbf{H}(f)] \mathbf{H}^{-1}(f)$. Esta aproximação implica que a transformação em (3.5) gera uma matriz $\mathbf{W}(f)$ com um escalamento razoável, e não com um escalamento completamente arbitrário.

Já a dificuldade (b) é normalmente contornada por meio de uma restrição do número de coeficientes não-nulos dos filtros de separação no domínio do tempo. Por exemplo, em [38], caso se utilize uma DFT com 1024 raias, a cada iteração os filtros no domínio da freqüência são convertidos para o domínio do tempo e todos os coeficientes, exceto os primeiros 128, são zerados antes de nova conversão para o domínio da freqüência (a proposta consiste em um número de coeficientes não

zerados igual a $1/8$ do número total de raias na frequência). Neste mesmo artigo, este procedimento, ao tornar mais suaves as respostas na frequência dos filtros de separação, serviu para contornar o problema da permutação, visto a seguir.

A dificuldade (c) é mais complexa que as outras, pois sem um tratamento adequado as componentes no domínio da frequência podem permutar livremente nas saídas estimadas, fazendo com que aproximadamente metade das raias de uma das saídas corresponda às raias de uma das fontes, enquanto que o restante das raias corresponda a estimativas da outra fonte (no caso de duas fontes). As primeiras propostas para esta dificuldade são: restrição do comprimento do filtro no domínio do tempo (como visto acima, em [38]), minimização de correlação (ou uma versão modificada da mesma, eventualmente com critérios que contemplem simultaneamente diversas frequências) [39],[40], estimativa de direção de chegada de cada fonte [41],[42] e até mesmo uma conjunção destas duas últimas técnicas [3], almejando atenuar as desvantagens de uma abordagem com as vantagens da outra.

Não é muito fácil classificar as diferentes abordagens para CBSS existentes, embora grande esforço neste sentido tenha sido efetuado em [43],[44]. Primeiramente, os coeficientes do sistema de separação podem se encontrar tanto no domínio do tempo quanto no da frequência. Independentemente desta escolha, o critério utilizado para separação pode ser aplicado em ambos os domínios (um método que almeja encontrar coeficientes no domínio do tempo e utiliza um critério no domínio da frequência exige constantes conversões dos filtros para ambos os domínios durante as iterações).

Nesta dissertação, contemplaremos métodos no domínio do tempo ² e no da frequência, sendo que nestes últimos enfocaremos métodos mais recentes do tipo banda larga. Um método no domínio da frequência é classificado de banda larga se o critério utilizado para separação contempla todas as frequências simultaneamente; em outras palavras, as raias não são tratadas de forma independente. Estes métodos costumam apresentar resultados melhores, contornando facilmente a dificuldade (c) supracitada.

²Há uma sutileza a notar aqui: existem métodos que utilizam critérios no domínio do tempo e que recorrem ao domínio da frequência apenas para reduzir a carga computacional; estes métodos serão considerados pertencentes à classe dos que atuam no domínio do tempo.

3.2 Generalização de Algoritmos de Separação Cega de Fontes para Misturas Convolutivas Utilizando Estatísticas de Segunda Ordem (GASFC)

Esta seção dedica-se a detalhar um dos mais bem-sucedidos métodos de separação de fontes no caso convolutivo, denominado, nesta dissertação, GASFC [45],[11]. Este método pode explorar simultaneamente três propriedades dos sinais a separar. Embora cada uma dessas propriedades já tenha sido utilizada em algum algoritmo anterior, o GASFC foi o primeiro algoritmo a empregar as três. A Tabela 3.1 mostra estas três propriedades, bem como sua respectiva influência na função custo a otimizar.

A terceira propriedade (não-gaussianidade) não será contemplada nesta

Tabela 3.1: Propriedades exploradas e suas respectivas influências na definição da função custo.

Propriedade	Influência
Espectro colorido	diagonalização simultânea das matrizes de correlação das saídas, com diferentes intervalos de tempo (<i>lags</i>) [46].
Não-estacionariedade	diagonalização simultânea de matrizes de correlação das saídas em diferentes instantes de tempo [47], [48].
Não-gaussianidade	emprego de estatísticas de ordem mais alta [15].

seção, dedicada ao caso que utiliza estatísticas de segunda ordem. A inserção de estatísticas de mais alta ordem exige modificações mínimas no algoritmo, as quais não são exploradas nesta dissertação.

3.2.1 Formulação Matricial para Misturas Convolutivas com Extensão para Múltiplos Intervalos Temporais

O fato de os sinais a separar não serem considerados brancos, como vimos, permite-nos explorar as correlações entre intervalos de tempo na função custo. O parâmetro D consiste no número de intervalos de tempo a considerar, devendo ser, no máximo, igual a L (comprimento dos filtros de separação).

Já a propriedade de não-estacionariedade é explorada por meio de matrizes de correlação das saídas em diferentes instantes de tempo. Para o cálculo dessas matrizes, devemos particionar a saída em blocos, cada qual de comprimento N , com $N \geq PD$ (P , como visto no início do capítulo, é o número de fontes, suposto igual ao de misturas).

Para facilitar a dedução do algoritmo, bem como para facilidade de notação, cumpre apresentar a estrutura de mistura e de separação num formato matricial. Esta formulação permite converter as convoluções em produtos matriciais. Sendo $x_p(n)$ a n -ésima amostra da p -ésima mistura, podemos definir o vetor $\mathbf{x}_p(n)$, contendo as L últimas amostras de x_p (em relação ao instante n) da forma a seguir:

$$\mathbf{x}_p(n) = [x_p(n) \ x_p(n-1) \ \dots \ x_p(n-L+1)]^T. \quad (3.7)$$

Convertendo em vetores os filtros w_{pq} (que filtram a p -ésima mistura e contribuem para a q -ésima saída), definimos:

$$\mathbf{w}_{pq} = [w_{pq}(0) \ w_{pq}(1) \ \dots \ w_{pq}(L-1)]^T. \quad (3.8)$$

As definições de $\mathbf{x}_p(n)$ e \mathbf{w}_{pq} nos permitem escrever como produto vetorial a q -ésima saída do sistema de separação no instante n , ou seja:

$$y_q(n) = \sum_{p=1}^P \mathbf{x}_p^T(n) \mathbf{w}_{pq}. \quad (3.9)$$

A primeira extensão da formulação acima consiste em definir um bloco de comprimento N da q -ésima saída, como:

$$\mathbf{y}_q(m) = [y_q(mL) \ y_q(mL+1) \ \dots \ y_q(mL+N-1)]^T. \quad (3.10)$$

O vetor $\mathbf{y}_q(m)$, vetor que contém um bloco da q -ésima saída pode ser facilmente escrito em função das amostras oriundas dos sensores e dos filtros de separação

como um produto entre a matriz $\mathbf{U}_p^T(m)$ e o vetor \mathbf{w}_{pq} :

$$\mathbf{y}_q(m) = \sum_{p=1}^P \mathbf{U}_p^T(m) \mathbf{w}_{pq}, \quad (3.11)$$

com $\mathbf{U}_p^T(m)$ dado por:

$$\mathbf{U}_p^T(m) = \begin{bmatrix} x_p(mL) & \cdots & x_p(mL - L + 1) \\ x_p(mL + 1) & \cdots & x_p(mL - L + 2) \\ \vdots & \ddots & \vdots \\ x_p(mL + N - 1) & \cdots & x_p(mL - L + N) \end{bmatrix}. \quad (3.12)$$

A segunda extensão a efetuar contempla os D intervalos de tempo utilizados na matriz de correlação (a qual será definida mais à frente). Para utilizar estes intervalos, cumpre recorrer à matriz $\mathbf{Y}_q(m)$ (aumentada em relação a $\mathbf{y}_q(m)$), definida como:

$$\mathbf{Y}_q(m) = \begin{bmatrix} y_q(mL) & \cdots & y_q(mL - D + 1) \\ y_q(mL + 1) & \cdots & y_q(mL - D + 2) \\ \vdots & \ddots & \vdots \\ y_q(mL + N - 1) & \cdots & y_q(mL - D + N) \end{bmatrix}. \quad (3.13)$$

De forma semelhante à Equação (3.11), podemos expressar $\mathbf{Y}_q(m)$ como:

$$\mathbf{Y}_q(m) = \sum_{p=1}^P \mathbf{X}_p(m) \mathbf{W}_{pq}, \quad (3.14)$$

sendo:

$$\mathbf{X}_p(m) = [\mathbf{U}_p^T(m) \quad \mathbf{U}_p^T(m - 1)], \quad (3.15)$$

e

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq}(0) & 0 & \cdots & 0 \\ w_{pq}(1) & w_{pq}(0) & \ddots & \vdots \\ \vdots & w_{pq}(1) & \ddots & 0 \\ w_{pq}(L - 1) & \vdots & \ddots & w_{pq}(0) \\ 0 & w_{pq}(L - 1) & \ddots & w_{pq}(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq}(L - 1) \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (3.16)$$

A concatenação de duas matrizes $\mathbf{U}_p(m)$ (gerando a matriz $\mathbf{X}_p(m)$) é necessária, pois os filtros FIR de separação utilizam amostras anteriores ao instante n para produzir uma saída no mesmo instante.

Antes de determinar as dimensões das matrizes acima definidas, expomos a última extensão da notação matricial, a qual passará a combinar todos os canais de forma concisa. Seja a matriz $\mathbf{Y}(m)$ que contém os blocos de todas as saídas do sistema de separação, incluindo os D intervalos de tempo:

$$\mathbf{Y}(m) = [\mathbf{Y}_1(m) \ \cdots \ \mathbf{Y}_P(m)]. \quad (3.17)$$

Esta matriz pode ser obtida a partir das misturas e dos filtros de separação através de um mero produto matricial, ou seja:

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W}, \quad (3.18)$$

com as matrizes $\mathbf{X}(m)$ e \mathbf{W} definidas como:

$$\mathbf{X}(m) = [\mathbf{X}_1(m) \ \cdots \ \mathbf{X}_P(m)], \quad (3.19)$$

e

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}. \quad (3.20)$$

Para facilidade de consulta, a Tabela 3.2 mostra o significado das variáveis e a Tabela 3.3 mostra as dimensões das matrizes acima definidas.

Tabela 3.2: Propriedades exploradas e suas respectivas influências na definição da função custo.

Variável	Significado
P	número de fontes, igual ao número de misturas
Q	comprimento dos filtros de mistura
L	comprimento dos filtros de separação
N	comprimento do bloco de saída utilizado para estimativas de correlação
D	número de intervalos temporais a considerar nas correlações

Tabela 3.3: Matrizes utilizadas e suas respectivas dimensões.

Matriz	Dimensões
$\mathbf{U}_p(m)$	$L \times N$
$\mathbf{Y}_q(m)$	$N \times D$
$\mathbf{X}_p(m)$	$N \times 2L$
\mathbf{W}_{pq}	$2L \times D$
$\mathbf{Y}(m)$	$N \times PD$
$\mathbf{X}(m)$	$N \times 2PL$
\mathbf{W}	$2PL \times PD$

3.2.2 Função Custo

O método GASFC utiliza, como vimos, estimativas de correlação das saídas. No m -ésimo bloco, definimos as matrizes de correlação de tempo curto como:

$$\mathbf{R}_{\mathbf{xx}}(m) = \mathbf{X}^H(m)\mathbf{X}(m), \quad (3.21)$$

$$\mathbf{R}_{\mathbf{yy}}(m) = \mathbf{Y}^H(m)\mathbf{Y}(m), \quad (3.22)$$

cujas dimensões são, respectivamente, $2PL \times 2PL$ e $PD \times PD$, com $(\cdot)^H$ indicando o operador hermitiano. Cumpre notar que as matrizes $\mathbf{R}_{\mathbf{yy}}(m)$ já contemplam, implicitamente, as correlações em diferentes intervalos de tempo e que apresentam posto completo, dado que $N \geq PD$ e que as mesmas são obtidas via produtos de uma matriz de dimensão $PD \times N$ por outra de dimensão $N \times PD$.

Sejam os ponderadores $\beta(i, m) \geq 0$, com $\sum_{i=0}^{\infty} \beta(i, m) = 1$. A concisa formulação matricial dada na Equação (3.18) permite expressar economicamente a função custo, da forma abaixo:

$$\mathfrak{S}(m) = \sum_{i=0}^{\infty} \beta(i, m) \{ \log \det \text{bdiag } \mathbf{Y}^H(i)\mathbf{Y}(i) - \log \det \mathbf{Y}^H(i)\mathbf{Y}(i) \}, \quad (3.23)$$

que introduz o operador $\text{bdiag}(\cdot)$, o qual interpreta uma matriz como uma matriz de submatrizes (cada qual com uma origem claramente distinta), e então zera todas as submatrizes que não estão situadas na diagonal principal (da matriz de submatrizes). Para melhor entendimento deste operador, seja o caso de 3 fontes. Neste caso, a

matriz $\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)$ pode ser descrita como:

$$\begin{aligned} \mathbf{R}_{\mathbf{y}\mathbf{y}}(m) &= \begin{bmatrix} \mathbf{Y}_1^H(m) \\ \mathbf{Y}_2^H(m) \\ \mathbf{Y}_3^H(m) \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1(m) & \mathbf{Y}_2(m) & \mathbf{Y}_3(m) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Y}_1^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_1^H(m)\mathbf{Y}_2(m) & \mathbf{Y}_1^H(m)\mathbf{Y}_3(m) \\ \mathbf{Y}_2^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_2^H(m)\mathbf{Y}_2(m) & \mathbf{Y}_2^H(m)\mathbf{Y}_3(m) \\ \mathbf{Y}_3^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_3^H(m)\mathbf{Y}_2(m) & \mathbf{Y}_3^H(m)\mathbf{Y}_3(m) \end{bmatrix}. \end{aligned} \quad (3.24)$$

As matrizes $\mathbf{Y}_p^H(m)\mathbf{Y}_p(m)$, $p = 1, \dots, 3$ são as matrizes de autocorrelação da q -ésima saída, enquanto que as matrizes $\mathbf{Y}_p^H(m)\mathbf{Y}_q(m)$, $p \neq q$ são as matrizes de correlação cruzada entre a p -ésima e a q -ésima saídas. É natural subdividir a matriz $\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)$ em submatrizes, sendo as submatrizes de autocorrelação as pertencentes à diagonal principal da matriz de submatrizes. Assim, $\text{bdiag}\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)$ produz o seguinte resultado:

$$\text{bdiag}\mathbf{R}_{\mathbf{y}\mathbf{y}}(m) = \begin{bmatrix} \mathbf{Y}_1^H(m)\mathbf{Y}_1(m) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_2^H(m)\mathbf{Y}_2(m) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Y}_3^H(m)\mathbf{Y}_3(m) \end{bmatrix}, \quad (3.25)$$

onde $\mathbf{0}$ é uma matriz nula, com as mesmas dimensões que $\mathbf{Y}_p^H(m)\mathbf{Y}_q(m) = \mathbf{R}_{\mathbf{y}_p\mathbf{y}_q}(m)$.

A função custo - Equação (3.23) - pode ser obtida mediante uma abordagem geral baseada na Teoria da Informação [49]. Podemos interpretá-la como uma generalização da função custo introduzida em [50], sendo pela primeira vez apresentada em [51].

Não é difícil provar que a função-custo é não-negativa. Detalhando os passos elididos em [11], temos:

$$\begin{aligned} \mathfrak{S}(m) &= \sum_{i=0}^{\infty} \beta(i, m) \left\{ \log \det \text{bdiag}\mathbf{Y}^H(i)\mathbf{Y}(i) - \log \det \mathbf{Y}^H(i)\mathbf{Y}(i) \right\} \\ &= \sum_{i=0}^{\infty} \beta(i, m) \left\{ \log \prod_{q=1}^N \det \mathbf{Y}_q^H(i)\mathbf{Y}_q(i) - \log \det \mathbf{Y}^H(i)\mathbf{Y}(i) \right\} \\ &= \sum_{i=0}^{\infty} \beta(i, m) \underbrace{\left\{ \sum_{q=1}^N \log \det \mathbf{Y}_q^H(i)\mathbf{Y}_q(i) - \log \det \mathbf{Y}^H(i)\mathbf{Y}(i) \right\}}_{\geq 0}, \end{aligned} \quad (3.26)$$

onde $\sum_{q=1}^N \log \det \mathbf{Y}_q^H(i)\mathbf{Y}_q(i) - \log \det \mathbf{Y}^H(i)\mathbf{Y}(i) \geq 0$ pela desigualdade de Oppenheim [52].

3.2.3 Minimização da função custo

A minimização da função custo proposta em [11] utiliza gradientes de primeira ordem. Omitiremos, por simplicidade, o índice m (referente ao m -ésimo bloco) no desenvolvimento a seguir. A otimização visa encontrar os filtros de separação \mathbf{w}_{pq} , os quais estão dispostos na matriz \mathbf{W} da forma mostrada nas equações (3.16) e (3.20). Assim, a atualização é efetuada via adaptação por gradiente descendente. Para obter a equação de atualização, cumpre calcular o gradiente, em relação a \mathbf{W} , da função custo. Este gradiente é dado por:

$$\begin{aligned} \nabla_{\mathbf{W}} \mathfrak{S}(m) &= 2 \frac{\partial \mathfrak{S}(m)}{\partial \mathbf{W}^*} = 2 \frac{\partial \left\{ \sum_{i=0}^{\infty} \beta(i, m) \left[\log \det \text{bdiag} \mathbf{Y}^H(i) \mathbf{Y}(i) - \log \det \mathbf{Y}^H(i) \mathbf{Y}(i) \right] \right\}}{\partial \mathbf{W}^*} \\ &= 2 \left\{ \sum_{i=0}^{\infty} \beta(i, m) \left[\underbrace{\frac{\partial \log \det \text{bdiag} \mathbf{Y}^H(i) \mathbf{Y}(i)}{\partial \mathbf{W}^*}}_{\mathbf{G}_1} - \underbrace{\frac{\partial \log \det \mathbf{Y}^H(i) \mathbf{Y}(i)}{\partial \mathbf{W}^*}}_{\mathbf{G}_2} \right] \right\}, \end{aligned} \quad (3.27)$$

o que torna necessário o cálculo dos gradientes \mathbf{G}_1 e \mathbf{G}_2 .

Por ser mais simples, comecemos pelo cálculo de \mathbf{G}_2 :

$$\begin{aligned} \mathbf{G}_2 &= \frac{\partial \log \det [\mathbf{Y}^H(i) \mathbf{Y}(i)]}{\partial \mathbf{W}^*} = \frac{\partial \log \det \left\{ [\mathbf{X}(i) \mathbf{W}]^H \mathbf{X}(i) \mathbf{W} \right\}}{\partial \mathbf{W}^*} \\ &= \frac{\partial \log \det [\mathbf{W}^H \mathbf{X}^H(i) \mathbf{X}(i) \mathbf{W}]}{\partial \mathbf{W}^*} = \frac{\partial \log \det [\mathbf{W}^H \mathbf{R}_{\mathbf{xx}} \mathbf{W}]}{\partial \mathbf{W}^*} \end{aligned} \quad (3.28)$$

Do apêndice de [22], temos:

$$\frac{\partial \log \det (\mathbf{X}^T \mathbf{C} \mathbf{X})}{\partial \mathbf{X}} = 2 \mathbf{C} \mathbf{X} (\mathbf{X}^T \mathbf{C} \mathbf{X})^{-1}. \quad (3.29)$$

Esta expressão é facilmente estendida para o caso complexo [11], o que permite obter \mathbf{G}_2 da forma a seguir:

$$\mathbf{G}_2 = \mathbf{R}_{\mathbf{xx}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\mathbf{xx}} \mathbf{W})^{-1}. \quad (3.30)$$

Calculado \mathbf{G}_2 , passemos ao cálculo de \mathbf{G}_1 , repetido por conveniência a seguir:

$$\mathbf{G}_1 = \frac{\partial \log \det \text{bdiag} \mathbf{Y}^H(i) \mathbf{Y}(i)}{\partial \mathbf{W}^*}. \quad (3.31)$$

Para dividir o problema, [11] efetua uma regra da cadeia matricial. Primeiramente, seja $\mathfrak{S}_1 = \log \det \text{bdiag} \mathbf{Y}^H(i) \mathbf{Y}(i)$. Isso permite-nos expressar \mathbf{G}_1 como:

$$\mathbf{G}_1 = \frac{\partial \mathfrak{S}_1}{\partial \mathbf{W}^*}. \quad (3.32)$$

Seja a matriz \mathbf{A} definida por:

$$\mathbf{A} = \text{bdiag}(\mathbf{W}^H \mathbf{R} \mathbf{W}), \quad (3.33)$$

e w_{mn}^{MN} o elemento situado na m -ésima linha e n -ésima coluna de \mathbf{W}_{MN} , onde \mathbf{W}_{MN} é a submatriz de \mathbf{W} situada na M -ésima linha e na N -ésima coluna da sua partição em blocos; vide Equação (3.16). A mesma nomenclatura se aplica em a_{kj}^{KJ} , em relação aos elementos da matriz \mathbf{A} . Expressando \mathbf{G}_1 em relação a cada um de seus elementos e utilizando a regra da cadeia, obtemos:

$$\mathbf{G}_1|_{w_{mn}^{MN}} = \sum_{k,j,K,J} \frac{\partial \mathfrak{S}_1}{\partial a_{kj}^{KJ}} \frac{\partial a_{kj}^{KJ}}{\partial (w_{mn}^{MN})^*} \quad (3.34)$$

O termo $\frac{\partial \mathfrak{S}_1}{\partial a_{kj}^{KJ}}$ pode ser exposto economicamente numa formulação matricial (novamente estendendo o caso real apresentado em [22]):

$$\frac{\partial \log \det \mathbf{A}}{\partial \mathbf{A}} = (\mathbf{A}^H)^{-1} = \mathbf{A}^{-1}, \quad (3.35)$$

onde foi utilizado o fato de que $\mathbf{A}^H = \mathbf{A}$. Já o segundo termo exige um desenvolvimento maior. Seja b_{ij}^{IJ} o elemento de $\mathbf{R} \mathbf{W}$ que segue a mesma nomenclatura de a_{kj}^{KJ} e w_{mn}^{MN} . Expressando a_{kj}^{KJ} em função de w_{ki}^{KI} e b_{ij}^{IJ} , obtemos:

$$a_{kj}^{KJ} = \sum_i \sum_I (w_{ki}^{KI})^H b_{ij}^{IJ} \delta_{KJ} = \sum_i \sum_I (w_{ik}^{IK})^* b_{ij}^{IJ} \delta_{KJ}, \quad (3.36)$$

onde o símbolo δ_{KJ} é zero para $K \neq J$ e 1 para $K = J$, correspondendo ao operador $\text{bdiag}(\cdot)$. Esta formulação permite-nos escrever [11]:

$$\begin{aligned} \frac{\partial a_{kj}^{KJ}}{\partial (w_{mn}^{MN})^*} &= \frac{\partial}{\partial (w_{mn}^{MN})^*} \sum_i \sum_I (w_{ik}^{IK})^* b_{ij}^{IJ} \delta_{KJ} \\ &= \sum_i \sum_I \delta_{im} \delta_{kn} \delta_{IM} \delta_{KN} \delta_{KJ} b_{ij}^{IJ} \\ &= \delta_{kn} \delta_{KN} \delta_{KJ} b_{mj}^{MJ}, \end{aligned} \quad (3.37)$$

onde foi omitida a constante 2 (o mesmo ocorrendo no cálculo de \mathbf{G}_1), que é a origem do fator 2 que consta na Equação (2.27).

Reunindo as Eqs. (3.33), (3.35) e (3.37) na regra da cadeia (3.34) e desenvolvendo, temos:

$$\mathbf{G}_1 = \sum_{k,j,K,J} [\text{bdiag}^{-1} \mathbf{W}^H \mathbf{R} \mathbf{W}]_{kj}^{KJ} \delta_{kn} \delta_{KN} \delta_{KJ} [\mathbf{R} \mathbf{W}]_{mj}^{MJ}$$

$$\begin{aligned}
&= \sum_j \left[(\text{bdiag} \mathbf{W}^H \mathbf{R} \mathbf{W})^{-1} \right]_{nj}^{NN} [\mathbf{R} \mathbf{W}]_{mj}^{MN} \\
&= \left[\mathbf{R} \mathbf{W} (\text{bdiag} \mathbf{W}^H \mathbf{R} \mathbf{W})^{-1} \right]_{mn}^{MN}. \tag{3.38}
\end{aligned}$$

O cálculo de \mathbf{G}_1 e \mathbf{G}_2 permite-nos, por fim, fornecer o gradiente da função custo em relação a \mathbf{W} (com \mathbf{I} sendo a matriz identidade de dimensões $PD \times PD$):

$$\begin{aligned}
\nabla_{\mathbf{W}} \mathfrak{S}(m) &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[-\mathbf{R}_{xx} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{xx} \mathbf{W})^{-1} + \mathbf{R}_{xx} \mathbf{W} (\text{bdiag} \mathbf{W}^H \mathbf{R}_{xx} \mathbf{W})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[-\mathbf{R}_{xx} \mathbf{W} \mathbf{R}_{yy}^{-1} + \mathbf{R}_{xx} \mathbf{W} (\text{bdiag} \mathbf{R}_{yy})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{xx} \mathbf{W} \left[-\mathbf{R}_{yy}^{-1} + (\text{bdiag} \mathbf{R}_{yy})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{X}^H \mathbf{X} \mathbf{W} \left[-\mathbf{R}_{yy}^{-1} + (\text{bdiag} \mathbf{R}_{yy})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \underbrace{\mathbf{X}^H \mathbf{Y}}_{\mathbf{R}_{xy}} \left[-\mathbf{R}_{yy}^{-1} + (\text{bdiag} \mathbf{R}_{yy})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \left[-\mathbf{I} + \mathbf{R}_{yy} (\text{bdiag} \mathbf{R}_{yy})^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \left[\mathbf{R}_{yy} - \text{bdiag} \mathbf{R}_{yy} \right] \text{bdiag}^{-1} \mathbf{R}_{yy}. \tag{3.39}
\end{aligned}$$

Sendo encontrado o gradiente, a atualização da matriz \mathbf{W} é efetuada por:

$$\mathbf{W}(m) = \mathbf{W}(m-1) - \mu \nabla_{\mathbf{W}} \mathfrak{S}(m). \tag{3.40}$$

Cada diagonal das matrizes \mathbf{W}_{pq} que compõem \mathbf{W} - vide Equação (3.16) - possui elementos iguais; logo, podemos classificar as matrizes \mathbf{W}_{pq} como Toeplitz. O uso indiscriminado de um gradiente que atue sobre toda a matriz pode destruir esta característica, retirando a redundância que permite uma relação biunívoca entre as matrizes \mathbf{W}_{pq} e os filtros correspondentes (\mathbf{w}_{pq}). Por isso, devemos aplicar uma restrição ao gradiente (o qual é uma matriz) obtido, de forma a projetá-lo para o conjunto das matrizes Toeplitz. Em [11], mostra-se que a escolha dos L primeiros elementos da primeira coluna de \mathbf{W}_{pq} (ou da L -ésima linha da mesma matriz) é a melhor, para propósitos de otimização.

3.2.4 Gradiente Natural

O gradiente $\nabla_{\mathbf{W}}\mathfrak{S}(m)$ calculado na seção anterior não é utilizado, mas sua versão mais robusta, oriunda de uma formulação natural, já mostrada no Capítulo 2³. No caso em questão, o gradiente natural pode ser colocado em função do gradiente padrão da forma abaixo [11]:

$$\begin{aligned}
\nabla_{\mathbf{W}}^N \mathfrak{S} &= \mathbf{W}\mathbf{W}^H \nabla_{\mathbf{W}} \mathfrak{S} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W}\mathbf{W}^H \mathbf{R}_{\mathbf{x}\mathbf{y}} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W}\mathbf{W}^H \mathbf{X}^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W} (\mathbf{X}\mathbf{W})^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W}\mathbf{Y}^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W} \mathbf{R}_{\mathbf{y}\mathbf{y}} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W} [\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}] \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}}, \tag{3.41}
\end{aligned}$$

onde verificamos uma significativa redução da complexidade computacional: a inversa de $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ (matriz de dimensões $PD \times PD$) não precisa mais ser calculada. O termo $\text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}}$, devido à operação bdiag , para ser calculado só necessita da inversão de P matrizes de dimensões $D \times D$ (além disso, N não precisa ser maior ou igual a PD , mas apenas a D).

Para o caso $P = 2$, o gradiente natural pode ser escrito como:

$$\nabla_{\mathbf{W}}^N = 2 \sum_{i=0}^{\infty} \beta(i, m) \begin{bmatrix} \mathbf{W}_{12} \mathbf{R}_{\mathbf{y}2\mathbf{y}1} \mathbf{R}_{\mathbf{y}1\mathbf{y}1}^{-1} & \mathbf{W}_{11} \mathbf{R}_{\mathbf{y}1\mathbf{y}2} \mathbf{R}_{\mathbf{y}2\mathbf{y}2}^{-1} \\ \mathbf{W}_{22} \mathbf{R}_{\mathbf{y}2\mathbf{y}1} \mathbf{R}_{\mathbf{y}1\mathbf{y}1}^{-1} & \mathbf{W}_{21} \mathbf{R}_{\mathbf{y}1\mathbf{y}2} \mathbf{R}_{\mathbf{y}2\mathbf{y}2}^{-1} \end{bmatrix}, \tag{3.42}$$

³Rigorosamente falando, o gradiente apresentado não é “natural”, dado que não se baseia, por exemplo, nas propriedades de grupo de Lie [21] deste tipo de gradiente. Porém, a formulação apresentada foi largamente inspirada no gradiente natural para misturas instantâneas, além de apresentar a importante propriedade de se tornar praticamente independente do sistema de mistura [11].

onde as operações de inversão são precedidas de uma regularização, de forma a evitar matrizes próximas à singularidade. Um procedimento de regularização comum (e o utilizado nesta dissertação) é o de adicionar uma pequena constante positiva à diagonal principal da matriz a ser invertida [11]. Em [53], uma alternativa é proposta. Seja \mathbf{Z} a matriz a inverter e $\text{diag}(\mathbf{Z})$ uma matriz diagonal cuja diagonal principal coincide com a de \mathbf{Z} e ρ um parâmetro entre 0 e 1. A alternativa citada consiste em efetuar o seguinte procedimento:

$$\mathbf{Z}' \leftarrow \rho\mathbf{Z} + (1 - \rho)\text{diag}(\mathbf{Z}). \quad (3.43)$$

3.2.5 Implementações *On-Line*, *Block On-Line* e *Off-Line*

O parâmetro $\beta(i, m)$ na Equação (3.41) pode ser facilmente alterado de modo a permitir implementações *on-line*, *block on-line* e *off-line* do método GASFC.

Seja um sinal com um total de B blocos. Para uma iteração *off-line*, basta assumir $\beta(i, m) = 1/B$. Para uma iteração *on-line*, a alteração é feita à medida em que temos acesso a um novo bloco de dados, bastando fazer $\beta(i, m) = 1$ para o novo bloco e $\beta(i, m) = 0$ para os outros blocos (uma alternativa também utilizada seria $\beta(i, m) = (1 - \lambda)\lambda^{m-i}$ para $0 \leq i \leq m$ e $\beta(i, m) = 0$ no resto; esta alternativa pode ser expressa de forma recursiva, o que reduz enormemente o custo computacional; vide [56] para maiores detalhes). Uma iteração *off-line* costuma ser muito custosa computacionalmente, enquanto que uma *on-line* pode não obter bons resultados em casos de filtros de separação com muitos coeficientes. Por isso, os autores de [11] propõem uma abordagem híbrida, por eles intitulada de *block on-line*, na qual apenas os últimos K blocos influenciam o gradiente, com $\beta(i, m) = 1/K$ somente nestes blocos, sendo 0 nos restantes.

3.2.6 Avaliação do Método GASFC

Para avaliar o desempenho da separação por meio do método GASFC, normalmente é utilizado o parâmetro SIR médio. Para processos de mistura e de separação convolutivos, podemos expressar o SIR (em dB) da i -ésima fonte na j -ésima

saída por:

$$\text{SIR}_i^j = 10 \log_{10} \frac{\text{E} \left[\left(\sum_{p=1}^P (s_i * h_{ip} * w_{pj})(n) \right)^2 \right]}{\text{E} \left[\left(\sum_{p=1}^P \sum_{k=1, k \neq i}^P (s_k * h_{kp} * w_{pj})(n) \right)^2 \right]} \quad (3.44)$$

3.3 Método II: Separação Cega de Fontes Explorando as Dependências de Alta Ordem entre Raias

Um método recente de separação cega de fontes (proposto em [12], e daqui em diante referenciado por EDAO) apresenta um novo paradigma para métodos de separação do tipo banda larga. Por meio de uma redefinição da probabilidade *a priori* dos sinais das fontes, um método de banda estreita é facilmente convertido num método do tipo banda larga.

Seja a função q definida na Equação (2.20). Como vimos, este “parâmetro de *nuisance*” pode ser entendido como uma função distribuição de probabilidade do sinal a ser estimado (a fonte). Comumente, ele é unidimensional: supõe-se que cada amostra dos sinais das fontes é obtida segundo a distribuição q (em geral supergaussiana), de forma independente e identicamente distribuída. No Capítulo 2, notamos que uma eventual dependência entre amostras - comum, por exemplo, em sinais de áudio - não inutiliza esta hipótese e não costuma degradar os resultados.

O método EDAO tem por primeiro passo a transformação das misturas para o domínio da frequência. Esta transformação (por meio de uma STFT) é precedida pela divisão dos sinais das misturas em quadros (janelados com janela de Hanning e com sobreposição).

O acima exposto permite-nos classificar o método EDAO qual um método no domínio da frequência; devido a esta característica, este método apresenta rápida convergência, aliada a um baixo custo computacional.

Uma abordagem do tipo banda estreita atenta para cada raia de forma independente, aplicando normalmente algum método de ICA instantâneo. A inovação do método EDAO consiste em inserir na função custo uma dependência entre todas as raias. Para tal fim, ela torna a distribuição q multivariada.

No Capítulo 2, chegamos à função custo expressa na Equação (2.26) por uma abordagem bayesiana. Porém, a forma mais comum de se obter a função custo supracitada consiste em explicitar a divergência de Kullback-Leibler entre duas variáveis aleatórias (uma variável aleatória é representada pela estimativa das fontes e a outra é regida pela distribuição *a priori* arbitrariamente eleita para representar a distribuição das fontes). O objetivo então residiria numa minimização desta divergência, minimização que resultaria em estimativas das fontes cada vez menos gaussianas. Como a soma de duas variáveis aleatórias tende a ser mais gaussiana do que cada variável em separado, maximizar a não-gaussianidade das saídas torna-se um procedimento justificável. Diferentes abordagens (bayesiana, redução da divergência de Kullback-Leibler ou maximização da não-gaussianidade) levam à função custo C_{bl} (já expressa na Equação (2.26), adaptada aqui para o caso de diferentes matrizes em cada raia):

$$C_{bl} = \text{const.} - \sum_{l=1}^L \log \left| \det \mathbf{G}^{(l)} \right| - \sum_{p=1}^P \text{E} [\log q(\tilde{\mathbf{s}}_p)], \quad (3.45)$$

onde L é o comprimento dos quadros (ou o número de raias, equivalente ao comprimento dos filtros de separação), P é o número de fontes e $\tilde{\mathbf{s}}_p$ é o vetor gerado pela STFT (um vetor é obtido a partir de cada quadro). Já a matriz $\mathbf{G}^{(l)}$ é a matriz de separação aplicada à l -ésima raia das misturas, no mesmo formato do presente na Equação (2.5). Notamos, portanto, a existência de L diferentes matrizes a estimar. O índice bl faz referência a “banda larga”.

Se definirmos a função q como unidimensional (associada a cada raia) e as raias como independentes e identicamente distribuídas, podemos reescrever a função custo da forma a seguir:

$$\begin{aligned} C_{be} &= \text{const.} - \sum_{l=1}^L \log \left| \det \mathbf{G}^{(l)} \right| - \sum_{p=1}^P \sum_{l=1}^L \text{E} [\log q(\tilde{s}_p^{(l)})] \\ &= \text{const.} - \sum_{l=1}^L \left\{ \log \left| \det \mathbf{G}^{(l)} \right| - \sum_{p=1}^P \text{E} [\log q(\tilde{s}_p^{(l)})] \right\}, \end{aligned} \quad (3.46)$$

onde $\tilde{s}_p^{(l)}$ é a l -ésima raia oriunda da p -ésima fonte.

O problema da função custo C_{be} (o índice be é oriundo de “banda estreita”) é que sua minimização pode, sem perda da otimalidade da solução, ser desmembrada em diversos subproblemas, cada qual visando a minimização de uma “sub-

função” custo em cada raia l . Devido à independência assim presumida entre raias, mesmo que obtivéssemos uma separação razoável em cada raia, caberia posteriormente resolver a questão da permutação.

A grande contribuição do método EDAO consiste numa densidade q multidimensional. Esta densidade descreve simultaneamente *todas* as raias. Esta descrição conjunta permite que, durante o processo de otimização, dependências estatísticas entre raias distintas sejam devidamente contempladas. Isto permite que, após a convergência, as matrizes $\mathbf{G}^{(l)}$ não apresentem permutações das estimativas.

Como somente a função custo C_{bl} tem o potencial de evitar o problema da permutação, ela será a abordada nesta seção. Expressando seu gradiente (natural) em função de cada elemento $g_{ij}^{(l)}$, temos:

$$\Delta g_{ij}^{(l)} = -\frac{\partial^N C_{bl}}{\partial g_{ij}^{(l)}} = \sum_{p=1}^L \left(I_{ip} - \mathbb{E} \left[\Phi^{(l)} \left(\tilde{s}_i^{(1)} \dots \tilde{s}_i^{(L)} \right) \right] \right) g_{pj}^{(l)}, \quad (3.47)$$

onde $\Phi^{(l)}(\cdot)$ é a função *score* (já vista no Capítulo 2) na l -ésima raia, sendo agora dada por:

$$\Phi^{(l)} \left(\tilde{s}_i^{(1)} \dots \tilde{s}_i^{(L)} \right) = -\frac{\partial \log q \left(\tilde{s}_i^{(1)} \dots \tilde{s}_i^{(L)} \right)}{\partial \tilde{s}_i^{(l)}}. \quad (3.48)$$

Nota-se na função acima que a função *score* numa dada raia é influenciada pelos valores de todas as outras, como desejado.

A distribuição *a priori* das fontes proposta pelo método EDAO é:

$$q(\tilde{\mathbf{s}}_i) = \alpha e^{-\sqrt{(\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)^H \Sigma_i^{-1} (\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)}}, \quad (3.49)$$

onde α é um fator de normalização e $\boldsymbol{\mu}_i$ e Σ_i são respectivamente, o vetor média e a matriz de covariância do sinal da i -ésima fonte.

Admitindo que as raias na frequência são descorrelacionadas entre si (o que não implica independência), podemos tornar diagonal o termo de covariância Σ_i . Supondo média nula dos coeficientes na frequência, podemos então reescrever a função *score* da forma a seguir:

$$q(\tilde{\mathbf{s}}_i) = \alpha e^{-\sqrt{\sum_k \left| \frac{s_i^{(l)}}{\sigma_i^{(l)}} \right|^2}}, \quad (3.50)$$

onde $\sigma_i^{(l)}$ é a variância da i -ésima fonte na l -ésima raia, sendo um fator de escala de cada elemento do vetor dos coeficientes da STFT de cada fonte. Como este fator

é desconhecido, o método EDAO o supõe unitário; o escalamento é resolvido pelo método da mínima distorção, já detalhado no início deste capítulo. Assim, a função *score* multidimensional é dada por:

$$\Phi^{(l)}(\tilde{s}_i^1 \dots \tilde{s}_i^L) = \frac{\partial \sqrt{\sum_{l=1}^L |\tilde{s}^{(l)}|^2}}{\partial \tilde{s}_i^{(l)}} = \frac{\tilde{s}_i^{(l)}}{\sqrt{\sum_{l=1}^L |\tilde{s}_i^{(l)}|^2}}. \quad (3.51)$$

A inserção desta função *score* na Equação (3.47) fornece-nos o método iterativo de otimização do algoritmo EDAO, da forma a seguir:

$$g_{ij}^{(l)} = g_{ij}^{(l)} + \mu \Delta g_{ij}^{(l)}, \quad (3.52)$$

onde μ é um fator de aprendizagem, geralmente menor que 0,1.

3.3.1 Avaliação do Método EDAO

A avaliação do método EDAO é realizada mediante a razão sinal-interferência. A medida de SIR é efetuada no domínio da frequência, sendo calculada por [12]:

$$\text{SIR}_{\text{inicial}} = 10 \log_{10} \left(\frac{\sum_{n,l} \left| \sum_i h_{iq(i)}^{(l)} \tilde{s}_{q(i)}^{(l)}[n] \right|^2}{\sum_{n,l} \left| \sum_{i \neq j} h_{iq(j)}^{(l)} \tilde{s}_{q(j)}^{(l)}[n] \right|^2} \right), \quad (3.53)$$

$$\text{SIR}_{\text{final}} = 10 \log \left(\frac{\sum_{n,l} \left| \sum_i c_{iq(i)}^{(l)} \tilde{s}_{q(i)}^{(l)}[n] \right|^2}{\sum_{n,l} \left| \sum_{i \neq j} c_{iq(j)}^{(l)} \tilde{s}_{q(j)}^{(l)}[n] \right|^2} \right), \quad (3.54)$$

onde $q(i)$ indica a saída onde a fonte i aparece e $c_{iq(i)}^{(l)}$ é a resposta ao impulso global (ou seja, dos sistemas de mistura e de separação), definida por:

$$c_{iq(j)}^{(l)} = \sum_m g_{im}^{(l)} h_{mq(j)}^{(l)}. \quad (3.55)$$

3.4 Estimativa do Ângulo de Chegada

Os métodos de separação de fontes em misturas convolutivas que utilizam a abordagem de banda estreita têm que lidar, como vimos, com o problema da permutação. Uma estimativa, em cada raia, da direção de chegada do sinal de cada fonte permitiria, a princípio, distinguir as componentes de fontes distintas. Nesta seção, detalharemos um método, utilizado em [3], para a estimativa da direção de

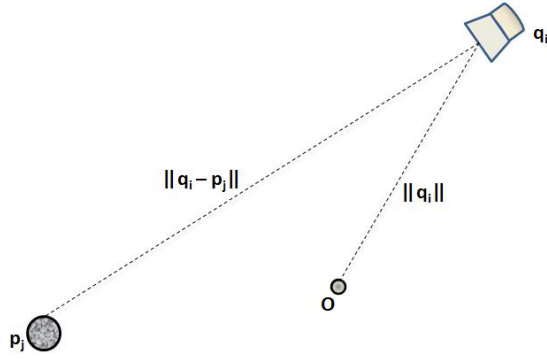


Figura 3.1: Ilustração das posições do sensor j e da fonte i , com a origem no ponto O .

chegada (DOA, do inglês *Direction of Arrival*), aplicável em cada raia na frequência.

Admitindo um modelo de mistura instantânea em cada raia, temos diversas matrizes de mistura $\mathbf{H}(f)$ (uma para cada raia), correspondendo à resposta em frequência do sistema de mistura. Definindo os vetores de posição \mathbf{p}_j e \mathbf{q}_i do j -ésimo sensor e da i -ésima fonte (vide a Figura 3.1 para melhor visualização), uma modelagem do campo próximo (a qual assume uma atenuação inversamente proporcional à distância) permite-nos aproximar o coeficiente $h_{ji}(f)$ da matriz $\mathbf{H}(f)$ da forma a seguir ⁴:

$$h_{ji}(f) \approx \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j\frac{2\pi f}{c}(\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i\|)}, \quad (3.56)$$

onde c é a velocidade do som (supondo sinais de áudio).

Na prática, não temos acesso às matrizes de mistura $\mathbf{H}(f)$; porém, as inversas das matrizes de separação $\mathbf{W}(f)$ (normalmente obtidas com métodos de análise de componentes independentes) permitem-nos estimar as matrizes $\mathbf{H}(f)$. Sejam $\mathbf{A}(f)$ as matrizes $\mathbf{H}(f)$ estimadas mediante este procedimento. Supondo estimativas razoavelmente precisas e efetuando a razão entre dois elementos de $\mathbf{A}(f)$ que estejam na mesma coluna i e em linhas distintas (j e j'), obtemos:

$$\frac{a_{ji}(f)}{a_{j'i}(f)} \approx \frac{\alpha_i h_{ji}(f)}{\alpha_i h_{j'i}(f)} = \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j\frac{2\pi f}{c}(\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|)}, \quad (3.57)$$

onde α_i é uma constante de proporcionalidade oriunda da ambigüidade de escalamento. Atentando para a fase do primeiro e terceiro termos da equação acima,

⁴A ambigüidade de escalamento permite-nos elidir uma constante de proporcionalidade.

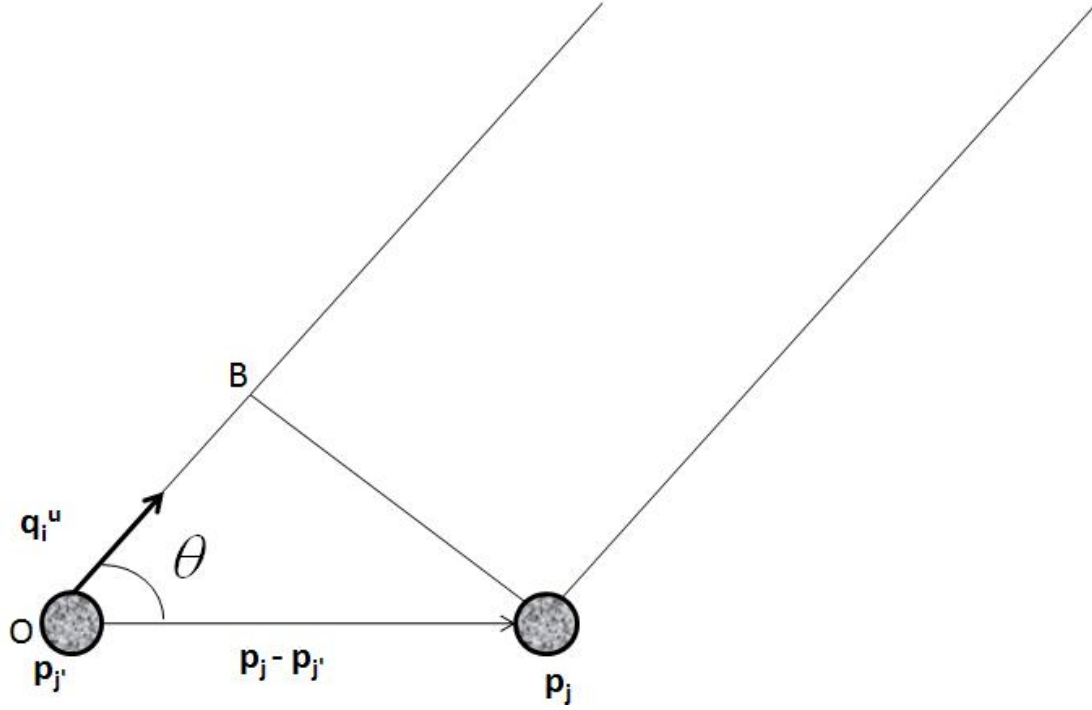


Figura 3.2: Ilustração da suposição de paralelismo do modelo de campo distante.

encontramos:

$$\angle \left(\frac{a_{ji}(f)}{a_{j'i}(f)} \right) \approx \frac{2\pi f}{c} (\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|), \quad (3.58)$$

de onde se obtém:

$$\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\| \approx \frac{\angle \left(\frac{a_{ji}(f)}{a_{j'i}(f)} \right)}{\frac{2\pi f}{c}}. \quad (3.59)$$

Para estimarmos a direção de chegada dos sinais da i -ésima fonte, utilizaremos um modelo de campo distante. Neste modelo, os vetores $\|\mathbf{q}_i - \mathbf{p}_{j'}\|$ e $\|\mathbf{q}_i - \mathbf{p}_j\|$ são considerados paralelos (baseando-se na suposição de que as fontes encontram-se muito mais distantes dos sensores do que estes entre si). Arbitrando a posição do j' -ésimo sensor como sendo a origem, a Figura 3.2 mostra-nos que podemos associar a direção de chegada da i -ésima fonte a um ângulo θ (nesta figura, \mathbf{q}_i^u é o vetor unitário na direção de $\mathbf{q}_i - \mathbf{p}_{j'} = \mathbf{q}_i$; assim, $\mathbf{q}_i^u = \frac{\mathbf{q}_i}{\|\mathbf{q}_i\|}$). Admitindo que os vetores supracitados sejam paralelos, a diferença entre eles - o primeiro termo da Equação (3.59) - pode ser aproximada pela distância do ponto \mathbf{O} ao ponto \mathbf{B} , o qual também é a projeção de $\mathbf{p}_j - \mathbf{p}_{j'}$ na direção \mathbf{q}_i^u . Essas considerações nos permitem reescrever a Equação (3.59) como:

$$[\mathbf{p}_j - \mathbf{p}_{j'}]^T \frac{\mathbf{q}_i}{\|\mathbf{q}_i\|} \approx \frac{\angle \left(\frac{a_{ji}(f)}{a_{j'i}(f)} \right)}{\frac{2\pi f}{c}}, \quad (3.60)$$

e, já que a suposição de paralelismo implica um ângulo de 90 graus no ponto \mathbf{B} , o cosseno de θ pode ser estimado como:

$$\cos\theta = \frac{[\mathbf{p}_j - \mathbf{p}_{j'}]^T \mathbf{q}_i}{\|\mathbf{p}_j - \mathbf{p}_{j'}\| \|\mathbf{q}_i\|} \approx \frac{\angle \left(\frac{a_{ji}(f)}{a_{j'i}(f)} \right)}{\frac{2\pi f}{c} \|\mathbf{p}_j - \mathbf{p}_{j'}\|}. \quad (3.61)$$

Admitindo-se P fontes, cada matriz $\mathbf{A}(f)$ permite a estimativa de P valores de θ . Como vimos, estas estimativas podem ser utilizadas para resolver o problema da permutação nas abordagens do tipo banda estreita.

3.5 Propostas

Já que as aplicações *on-line* do método GASFC comumente exigem que este convirja num intervalo de tempo muito reduzido (em torno de décimos de segundo), neste capítulo estudamos propostas para aceleração da convergência do método GASFC, na sua formulação natural e *on-line*.

São também avaliadas as possibilidades de se utilizar filtros de separação de comprimento superior ao comprimento dos filtros de mistura, bem como a possibilidade de se recorrer a filtros IIR.

Por fim, contemplaremos duas propostas para o método EDAO: uma que incorpora a ele a estimativa de direção de chegada e outra que introduz uma modificação na sua distribuição *a priori* das fontes.

As simulações efetuadas neste capítulo contemplam o caso de duas fontes e duas misturas ($P = 2$).

3.5.1 Adoção de uma Taxa de Aprendizagem Distinta para Cada Parâmetro

Na abordagem proposta em [11], o parâmetro μ - o fator de aprendizagem, utilizado na equação de atualização (3.40) - é considerado fixo (invariante no tempo) e idêntico para todos os parâmetros $w_{pq}(k)$. Como freqüentemente são necessários centenas de parâmetros a estimar (coeficientes de filtros FIR), deve-se indagar se um fator de aprendizagem constante para todos os parâmetros é uma boa estratégia. Com efeito, podemos empregar algumas heurísticas para o fator de aprendizagem

que podem resultar numa aceleração da convergência do método [54],[55]. Eis, em resumo, quatro heurísticas comumente utilizadas:

- (I) cada parâmetro a ajustar deve possuir seu próprio fator de aprendizagem;
- (II) o fator de aprendizagem deve variar a cada iteração;
- (III) quando as alterações em um determinado parâmetro possuírem o mesmo sinal em iterações consecutivas, seu fator de aprendizagem deve ser aumentado;
- (IV) quando o sinal das alterações em um determinado parâmetro em iterações consecutivas se alternar, seu fator de aprendizagem correspondente deve ser reduzido.

Nossa proposta inicial (GASFC-VM-I) consiste em verificar as alterações em cada parâmetro $w_{pq}(k)$ entre duas iterações consecutivas. Caso estas alterações possuam o mesmo sinal, o fator de aprendizagem correspondente - denominado $\mu_{pq}(k)$ - será multiplicado por α . Caso contrário, $\mu_{pq}(k)$ será dividido por α , onde α é ligeiramente maior do que 1.

Para simular os efeitos da nossa proposta, seja uma configuração com duas fontes, onde os filtros de mistura tenham comprimento 8 e sejam dados por:

$$\begin{aligned}
 h_{11} &= [0,7638 \quad 0,6216 \quad -0,0261 \quad -0,8595 \quad -0,1722 \quad -0,4489 \quad 0,0483 \quad 0,0731] \\
 h_{12} &= [0,6139 \quad 0,1168 \quad 0,1443 \quad -0,5389 \quad -0,8738 \quad 0,0462 \quad 0,1373 \quad 0,3542] \\
 h_{21} &= [0,3451 \quad 1,0051 \quad -0,0896 \quad -0,5774 \quad -0,2202 \quad -0,2143 \quad -0,1536 \quad -0,0954] \\
 h_{22} &= [1,2571 \quad 0,0762 \quad 0,0898 \quad -0,1719 \quad -1,0118 \quad -0,2630 \quad -0,1434 \quad 0,1670]
 \end{aligned}$$

Utilizando filtros de comprimento 6 para separar dois segundos de sinais de voz (uma feminina, outra masculina, amostrados em 8 kHz)⁵, temos nas Figuras 3.3 e 3.4 a razão sinal-ruído (em dB) do algoritmo original e do proposto, com diferentes parâmetros μ e α (em todos os experimentos com o método GASFC deste capítulo, a constante adicionada às matrizes para sua regularização antes do cálculo de suas inversas foi de 10^{-9}). Inicialmente, todos os coeficientes $\mu_{pq}(k)$ são idênticos ao μ utilizado no método GASFC. A primeira constatação que as Figuras 3.3 e 3.4 nos

⁵Em todas as simulações desta subseção, esta configuração foi utilizada.

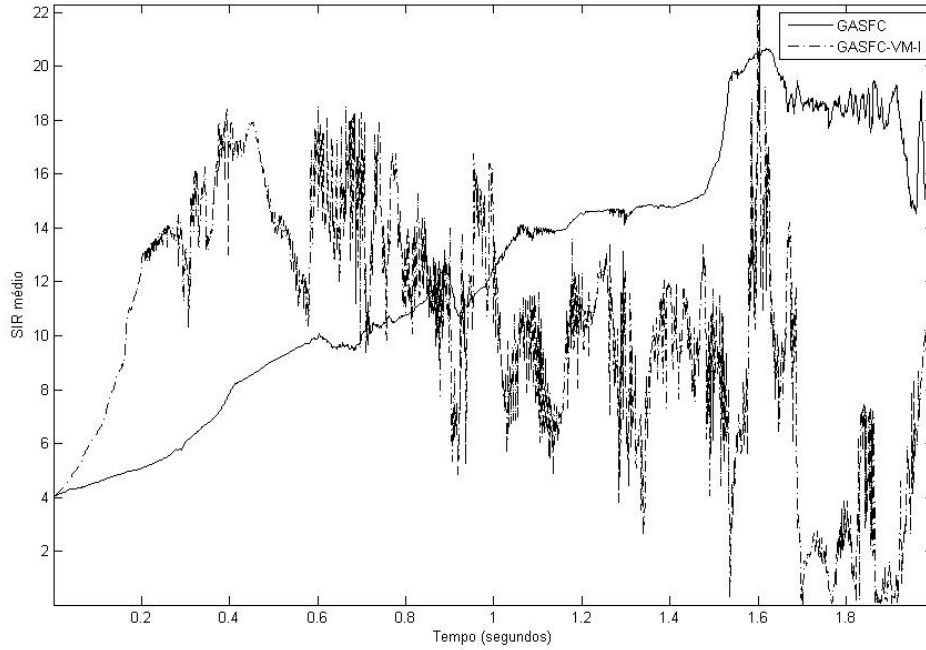


Figura 3.3: SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-I (proposto), para $\alpha = 1,01$, $L = 6$ e $\mu = 3 \times 10^{-4}$.

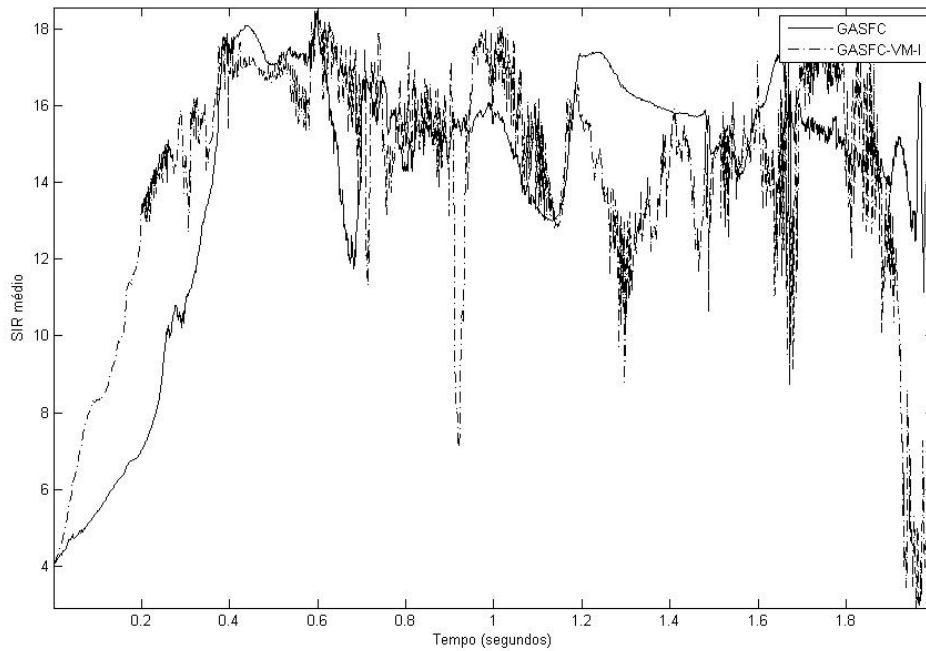


Figura 3.4: SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-I (proposto), para $\alpha = 1,005$, $L = 6$ e $\mu = 7 \times 10^{-4}$.

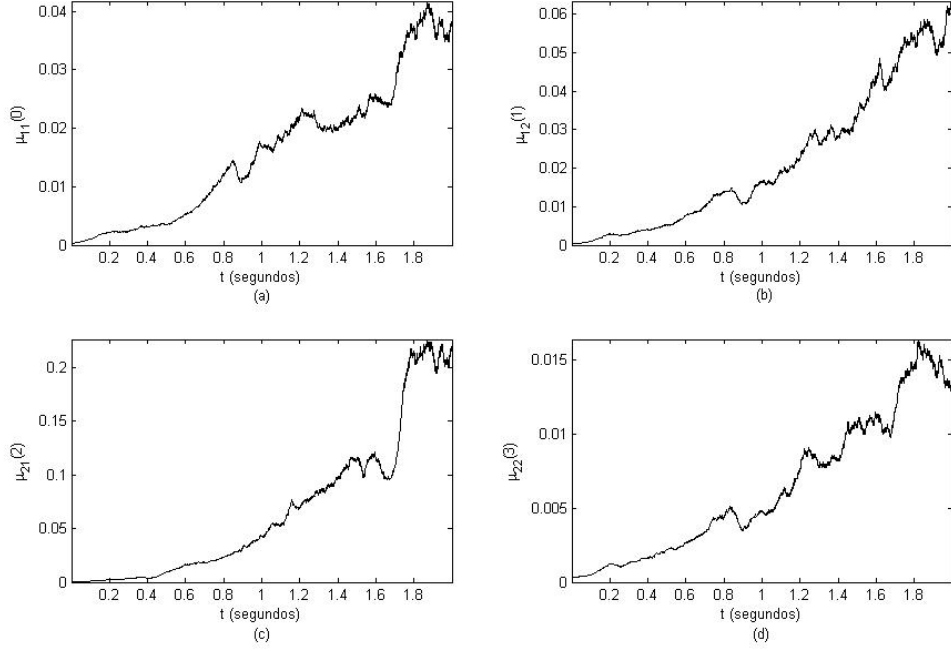


Figura 3.5: Evolução dos parâmetros $\mu_{pq}(k)$, em função do tempo t (em segundos). (a) $\mu_{11}(0)$; (b) $\mu_{12}(1)$; (c) $\mu_{21}(2)$ e (d) $\mu_{22}(3)$.

fornece é que o algoritmo proposto obtém uma taxa de convergência inicial significativamente superior à do algoritmo convencional. Porém, após a convergência, verificamos um grande decréscimo do desempenho (especialmente na Figura 3.3).

Para entender este comportamento, cumpre recorrer à evolução temporal de $\mu_{pq}(k)$. A Figura 3.5 evidencia os valores de $\mu_{11}(0)$, $\mu_{12}(1)$, $\mu_{21}(2)$ e $\mu_{22}(3)$ em relação ao tempo. Cabe ressaltar que todos os parâmetros $\mu_{pq}(k)$ seguem uma evolução estreitamente semelhante, tendendo a valores muito altos, o que acaba por provocar instabilidade. Esta instabilidade é basicamente a causa da queda de desempenho do método GASFC-VM-I. Para evitar esta instabilidade, basta arbitrarmos um limite superior para todos os coeficientes $\mu_{pq}(k)$. Se μ_0 é o valor inicial de todos os $\mu_{pq}(k)$, nossa proposta de modificação no método GASFC-VM-I consiste em estabelecer um limite superior igual a $C\mu_0$ (onde C é uma constante positiva maior que 1) para todos $\mu_{pq}(k)$. Essa modificação gera uma variante do GASFC-VM-I, denominada GASFC-VM-II.

As Figuras 3.6 e 3.7 mostram o desempenho dos métodos GASFC e GASFC-VM-II para $C = 8$ e $C = 2$, utilizando os mesmos parâmetros dos que constam na

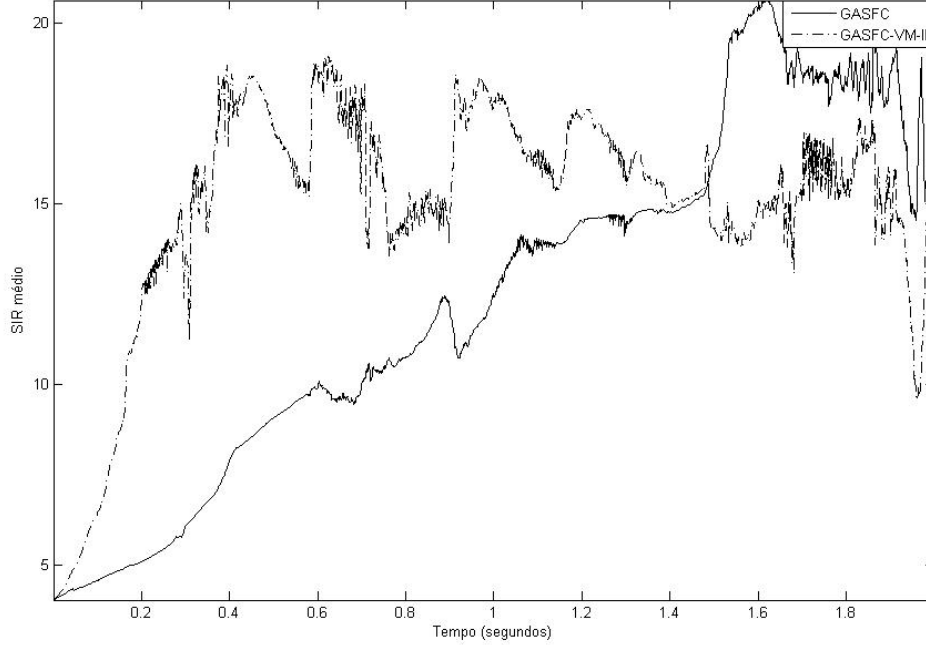


Figura 3.6: SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-II (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $L = 6$ e $C = 8$.

Figura 3.3. A comparação destas figuras nos mostra que o valor mais baixo de C apresenta uma convergência mais regular, embora a taxa de convergência seja ligeiramente menor. Notamos novamente que, após a convergência, ambos os métodos apresentam uma redução de desempenho, caso comparados ao método original. Quanto mais próximo da unidade for C , menor esta tendência de degradação. Embora melhor do que o método GASFC-VM-I, a alternativa GASFC-VM-II ainda possui uma tendência - bastante atenuada - à degradação após a convergência.

Embora os valores de μ estejam razoáveis (no sentido de não serem muito grandes) no método GASFC-VM-II, sua tendência de degradação (após a convergência) pode ser explicada por meio de uma analogia com o célebre método LMS em filtragem adaptativa [57], onde um fator de aprendizagem mais elevado tende a aumentar o erro quadrático médio após a convergência.

Estas considerações motivam uma modificação no GASFC-VM-II, a qual permitiria que o limite superior de $\mu_{pq}(k)$ seja variável, sendo alto no início do algoritmo (de forma a garantir uma convergência mais rápida) e menor após um determinado tempo (para evitar a degradação de desempenho oriunda de um fator μ

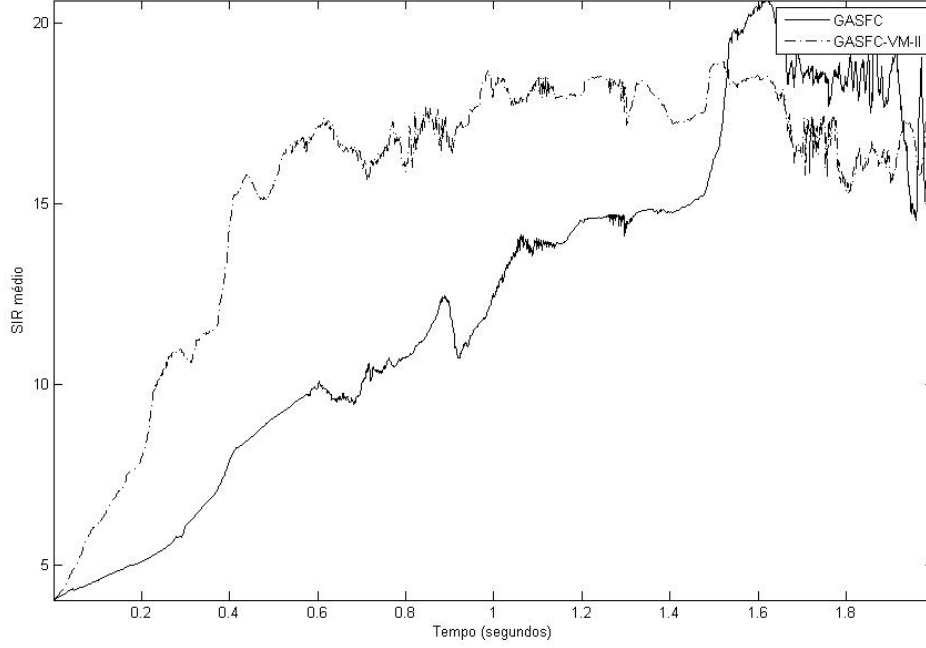


Figura 3.7: SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-II (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $L = 6$ e $C = 2$.

relativamente alto). Assim, o novo algoritmo (GASFC-VM-III) proposto apresenta C no i -ésimo bloco de dados expresso por:

$$C(i) = 1 + (C_0 - 1)e^{-\lambda(i-1)}, \quad (3.62)$$

onde λ é um novo parâmetro, que ajusta a taxa com que $C(i)$ tende a 1. A equação acima nos permite observar que $C(1) = C_0$ e que $\lim_{i \rightarrow \infty} C(i) = 1$ (ou seja, o algoritmo, após acelerar a convergência, tende a se confundir com o método GASFC padrão). Utilizando os mesmos parâmetros das Figuras 3.6 e 3.7, o desempenho comparado do algoritmo GASFC-VM-III é mostrado na Figura 3.8. Como esperado, o método GASFC-VM-III mantém uma alta taxa de convergência (semelhante ao do método GASFC-VM-I), porém sem degradação do desempenho após a convergência (como ocorre com os métodos GASFC-VM-I e GASFC-VM-II). Em nossas simulações, a performance do algoritmo GASFC-VM-III tende a ser melhor do que a do original, mesmo mediante grandes variações dos parâmetros C_0 e λ .

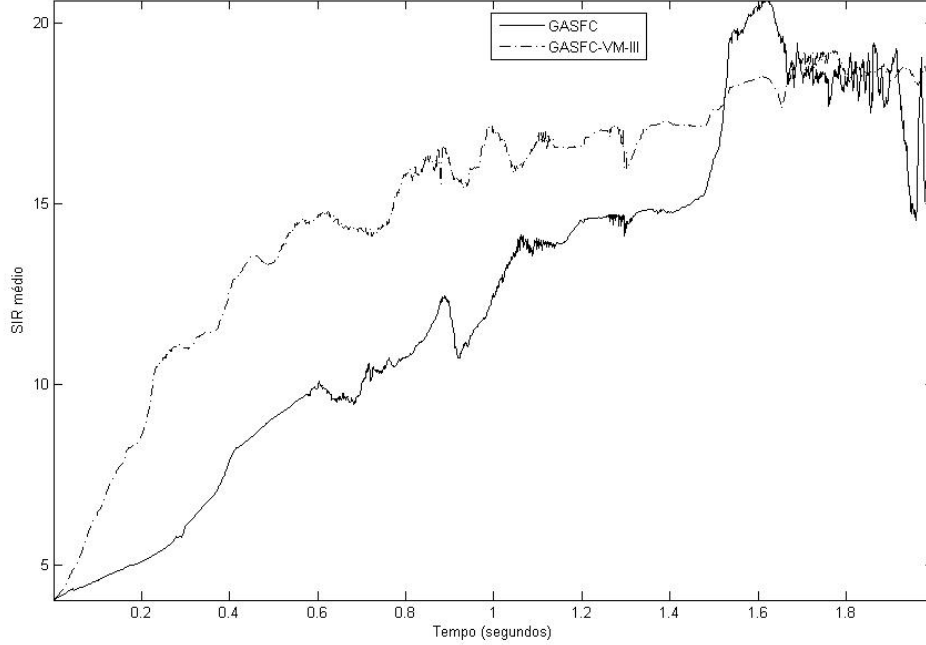


Figura 3.8: SIR médio (em dB) dos métodos GASFC (original) e GASFC-VM-III (proposto), para $\alpha = 1,01$, $\mu = 3 \times 10^{-4}$, $C_0 = 10$, $L = 6$ e $\lambda = 0,01$.

3.5.2 Atualização Recursiva das Matrizes

Na configuração *on-line* do método GASFC, as matrizes dos dados a cada iteração são obtidas apenas a partir do bloco de dados atual. As estimativas de correlação oriundas destes blocos são muito ruidosas, por serem estocásticas. Uma forma (que acarreta reduzido acréscimo de carga computacional) de se inserir informações oriundas de blocos anteriores (não necessitando das formulações *block on-line* ou *off-line*), seria atualizar as matrizes de correlação utilizadas no gradiente de forma recursiva.

Estas considerações motivaram-nos o recurso a uma estimativa recursiva das matrizes $\mathbf{R}_{\mathbf{y}_p \mathbf{y}_q}$ em (2.42), da forma a seguir:

$$\mathbf{R}_{\mathbf{y}_p \mathbf{y}_q}^{nova} = \gamma \mathbf{R}_{\mathbf{y}_p \mathbf{y}_q}^{antiga} + (1 - \gamma) \mathbf{R}_{\mathbf{y}_p \mathbf{y}_q}^{atual}, \quad (3.63)$$

onde γ , denominado *fator de lembrança*, é geralmente um número ligeiramente inferior a 1.

Denominamos esta proposta de GASFC_{rec} , a qual possibilita o reuso das informações contidas em blocos anteriores (por meio da recursão), permitindo um

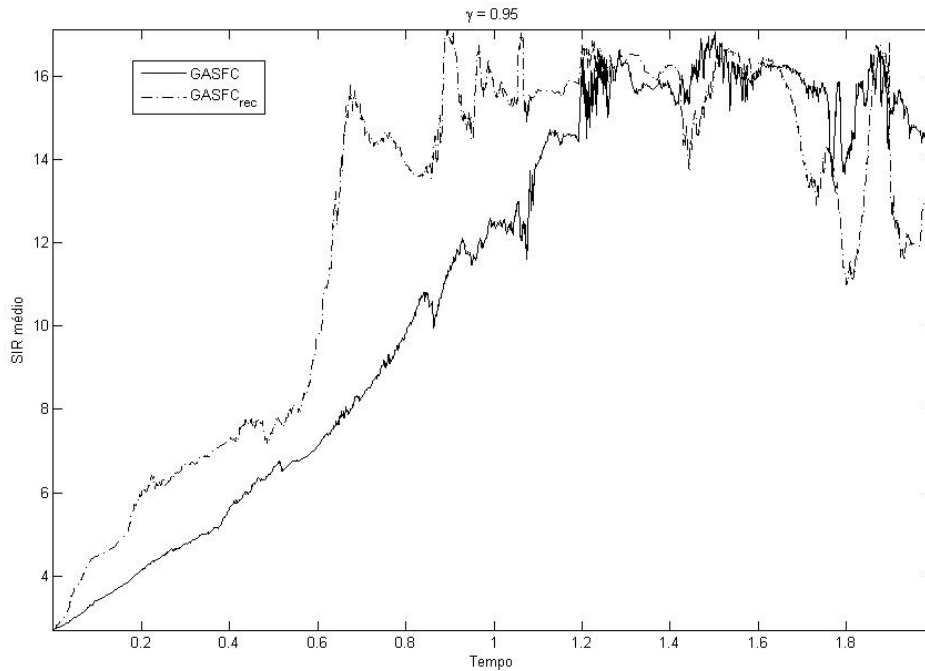


Figura 3.9: SIR médio (em dB) dos métodos GASFC (original) e GASFC_{rec} (proposto), para $\lambda = 0,95$, $\mu = 3 \times 10^{-4}$ e $L = 5$.

aumento da convergência do método GASFC.

Para efeitos de comparação, simulamos os métodos GASFC e GASFC_{rec} para os mesmos sinais e filtros de mistura utilizados na proposta anterior (com uma única diferença: os filtros de separação agora possuem comprimento 5, ao invés de 6).

As Figuras 3.9 e 3.10 comparam ambos os métodos (em relação à razão sinal-interferência) para dois valores distintos de γ . A Figura 3.10 mostra-nos que o método proposto pode acelerar enormemente a convergência, com um pequeno acréscimo de custo computacional. Eis um resultado muito importante, já que o tempo de convergência é um parâmetro muito crítico na prática.

Porém, estes resultados evidenciam uma queda de desempenho do método proposto (em relação ao método original) após a convergência, de forma muito semelhante à que ocorreu na proposta anterior.

Esta queda pode ser atribuída ao fato de que, após a convergência, as perturbações no gradiente natural introduzidas pela recursão se tornam fonte de ruído. Uma possibilidade de se contornar este problema seria efetuar algo semelhante à proposta anterior (GASFC-VM-III), de variar γ de forma que, no início da atual-

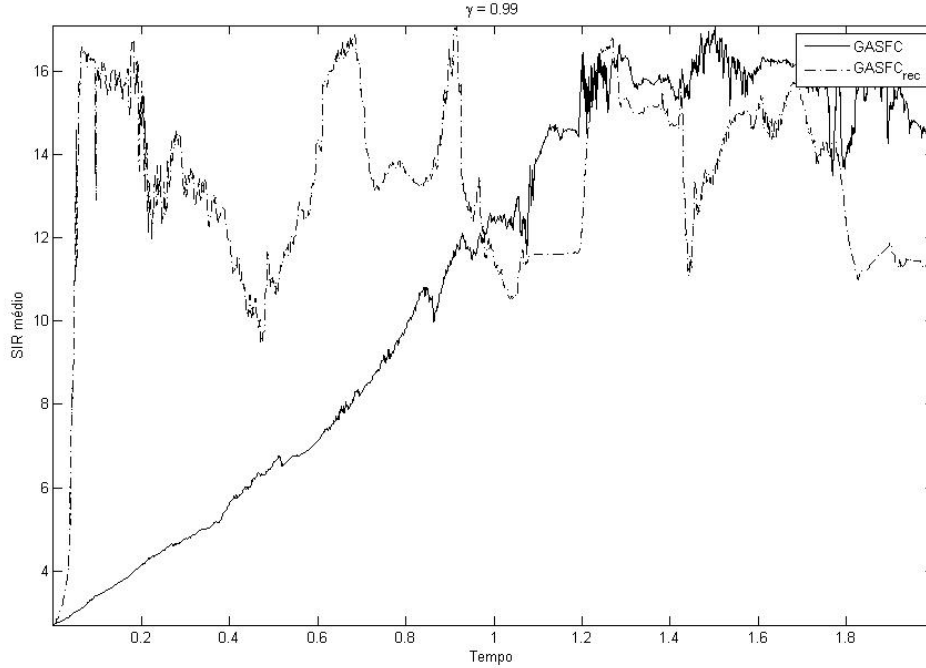


Figura 3.10: SIR médio (em dB) dos métodos GASFC (original) e GASFC_{rec} (proposto), para $\lambda = 0,99$, $\mu = 3 \times 10^{-4}$ e $L = 5$.

ização, γ seja alto, sendo então lentamente reduzido à medida em que o sistema de separação se estabilize. Não efetuamos simulações neste sentido.

3.5.3 Aumento do Comprimento dos Filtros de Separação

Os métodos de separação cega de fontes utilizam filtros FIR de separação cujo comprimento não excede o comprimento dos filtros de mistura ⁶. A exceção por nós encontrada (a referência [58]) argumenta que, no método de separação do tipo banda estreita utilizado, filtros de separação de comprimento maior do que os filtros de mistura degradam o resultado. Em [11], há uma argumentação, fundamentada no artigo [59], que considera, para efeitos de deconvolução, idealmente, $L \geq \frac{M(Q-1)}{P-M}$ (onde P é o número de fontes, M o de misturas e Q é o comprimento dos filtros de mistura). Para o caso $P = M = 2$, esta restrição gera $L \geq Q - 1 \approx Q$;

⁶Esta restrição supõe um conhecimento *a priori* dos comprimentos dos filtros; na realidade, a literatura defende que, caso sejam conhecidos estes comprimentos, os comprimentos dos filtros de separação não devem excedê-los.

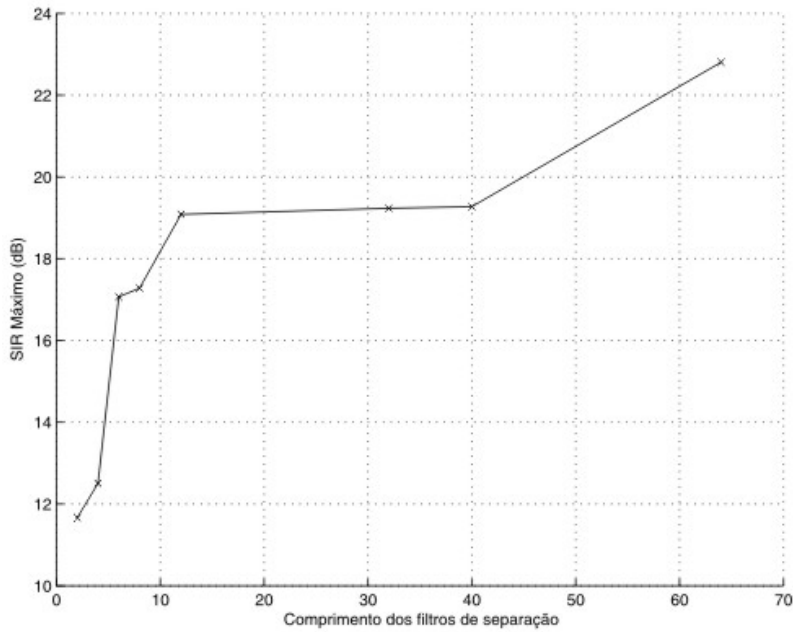


Figura 3.11: SIR máximo obtido para diferentes comprimentos L dos filtros de separação.

em geral, utiliza-se $L \leq Q$. Não temos conhecimento de artigos que recomendem a condição $L > Q$.

Embora o valor $L = \frac{M(Q-1)}{P-M}$ possa obter, em alguns casos, a deconvolução, na prática comprimentos de filtros maiores podem apresentar desempenho superior na separação.

Para avaliar esta possibilidade, utilizamos os mesmos filtros de mistura das propostas anteriores ($Q = 8$) e o método GASFC na sua formulação *off-line* com $\mu = 5 \times 10^{-3}$. Os sinais de voz utilizados (um de voz masculina, outro de feminina, amostrados em 8 kHz) apresentam duração de 10 segundos. A Figura 3.11 exhibe o SIR máximo da saída em função do comprimento dos filtros de separação (estes resultados já foram apresentados em nosso artigo [2]). Nesta figura, verificamos que utilizar filtros de separação com comprimentos maiores do que os de mistura ($L > Q$) pode resultar em ganhos significativos em termos de SIR. O valor $L = 64 = 8Q$, por exemplo, implicou uma melhora de 6 dB no SIR. Este resultado representa uma alternativa interessante, a despeito da possibilidade de a igualdade $L = \frac{M(Q-1)}{P-M}$, em alguns casos, permitir uma separação (quase) perfeita.

Embora haja desvantagem de maior custo computacional ao se utilizar $L > Q$, verificamos outra vantagem desta configuração: uma maior estabilidade

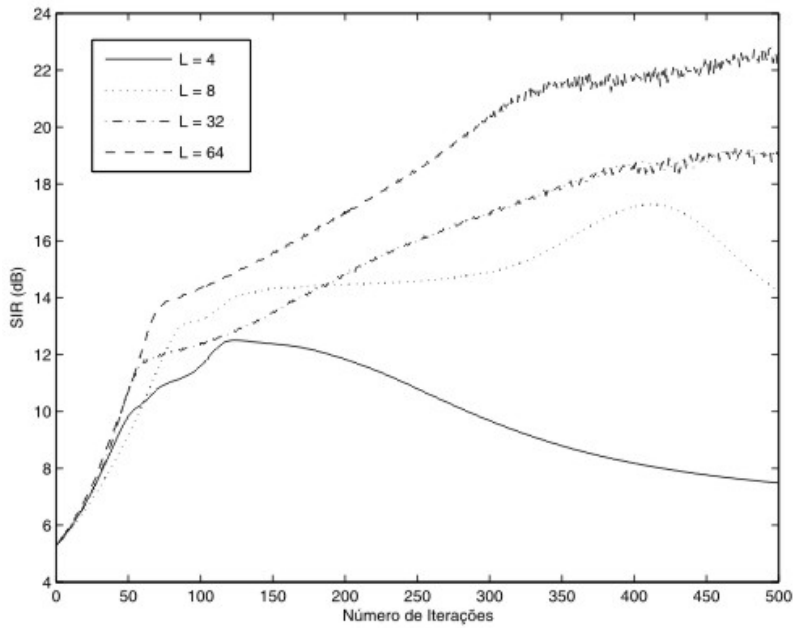


Figura 3.12: Evolução do SIR para diferentes comprimentos L dos filtros de separação.

do SIR após a convergência, conforme mostrado na Figura 3.12, onde filtros de separação de comprimentos reduzidos tendem a não se manter no melhor patamar de separação, ocasionando uma degradação do SIR. Por alguma razão a ser melhor estudada, o desempenho do sistema de separação muitas vezes não se estabiliza nos pontos ótimos.

Estas considerações supõem que a quantidade de amostras disponíveis possui informação suficiente para a determinação de filtros de separação maiores. A depender da quantidade de amostras, utilizar filtros menores pode ser benéfico.

3.5.4 Emprego de Filtros IIR

Os algoritmos de CBSS costumam utilizar filtros FIR no sistema de separação. Em primeiro lugar, porque filtros com resposta ao impulso finita são inerentemente estáveis, não havendo, portanto, preocupação quanto à estabilidade dos filtros de separação. Outra razão é que a obtenção de gradientes em relação aos parâmetros de filtros FIR é relativamente simples, caso comparemo-la com o mesmo procedimento em relação aos filtros IIR.

Nesta subseção, a proposta é avaliar se a função custo do método GASFC

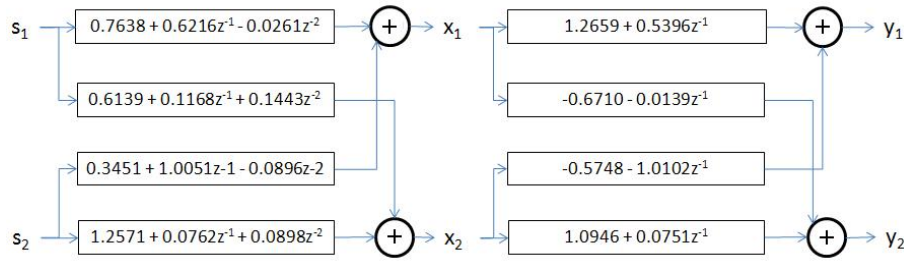


Figura 3.13: Configuração de mistura e os filtros de separação obtidos pelo algoritmo GASFC.

permite a seleção de filtros IIR que efetuem a separação de forma razoável, comparando o resultado desta separação com a obtida com filtros FIR. No momento, propomos um método muito simples de otimização para encontrar os parâmetros de um filtro IIR. A busca será feita por meio de uma “grade”, sendo selecionados os parâmetros de um filtro IIR que redundem numa função custo mínima.

Para uma avaliação de o quão bom é o desempenho quando empregamos filtros IIR, efetuamos uma comparação da separação por eles obtida com a separação obtida por filtros FIR (utilizando o método GASFC). A partir de 10 s de sinais de voz amostrados em 8 kHz, geramos as misturas segundo o sistema de mistura apresentado na Figura 3.13. Os filtros de separação, de comprimento 2, foram encontrados por meio do método GASFC, na sua formulação *block-online*, com os parâmetros $\mu = 10^{-4}$, $L = 2$ e 50 blocos por iteração ($K = 50$). A evolução do SIR nesta configuração é apresentada na Figura 3.14.

Nesta configuração de duas misturas e duas fontes, cumpre encontrar 4 filtros de separação: w_{11} , w_{12} , w_{21} e w_{22} . Escolhemos utilizar filtros de separação IIR de primeira ordem, com o formato a seguir:

$$W_{pq}(z) = \frac{b_{pq}}{1 + a_{pq}z^{-1}} \quad (3.64)$$

A realização na forma direta da estrutura IIR utilizada é mostrada na Figura 3.15.

A busca do parâmetro a_{pq} ocorre no intervalo (arbitrariamente escolhido) $[-3, 3]$; o espaçamento entre pontos testados é de 0,05 (gerando um total de 121 pontos no intervalo considerado). Assim, o parâmetro a_{pq} será testado em $-3; -2,95; -2,9; \dots; 3$. Já o parâmetro b_{pq} , por motivos de estabilidade, não pode possuir módulo maior do que a unidade. Assim, seu intervalo de busca será $[-0,99; 0,99]$,

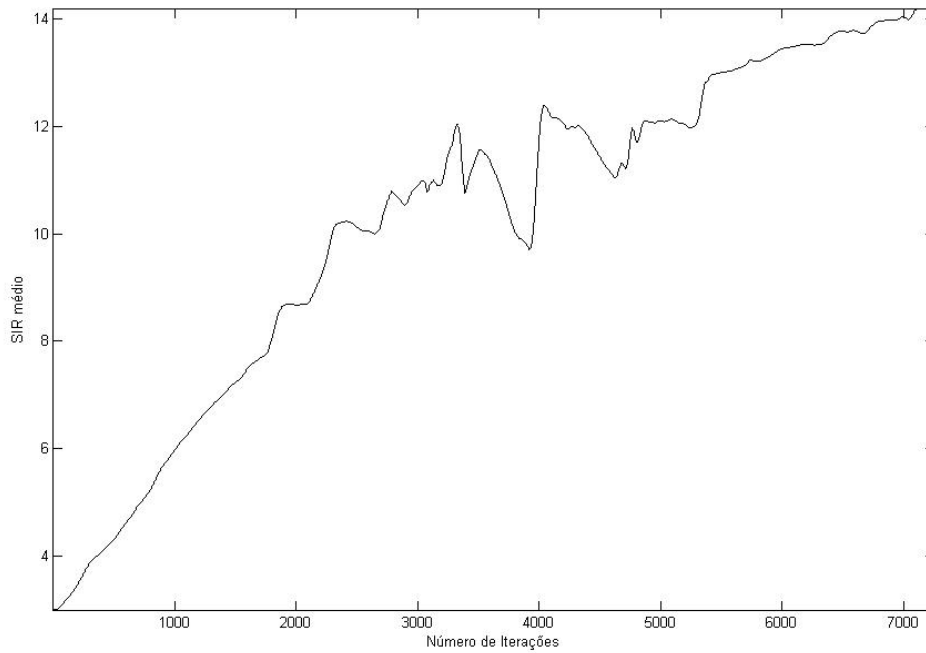


Figura 3.14: Evolução do SIR utilizando algoritmo GASFC no formato *block on-line*, $\mu = 10^{-4}$, $L = 2$ e número de blocos igual a 50.

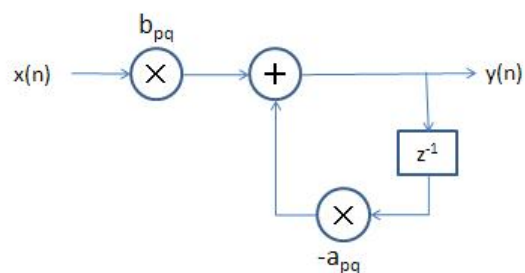


Figura 3.15: Realização na forma direta da estrutura IIR utilizada, utilizando entrada $x(n)$ com saída $y(n)$.

com o mesmo espaçamento (gerando um total de 40 pontos no intervalo considerado). A combinação destas possibilidades gera um total de $40^4 \times 121^4 = 5,4876.10^{14}$ diferentes configurações a analisar. Obviamente, testar todas estas possibilidades de filtros IIR nos intervalos considerados é uma tarefa enormemente custosa.

Desta forma, optamos por manter a estrutura de separação obtida pelo método GASFC, analisando o efeito da substituição de cada filtro FIR por um IIR de primeira ordem (ambos com dois parâmetros ajustáveis). Desta forma, o objetivo é verificar o novo desempenho obtido pelo sistema “híbrido”, com apenas um dos filtros FIR substituído por um IIR. Como agora há apenas um filtro IIR a otimizar, testamos apenas $40 \times 121 = 4840$ diferentes possibilidades.

A função custo contemplará toda a extensão do sinal (logo, estaremos utilizando uma formulação *off-line*). Caso arbitremos o parâmetro D (número de intervalos temporais a considerar nas correlações de tempo curto) como $D = N/2$ (abordagem comum no método GASFC), a questão de se escolher os parâmetros da função custo se resume na escolha de N . Nosso objetivo também consiste em avaliar a influência do parâmetro N sobre a seleção do filtro IIR.

Utilizando misturas de duração de 1s e sinais de voz amostrados em 8 kHz, escolhendo o filtro IIR que apresenta menor valor da função custo e calculando o SIR correspondente, formamos a Tabela 3.4. Nesta tabela, o maior SIR obtido entre os valores de a_{pq} e b_{pq} analisados é denominado “SIR máximo” (calculado utilizando o conhecimento dos filtros de mistura). O SIR final obtido pelo método GASFC foi de 16,2776 dB.

A Tabela 3.4 não nos permite extrair um critério simples de escolha de N . O aumento de N revelou-se proveitoso para o filtro w_{11} , ligeiramente benéfico para os filtros w_{21} e w_{22} , porém muito daninho para o filtro w_{12} . A função custo utilizada permitiu uma escolha dos filtros w_{11} e w_{21} muito próxima da opção que resulta num SIR máximo, enquanto que para os filtros w_{12} e w_{22} , o filtro escolhido apresenta desempenho muito aquém ao da melhor escolha.

As Figuras 3.16 a 3.25 mostram o SIR em função dos parâmetros a_{pq} e b_{pq} para diferentes filtros FIR substituídos. Estas figuras permitem explicar de forma simples os resultados da Tabela 3.4.

O filtro w_{11} apresenta um mínimo global da função custo próximo do máximo

Tabela 3.4: SIR obtido com diferentes parâmetros N e para filtros de separação distintos, admitindo o menor valor da função custo do método GASFC.

	w_{11}	w_{12}	w_{21}	w_{22}
$N = 4$	15,1246	16,4175	15,0638	11,8144
$N = 6$	15,1269	16,4175	15,0763	11,8386
$N = 8$	15,1289	14,4490	15,1533	11,8144
$N = 10$	19,1621	13,8059	15,1499	11,8236
$N = 12$	19,1621	13,6065	15,1581	11,8386
$N = 14$	19,1621	13,1683	15,1581	11,8698
$N = 16$	19,1621	12,5793	15,1611	11,8316
$N = 18$	19,1621	13,0787	15,1581	11,8702
$N = 20$	19,1621	12,8989	15,1624	11,8236
SIR máximo	19,3096	21,5293	15,4347	24,8765

global do SIR; o aumento do N claramente aproxima o mínimo da função custo do máximo do SIR - daí a melhora obtida com o aumento de N .

O filtro w_{21} apresenta um mínimo da função custo perto do máximo global do SIR e o formato da superfície da função custo altera-se muito pouco com alterações do valor de N . Porém, trata-se de um mínimo *local* da função custo; o mínimo global (o qual ocorre aproximadamente com os parâmetros $b_{21} = 0,5$ e $a_{21} = 0,6$) encontra-se distante do máximo global do SIR. Já os filtros w_{12} e w_{22} apresentam um mínimo global muito perto de um máximo da função SIR; porém os máximos da função SIR em questão são locais, e não globais.

Sendo assim, podemos verificar que a função custo do método GASFC não estima de forma confiável os parâmetros de um filtro IIR para propósitos de separação. É possível que um mínimo global da função custo coincida (aproximadamente) com um máximo global da função SIR, o que é um evento desejável. Porém, como vimos, isto não ocorre sempre; o mínimo global da função custo pode coincidir com um máximo local da função SIR ou o máximo global da função SIR é associado a um mínimo local da função custo. Isto nos leva a concluir que somente a função custo em questão não é suficiente para escolha de parâmetros de filtros IIR na sua

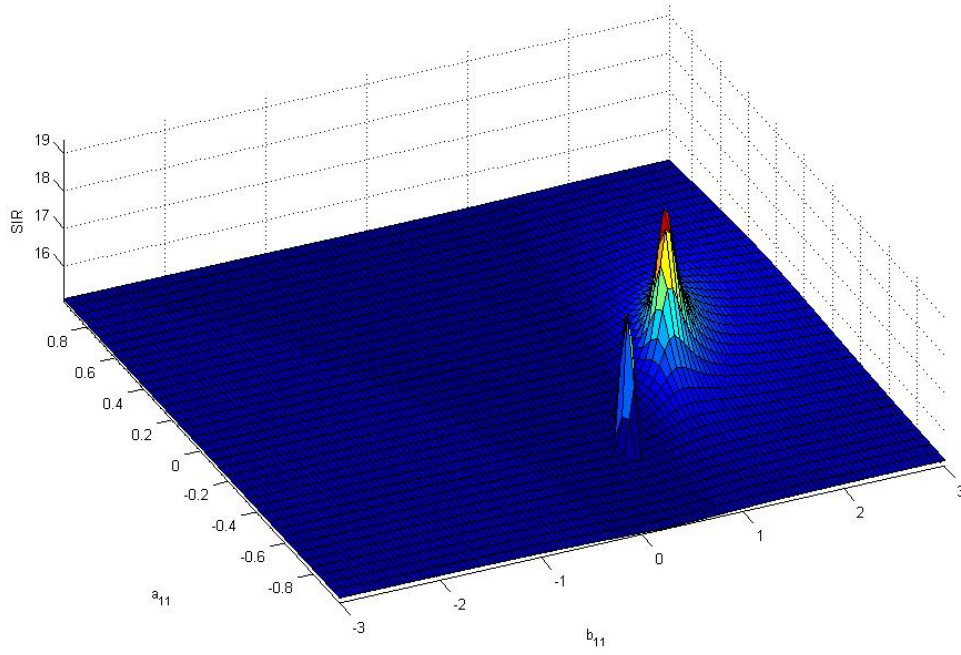


Figura 3.16: SIR em função dos parâmetros a_{11} e b_{11} .

formulação direta.

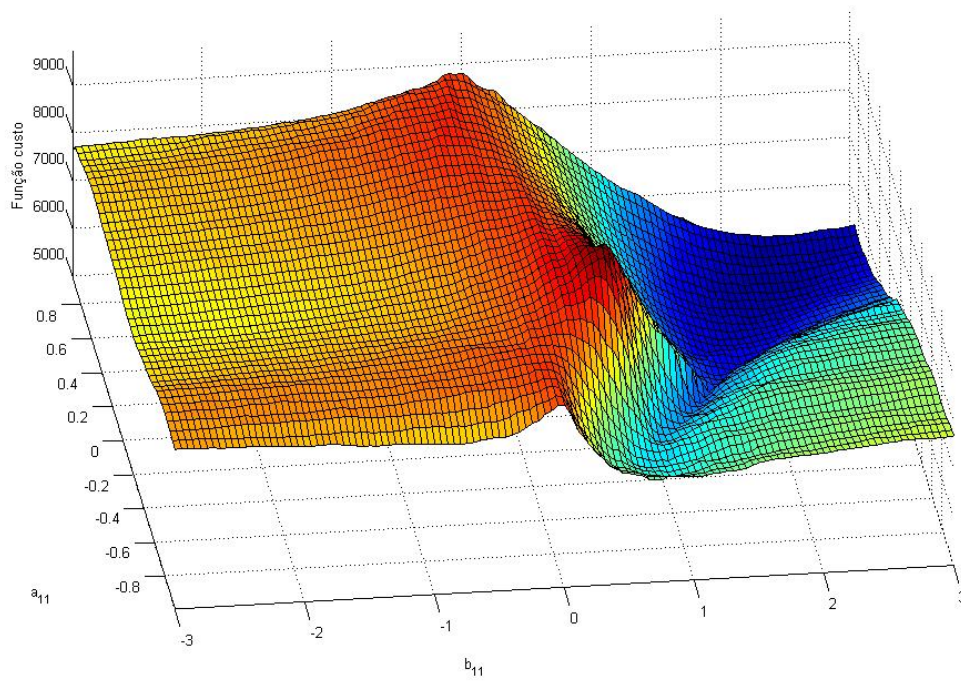


Figura 3.17: Função custo em função dos parâmetros a_{11} e b_{11} , com $N = 4$.

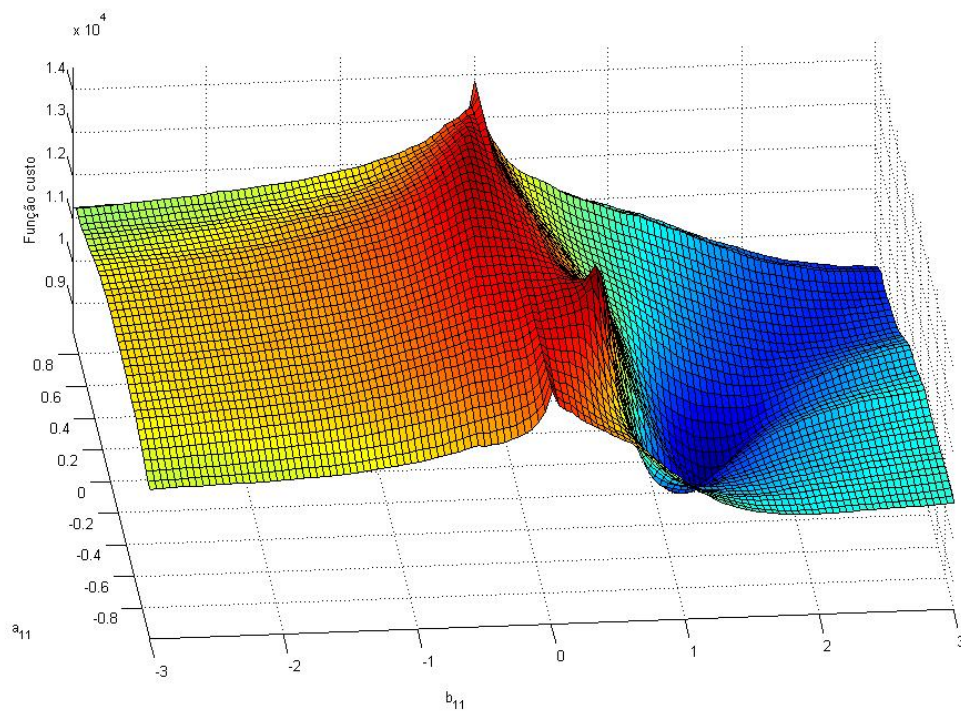


Figura 3.18: Função custo em função dos parâmetros a_{11} e b_{11} , com $N = 14$.

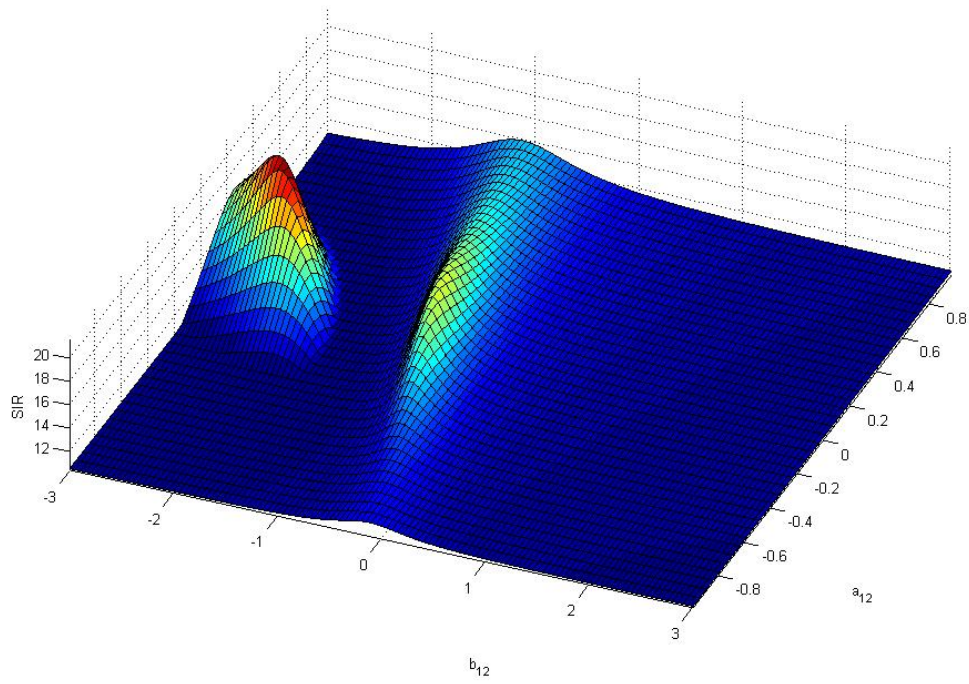


Figura 3.19: SIR em função dos parâmetros a_{12} e b_{12} .

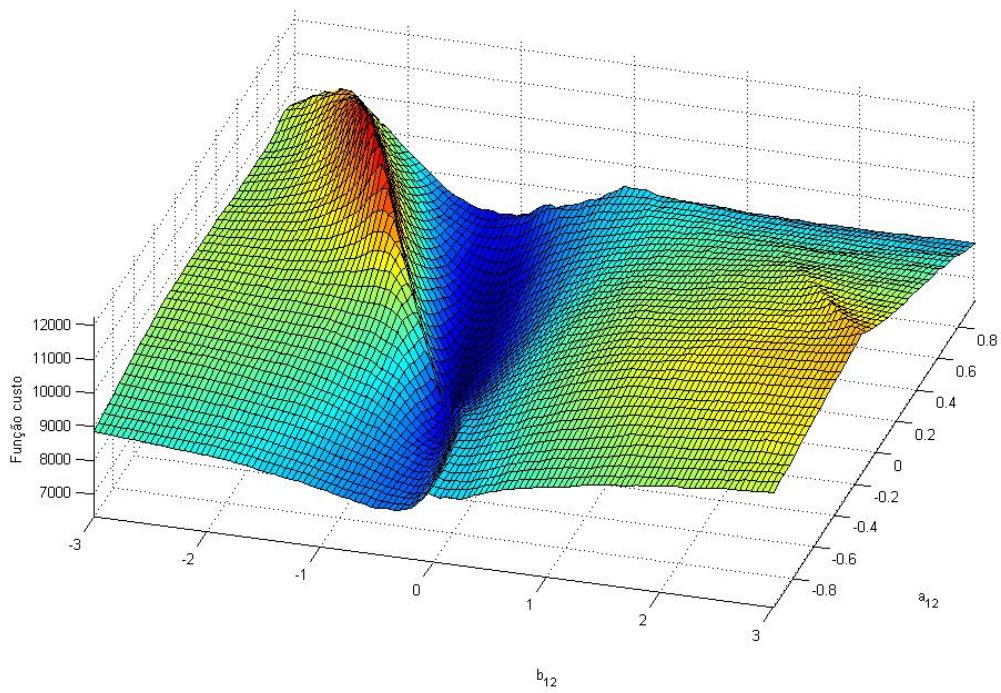


Figura 3.20: Função custo em função dos parâmetros a_{12} e b_{12} , com $N = 6$.

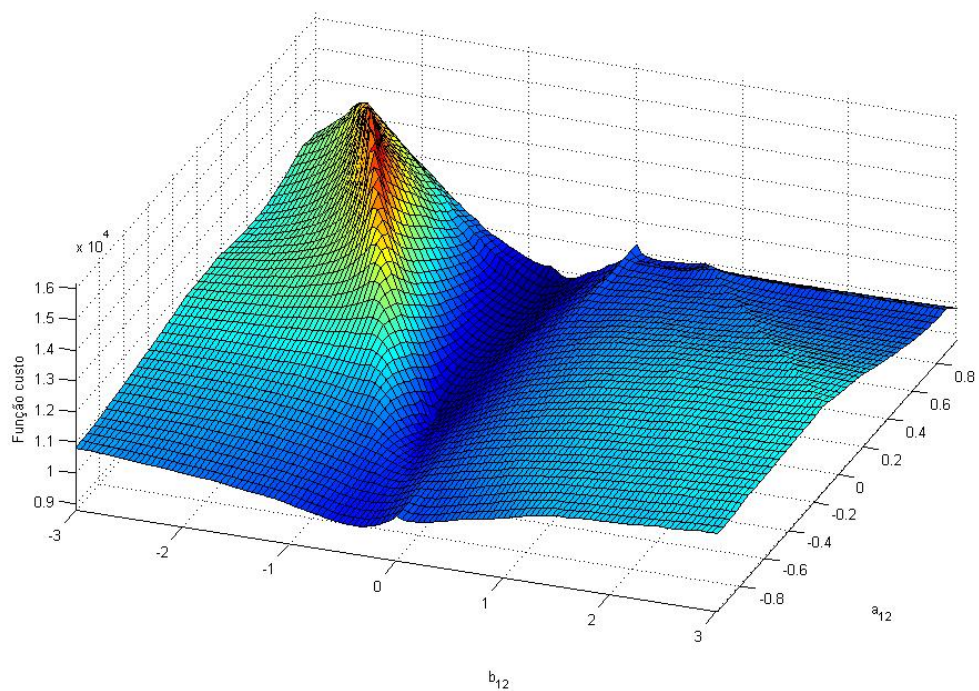


Figura 3.21: Função custo em função dos parâmetros a_{12} e b_{12} , com $N = 16$.

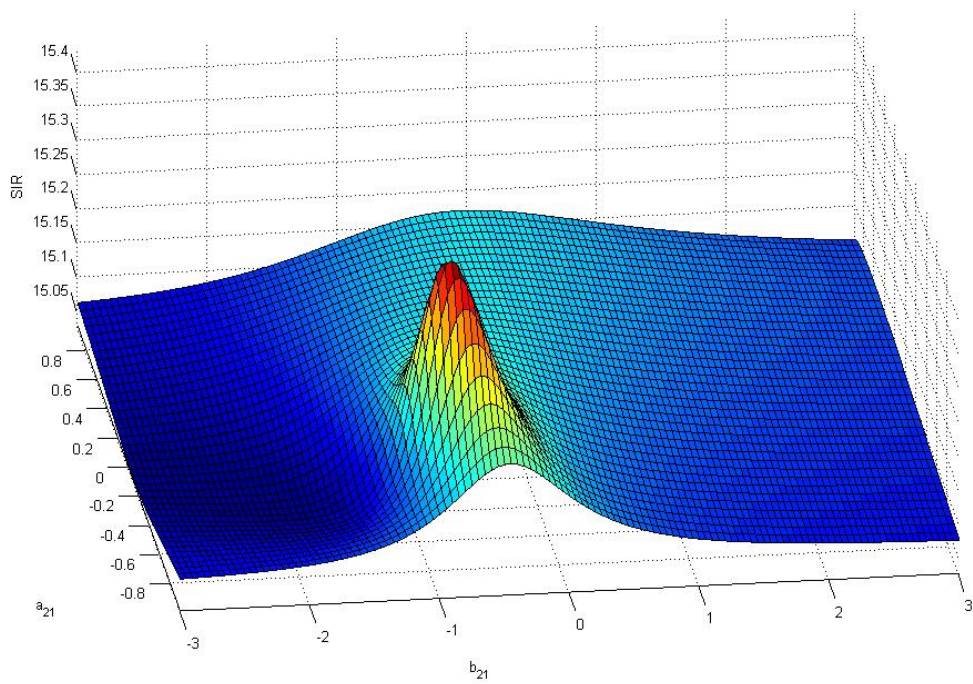


Figura 3.22: SIR em função dos parâmetros a_{21} e b_{21} .

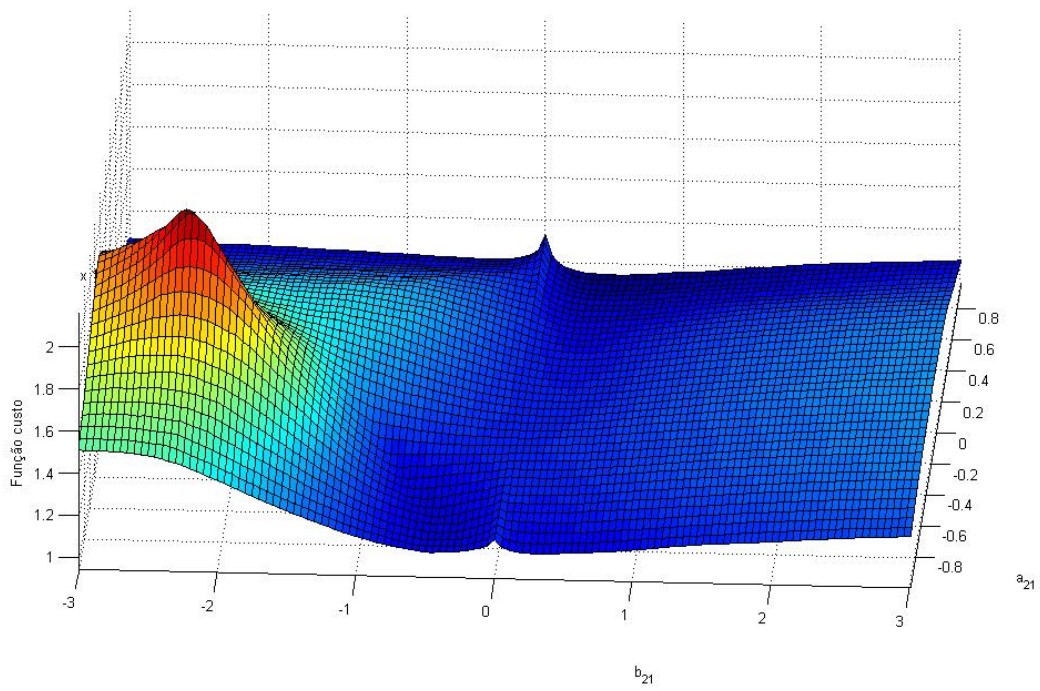


Figura 3.23: Função custo em função dos parâmetros a_{21} e b_{21} , com $N = 12$.

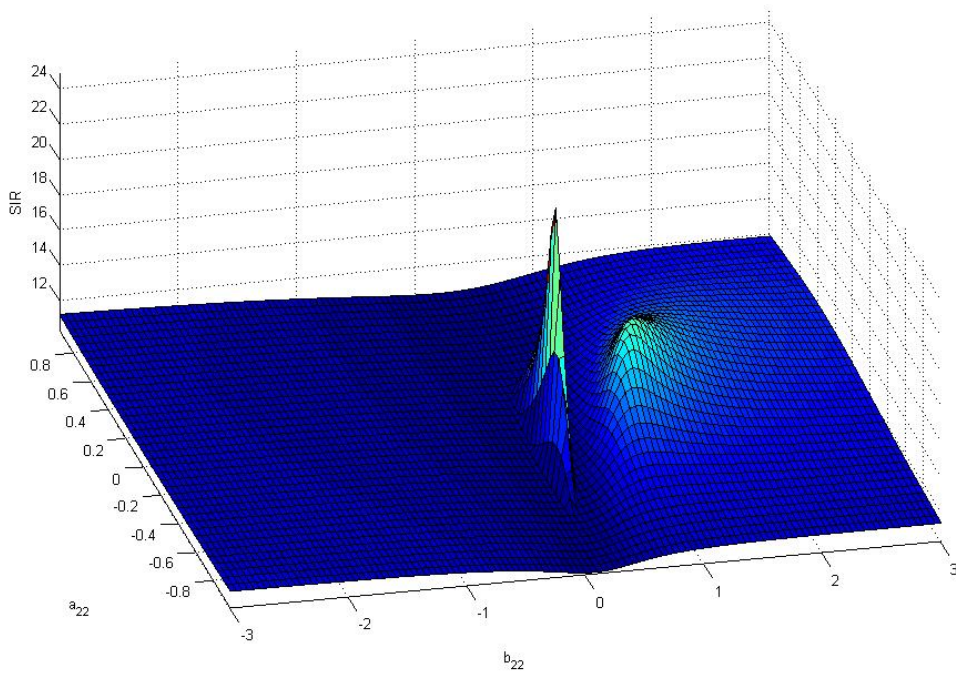


Figura 3.24: SIR em função dos parâmetros a_{22} e b_{22} .

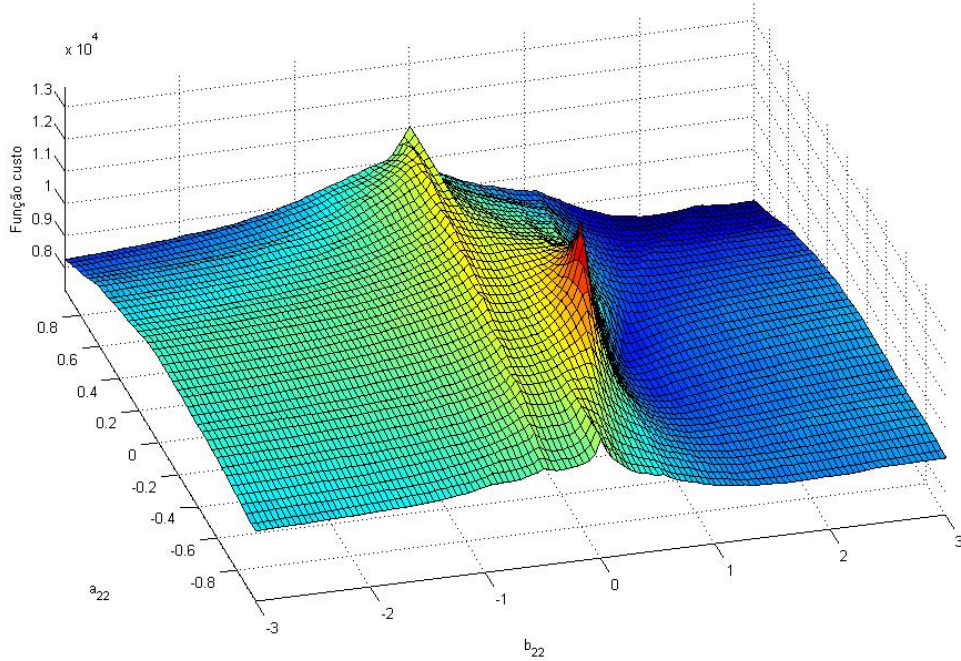


Figura 3.25: Função custo em função dos parâmetros a_{22} e b_{22} , com $N = 10$.

3.5.5 Avaliação de Eventuais Permutações por meio da Direção de Chegada

O método EDAO (preconizado em [12]) é considerado livre do problema de permutação. Nesta seção, propomos uma forma de se avaliar esta premissa, verificando se as direções de chegada de cada fonte estimada numa determinada raia coincidem com as direções das outras raias. Desta forma, nossa proposta consiste em aliar a estimativa da direção de chegada da fonte em [3] ao método EDAO.

Nas simulações, utilizamos gravações numa sala, com microfones distanciados de 5 cm, com vozes (uma masculina, outra feminina) amostradas em 16 kHz. A função de transferência medida da sala apresenta em torno de 4000 coeficientes. Utilizamos 1024 raias (o equivalente a 1024 coeficientes para separação), utilizando o método EDAO com $\mu = 0,05$ e avaliando, em cada frequência, as estimativas da direção de chegada.

A Figura 3.26 apresenta um destes resultados, para uma determinada posição das fontes (a diferença gráfica dos pontos indica estimativas de diferentes fontes). Em baixas frequências, a estimativa da direção de chegada não é precisa,

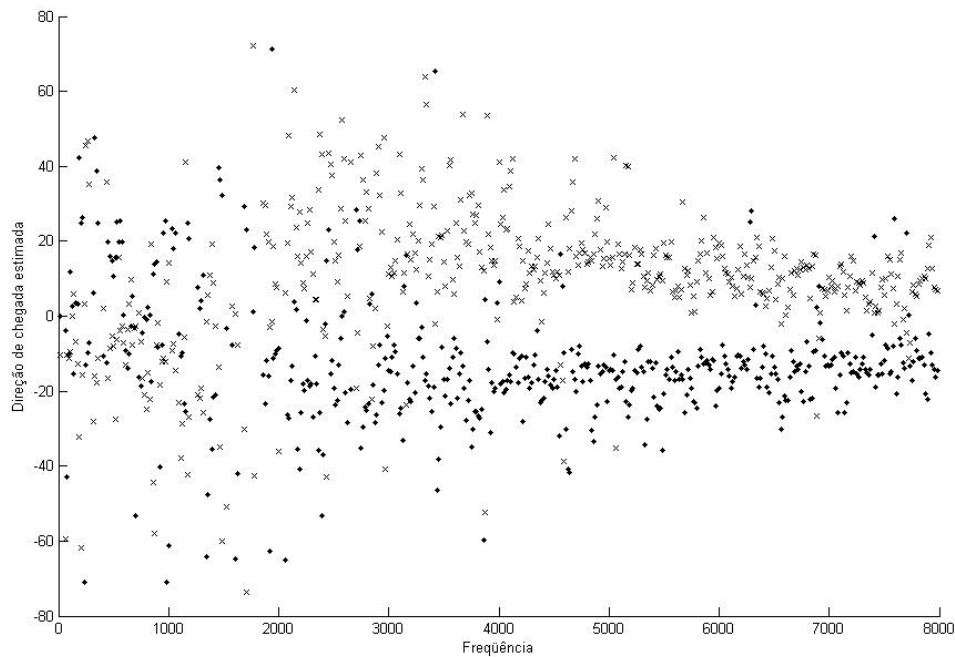


Figura 3.26: Estimativa da direção de chegada de cada fonte em função da frequência.

pois a diferença de fase provocada pela distância entre sensores é muito pequena. Porém, em médias e altas frequências notamos que as direções de chegada de cada fonte estimada são muito coerentes, se concentrando em dois ângulos distintos.

Este resultado confirma a premissa de ausência de permutação supracitada. Porém, interessante, modificando a posição das fontes, o resultado não foi o mesmo. Como exemplo, temos a Figura 3.27, a qual indica a existência de permutação, sendo trocadas as componentes de alta frequência nas fontes estimadas.

Este resultado revela que o método EDAO pode não resolver completamente o problema da permutação, mesmo contemplando, na sua função custo, uma dependência estatística entre raias distintas.

Uma utilização do critério de direção de chegada, neste caso, ajudaria a corrigir este problema, incrementando o desempenho do método EDAO, por meio de uma mera permutação das linhas das matrizes de separação estimadas. É importante notar que a estimativa da direção de chegada é uma medida “cega”, ou seja, que não depende do conhecimento exato do sistema de mistura, mas apenas de uma estimativa do mesmo. Esta característica torna a conjunção dos métodos valiosa.

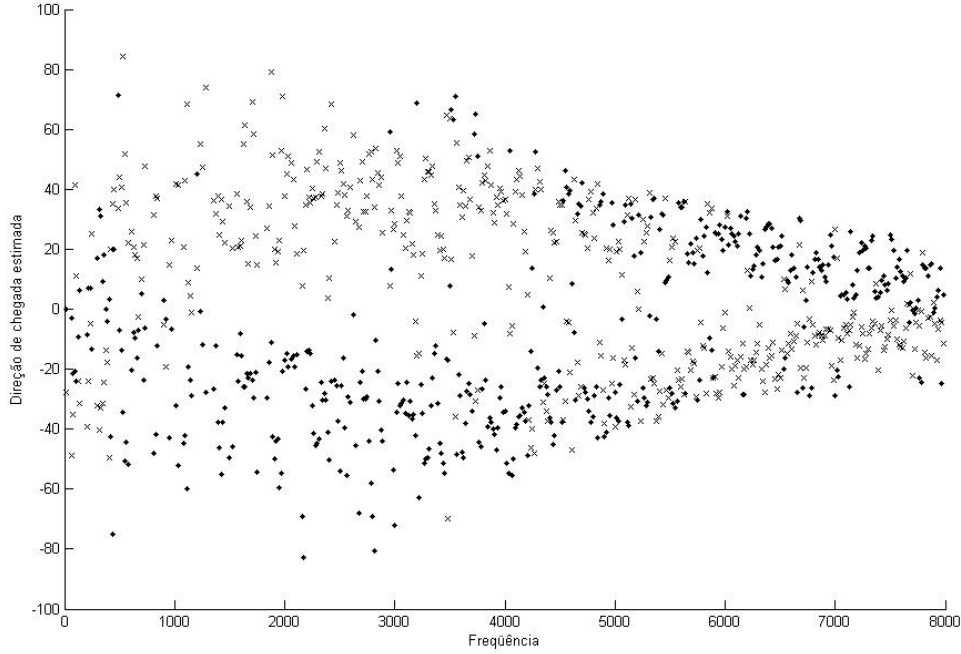


Figura 3.27: Estimativa da direção de chegada de cada fonte em função da frequência.

3.5.6 Mudança no Método EDAO por Modificações na Função Custo

Como vimos, o método EDAO utiliza uma distribuição *a priori* das fontes dada por:

$$q(\tilde{\mathbf{s}}_i) = \alpha e^{-\sqrt{(\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)^H \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)}}. \quad (3.65)$$

Esta distribuição contempla, simultaneamente, diversas raias. Porém, não é necessariamente a distribuição que apresenta o melhor desempenho. No artigo onde o método EDAO é proposto [12], afirma-se que outras distribuições q podem apresentar desempenho superior na separação. Por isso, verificamos nesta seção se uma ligeira modificação de q nos permitiria obter uma separação melhor. Seja a distribuição q' :

$$q'(\tilde{\mathbf{s}}_i) = \alpha' e^{-[(\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)^H \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{s}}_i - \tilde{\boldsymbol{\mu}}_i)]^{\frac{1}{D}}}, \quad (3.66)$$

onde α' é uma outra constante de normalização (a qual não influencia o gradiente do logaritmo de q') e D é um parâmetro arbitrário (podendo ser adaptativo). A condição $D = 2$ implica $q' = q$. Porém, não é necessário restringirmo-nos a este valor de D ⁷.

Interpretando $\left[\sum_{l=1}^L |\tilde{s}^{(l)}|^2\right]^{\frac{1}{D}}$ como a distância euclidiana entre a origem e o vetor $(\tilde{s}_i^1 \dots \tilde{s}_i^L)$, uma outra generalização possível, por meio da extensão da métrica euclidiana, seria:

$$\Phi^{(l)}(\tilde{s}_i^1 \dots \tilde{s}_i^L) = \frac{\partial \left\{ \left[\sum_{l=1}^L |\tilde{s}^{(l)}|^D \right]^{\frac{1}{D}} \right\}}{\partial \tilde{s}_i^{(l)}}, \quad (3.67)$$

escolha que não será contemplada nesta dissertação.

Calculando a nova função *score* (empregando as mesmas suposições de decorrelação entre raias e média zero das mesmas), temos:

$$\Phi^{(l)}(\tilde{s}_i^1 \dots \tilde{s}_i^L) = \frac{\partial \left\{ \left[\sum_{l=1}^L |\tilde{s}^{(l)}|^2 \right]^{\frac{1}{D}} \right\}}{\partial \tilde{s}_i^{(l)}} = \frac{2}{D} \frac{\tilde{s}_i^{(l)}}{\left(\sum_{l=1}^L |\tilde{s}_i^{(l)}|^2 \right)^{\frac{D-1}{D}}}. \quad (3.68)$$

A nova função *score* obtida pode ser inserida na equação de atualização do método EDAO. Truncando a função de transferência (medida) da sala em 512 coeficientes, utilizamos 1024 coeficientes para separação. Variando o parâmetro D e calculando o SIR na saída, verificamos na Figura 3.28 que é possível angariar melhoras no desempenho do algoritmo, as quais podem chegar a mais de 0,5 dB. Interessantemente, dos valores de D testados, a possibilidade $D = 2$ foi a segunda pior; o desempenho se revela melhor para valores superiores e inferiores a 2. Os valores inferiores a 1,9 foram os que apresentaram melhores resultados. Não é viável reduzir D a um valor inferior a 1,4 já que nesta configuração costuma ocorrer divergência na estimativa dos coeficientes.

Uma possível abordagem seria o recurso a um valor de D adaptativo, cujo valor ótimo seja calculado de forma cega.

⁷Mesmo que alguns valores de D gerem uma distribuição imprópria (no sentido de a integral de todos os seus argumentos de $-\infty$ a $+\infty$ ser infinita, e não unitária), isso não é particularmente problemático, já que a distribuição não-informativa de Jeffrey também o é, sendo utilizada de forma ubíqua em análise bayesiana.

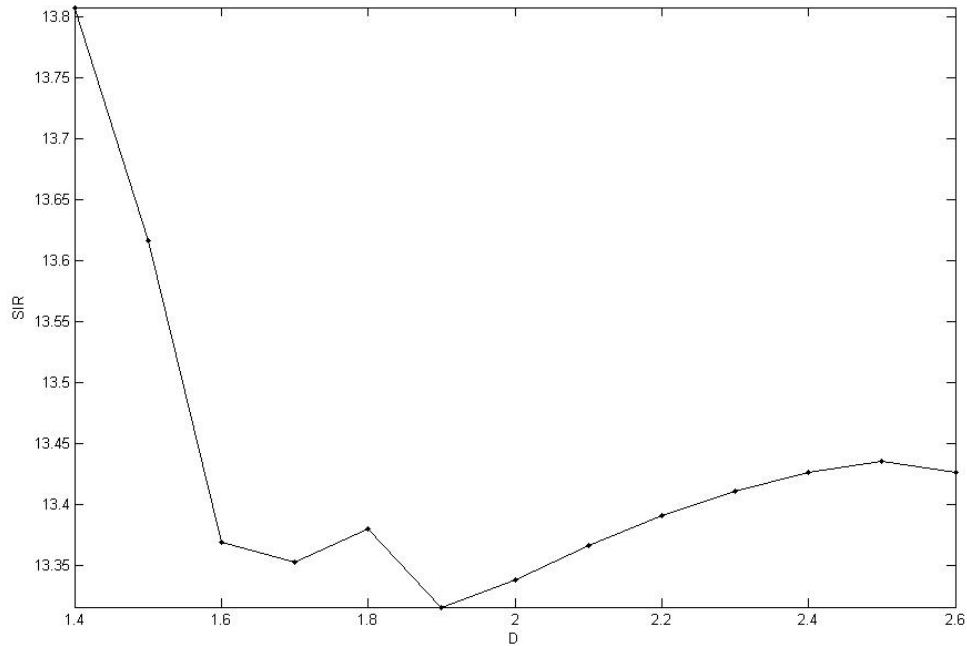


Figura 3.28: Razão sinal interferência em função do parâmetro D .

3.6 Considerações Finais

Este capítulo abordou métodos de separação em misturas convolutivas, as quais apresentam um nível de dificuldade (no tocante à separação) muito superior ao das misturas instantâneas. Foram contemplados dois métodos dos mais bem-sucedidos para estes casos, a saber: GASFC e EDAO.

Para o método GASFC, avaliamos o efeito de três propostas. A primeira destas consistiu em adotar taxas de aprendizagem distintas para cada parâmetro a estimar. Após algumas adaptações, obtemos uma variante do método GASFC (denominada GASFC-VM-III) a qual apresenta uma taxa de convergência superior à do método original, sem queda no desempenho após a estabilização dos parâmetros.

A segunda proposta recorreu a estimativas recursivas de matrizes de correlação curta, o que acabou por acelerar enormemente a convergência inicial do sistema de separação, à custa de uma degradação após certo tempo (degradação esta que pode ser anulada mediante a eliminação do fator de lembrança).

A terceira proposta avaliou o efeito de se utilizar filtros de separação de comprimentos maiores do que os comprimentos dos filtros de mistura, com efeitos benéficos tanto em relação ao desempenho do sistema quanto em relação à estabili-

dade do ponto de melhor separação.

Analisamos também a conveniência de se utilizar a função custo do método GASFC para estimar os parâmetros de filtros com resposta ao impulso infinita. Observamos que não há garantia de que os filtros dessa forma selecionados estejam localizados próximos do máximo global da função SIR.

Em relação ao método EDAO, foram apresentadas duas propostas. A primeira consistia em aliar a este método uma estimativa da direção de chegada das fontes, o que permitiu avaliar de forma cega a existência de permutações nas estimativas das fontes.

A segunda proposta para o método EDAO consistiu em modificar levemente a sua distribuição *a priori* das fontes, o que acabou por gerar ganhos de desempenho na separação.

Capítulo 4

Misturas Indeterminadas

4.1 Introdução

Comumente, os métodos de separação de fontes exigem um número de misturas pelo menos igual ao número de fontes. No entanto, na prática, esta restrição não é atendida com frequência. Neste caso, defrontamo-nos com o problema de separação indeterminada de misturas (em inglês, *underdetermined*). Como não se pode recompor as fontes a partir de uma mera estimativa da matriz de misturas (ou de sua inversa), podemos inferir que o caso indeterminado é bem mais difícil do que o determinado (ou o sobredeterminado, o qual ocorre quando o número de misturas é maior do que o número de fontes).

Como vimos no Capítulo 2, a Análise de Componentes Independentes não costuma ser muito útil no caso indeterminado, sendo em geral exploradas neste caso outras formas de diversidade entre as fontes, notadamente a *esparsidade* (daí o nome, muito comum, *Análise de Componentes Esparsos*, comumente abreviada por SCA, do inglês *Sparse Component Analysis*). Como esta abordagem também se aplica aos casos (sobre)determinados, alguns autores até afirmam ser a esparsidade “mais poderosa” quanto à capacidade de separação do que a independência [60]. No contexto de separação de fontes, esparsidade significa que os coeficientes das fontes (ou, como ocorre mais frequentemente, de uma decomposição das fontes mediante uma transformada) são, em sua maior parte, próximos de zero. Esta esparsidade, portanto, implica que poucos coeficientes das fontes possuem módulo muito grande (onde “módulo grande” é, obviamente, um termo relativo). Visto de outro modo,

poderíamos esperar que a representação obtida das fontes apresente uma baixa porcentagem dos coeficientes capturando grande porcentagem da energia dos sinais de interesse [61]. Assim, um histograma dos coeficientes das fontes se assemelharia a uma distribuição super-gaussiana, com grande freqüência de módulos pequenos e baixa freqüência dos grandes. Frequentemente, escolhe-se a distribuição exponencial para a modelagem da distribuição dos coeficientes de cada fonte.

Os métodos avaliados neste capítulo exploram a *diversidade espacial* das misturas, obtida através do fato de que a mesma fonte contribui de forma distinta em cada uma das misturas. Para que isso seja possível, deve existir mais do que uma mistura. Isso possibilita a subdivisão do caso indeterminado em dois: quando há apenas uma mistura (caso ainda mais difícil) e quando há mais do que uma mistura. O próximo capítulo tratará do caso extremo de termos apenas uma mistura, onde a diversidade espacial entre as misturas não constitui auxílio para as estimativas.

A resolução do problema de separação indeterminada pela maioria das técnicas que utilizam Análise de Componentes Esparsos pode ser caracterizada por quatro passos na ordem dada a seguir [62]:

- (I) aplicação de uma transformada linear esparsificadora às misturas;
- (II) estimativa da matriz de misturas ¹ a partir dos coeficientes da decomposição obtida via transformada;
- (III) estimativa da representação de cada uma das fontes no domínio da transformada;
- (IV) reconstrução das fontes.

A maioria dos artigos resume o problema de separação indeterminada em duas etapas, a saber: a) estimativa da matriz de misturas (a partir de uma transformada) e b) reconstrução das fontes. Cabe notar que (a) condensa as etapas (I) e (II) supracitadas, bem como (b) condensa (III) e (IV). Em geral, o subproblema (b) é resolvido após a estimativa encontrada em (a). Há métodos, entretanto, que estimam tanto a matriz de misturas quanto as fontes de forma iterativa: a estimativa de (a) é

¹Neste capítulo, não trataremos de misturas convolutivas. Logo, só haverá uma única matriz a ser estimada.

utilizada para melhorar a estimativa de (b) e vice-versa, em diversas iterações; estes métodos em geral tendem a ser mais pesados computacionalmente. Denominaremos estes métodos de “simultâneos”, em contraposição aos métodos mais frequentes, ditos “seqüenciais”. Neste capítulo contemplaremos algumas das principais técnicas seqüenciais, todas centradas no caso instantâneo (embora não seja difícil estendê-las para o caso convolutivo - vide, por exemplo, [74]). Interessante observar que no caso da Análise de Componentes Esparsos, o objetivo é estimar a matriz de mistura, enquanto que nos métodos baseados em ICA em geral estima-se a *inversa* da matriz de misturas (dado que ela em geral existe nos casos não indeterminados). Daí é possível fazer uma distinção entre métodos de busca no espaço de mistura (SCA) e no espaço de separação (ICA).

Muitas das técnicas aqui apresentadas são dedicadas ao caso estéreo (duas misturas), e tendem a apresentar uma “maldição da dimensionalidade” caso sejam adaptadas para um número maior de misturas.

4.2 Problema de Estimativa da Matriz de Misturas

Os métodos seqüenciais a descrever são:

- Varredura de Zibulevsky
- Modelos de misturas de exponenciais (sigla em inglês, LMM - *Laplacian Mixture Models*)
- Clusterização *K-means*
- Clusterização *fuzzy-C*
- Clusterização hierárquica
- Método DUET
- Extensão dos métodos DUET e TIFROM
- Proposta: Modelos de misturas de gaussianas (sigla em inglês, GMM - *Gaussian Mixture Models*)

4.2.1 Varredura de Zibulevsky

Este foi o primeiro método [63] que utilizou a hipótese de esparsidade entre as fontes para estimar a matriz de misturas, sendo até hoje um dos mais competitivos. Ele também permite estimar o número N de fontes.

Denominando \mathbf{h}^j a j -ésima coluna da matriz \mathbf{H} , podemos expressar $\mathbf{x}(n)$, $n = 1, \dots, K$ por:

$$\mathbf{x}(n) = \sum_{j=1}^N \mathbf{h}^j s_j(n), \quad (4.1)$$

donde se vê que, na ausência de ruído, podemos interpretar $\mathbf{x}(n)$ como uma combinação linear dos vetores-coluna da matriz \mathbf{H} . Admitida a esparsidade das fontes, espera-se que em geral alguns coeficientes sejam muito próximos de zero. Isto nos permite esperar que não raro os coeficientes $s_j(n)$ na mistura sejam iguais a zero e apenas um a cada instante n - o “dominante” - seja não-nulo. Se isto realmente ocorresse, notar-se-ia uma *tendência de aglomeração* dos pontos $\mathbf{x}(n)$ em torno das direções dos vetores \mathbf{h}^j .

Não é difícil ver uma equivalência entre estimar as direções dos vetores \mathbf{h}^j e estimar a matriz de mistura \mathbf{H} (a menos de um escalamento das colunas).

Para melhor compreensão deste ponto, veja a Figura 4.1, a qual apresenta fontes maximamente esparsas (quando uma é não-nula, todas as outras são nulas). Quando plotamos o gráfico das misturas ($x_1(n) \times x_2(n)$), há a formação de três retas, cada qual associada a uma das colunas da matriz de misturas.

Para motivar uma relação formal entre a direção das retas e as colunas da matriz \mathbf{H} , eis um exemplo. Seja o caso $M = 2$; admitindo somente a i -ésima fonte não-nula no instante n :

$$\begin{aligned} x_1(n) &= \sum_{j=1}^N h_{1j} s_j(n) = h_{1i} s_i(n) \\ x_2(n) &= \sum_{j=1}^N h_{2j} s_j(n) = h_{2i} s_i(n) \\ \frac{x_2(n)}{x_1(n)} &= \frac{h_{2i}}{h_{1i}} \implies x_2(n) = \frac{h_{2i}}{h_{1i}} x_1(n). \end{aligned} \quad (4.2)$$

Devido ao já citado problema do escalamento, não há problema em igualar a 1 todos os coeficientes da primeira linha da estimativa da matriz de mistura (abordagem feita, por exemplo, em [16]). Dessa forma, chegamos a:

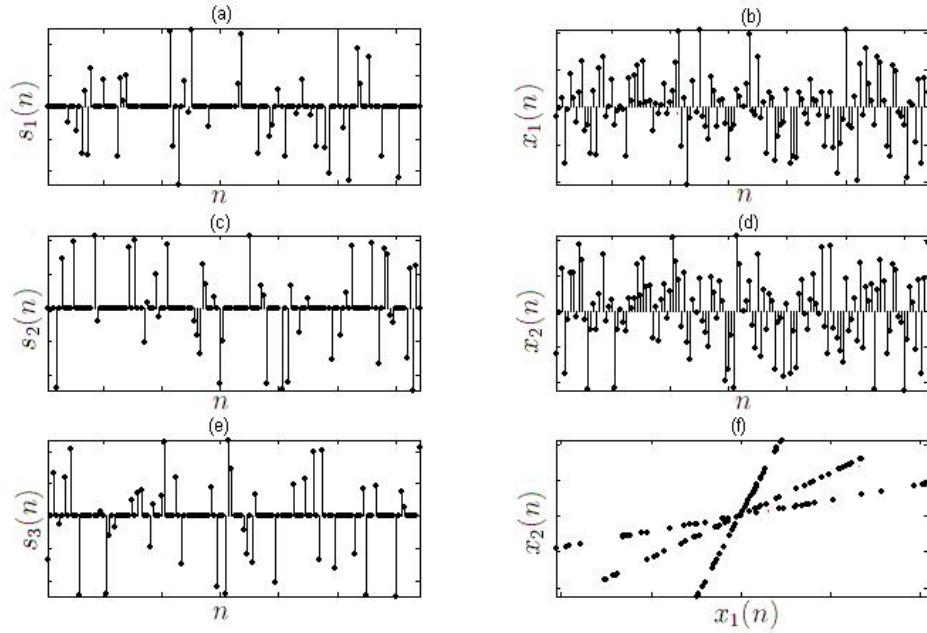


Figura 4.1: Exemplo de fontes maximamente esparsas: (a) fonte 1; (b) mistura 1; (c) fonte 2; (d) mistura 2; (e) fonte 3; (f) gráfico de $x_2(n) \times x_1(n)$.

$$x_2(n) = h_{2i}x_1(n), \quad (4.3)$$

daí concluímos que o fato de somente a fonte s_i ser não-nula no instante n implica $x_2(n)$ ser proporcional a $x_1(n)$, com uma constante de proporcionalidade determinada pela segunda linha da i -ésima coluna da matriz (estimada) de mistura (e que independe de n). Como o vetor $\mathbf{x}(n)$ possui componentes $x_1(n)$ e $x_2(n)$, a tangente do seu ângulo é:

$$\text{tg}(\theta_n) = \frac{x_2(n)}{x_1(n)} = \frac{h_{2i}x_1(n)}{x_1(n)} = h_{2i}, \quad (4.4)$$

daí podemos estimar a matriz de mistura tornando os coeficientes da primeira coluna iguais a 1 e os da segunda, às tangentes dos ângulos das direções de maior densidade de pontos.

Antes de continuar, cabe responder a uma objeção comum à Análise de Componentes Esparsos, que tem por alvo sua hipótese principal, a saber: a esparsidade das fontes. Sinais de voz, por exemplo, não são exemplos de sinais esparsos. E comumente desejamos separá-los. Poderia a Análise de Componentes Esparsos, neste caso, vir em nosso auxílio?

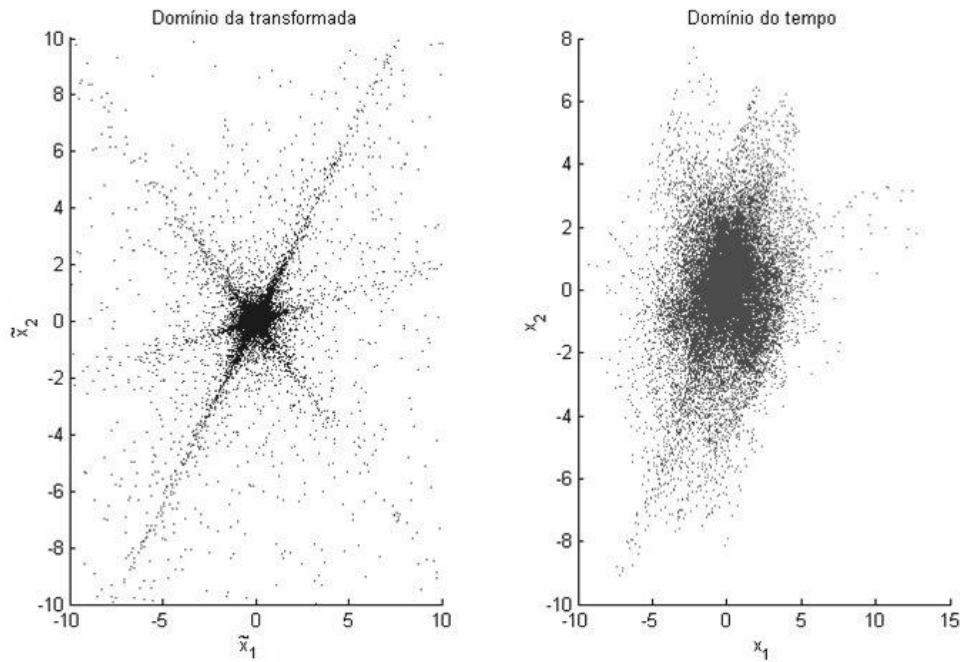


Figura 4.2: Gráfico de $x_2(n) \times x_1(n)$; à esquerda, coeficientes numa transformada *wavelet packet* com 7 níveis de decomposição e à direita, amostras no domínio do tempo.

A resposta é afirmativa. A maioria dos métodos exige uma etapa de pré-processamento, de forma a efetuar uma transformada nos sinais das misturas, almejando torná-los os mais esparsos possíveis (Passo I já citado anteriormente). Os meios mais utilizados para esse objetivo são: transformadas discretas de Fourier [63], transformadas *wavelet packet* [60],[64] e algoritmos derivados do *Matching Pursuit* [65]. O uso destas transformadas, obviamente, não implicará que o vetor $\tilde{\mathbf{x}}$ das misturas a cada instante n (com $\tilde{\mathbf{x}}$ implicando o uso de uma transformada) se encontre exatamente numa das retas associadas a uma fonte, porém o resultado é muito mais perto disto do que no domínio do tempo, como vê-se na Figura 4.2 para o caso 2×3 e nas Figuras 4.3 e 4.4 para o caso 3×4 , obtidas a partir de sinais de voz (a nomenclatura $M \times N$ implica M misturas e N fontes).

A idéia do padrão de varredura de Zibulevsky reside em efetuar uma espécie de histograma angular, procurando os ângulos do gráfico $x_2(n) \times x_1(n)$ onde há maior densidade de pontos. Para tal, o método escolhe ângulos-teste (tipicamente em torno de 200) uniformemente espaçados no intervalo $[0, \pi]$; não é necessário contemplar o intervalo $[\pi, 2\pi]$, já que uma reta definida por um ângulo θ equivale à outra reta

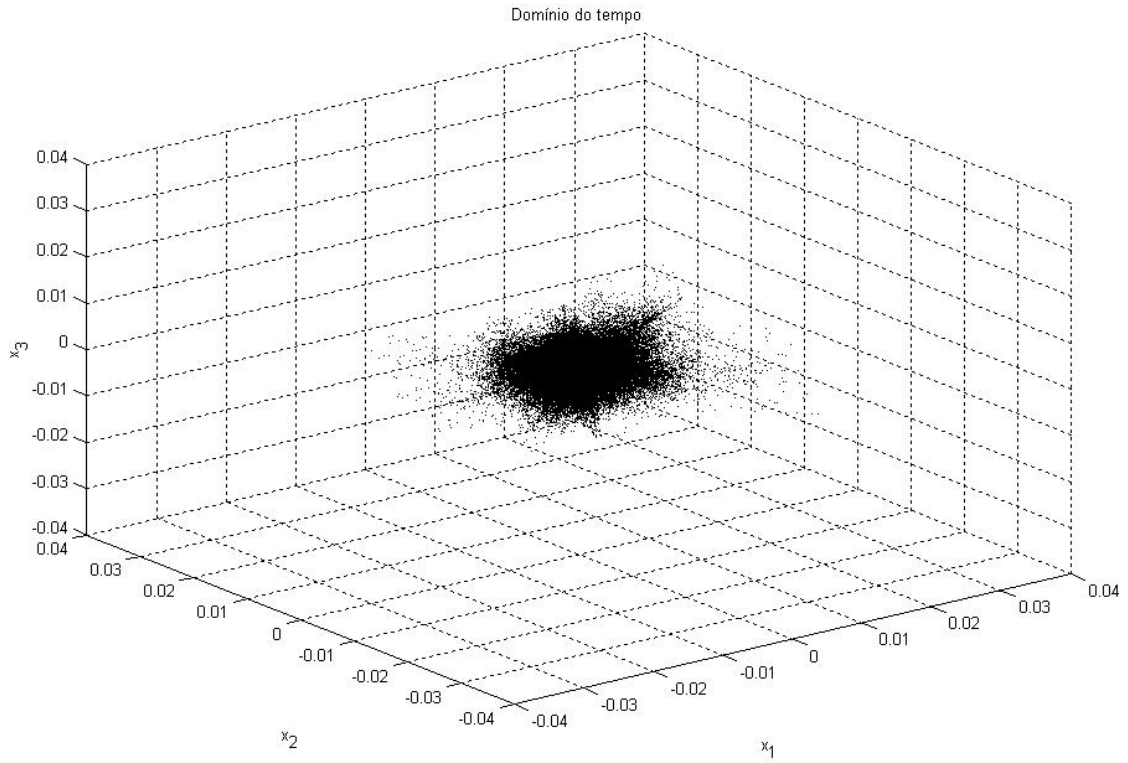


Figura 4.3: Gráfico de $x_3(n) \times x_2(n) \times x_1(n)$, amostras no domínio do tempo.

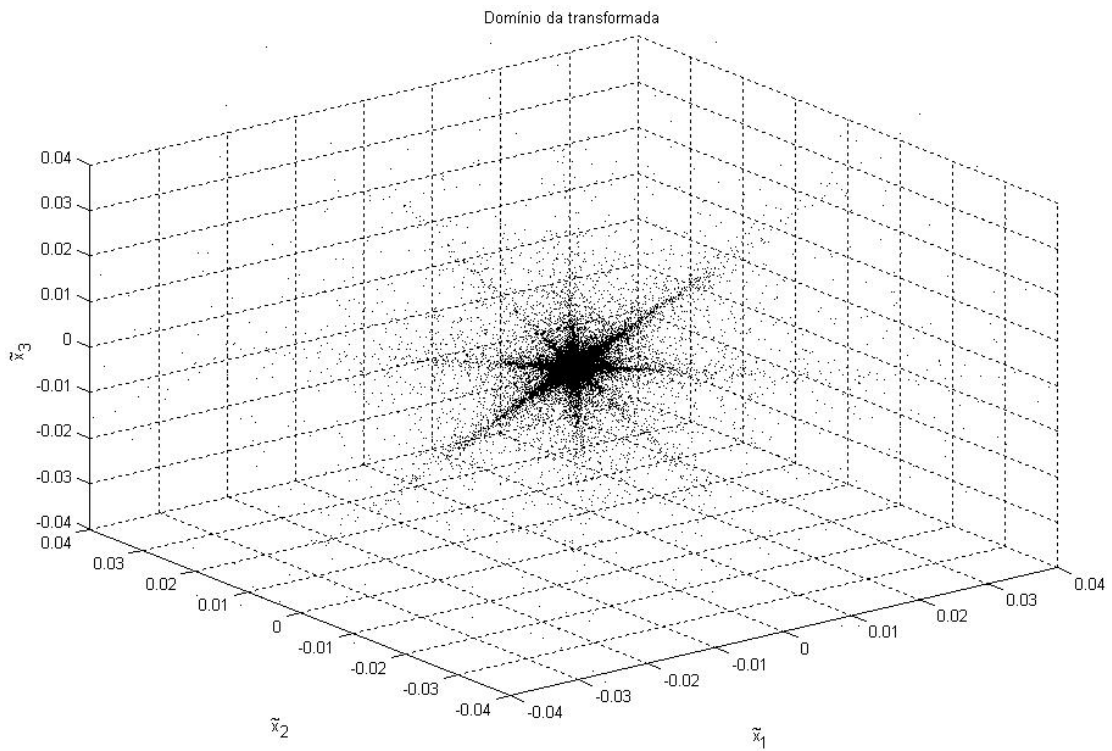


Figura 4.4: Gráfico de $\tilde{x}_3(n) \times \tilde{x}_2(n) \times \tilde{x}_1(n)$; coeficientes numa transformada *wavelet packet* com 7 níveis de decomposição.

definida pelo ângulo $\pi + \theta$, já que consideramos apenas retas que contém a origem. Então o método calcula uma espécie de histograma de pontos em cada ângulo-teste, da forma a seguir.

Seja θ_t o ângulo de teste. O método efetua uma varredura em todos os ângulos θ_n associados aos pontos de $\tilde{x}_2(n) \times \tilde{x}_1(n)$, calculando a contribuição destes a θ_t . Os θ_t associados aos N picos mais altos do “histograma” (um histograma típico, com três fontes, é mostrado na Figura 4.5) são considerados os ângulos das retas de cada fonte. A contribuição de $\mathbf{x}(n)$ (e seu respectivo θ_n) a θ_t é avaliada segundo a “função potencial” abaixo:

$$\Phi(\theta_t, \theta_n, \mathbf{x}(n)) = \begin{cases} \lambda |\mathbf{x}(n)| \left(1 - \frac{|\Delta|}{\pi/4}\right), & |\Delta| < \frac{\pi}{4} \\ 0, & |\Delta| \geq \frac{\pi}{4} \end{cases}, \quad (4.5)$$

onde $\Delta = \theta_t - \theta_n$ e λ é um parâmetro arbitrário, tipicamente da ordem de dezenas. A função potencial acima dá um peso para a proximidade de θ_n (sendo nula a contribuição caso esteja muito distante) e atribui uma confiança maior aos $\mathbf{x}(n)$ de maior módulo (pois o erro angular destes é menos sensível a ruído). O método acima descrito despreza os pontos $(\tilde{x}_1(n), \tilde{x}_2(n))$ de módulos muito baixos, pois estes costumam ser muito suscetíveis a ruído e interferências. Este procedimento possui a vantagem de agilizar bastante o processamento e tornou-se comum na literatura. Tipicamente, desprezam-se os pontos cujo módulo seja inferior a um décimo do maior módulo de $\tilde{\mathbf{x}}(n)$ (o que equivale a retirar os componentes menos super-gaussianos [67]).

4.2.2 Modelos de Misturas de Exponenciais

Um método mais recente [66], [67] utiliza a técnica denominada *Expectation Maximization* (EM) para encontrar os parâmetros de um modelo de mistura de exponenciais. Antes de tratar deste modelo, importa condensar o porquê de o método EM convergir para um máximo (local) de uma função, algo nem sempre bem entendido.

A idéia básica do algoritmo EM pode ser melhor compreendida caso comparemo-lo com o método de Newton, já que ambos recorrem, a cada iteração, a uma maximização de uma função na vizinhança da estimativa mais recente. Este procedimento foi adotado em [68], fonte da abordagem que vem a seguir.

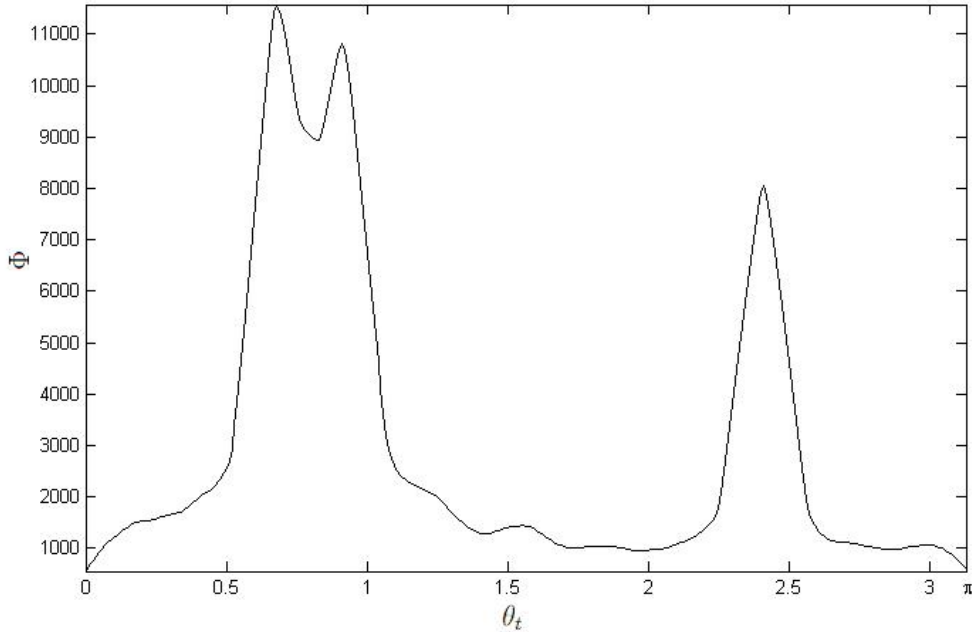


Figura 4.5: Função potencial típica, com 3 fontes e $\lambda = 5$.

Tendo por objetivo maximizar uma função escalar $f(\theta)$ de um vetor de variáveis θ , o método de Newton efetua sucessivos refinamentos, rumo a um máximo local de f , a partir de uma estimativa inicial $\theta^{(0)}$. O refinamento consiste em aproximar f na vizinhança de $\theta^{(0)}$ por um parabolóide, para então obter o máximo deste por meio de um sistema linear de equações. A solução deste sistema é $\theta^{(1)}$, resultado do primeiro refinamento, o qual é repetido de forma iterativa até que $\theta^{(i)}$ (obtido a partir do i -ésimo refinamento) seja muito próximo de $\theta^{(i-1)}$ (condição de parada).

O método EM, assim como o de Newton, é iterativo; fundamenta-se numa função $b_i(\theta)$ a qual nunca supera f na vizinhança de $\theta^{(i)}$ (sendo portanto uma espécie de limite inferior - *lower bound*, em inglês). A cada iteração encontra-se o vetor $\theta^{(i)}$ que maximiza a função b_i . Cumpre ressaltar que tanto o parabolóide do método de Newton quanto a função b_i do método EM coincidem com f no ponto $\theta^{(i)}$. Para o EM, não há necessidade de b_i admitir a forma de um parabolóide; já o método de Newton não exige que o parabolóide seja sempre inferior ou igual a f na vizinhança da estimativa $\theta^{(i)}$. Daí não podermos afirmar que um método seja uma extensão do outro.

Resumindo, temos as seguintes propriedades de b_i : (1) $b_i(\theta) \leq f(\theta)$ na vizinhança de $\theta^{(i)}$ e (2) $b_i(\theta^{(i)}) = f(\theta^{(i)})$. Estas propriedades garantem $f(\theta^{(i+1)}) \geq$

$f(\theta^{(i)})$, já que:

$$\begin{aligned} f(\theta^{(i+1)}) &\geq b_i(\theta^{(i+1)}) \text{ por (1),} \\ b_i(\theta^{(i+1)}) &\geq b_i(\theta^{(i)}) \text{ devido à maximização de } b_i(\theta), \\ b_i(\theta^{(i)}) &= f(\theta^{(i)}) \text{ por (2).} \end{aligned} \tag{4.6}$$

Não há como prever de forma genérica qual dos dois métodos obterá melhores resultados. O desempenho do método de Newton depende fortemente da função f ; a acurácia da estimativa via EM não depende somente de f , mas também da função b_i adotada (o que implica que esta deve ser escolhida de forma apropriada). Ambos os métodos são sensíveis à escolha de $\theta^{(0)}$, apresentando em geral resultados distintos para diferentes estimativas iniciais. A importância do método EM para esta dissertação reside na sua utilidade em problemas de estimativa dos parâmetros de uma distribuição.

Nesta seção, o método EM será aplicado a uma mistura de distribuições exponenciais. Cumpre lembrar que a distribuição exponencial é muito utilizada para modelar a esparsidade. Uma distribuição exponencial (tendo por suporte todos os números reais) é definida por:

$$p_e(\theta, c, m) = \frac{1}{2\mu} e^{-\frac{|\theta-m|}{\mu}} = ce^{-2c|\theta-m|}, \tag{4.7}$$

onde $c = 1/(2\mu) > 0$.

Uma distribuição denominada “mistura de exponenciais” pode ser expressa como:

$$p(\theta) = \sum_{i=1}^N \alpha_i c_i e^{-2c_i|\theta-m_i|}, \tag{4.8}$$

onde α_i , m_i e c_i representam a ponderação, a média e uma medida da largura de cada exponencial, respectivamente. Todas as ponderações são positivas e apresentam soma $\sum_{i=1}^N \alpha_i$ unitária.

O artigo [67] apresenta um algoritmo EM que maximiza a verossimilhança dos parâmetros α_i , m_i e c_i de uma mistura de exponenciais. Esta mistura de exponenciais modela a distribuição dos ângulos θ_n definidos na seção anterior. Ao final do processo, os parâmetros m_i devem estar próximos dos ângulos de cada fonte (os quais dependem basicamente da matriz de misturas).

Se a probabilidade de θ_n pertencer à i -ésima exponencial for dada por:

$$p(i|\theta_n) = \frac{\alpha_i c_i e^{-2c_i|\theta-m_i|}}{\sum_{i=1}^N \alpha_i c_i e^{-2c_i|\theta-m_i|}}, \quad (4.9)$$

podemos escrever as equações de atualização na forma a seguir [67]:

$$\begin{aligned} \alpha_i^+ &\leftarrow \frac{1}{K} \sum_{n=1}^K p(i|\theta_n), \\ m_i^+ &\leftarrow \frac{\sum_{n=1}^K \frac{\theta_n}{|\theta_n-m_i|} p(i|\theta_n)}{\sum_{n=1}^K \frac{1}{|\theta_n-m_i|} p(i|\theta_n)}, \\ c_i^+ &\leftarrow \frac{\sum_{n=1}^K p(i|\theta_n)}{2 \sum_{n=1}^K |\theta_n - m_i| p(i|\theta_n)}. \end{aligned} \quad (4.10)$$

O método acima apresentado, além de necessitar conhecer *a priori* o número N de fontes, necessita de alguns cuidados no tocante à inicialização (já observamos que o método EM é sensível às estimativas iniciais). Este é um tópico importante para evitar, por exemplo, que duas exponenciais distintas convirjam para o mesmo ângulo. Uma possibilidade seria utilizar um passo de inicialização baseado na clássica técnica de clusterização *K-means*. O artigo [67] pondera que uma inicialização dos ângulos em intervalos uniformes em $[-\pi/2, \pi/2]$ é mais razoável.

Há uma outra dificuldade neste método, oriunda do fato de a mistura exponencial ter por domínio o intervalo $(-\infty, +\infty)$ e os ângulos θ_n serem restritos ao intervalo $[-\pi/2, \pi/2]$, cujos limites $-\pi/2$ e $\pi/2$ são na verdade conectados, devido à função atan (vide a Equação 4.4). Esta diferença entre os valores reais e o modelo teórico acarreta um desvio significativo nas estimativas, ocorrendo o caso crítico quando pelo menos um dos ângulos das fontes se aproxima de $\pi/2$ ou de $-\pi/2$. Ainda não há solução satisfatória para este problema.

Para o caso de uma matriz de mistura que varia lentamente no tempo, este método pode atuar num esquema adaptativo, porém são necessárias algumas modificações para evitar instabilidade. Este caso não será abordado nesta dissertação.

4.2.3 Clusterização *K-means*

Como vimos, há uma tendência dos ângulos θ_n agruparem-se em torno de alguns valores, os quais são determinados pelas colunas da matriz de misturas.

Esta configuração torna natural o recurso a algoritmos de clusterização, os quais automaticamente separam dados em grupos, cada qual destes representando uma “classe” (no nosso caso, cada grupo é associado a uma fonte). Os algoritmos de clusterização constituem uma parte de um grupo mais geral de técnicas de aprendizagem não-supervisionada [69].

O método de clusterização *K-means* é um dos mais simples métodos de clusterização. Supondo que se queira dividir os ângulos em K grupos, na inicialização basta efetuar uma partição arbitrária dos ângulos, computando então a média (centróide) de cada *cluster*. A recursão consiste em (I) associar cada um dos ângulos ao *cluster* de centróide mais próximo (onde o conceito de “próximo” deriva de uma medida de distância, comumente a euclidiana) e (II) recalculer os centróides de cada *cluster*, retornando ao passo (I) caso haja modificação em algum dos centróides.

A clusterização *K-means* foi empregada, para propósitos de estimativa da matriz de misturas, em [70]. Em [64], há a observação de que é difícil estimar com precisão a matriz de mistura, utilizando a clusterização *K-means*, quando a esparsidade das fontes é insuficiente. Em outras palavras, a clusterização *K-means* apresenta grande sensibilidade no tocante à esparsidade obtida mediante transformação, podendo apresentar alta degradação nos resultados.

4.2.4 Clusterização *Fuzzy-C*

Podendo ser interpretada como uma generalização da clusterização *K-means*, a clusterização *fuzzy-C* (desenvolvida em [71] e melhorada em [72]) é utilizada, com o objetivo de estimar a matriz de misturas, em [73]. Nesta técnica, uma amostra pode pertencer, em diferentes graus, a *clusters* distintos.

Supondo dados unidimensionais para simplificar a nomenclatura, podemos descrever a clusterização *fuzzy-C* de ângulos θ_n , $n = 1, \dots, N$, como um processo de minimização da função-custo:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|\theta_i - c_j\|^2, \quad (4.11)$$

onde m é um número real igual ou superior a 1, c_j é o centróide do j -ésimo *cluster*, C é o número de *clusters* e u_{ij} reflete o grau em que o ângulo θ_i pertence ao j -ésimo

cluster; u_{ij} e c_j são calculados por:

$$u_{ij} = \frac{1}{\sum_{n=1}^C \left(\frac{\|\theta_i - c_j\|}{\|\theta_i - c_n\|} \right)^{\frac{2}{m-1}}} \quad (4.12)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m c_i}{\sum_{i=1}^N u_{ij}^m} \quad (4.13)$$

O método é iterativo, e o critério de parada ocorre quando $\max_{ij} \{|u_{ij}^{(k+1)} - u_{ij}^{(k)}|\} < \epsilon$, sendo ϵ um parâmetro arbitrário e k o número da iteração. O procedimento pode convergir para um mínimo local da função-custo.

As equações de iteração (a partir de centróides arbitrariamente escolhidos para inicialização) são dadas por:

$$c_j^{(k+1)} = \frac{\sum_{i=1}^N (u_{ij}^{(k)})^m \theta_i}{\sum_{i=1}^N (u_{ij}^{(k)})^m} \quad (4.14)$$

$$u_{ij}^{(k+1)} = \frac{1}{\sum_{n=1}^C \left(\frac{\|\theta_i - c_j^{(k)}\|}{\|\theta_i - c_n^{(k)}\|} \right)^{\frac{2}{m-1}}}. \quad (4.15)$$

Após a convergência, caso quisermos associar a i -ésima amostra a apenas um dos *clusters*, basta associá-la ao *cluster* dado por $\max_j \{c_{ij}\}$.

4.2.5 Clusterização Hierárquica

Num contexto diferente², [74] utiliza a técnica de clusterização hierárquica para estimar os ângulos associados às colunas da matriz de mistura. O algoritmo da clusterização hierárquica é tão simples quanto o da clusterização *k-means*.

A estratégia consiste em, inicialmente, associar cada amostra a um *cluster*. Então, verifica-se quais destes *clusters* apresentam menor distância entre si (foi adotada a métrica euclidiana padrão para a medida de distância). Estes dois são aglomerados em um único *cluster*. Então, repete-se a aglomeração entre os *clusters* mais próximos até que o número de *clusters* chegue a um número previamente arbitrado c (costumeiramente $c \gg N$). Quando pelo menos um dos *clusters* contiver

²O artigo trabalha no contexto de misturas indeterminadas convolutivas e efetua a separação no domínio da transformada de Fourier, onde são necessárias algumas adaptações para o caso de coeficientes complexos.

mais do que uma amostra, a distância entre eles é definida como a menor distância entre quaisquer de seus elementos.

Quando o número de *clusters* iguala c , são escolhidos como válidos os N *clusters* mais numerosos (caso, entre os *clusters* mais numerosos, haja *clusters* de médias muito próximas, devemos extrair os redundantes). A média de cada um destes representa a estimativa final do método.

4.2.6 Método DUET

O método DUET, apresentado em [61], apresenta semelhanças com a varredura de Zibulevsky. Uma característica importante que o distingue dos demais métodos vistos neste capítulo é a sua capacidade de tratar do caso (mais realista) de misturas anecóicas, onde ocorrem atrasos das fontes entre as misturas. Outra distinção notória é a necessidade de se utilizar a transformada de Fourier de tempo curto (STFT, do inglês *Short Time Fourier Transform*) para estimar o atraso entre as fontes. Definindo uma janela $v(n)$ (que pode ser, por exemplo, retangular, de Hanning, de Hamming ou triangular), a STFT (discreta) de um sinal $s_i(n)$ pode ser escrita como [75]³:

$$\tilde{s}_i(\tau, \omega) = \sum_{n=-\infty}^{\infty} s_i(n)v(n - \tau)e^{-j\omega n}, \quad (4.16)$$

onde $\tau = k\tau_0$ e $\omega = l\omega_0$, com k e l inteiros. Cabe notar que STFT apresenta inversa, sendo equivalente à sua formulação contínua, sob algumas condições facilmente alcançáveis [75].

Embora apresentemos integralmente o método DUET, somente utilizaremos uma extensão do mesmo e no caso instantâneo, para propósitos de comparação. No caso instantâneo, o método DUET não exige que a transformada a efetuar nas misturas seja a STFT (podendo ser, por exemplo, uma transformada *wavelet packet*).

As duas misturas de uma configuração anecóica podem ser descritas matematicamente da forma (equivalente) a seguir:

$$x_1(n) = \sum_{j=1}^N s_j(n), \quad (4.17)$$

³Embora tenhamos empregado a STFT em capítulos anteriores, julgamos melhor fornecer uma definição somente neste momento, pois o método DUET se baseia mais explicitamente nela.

$$x_2(n) = \sum_{j=1}^N a_j s_j(n - \delta_j). \quad (4.18)$$

A descrição acima utiliza simplificações oriundas da ambigüidade de escalamento (impondo que todas as fontes sejam somadas sem ponderações na primeira mistura) e de *deslocamento*, uma ambigüidade que somente existe no caso de misturas anecóicas. A ambigüidade de escalamento implica que a uma estimativa da i -ésima fonte $\hat{s}_i(n)$ seja avaliada, quanto ao desempenho da separação, de forma equivalente à $\hat{s}_i(n - \delta_i)$.

O método DUET inspira-se na propriedade denominada “ortogonalidade W-disjunta” (em inglês, *W-disjoint orthogonality*, daqui em diante denominada apenas “W-ortogonalidade”), a qual significa que as representações no espaço tempo-freqüência são disjuntas. Definindo o suporte de uma representação da i -ésima fonte como o conjunto de pares (τ, ω) onde os coeficientes da transformada $\tilde{s}_i(\tau, \omega)$ são não-nulos, podemos dizer que dois sinais (presumivelmente das fontes) são W-ortogonais quando os suportes de suas transformadas não possuem interseção.

O método DUET exige que as fontes a separar sejam, ao menos aproximadamente, W-ortogonais. Esta condição pode não ser obtida, por exemplo, quando há um número relativamente alto de fontes, ruído ou mesmo no caso trivial de as fontes apresentarem razoável sobreposição no espaço tempo-freqüência [64].

A condição de W-ortogonalidade entre os sinais s_1 e s_2 pode ser expressa concisamente da forma a seguir:

$$\tilde{s}_1(\tau, \omega)\tilde{s}_2(\tau, \omega) = 0, \forall \tau, \omega. \quad (4.19)$$

Sendo as fontes W-ortogonais, podemos associar cada coeficiente não-nulo $\tilde{x}_j(\tau, \omega)$ a uma das fontes. Isto é facilmente implementado mediante um *mascara-mento binário*, obtido mediante um produto pela função de mascaramento $M_j(\tau, \omega)$. A função de mascaramento (também chamada de função indicador do suporte) da j -ésima fonte é dada por:

$$M_j(\tau, \omega) = \begin{cases} 1, & \tilde{s}_j(\tau, \omega) \neq 0 \\ 0, & \tilde{s}_j(\tau, \omega) = 0 \end{cases} \quad (4.20)$$

Assim, sob a condição de W-ortogonalidade, é fácil perceber que podemos recuperar a j -ésima fonte mediante a operação:

$$\tilde{s}_j(\tau, \omega) = M_j(\tau, \omega)\tilde{x}(\tau, \omega), \forall \tau, \omega, \quad (4.21)$$

seguida da transformada STFT inversa.

O artigo [61] apresenta um índice de W-ortogonalidade (denominado WDO_M), cujo valor máximo (a unidade) é obtido quando as fontes são W-ortogonais. Antes de defini-lo, conceituaremos dois critérios: a razão de sinal preservado (PSR, do inglês *preserved-signal ratio*) e a razão sinal-interferência (SIR, do inglês *signal-to-interference ratio*).

Para uma função de mascaramento $M_j(\tau, \omega)$ tal que $0 \leq M(\tau, \omega) \leq 1$, a PSR_{M_j} da j -ésima fonte é dada por:

$$PSR_{M_j} = \frac{\|M_j(\tau, \omega)\tilde{s}_j(\tau, \omega)\|^2}{\|\tilde{s}_j(\tau, \omega)\|^2}, \quad (4.22)$$

a qual pode ser entendida como a porção da energia da j -ésima fonte que permanece após a separação por meio da função de mascaramento $M_j(\tau, \omega)$.

Definindo a soma das interferências na j -ésima fonte como:

$$y_j(n) = \sum_{k=1, j \neq k}^N s_k(n), \quad (4.23)$$

podemos expressar a razão sinal-interferência na saída (utilizando a máscara $M(\tau, \omega)$ para separação) por:

$$SIR_{M_j} = \frac{\|M_j(\tau, \omega)\tilde{s}_j(\tau, \omega)\|^2}{\|M_j(\tau, \omega)\tilde{y}_j(\tau, \omega)\|^2}. \quad (4.24)$$

O índice de WDO_M combina as medidas de SIR e de PSR da forma a seguir:

$$\begin{aligned} WDO_{M_j} &= \frac{\|M_j(\tau, \omega)\tilde{s}_j(\tau, \omega)\|^2 - \|M(\tau, \omega)\tilde{y}_j(\tau, \omega)\|^2}{\|\tilde{s}_j(\tau, \omega)\|^2} \\ &= PSR_{M_j} - \frac{PSR_{M_j}}{SIR_{M_j}} \end{aligned} \quad (4.25)$$

Sinais W-ortogonais apresentam $PSR_{M_j} = 1$ e $SIR_{M_j} = \infty$, implicando um WDO_{M_j} igual à unidade. A obtenção de um alto WDO_{M_j} implica alta supressão de interferência e significativa preservação da energia das fontes.

Podemos definir uma família Φ de funções de mascaramento dada por [61]:

$$\Phi_j^p = \begin{cases} 1, & 20 \log \left(\frac{|\tilde{s}_j(\tau, \omega)|}{|\tilde{y}(\tau, \omega)|} \right) \geq p \\ 0, & \text{caso contrário} \end{cases} \quad (4.26)$$

onde Φ_j^p é a função indicador para os pontos no espaço tempo \times frequência onde s_j domina a interferência da mistura por p dB. Pode-se provar que o valor $p = 0$

maximiza WDO_{M_j} . Após diversos experimentos, [61] conclui que dois sinais de voz, amostrados em 16 kHz, com janelamento de Hamming, quadros com 1024 amostras e transformada STFT, são em média 93,6% W-ortogonais (em outras palavras, apresentam $WDO_{M_j} = 0,936$). Foram feitos experimentos com até 10 sinais de voz. Com a configuração de 10 fontes, o parâmetro WDO_{M_j} se manteve superior a 0,6, o que mostra uma razoável esparsificação de sinais de voz mediante a STFT.

Supondo que só a j -ésima fonte seja ativa (ou não-nula) em um ponto (τ, ω) , podemos calcular a_j e δ_j (vide Equação 4.18) por:

$$a_j = F(\tilde{x}_1(\tau, \omega), \tilde{x}_2(\tau, \omega)) = \frac{\tilde{x}_2(\tau, \omega)}{\tilde{x}_1(\tau, \omega)}, \quad (4.27)$$

$$\delta_j = G(\tilde{x}_1(\tau, \omega), \tilde{x}_2(\tau, \omega)) = -1/\omega \angle \left(\frac{\tilde{x}_2(\tau, \omega)}{\tilde{x}_1(\tau, \omega)} \right) \quad (4.28)$$

onde $\angle z$ é a fase do número complexo z .

Calculando a e δ em cada ponto (τ, ω) da representação de fontes aproximadamente W-ortogonais, notamos, da mesma forma que na varredura de Zibulevsky, uma tendência de aglomeração da densidade do histograma (definido mais adiante) em torno dos parâmetros $(a_j$ e $\delta_j)$ de cada fonte. O método DUET recorre a um histograma bidimensional para estimar a_j e δ_j , para $j = 1, \dots, N$. Como a varredura de Zibulevsky recorria a um histograma unidimensional para estimar a atenuação, o histograma utilizado no método DUET (o qual contempla tanto a atenuação quanto o atraso) faz com que interpretemos o método DUET como uma extensão da varredura de Zibulevsky.

O histograma utilizado é uma espécie de *histograma ponderado*, o qual apresenta grande apelo do ponto de vista da estimativa de máxima verossimilhança (ML, do inglês *maximum likelihood*). A estimativa ML utilizada exige algumas hipóteses simplificadoras, como a de interferências entre misturas sendo gaussianas e iid - independentes e identicamente distribuídas - o que não corresponde à realidade, já que há uma correlação entre as interferências em cada mistura. Para o cálculo da estimativa ML, é mais adequado utilizar, ao invés do parâmetro a_j , o parâmetro α_j (denominado de *atenuação simétrica*), definido por:

$$\alpha_j = a_j - \frac{1}{a_j} \quad (4.29)$$

Sejam A e D , respectivamente, os parâmetros de resolução para α e para δ . As funções indicadoras dos pontos no espaço tempo \times frequência que apresentam

atenuação simétrica (estimada) distanciada no máximo de A (em relação a um α considerado) e atraso (estimado) a uma distância de no máximo D de um δ considerado são definidas por:

$$M_{\alpha,A}[k,l] = \begin{cases} 1, & |\hat{\alpha}[k,l] - \alpha| < A \\ 0, & \text{caso contrário} \end{cases} \quad (4.30)$$

$$M_{\delta,D}[k,l] = \begin{cases} 1, & |\hat{\delta}[k,l] - \delta| < D \\ 0, & \text{caso contrário} \end{cases} \quad (4.31)$$

Utilizando larguras de resolução β e Δ e expoente p , definimos o histograma do método DUET, em função de α e δ como:

$$h(\alpha, \delta) = \sum_{k,l} M_{\alpha,\beta/2}[k,l] M_{\delta,\Delta/2}[k,l] |\tilde{x}_1[k,l] \tilde{x}_2[k,l]|^p, \quad (4.32)$$

o qual é suavizado mediante uma convolução bidimensional com um kernel retangular r abaixo definido:

$$r(\alpha, \delta) = \begin{cases} \frac{1}{AD}, & (\alpha, \delta) \in \left[-\frac{A}{2}, \frac{A}{2}\right] \times \left[-\frac{D}{2}, \frac{D}{2}\right] \\ 0, & \text{caso contrário} \end{cases} \quad (4.33)$$

Sendo H o histograma suavizado, podemos descrevê-lo sucintamente da forma a seguir:

$$H(\alpha, \delta) = [h * r](\alpha, \delta). \quad (4.34)$$

Eis, por fim, uma sucinta descrição do método DUET:

- (I) a partir das misturas $x_1(n)$ e $x_2(n)$, construa as representações $\tilde{x}_1[k,l]$ e $\tilde{x}_2[k,l]$.
- (II) para cada ponto no espaço tempo \times frequência, calcule $(\hat{\alpha}[k,l], \hat{\delta}[k,l])$.
- (III) construa um histograma e localize seus picos:
 - a) construa um histograma por meio da Equação (4.32);
 - b) suavize-o por meio da Equação (4.34);
 - c) localize os N picos no histograma, cada qual associado a uma fonte e localizado na vizinhança dos parâmetros reais $(a_j, \delta_j), j = 1, \dots, N$.

4.2.7 Extensão dos Métodos DUET e TIFROM

Inspirado nos métodos do tipo DUET [76],[61] e do tipo TIFROM [77],[78],[79], o artigo [64] apresenta um método, doravante denominado EXT-DUETIF, que pode ser considerado uma extensão de ambos os tipos (por utilizar técnicas semelhantes de razão entre os dados das misturas), com menos restrições no que diz respeito à esparsidade. Este método pode ser aplicado ao caso de mais de 2 misturas, também se prestando à estimativa do número de fontes. Estas duas possibilidades não serão exploradas nesta dissertação.

Antes de expor o método, exporemos suas idéias subjacentes. Seja a matriz das misturas (supondo fontes esparsas ou convenientemente esparsificadas mediante uma transformada):

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{x}_1(1) & \cdots & \tilde{x}_1(K) \\ \vdots & \ddots & \vdots \\ \tilde{x}_M(1) & \cdots & \tilde{x}_M(K) \end{bmatrix}. \quad (4.35)$$

Dividir cada linha da matriz das misturas pela sua q -ésima linha (o caso de coeficientes nulos será contemplado no algoritmo) gera uma *matriz de razões* $\tilde{\mathbf{R}}_q$, da forma a seguir:

$$\tilde{\mathbf{R}}_q = \begin{bmatrix} \frac{\tilde{x}_1(1)}{\tilde{x}_q(1)} & \cdots & \frac{\tilde{x}_1(K)}{\tilde{x}_q(K)} \\ \vdots & \ddots & \vdots \\ \frac{\tilde{x}_M(1)}{\tilde{x}_q(1)} & \cdots & \frac{\tilde{x}_M(K)}{\tilde{x}_q(K)} \end{bmatrix}. \quad (4.36)$$

Quando, num instante $n = i$, apenas a j -ésima fonte for ativa (ou não-nula), a coluna correspondente das matriz das razões \mathbf{R}_q apresentará o seguinte formato:

$$\tilde{\mathbf{r}}_q^i = \begin{bmatrix} \frac{\tilde{x}_1(i)}{\tilde{x}_q(i)} \\ \vdots \\ \frac{\tilde{x}_M(i)}{\tilde{x}_q(i)} \end{bmatrix} = \begin{bmatrix} \frac{h_{1j}}{h_{qj}} \\ \vdots \\ \frac{h_{Mj}}{h_{qj}} \end{bmatrix} = \frac{\mathbf{h}^j}{h_{qj}} \quad (4.37)$$

No caso de uma fonte dominante, a igualdade acima transforma-se numa aproximação. Havendo M fontes, q varia de 1 a M . Nos instantes em que a j -ésima fonte é a única ativa, as matrizes de razões $\tilde{\mathbf{R}}_q$ apresentam as colunas correspondentes proporcionais à j -ésima coluna da matriz de misturas, com o fator de proporcionalidade dependente de q . Este fator de proporcionalidade ($1/h_{qj}$) não constitui problema nas estimativas devido à ambigüidade de escalamento, bastando normalizar as colunas da estimativa da matriz de misturas. Cabe notar que após esta

operação permanece uma *ambigüidade de sinal* facilmente contornável (impondo, por exemplo, que os coeficientes da primeira linha sejam todos positivos).

Admitindo que cada matriz de razões $\tilde{\mathbf{R}}_q$ apresenta colunas onde apenas uma fonte é a dominante, podemos encontrar nestas uma estimativa das colunas (devidamente normalizadas) da matriz de mistura. A principal contribuição do método EXT-DUETIF constitui encontrar estas colunas de forma cega. A estratégia é encontrar submatrizes da matriz de mistura $\tilde{\mathbf{R}}_q$, de dimensões $M \times L$, onde $L \ll K$, as quais apresentam colunas praticamente idênticas. Estas submatrizes podem fornecer uma estimativa de uma das colunas da matriz de misturas por meio de uma média de cada linha (a estimativa obtida deve ser normalizada). Como há mais do que uma matriz $\tilde{\mathbf{R}}_q$, uma coluna da matriz de mistura acaba por ser estimada mais de uma vez, exigindo uma posterior aglutinação de estimativas muito próximas.

A partir do acima explicado, podemos descrever o algoritmo EXT-DUETIF (simplificado para o caso de duas misturas). Alguns parâmetros utilizados no método e ainda não comentados serão logo a seguir definidos.

Algoritmo EXT-DUETIF

- (I) Aplique uma transformada esparsificadora (é proposta uma *wavelet packet*). Uma matriz $\tilde{\mathbf{X}}$ é obtida a partir desta operação (no caso de *wavelet packets*, cumpre concatenar os coeficientes da decomposição num arranjo unidimensional).
- (II) Encontre uma submatriz $\hat{\mathbf{X}}$ de $\tilde{\mathbf{X}}$ de forma que a norma de cada uma de suas colunas seja maior do que ξ_1 .
- (III) Para $m_1 = 1$ até 2, faça (laço incluindo os Passos III.1 e III.2):

(III.1) Se o módulo de uma entrada da q -ésima linha de $\hat{\mathbf{X}}$ é menor do que uma constante positiva ξ_2 , remova a coluna correspondente de $\hat{\mathbf{X}}$. Restando ainda K_1 colunas em $\hat{\mathbf{X}}$, de índices q_1, \dots, q_{K_1} , podemos escrever uma nova matriz de razões:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \frac{\hat{x}_1(q_1)}{\hat{x}_{m_1}(q_1)} & \dots & \frac{\hat{x}_1(q_{K_1})}{\hat{x}_{m_1}(q_{K_1})} \\ \vdots & \ddots & \vdots \\ \frac{\hat{x}_M(q_1)}{\hat{x}_{m_1}(q_1)} & \dots & \frac{\hat{x}_M(q_{K_1})}{\hat{x}_{m_1}(q_{K_1})} \end{bmatrix} \quad (4.38)$$

(III.2) $m_2 = -m_1 + 3$ (ou seja, $m_2 = 2$ quando $m_1 = 1$ e $m_2 = 1$ quando $m_1 = 2$).

(III.2.1) Encontre os valores mínimo e máximo (\tilde{r}_{m_2} e \tilde{R}_{m_2} , respectivamente), da m_2 -ésima linha de $\tilde{\mathbf{X}}$. Divida o intervalo $[\tilde{r}_{m_2}, \tilde{R}_{m_2}]$ uniformemente em M_0 raias. Então, divida a matriz $\tilde{\mathbf{X}}$ em M_0 submatrizes ($\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_{M_0}$) de forma que todas as entradas da m_2 -ésima linha de $\tilde{\mathbf{X}}_k$ pertençam à k -ésima raia ($k = 1, \dots, M_0$).

(III.2.2) A partir do conjunto de submatrizes $\{\tilde{\mathbf{X}}_k, k = 1, \dots, M_0\}$, apague as submatrizes que tiverem um número de colunas inferior a J_1 . O novo conjunto de submatrizes é denotado por $\{\tilde{\mathbf{X}}_{j_k}, k = 1, \dots, M_1\}$.

(IV) Calcule a média das linhas de todas as matrizes $\tilde{\mathbf{X}}_{j_k}$, obtidas no passo III.2.2, normalizando o vetor-coluna \mathbf{e}_i resultante. O resultado é uma estimativa de uma das colunas da matriz de mistura \mathbf{H} .

(V) Após todos os passos acima, obtemos um conjunto de colunas estimadas, contidos na matriz $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_{M_0}]$. Cumpra então remover as estimativas duplicadas (da mesma coluna), sendo as estimativas muito semelhantes aglutinadas em uma só.

Cabe esclarecer alguns dos passos cujos propósitos podem não ser claros. O passo (II), também utilizado em métodos já apresentados, reduz a carga computacional bem como remove as colunas muito afetadas por ruído. O Passo (III) é o essencial do método: gera matrizes semelhantes à $\tilde{\mathbf{R}}_q$, extraíndo as submatrizes onde apenas uma fonte é ativa de forma cega. A forma pela qual são extraídas estas submatrizes é muito simples: dividindo o intervalo de possíveis valores de uma linha e retendo os mais numerosos (indicando as direções onde se concentram as distribuições das fontes, de forma semelhante, por exemplo, à varredura de Zibulevsky).

Por fim, importa descrever como determinar os parâmetros ξ_1, ξ_2 e M_0 . Denominando Q_1 o maior módulo de todas as colunas da matriz das misturas e Q_2 a maior amplitude das entradas da mesma matriz, o método propõe $\xi_1 = 0,3Q_1$, $\xi_2 = 0,1Q_2$ e $M_0 = 400$. Já o parâmetro J_1 não apresenta um critério claro e geral de escolha.

4.2.8 Método Proposto: Modelos de Misturas de Gaussianas

O método por nós proposto para tratar do problema de estimativa das fontes reside num modelo de misturas gaussianas (sigla em inglês, GMM - *Gaussian Mixture Models*). Uma função de densidade de probabilidade (distribuição) de D dimensões que segue o modelo de misturas de K gaussianas, doravante denominada GMM_K , pode ser expressa por ⁴:

$$f(\mathbf{x}; \theta) = \sum_{k=1}^K p_k g(\mathbf{x}; \mathbf{m}_k, \sigma_k), \quad (4.39)$$

onde

$$g(\mathbf{x}; \mathbf{m}_k, \sigma_k) = \frac{1}{(\sqrt{2\pi}\sigma_k)^D} e^{-\frac{1}{2}\left(\frac{\|\mathbf{x}-\mathbf{m}_k\|}{\sigma_k}\right)^2}, \quad (4.40)$$

com

$$p_k \geq 0, \forall k \text{ e} \\ \sum_{k=1}^K p_k = 1, \quad (4.41)$$

onde \mathbf{m}_k e σ_k^2 são, respectivamente, a média e a variância da k -ésima gaussiana. Consideraremos nesta dissertação distribuições GMMs unidimensionais. Cabe lembrar que as GMMs são adequadas para a modelagem de *clusters*, sendo cada *cluster* associado a uma gaussiana, com a média desta de alguma forma situada no centro do *cluster* e dotada de uma variância, a qual é uma medida do espalhamento médio do *cluster*.

Nossa proposta consiste em recorrer a GMMs para modelar a distribuição dos ângulos θ_n (definidos da mesma forma que na varredura de Zibulevsky). Sabemos que estes se concentram em tantas direções quantas forem as fontes. Assim, a média da gaussiana associada a uma reta seria uma estimativa do ângulo da reta. No entanto, sabemos que a distribuição dos ângulos das fontes está muito longe de ser uma GMM_N , de onde se concluiria que não é muito útil recorrer a GMMs. Porém, caso estimássemos os parâmetros (média e variância) de L gaussianas, onde $L > N$ (tipicamente $L \geq 3N$), a aproximação seria muito menos grosseira. Esta abordagem

⁴Na realidade, este é um caso particular de uma GMM de D -dimensões, onde a matriz de covariância é entendida como a matriz identidade de ordem D multiplicada por σ_k^2 .

possui o mérito de contornar o fato de os ângulos não apresentarem uma distribuição igual a GMM_N , à custa de tornar necessário um pós-processamento destinado a estimar quais gaussianas possuem médias que representam colunas legítimas da matriz de mistura. Todo este custo computacional pode ser justificado caso as estimativas sejam mais exatas. O pós-processamento por nós proposto possui duas etapas. Primeiro, verificamos que, mais do que a probabilidade p_k de cada gaussiana, a variância é um fator relevante. Quanto menor a variância de uma GMM, maior a probabilidade de a mesma representar uma coluna legítima da matriz de misturas. Assim, a primeira parte do pós-processamento consiste em classificar as gaussianas em ordem crescente de variância. A segunda parte exclui as gaussianas ditas redundantes, ou seja, as que apresentam médias muito próximas. Quando há um grupo de gaussianas em torno de uma determinada média, todas as gaussianas pertencentes a este grupo são excluídas, exceto a gaussiana de menor variância. As N gaussianas restantes de menor variância são escolhidas para a estimativa da matriz de mistura (gaussianas “vencedoras”). Em alguns casos, pouco freqüentes, uma coluna da matriz de mistura é calculada de forma equivocada. Este fenômeno é comum na literatura [64],[66],[67]. Em geral, a gaussiana vencedora nestes casos possui uma probabilidade p_k muito baixa; assim, é possível refinar nosso método para reduzir a proporção destes casos.

Usualmente, a estimativa dos parâmetros da GMM é efetuada de forma iterativa (por meio do algoritmo EM) a partir de valores iniciais arbitrários, ou mesmo obtidos mediante métodos de clusterização (como por exemplo o célebre *K-means*). As equações de iteração (passo M do algoritmo) são [68], [80]:

$$m_k^{(i+1)} = \frac{\sum_{n=1}^K p^{(i)}(k|n)\theta_n}{\sum_{n=1}^K p^{(i)}(k|n)},$$

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D} \frac{\sum_{n=1}^K p^{(i)}(k|n) \|\theta_n - m_k^{(i+1)}\|^2}{\sum_{n=1}^K p^{(i)}(k|n)}}, \quad (4.42)$$

$$p_k^{(i+1)} = \frac{1}{K} \sum_{n=1}^K p^{(i)}(k|n), \quad (4.43)$$

onde $p^{(i)}(k|n) = \frac{p_k^{(i)} g(x_n; m_k^{(i)}, \sigma_k^{(i)})}{\sum_{k=1}^L p_k^{(i)} g(x_n; m_k^{(i)}, \sigma_k^{(i)})}$ e K significa o número total de pontos.

4.2.9 Simulações

A partir de três sinais de voz (duas femininas e uma masculina), amostrados em 16 kHz, efetuamos misturas com três diferentes matrizes, mostradas a seguir:

$$\mathbf{H}_1 = \begin{bmatrix} 1 & 1 & 1 \\ -0,9 & -0,5 & 1,3 \end{bmatrix}$$

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 & 1 \\ -0,9 & 0,5 & 1,3 \end{bmatrix}$$

$$\mathbf{H}_3 = \begin{bmatrix} 1 & 1 & 1 \\ -0,9 & 0,8 & 1,3 \end{bmatrix}$$

As dimensões das matrizes acima implicam uma configuração de três fontes e duas misturas. Este tipo de teste foi inspirado na avaliação efetuada em [16], onde foram comparados certos métodos simultâneos - não explorados nesta dissertação - para estimativa da matriz de misturas. A ambigüidade de escalamento é facilmente contornável assumindo que a primeira linha da matriz de mistura somente possui elementos iguais a 1. Esta configuração permite-nos expressar cada matriz \mathbf{H}_i como:

$$\mathbf{H}_i = \begin{bmatrix} 1 & 1 & 1 \\ \operatorname{tg} \psi_i^{(1)} & \operatorname{tg} \psi_i^{(2)} & \operatorname{tg} \psi_i^{(3)} \end{bmatrix}$$

A Figura 4.6 permite-nos visualizar graficamente as direções de cada coluna das matrizes \mathbf{H}_i .

Após efetuarmos as misturas, decompomo-las utilizando uma transformada do tipo *wavelet packet* com 7 níveis. A *wavelet* utilizada foi a Daubechies 44 (db44). Os métodos que utilizamos para estimar a matriz de mistura foram a varredura de Zibulevsky, o modelo de misturas de exponenciais, as clusterizações *K-means* e *fuzzy*, a extensão dos métodos DUET e TIFROM e o por nós proposto, baseado em GMMs.

Em [81], há argumentação no sentido de que utilizar a mediana dos *clusters*, ao invés da média, é um procedimento muito simples e que implica melhoras na estimativa. Devido a isto, as estimativas oriundas das clusterizações *K-means* e *fuzzy* foram obtidas tanto a partir de médias quanto de medianas dos *clusters*.

O rearranjo unidimensional da transformada *wavelet packet* de cada linha da

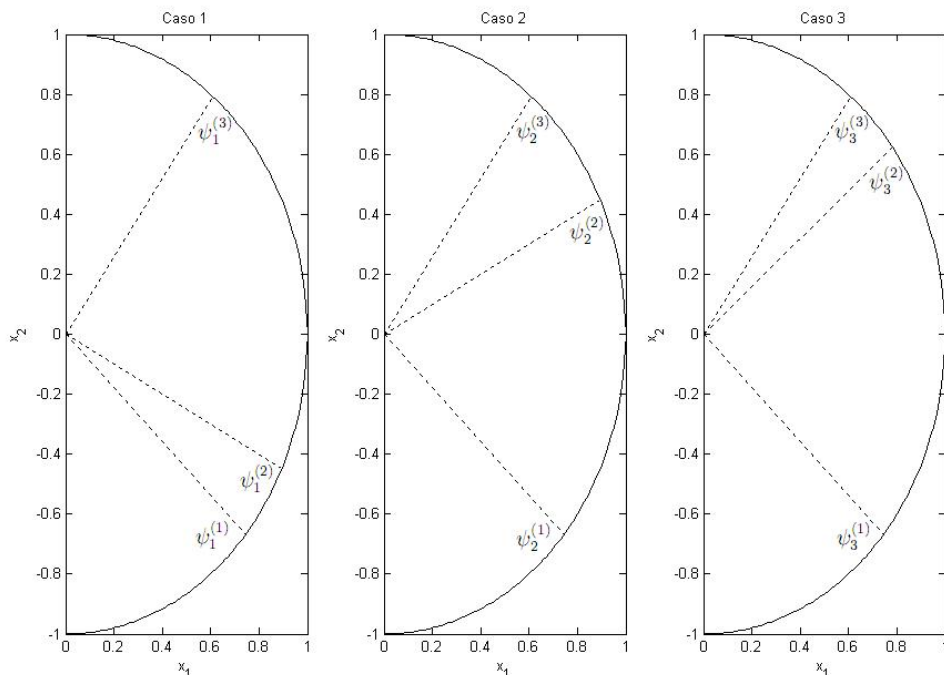


Figura 4.6: Direções geradas pelas matrizes \mathbf{H}_i nos três casos estudados.

matriz das misturas \mathbf{X} gera, como vimos, a matriz $\tilde{\mathbf{X}}$. Seja Q_1 o maior módulo dentre os módulos de $\tilde{\mathbf{x}}(n)$. A variável *exc* indica a fração de Q_1 acima da qual selecionam-se os vetores $\tilde{\mathbf{x}}(n)$ para estimativa (como vimos, a retirada dos vetores de menor módulo é justificada pelo fato de estes tenderem a ser mais ruidosos). Assim, por exemplo, o valor $exc = 0,2$ exclui todos os vetores $\tilde{\mathbf{x}}(n)$ de módulo inferior a 20% de Q_1 . Para todos os métodos testados, utilizamos o valor $exc = 0,05$. Duas foram as exceções. No caso da estimativa obtida a partir do modelo de misturas de exponenciais, foi utilizado o valor $exc = 0,01$; já para a extensão dos métodos DUET e TIFROM, utilizamos $exc = 0,3$.

Alguns dos métodos necessitam que arbitremos alguns parâmetros. Estes foram escolhidos de forma a otimizar o resultado de cada método. Para a varredura de Zibulevski, empregamos $\lambda = 10$, com um espaçamento entre ângulos de teste consecutivos de 0,01 radianos. No caso de misturas de exponenciais, utilizamos um limite superior de 3 para c_i (vide [67] para maiores detalhes). Já a extensão dos métodos DUET e TIFROM utilizou $M_0 = 400$, $\xi_1 = 0,3$, $\xi_2 = 0,01$ e $J_1 = 20$.

A Tabela 4.1 mostra as estimativas dos métodos escolhidos para as três matrizes de misturas. Nela, observamos que alguns métodos encontraram dificuldades

Tabela 4.1: Valores reais e estimativas oriundas de métodos distintos para os três casos abordados.

Caso 1			
valores reais	-0,9	-0,5	1,3
varredura de Zibulevsky	-0,8935	-0,5063	1,3028
mistura de exponenciais	-0,7932	-14,0721	1,2843
clusterização <i>K-Means</i> (média)	-0,7285	-15,3975	1,1843
clusterização <i>K-Means</i> (mediana)	-0,8226	-23,2922	1,2966
clusterização <i>Fuzzy</i> (média)	-0,7396	7,7369	1,1951
clusterização <i>Fuzzy</i> (mediana)	-0,8306	8,5181	1,2942
clusterização Hierárquica	-0,8995	-0,5062	1,2894
extensão de DUET-TIFROM	-0,9250	-0,6092	1,2890
GMM (proposto)	-0,89426	-0,5010	1,3046
Caso 2			
valores reais	-0,9	0,5	1,3
varredura de Zibulevsky	-0,8966	0,5062	1,3027
mistura de exponenciais	-0,8961	0,4963	1,2883
clusterização <i>K-Means</i> (média)	-0,8174	-70,2300	1,0931
clusterização <i>K-Means</i> (mediana)	-0,8782	72,2792	1,2818
clusterização <i>Fuzzy</i> (média)	-0,8695	0,5290	1,3470
clusterização <i>Fuzzy</i> (mediana)	-0,8959	0,5213	1,3052
clusterização Hierárquica	-0,8950	0,5096	1,2853
extensão de DUET-TIFROM	-0,82151	0,95761	1,3000
GMM (proposto)	-0,9003	0,5018	1,2994
Caso 3			
valores reais	-0,9	0,8	1,3
varredura de Zibulevsky	-0,8993	0,8103	1,3037
mistura de exponenciais	-0,9033	0,8008	1,2734
clusterização <i>K-Means</i> (média)	-0,8116	-17,4470	1,1500
clusterização <i>K-Means</i> (mediana)	-0,8795	-24,8988	1,2827
clusterização <i>Fuzzy</i> (média)	-0,8661	0,7730	1,3610
clusterização <i>Fuzzy</i> (mediana)	-0,8840	11,4023	1,2814
clusterização Hierárquica	-0,9321	0,8142	1,2975
extensão de DUET-TIFROM	-0,7939	0,93305	1,3306
GMM (proposto)	-0,8850	0,7998	1,3014

Tabela 4.2: Erros quadráticos médios oriundos de métodos distintos para os três casos abordados.

Caso 1	
varredura de Zibulevsky	2,9927.10⁻⁵
clusterização hierárquica	5,0675.10 ⁻⁵
extensão de DUET-TIFROM	4,2.10 ⁻³
GMM (proposto)	1,8369.10⁻⁵
Caso 2	
varredura de Zibulevsky	1,9097.10⁻⁵
mistura de exponenciais	5,5263.10 ⁻⁵
clusterização <i>Fuzzy</i> (média)	1,3.10 ⁻³
clusterização <i>Fuzzy</i> (mediana)	1,6585.10 ⁻⁴
clusterização hierárquica	1,1111.10 ⁻⁴
extensão de DUET-TIFROM	7,19.10 ⁻²
GMM (proposto)	1,2123.10⁻⁶
Caso 3	
varredura de Zibulevsky	4,009.10⁻⁵
mistura de exponenciais	2,3970.10 ⁻⁴
clusterização <i>Fuzzy</i> (média)	1,9.10 ⁻³
clusterização hierárquica	4,1221.10 ⁻⁴
extensão de DUET-TIFROM	10 ⁻²
GMM (proposto)	7,5271.10⁻⁵

em estimar o elemento central; por esta razão, as estimativas que divergem muito dos valores reais estão ressaltadas em negrito. O modelo de misturas de exponenciais falhou apenas no caso 1 neste quesito. Já a clusterização *K-Means* não logrou sucesso nos três casos, tanto nas estimativas oriundas de medianas quanto nas obtidas a partir de médias. A clusterização *Fuzzy* falhou no primeiro caso, enquanto que no terceiro a clusterização *Fuzzy* baseada em medianas também não obteve sucesso.

Na Tabela 4.2 apresentamos os erros quadráticos médios de cada método que, em cada um dos casos, obteve sucesso nas estimativas. Nos três casos estudados, o método proposto foi superior a todos nos casos 1 e 2, sendo o segundo melhor no caso 3. Os dois melhores métodos em cada caso encontram-se ressaltados em negrito.

O erro quadrático médio global (ou seja, de todos os três casos) é mostrado na Tabela 4.3. Seriam necessários testes estatísticos mais exaustivos para permitirem uma conclusão decisiva, mas esta tabela mostra que o método proposto apresenta erros compatíveis com os métodos concorrentes disponíveis na literatura (foi o se-

Tabela 4.3: Erros quadráticos médios globais dos quatro métodos que foram bem-sucedidos em todos os casos.

Método	Erro quadrático médio global
varredura de Zibulevsky	$2,9704.10^{-5}$
clusterização hierárquica	$1,9133.10^{-4}$
extensão de DUET-TIFROM	$2,87.10^{-2}$
GMM (proposto)	$3,1618.10^{-5}$

gundo melhor nos casos analisados).

4.3 Problema da reconstrução das fontes

Admitindo que as amostras de cada fonte são regidas por uma distribuição exponencial, a densidade de probabilidade conjunta da i -ésima fonte pode ser expressa como:

$$\prod_{n=1}^K \frac{1}{2\mu} e^{-\frac{|s_i(n)|}{\mu}} = \frac{1}{2^n \mu^n} e^{-\frac{1}{\mu} \sum_{n=1}^K |s_i(n)|}, \quad (4.44)$$

donde decorre, de forma trivial, que a solução de máxima probabilidade é a estimativa das fontes cuja soma dos módulos é menor. Esta constatação inspirou o método de reconstrução das fontes mais utilizado. Este método denomina-se “caminho mais curto” [63] (em inglês, *shortest path*), o qual estima a cada instante n o vetor $\mathbf{s}(n)$ que tenha norma l_1 mínima (sujeito à restrição $\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n)$). Pode-se provar que a solução deste problema possui $N - M$ zeros [74],[82],[83]. Daí sua solução ser particularmente rápida, como mostrado no exemplo ilustrativo abaixo (para o caso 2×3).

Seja uma matriz de misturas (estimada) igual a:

$$\hat{\mathbf{H}} = \begin{bmatrix} 1 & 1 & 1 \\ 0,2 & 1,1 & -2 \end{bmatrix}. \quad (4.45)$$

Primeiramente, cumpre tornar, por justiça, de norma unitária os vetores correspondentes à cada uma das direções, bastando normalizar cada coluna da matriz acima. Devido à ambigüidade de escalamento inerente aos métodos de separação

cega [84], esta operação não é problemática. Daí:

$$\hat{\mathbf{H}}_n = \begin{bmatrix} 0,98058 & 0,6726 & 0,44721 \\ 0,19612 & 0,73994 & -0,89443 \end{bmatrix}. \quad (4.46)$$

Suponhamos que num instante n tenhamos:

$$\mathbf{x}(n) = \begin{bmatrix} 0,91 \\ -0,356 \end{bmatrix}. \quad (4.47)$$

Supondo que apenas M fontes (considerando aqui $M = 2$) contribuem para o vetor acima (sendo a restante nula), efetuamos as seguintes estimativas:

$$\text{I: } \begin{bmatrix} \hat{s}_1(n) \\ \hat{s}_2(n) \end{bmatrix} = \begin{bmatrix} 0,9806 & 0,6726 \\ 0,1961 & 0,7399 \end{bmatrix}^{-1} \begin{bmatrix} 0,91 \\ -0,356 \end{bmatrix} = \begin{bmatrix} 1,5376 \\ -0,8887 \end{bmatrix}$$

$$\hat{s}_3(n) = 0 \text{ e } \sum_{i=1}^3 |\hat{s}_i(n)| = 2,4263$$

$$\text{II: } \begin{bmatrix} \hat{s}_1(n) \\ \hat{s}_3(n) \end{bmatrix} = \begin{bmatrix} 0,9806 & 0,4472 \\ 0,1961 & -0,8944 \end{bmatrix}^{-1} \begin{bmatrix} 0,91 \\ -0,356 \end{bmatrix} = \begin{bmatrix} 0,6786 \\ 0,5468 \end{bmatrix}$$

$$\hat{s}_2(n) = 0 \text{ e } \sum_{i=1}^3 |\hat{s}_i(n)| = 1,2255$$

$$\text{III: } \begin{bmatrix} \hat{s}_2(n) \\ \hat{s}_3(n) \end{bmatrix} = \begin{bmatrix} 0,6726 & 0,4472 \\ 0,7399 & -0,8944 \end{bmatrix}^{-1} \begin{bmatrix} 0,91 \\ -0,356 \end{bmatrix} = \begin{bmatrix} 0,702 \\ 0,9788 \end{bmatrix}$$

$$\hat{s}_1(n) = 0 \text{ e } \sum_{i=1}^3 |\hat{s}_i(n)| = 1,6809$$

Sendo a estimativa II a que possui menor soma dos módulos, ela seria a escolhida pelo método do caminho mais curto. Cabe observar que as inversões de matrizes necessárias acima só necessitam ser calculadas uma única vez, a partir da estimativa da matriz de mistura, pois independem de n . Quando M iguala N , o método acima coincide com o praticado na maioria dos algoritmos baseados em ICA.

4.3.1 Proposta I: Decomposição das fontes por meio de bancos de filtros não criticamente decimados

É sabido que podemos entender a transformada discreta de Fourier como oriunda de um banco de filtros de análise [75]. Já a transformada *wavelet*, no domínio

discreto, costuma ser *aproximada* mediante uma implementação com banco de filtros.

Interpretando a transformada *wavelet* (aproximada) como um banco de filtros, verificamos que o banco de filtros de síntese pode ser projetado de forma precisa para evitar efeitos de *aliasing* na saída [75]. Porém, qualquer modificação efetuada nos sinais antes de os mesmos serem introduzidos no banco de filtros de síntese (como por exemplo a quantização) destrói esse delicado equilíbrio, permitindo o surgimento de efeitos de *aliasing*. Quanto maior o fator de decimação, mais sensível a perturbações é o banco de filtros. Comumente, os artigos de análise de componentes esparsos que recorrem ao *wavelet packet* utilizam um banco de filtros criticamente decimado. Este fato, aliado ao fato de que a decomposição *wavelet* de cada fonte a que temos acesso na etapa de reconstrução não é a real (mas uma mera estimativa), pode tornar o *aliasing* um fator não desprezível no erro total. Se isso ocorre, conclui-se que parte do erro *não* se deve ao método de reconstrução das fontes, mas é oriundo de *aliasing*. Utilizar fatores de decimação/interpolação menores pode ajudar a reduzir a distorção por *aliasing*. Motivada por estas informações, nossa proposta consiste em recorrer a bancos de filtros com fatores de decimação inferiores ao fator crítico e averiguar se as razões sinal-interferência, sinal-ruído e sinal-artefato sofrem uma redução sensível. Em outras palavras, investigamos a influência do *aliasing* no erro de reconstrução.

4.3.2 Proposta II: Minimização de norma l_p , $p \neq 1$

A abordagem de minimização da norma l_1 pode ser deduzida mediante uma análise de máxima probabilidade *a posteriori*, admitindo que os coeficientes das fontes apresentam distribuições exponenciais [85]. Porém, é sabido que a distribuição destes coeficientes não é exatamente uma exponencial. Logo, não necessariamente a melhor solução seria a de mínima norma l_1 . Por isso, há a proposta [86] de minimizar a norma l_p ($p \leq 1$). Esta abordagem pode se traduzir em melhoras no caso de separação de sinais de áudio. Nossa proposta consiste em minimizar a norma l_p , onde p pode ser maior do que 1, caso surpreendentemente pouco explorado [87]. Nosso algoritmo de otimização, embora mais demorado, minimiza a norma l_p sem a restrição $p \leq 1$. Assim, podemos responder à questão: será que uma norma l_p com

$p > 1$, pode trazer benefícios à separação de sinais de áudio?

O problema em questão é minimizar $\|\hat{\mathbf{s}}(n)\|_p$, sujeito à restrição $\mathbf{x}(n) = \mathbf{H}\hat{\mathbf{s}}(n)$. Havendo M misturas, há M equações; já que o número de fontes N supera o número de misturas M (caso indeterminado), temos $N - M$ graus de liberdade. Suporemos que $N - M$ é igual a 1; é possível utilizar nosso algoritmo para os casos $N - M = c > 1$, porém caso c seja alto, pode advir uma “maldição da dimensionalidade” tornando intoleravelmente lento o processo de otimização.

Nossa proposta consiste primeiramente em arbitrar os limites de $\hat{s}_N(n)$ (a estimativa da fonte N no instante n). Se o intervalo $[a, b]$ for suficientemente extenso, espera-se que a solução ótima $\hat{s}_N^*(n)$ esteja dentro dele. Como há apenas um grau de liberdade (supondo $N - M = 1$), arbitrar um valor para $\hat{s}_N(n)$ implica arbitrar a solução $\hat{\mathbf{s}}(n)$. Variando o elemento $\hat{s}_N(n)$ no intervalo $[a, b]$, em L pontos uniformemente espaçados, calculamos a norma l_p da solução $\hat{\mathbf{s}}(n)$ correspondente. A solução com menor norma l_p será a estimativa inicial de nosso algoritmo. Seja Δ a distância entre os pontos $\hat{s}_N(n)$ utilizados no intervalo $[a, b]$. Podemos refinar a solução, dividindo o intervalo $[\hat{s}^* - \Delta + \epsilon, \hat{s}^* + \Delta - \epsilon]$ em L pontos novamente e repetindo o procedimento acima (ϵ é uma constante positiva próxima de zero). Este refinamento é efetuado diversas vezes. No caso $p = 1$, onde os valores ótimos são facilmente encontrados, a estimativa do nosso método apresenta um erro quadrático médio (comparada com a solução ótima), em nossas simulações, da ordem de 10^{-16} .

4.3.3 Proposta III: Introdução de um critério de continuidade

Os métodos de separação de fontes que recorrem a uma transformada partem da hipótese de que os coeficientes destas fontes no domínio da transformada são independentes. Isto, como sabemos, é falso. Na realidade, seria mais justo dizer que estes métodos, em geral, não utilizam a informação oriunda de uma eventual dependência estatística entre coeficientes próximos na decomposição [16]. Recorrer a esta dependência pode trazer benefícios ao processo de reconstrução das fontes, benefícios estes ainda mais explícitos no caso ruidoso [88],[89].

O artigo [90] utiliza um critério de minimização da norma quadrática da diferença entre coeficientes sucessivos, com relativamente bons resultados para o

difícil caso $M = 1$ (onde, como vimos, não se aplica a análise de componentes esparsos). Inspirada nesta abordagem, nossa proposta verifica se um critério de continuidade pode nos auxiliar quando no estágio de reconstrução exploramos a diversidade espacial.

No caso indeterminado, desejamos estimar \mathbf{S} sabendo que $\mathbf{X} = \mathbf{HS}$. Como há menos misturas do que fontes, há infinitas soluções. Admitir a esparsidade (e algumas hipóteses quanto às distribuições de probabilidade dos coeficientes) nos permite escolher a solução de norma mínima; da mesma forma, admitir a continuidade implica escolher a solução - também sujeita à restrição $\mathbf{X} = \mathbf{HS}$ - que apresenta menor norma quadrática da diferença entre sucessivas colunas de $\hat{\mathbf{S}}$.

Seja c_s a média da norma l_1 da estimativa $\hat{\mathbf{S}}$ e c_t dado por [90]:

$$c_t(\hat{\mathbf{S}}) = \sum_{i=1}^N \frac{1}{\sigma_i^2} \sum_{n=2}^K (\hat{s}_i(n) - \hat{s}_i(n-1))^2, \quad (4.48)$$

onde σ_i^2 é a variância da i -ésima linha de $\hat{\mathbf{S}}$ (estimativa da i -ésima fonte). Definindo a função custo por:

$$c(\hat{\mathbf{S}}) = c_s + \alpha c_t, \quad (4.49)$$

onde α é uma constante não-negativa, podemos alterar a solução $\hat{\mathbf{S}}$, no sentido de torná-la menos descontínua à medida em que aumentamos α .

No caso 2×3 (2 misturas e 3 fontes), a otimização consiste em estimar a terceira linha de \mathbf{S} . As duas primeiras linhas são determinadas a partir da terceira (e das matrizes \mathbf{X} e \mathbf{H}). O método de otimização é o analisado em [91]. Avaliamos se há certos valores de α que incrementam o processo de reconstrução das fontes.

4.3.4 Simulações

Em todos os experimentos abaixo, as fontes são sinais de áudio, amostrados em 16 kHz e com uma duração de aproximadamente 10s.

4.3.4.1 Avaliação da dificuldade da etapa de reconstrução

Espera-se que reduzir o número de misturas, mantendo fixo o de fontes, torne o problema de separação mais difícil, assim como aumentar o número de

fontes, mantendo o de misturas. Por outro lado, espera-se que manter o número de misturas igual ao de fontes menos 1 ($M = N - 1$) seja um processo que apresente mais facilidade à medida em que se aumente o número de fontes (o caso $2 \times 3 - 2$ misturas e 3 fontes - deve ser mais difícil que o caso 15×16). Claro que “facilidade” refere-se à capacidade de estimarmos as fontes, e não à uma carga computacional mais leve. Para analisarmos esta dificuldade, suporemos que a matriz de mistura foi estimada com perfeição (portanto, concentramo-nos sobre o problema da reconstrução). Quando o número de fontes iguala o de misturas, a reconstrução é idealmente perfeita (meramente uma inversão de matriz).

Para avaliar a dificuldade da reconstrução, optamos trabalhar com a medida logerr , definida como:

$$\text{logerr} = 10 \cdot \log_{10}(\text{err}) \text{ [dB]}, \quad (4.50)$$

onde err é o erro quadrático médio de reconstrução.

Em nossas simulações, quando o número de fontes iguala o de misturas, o logerr obtido foi da ordem de -300 dB, devido à presença de erros de aproximação inerentes ao processo computacional.

Apresentaremos resultados experimentais da medida logerr que permitem avaliar de forma quantitativa a dificuldade do problema $M \times N$, avaliação esta que não encontramos na literatura. Por exemplo: será que o caso $N - 1 \times N$ converge para um logerr de -300 dB com rapidez? Seria o caso 6×9 mais fácil ou mais difícil do que o caso 7×11 ? Estas questões são avaliadas experimentalmente.

Algumas das configurações testadas são custosas computacionalmente. Por exemplo, o caso $M = 7$ e $N = 14$ apresenta um total de 3.432 combinações possíveis, cada qual representando uma matriz que deve ser invertida e multiplicada por cada $\mathbf{x}(n)$, para escolha da estimativa das fontes de norma mínima.

Devido ao custo computacional, os resultados foram obtidos a partir de uma única matriz de mistura. Cabe ressaltar que alguns dos casos foram testados com outras matrizes, e as diferenças não foram grandes, verificando-se as mesmas tendências das aqui apresentadas.

A Tabela 4.4 mostra o logerr para $M = 2, 3, 4$ e 5 e com o número de fontes variando de $M + 1$ a 14 (foi utilizado o método do caminho mais curto para a reconstrução). Nesta tabela, verificamos que os valores de logerr estão sempre muito

Tabela 4.4: A medida logerr para diferentes números de fontes e de misturas.

N	$M = 2$	$M = 3$	$M = 4$	$M = 5$
3	-58,2053	-	-	-
4	-54,6925	-59,0981	-	-
5	-53,4984	-56,8474	-61,6515	-
6	-52,4642	-55,1108	-58,8246	-65,4921
7	-51,4316	-53,6193	-56,9674	-60,7552
8	-50,8565	-53,0169	-55,8057	-58,5670
9	-50,8025	-52,9024	-54,7776	-57,1363
10	-50,4775	-52,2491	-53,8465	-55,8809
11	-50,4327	-52,1100	-53,5353	-55,5539
12	-50,4540	-52,2785	-53,7092	-55,6336
13	-50,2494	-52,2430	-53,6964	-55,4514
14	-50,0018	-51,7977	-53,1781	-54,3524

distantes de -300 dB. Outro dado a destacar é que uma pequena diferença no logerr implica grande diferença nos resultados. Por exemplo: o caso 2×3 apresenta estimativas muito melhores do que o caso 2×14 ; porém, a diferença no logerr entre estes casos foi de 8,2035 dB.

Também é intuitivo esperar que aumentar o número M de misturas implica facilitar a reconstrução das fontes. As Figuras 4.7 e 4.8 mostram os resultados para $N = 7, \dots, 14$, os quais corroboram estas expectativas. O caso $M = N - 1$ é apresentado na Figura 4.9.

Alguns dados interessantes podem ser destacados. Pela pequena diferença nas medidas de logerr, podemos concluir que o problema 12×13 possui praticamente a mesma dificuldade que o 13×14 ; isto também ocorre com os conjuntos $(7 \times 8, 12 \times 14)$, $(8 \times 10, 9 \times 12)$, $(8 \times 12, 4 \times 5)$, $(7 \times 11, 8 \times 14)$, $(3 \times 4, 7 \times 12)$, $(4 \times 6, 7 \times 13)$, $(5 \times 9, 6 \times 11)$, $(6 \times 12, 7 \times 14)$, $(3 \times 5, 4 \times 7, 6 \times 13)$, $(4 \times 8, 5 \times 10, 6 \times 14)$ e $(2 \times 4, 4 \times 9)$.

O caso 13×14 ainda está bem longe de possuir um logerr da ordem de -300 dB, que seria aproximadamente o erro no caso 14×14 . Como na prática dificilmente contemplaremos muito mais do que 14 fontes, podemos dizer que o mesmo no caso $N - 1 \times N$ com N elevado o erro ainda é muito elevado no processo de reconstrução.

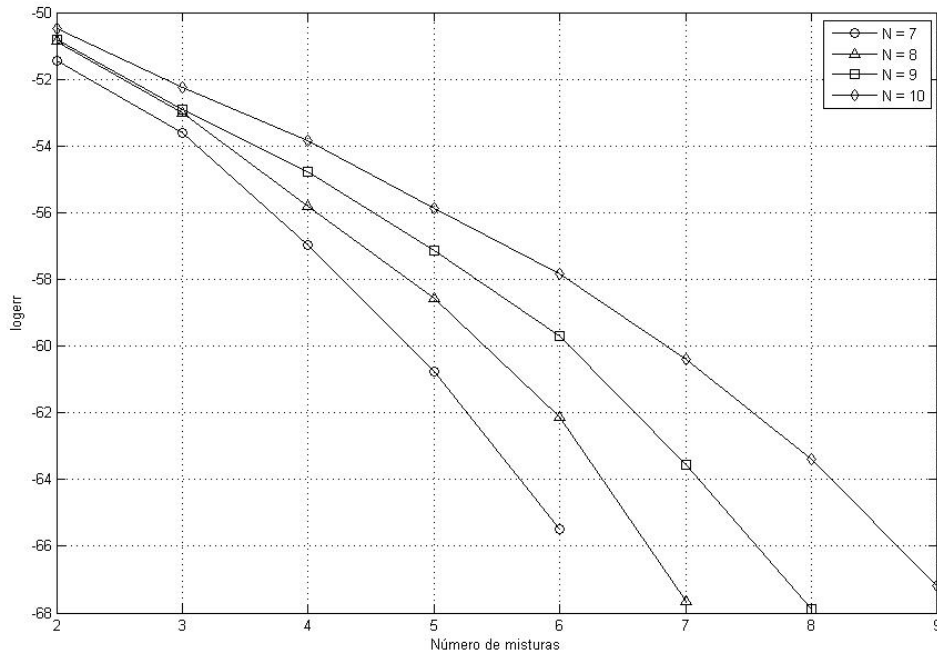


Figura 4.7: Medida logerr em função do número de misturas para $N = 7, 8, 9$ e 10 .

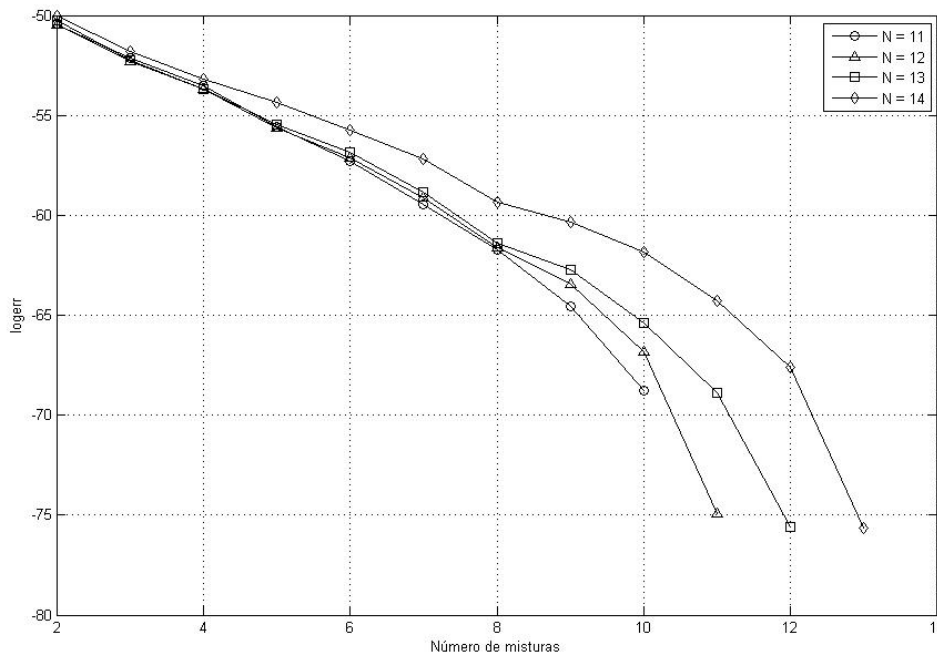


Figura 4.8: Medida logerr em função do número de misturas para $N = 11, 12, 13$ e 14 .

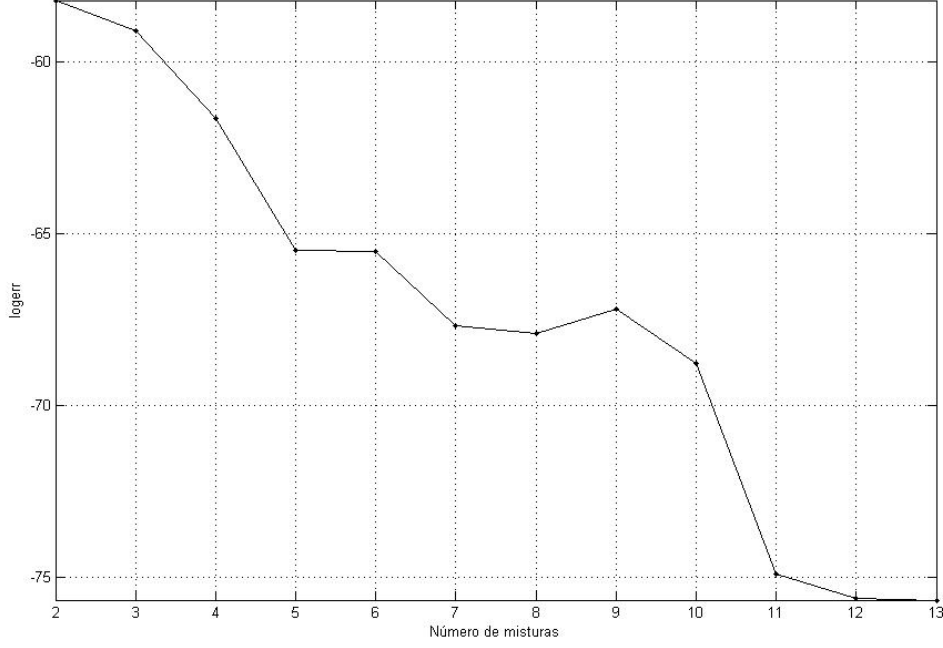


Figura 4.9: Medida logerr em função do número de misturas para o caso $M \times M + 1$.

4.3.4.2 Banco de filtros não criticamente decimados

Para avaliar os resultados da separação, aqui optamos por três medidas: SIR, SAR e SDR (respectivamente, razão sinal-ruído, razão sinal-artefato e razão sinal-distorção), calculadas a partir da proposta de [28]. Não foi utilizada apenas a medida SIR porque esta apenas mede a interferência oriunda de outras fontes, negligenciando o ruído musical (artefatos do tipo “*bubbling*”), o qual muitas vezes provoca mais desconforto na audição do que a interferência. Como a decimação não crítica alivia este tipo de artefatos, as medidas SAR e SDR também foram empregadas. Supondo fontes \mathbf{s}_i ($i = 1, \dots, M$) descorrelacionados e ausência de ruído de medição, podemos decompor $\hat{\mathbf{s}}_j$ (a estimativa da j -ésima fonte) como [28]:

$$\mathbf{s}_{\text{target}} = \frac{\langle \hat{\mathbf{s}}_j, \mathbf{s}_j \rangle \mathbf{s}_j}{\|\mathbf{s}_j\|^2}, \quad (4.51)$$

$$\mathbf{e}_{\text{interf}} = \sum_{j' \neq j} \frac{\langle \hat{\mathbf{s}}_j, \hat{\mathbf{s}}_{j'} \rangle \mathbf{s}_{j'}}{\|\mathbf{s}_{j'}\|^2}, \quad (4.52)$$

$$\mathbf{e}_{\text{artef}} = \hat{\mathbf{s}}_j - \mathbf{s}_{\text{target}} - \mathbf{e}_{\text{interf}}, \quad (4.53)$$

de modo que:

$$\mathbf{s}_j = \mathbf{s}_{\text{target}} + \mathbf{e}_{\text{interf}} + \mathbf{e}_{\text{artef}}. \quad (4.54)$$

Após esta decomposição, [28] calcula o SIR, o SAR e o SDR da forma a seguir:

$$\text{SIR} = 10\log_{10} \frac{\|\mathbf{s}_{\text{target}}\|^2}{\|\mathbf{e}_{\text{interf}}\|^2}, \quad (4.55)$$

$$\text{SAR} = 10\log_{10} \frac{\|\mathbf{s}_{\text{target}} + \mathbf{e}_{\text{interf}}\|^2}{\|\mathbf{e}_{\text{artef}}\|^2}, \quad (4.56)$$

$$\text{SDR} = 10\log_{10} \frac{\|\mathbf{s}_{\text{target}}\|^2}{\|\mathbf{e}_{\text{interf}} + \mathbf{e}_{\text{artef}}\|^2}. \quad (4.57)$$

Utilizamos, para comparação, 4 diferentes *wavelets*: bior6.8, coif5, db16 e sym9. Os níveis de decomposição foram variados de 1 a 7; os fatores de decimação D^5 foram variados em potências de 2 variando de 1 a 2^L , onde L é o número de níveis de decomposição. Os resultados constam nas Tabelas 4.5, 4.6, 4.7 e 4.8.

Nestas tabelas, vemos uma tendência de as três medidas de desempenho adotadas melhorarem com o aumento do número de níveis de decomposição. Na literatura em geral são utilizados 7 níveis de decomposição ($L = 7$). Este resultado é explicado pela tendência de os coeficientes da decomposição tornarem-se mais esparsos com o aumento de L , o que facilita a Análise de Componentes Esparsos.

Outra informação a ressaltar das tabelas é a relativa independência das medidas de desempenho em relação às *wavelets* escolhidas. As quatro *wavelets* apresentaram resultados bastante próximos de SIR, SAR e SDR. A *wavelet* que apresentou os melhores resultados (ligeiramente superiores) foi a db16.

Quanto ao fator de decimação, verificamos que sua redução implica melhoras nas três medidas de desempenho. O SIR apresenta, em alguns casos, algumas atipias. Por esta razão, enfocaremos as medidas de SAR e SDR. As melhoras citadas podem ser atribuídas à redução dos efeitos de *aliasing*. É possível que o aumento de informação disponível também possa contribuir para estas melhoras.

⁵Cabe ressaltar que neste capítulo o parâmetro D difere do capítulo anterior.

Tabela 4.5: Resultados finais para sinais de voz com wavelet bior6.8. L é o número de níveis de decomposição, e D é o fator de decimação.

<i>wavelet</i>	L	D	SIR (dB)	SAR (dB)	SDR (dB)
bior6.8	1	1	10.614	9.4051	13.7891
bior6.8	1	2	10.6054	9.3488	13.7574
bior6.8	2	1	10.7224	9.5189	13.9777
bior6.8	2	2	10.7153	9.5078	13.965
bior6.8	2	4	10.7063	9.3534	13.9244
bior6.8	3	1	10.8061	9.9541	13.9869
bior6.8	3	2	10.8036	9.9479	13.9919
bior6.8	3	4	10.8179	9.9088	13.9575
bior6.8	3	8	10.8333	9.4407	13.9178
bior6.8	4	1	11.2727	10.9013	14.0459
bior6.8	4	2	11.2747	10.9019	14.0479
bior6.8	4	4	11.2765	10.884	14.0433
bior6.8	4	8	11.2812	10.7109	14.0238
bior6.8	4	16	11.413	9.3979	13.9584
bior6.8	5	1	11.954	11.8158	14.8234
bior6.8	5	2	11.9515	11.8129	14.8183
bior6.8	5	4	11.9369	11.7955	14.8038
bior6.8	5	8	11.9506	11.7281	14.8169
bior6.8	5	16	11.9935	11.2541	14.7864
bior6.8	5	32	12.2216	10.0536	15.0324
bior6.8	6	1	12.4902	11.373	15.1104
bior6.8	6	2	12.49	11.3713	15.1092
bior6.8	6	4	12.4854	11.3648	15.0923
bior6.8	6	8	12.5447	11.3304	15.1415
bior6.8	6	16	12.5472	11.0488	15.1979
bior6.8	6	32	12.6725	10.4806	15.5375
bior6.8	6	64	12.6927	10.0779	15.7576
bior6.8	7	1	13.8874	12.1432	16.9442
bior6.8	7	2	13.888	12.1422	16.9453
bior6.8	7	4	13.8852	12.1337	16.9372
bior6.8	7	8	13.9437	12.1553	16.9526
bior6.8	7	16	13.961	11.9532	16.8611
bior6.8	7	32	14.0659	11.6105	16.9458
bior6.8	7	64	14.0405	11.4995	16.9284
bior6.8	7	128	14.0119	10.7405	16.6818

Tabela 4.6: Resultados finais para sinais de voz com wavelet *coif5*. L é o número de níveis de decomposição, e D é o fator de decimação.

<i>wavelet</i>	L	D	SIR (dB)	SAR (dB)	SDR (dB)
<i>coif5</i>	1	1	10.6217	9.3777	13.802
<i>coif5</i>	1	2	10.6136	9.3268	13.7722
<i>coif5</i>	2	1	10.7163	9.4755	13.9484
<i>coif5</i>	2	2	10.7098	9.4648	13.9448
<i>coif5</i>	2	4	10.7156	9.3477	13.9641
<i>coif5</i>	3	1	10.7877	9.8606	13.9879
<i>coif5</i>	3	2	10.7858	9.8546	13.9983
<i>coif5</i>	3	4	10.7718	9.8286	14.024
<i>coif5</i>	3	8	10.7593	9.2396	14.0404
<i>coif5</i>	4	1	11.1932	10.8924	14.0707
<i>coif5</i>	4	2	11.1936	10.8925	14.0576
<i>coif5</i>	4	4	11.1961	10.8836	14.041
<i>coif5</i>	4	8	11.2094	10.7444	13.9803
<i>coif5</i>	4	16	11.1812	9.3799	13.9569
<i>coif5</i>	5	1	12.113	11.9386	15.0574
<i>coif5</i>	5	2	12.1128	11.9374	15.0527
<i>coif5</i>	5	4	12.1211	11.9259	15.0673
<i>coif5</i>	5	8	12.1587	11.8676	15.0857
<i>coif5</i>	5	16	12.2306	11.2326	14.9991
<i>coif5</i>	5	32	12.3348	10.1954	15.0585
<i>coif5</i>	6	1	12.78	11.6045	15.5085
<i>coif5</i>	6	2	12.7799	11.6043	15.5043
<i>coif5</i>	6	4	12.7852	11.5925	15.5188
<i>coif5</i>	6	8	12.7777	11.4936	15.472
<i>coif5</i>	6	16	12.8424	11.0736	15.4012
<i>coif5</i>	6	32	12.8573	10.4992	15.3746
<i>coif5</i>	6	64	12.8221	10.0706	15.3882
<i>coif5</i>	7	1	14.3115	12.2876	17.3842
<i>coif5</i>	7	2	14.3117	12.2876	17.3816
<i>coif5</i>	7	4	14.3149	12.2809	17.384
<i>coif5</i>	7	8	14.3169	12.1864	17.3596
<i>coif5</i>	7	16	14.4128	11.8485	17.3366
<i>coif5</i>	7	32	14.4343	11.4529	17.1967
<i>coif5</i>	7	64	14.4208	11.3138	17.2939
<i>coif5</i>	7	128	14.4371	10.5526	17.1118

Tabela 4.7: Resultados finais para sinais de voz com wavelet db16. L é o número de níveis de decomposição, e D é o fator de decimação.

<i>wavelet</i>	L	D	SIR (dB)	SAR (dB)	SDR (dB)
db16	1	1	10.6223	9.38	13.8609
db16	1	2	10.6221	9.3378	13.8673
db16	2	1	10.7178	9.4718	14.1203
db16	2	2	10.7137	9.4611	14.1356
db16	2	4	10.7208	9.3212	14.1929
db16	3	1	10.8008	9.83	13.8317
db16	3	2	10.7987	9.8264	13.8261
db16	3	4	10.7826	9.7835	13.8897
db16	3	8	10.7715	9.2477	13.7353
db16	4	1	10.9159	10.7535	13.5088
db16	4	2	10.9156	10.7525	13.5403
db16	4	4	10.9103	10.7223	13.6043
db16	4	8	10.8736	10.4792	13.7483
db16	4	16	10.9076	9.1689	13.9359
db16	5	1	12.1011	11.9116	15.0594
db16	5	2	12.1015	11.909	15.0509
db16	5	4	12.1035	11.8929	15.0261
db16	5	8	12.0829	11.8007	14.964
db16	5	16	11.9938	11.2554	14.7882
db16	5	32	11.8019	9.8841	14.4262
db16	6	1	12.9863	11.746	15.6594
db16	6	2	12.986	11.7445	15.6583
db16	6	4	12.9886	11.7362	15.6646
db16	6	8	12.9532	11.6059	15.6326
db16	6	16	12.898	11.3308	15.4947
db16	6	32	12.8248	10.5439	15.2441
db16	6	64	12.8137	10.2239	15.1191
db16	7	1	14.6067	12.4074	17.4663
db16	7	2	14.6074	12.4066	17.4636
db16	7	4	14.6082	12.4021	17.4633
db16	7	8	14.5721	12.2857	17.4271
db16	7	16	14.499	12.1206	17.3258
db16	7	32	14.4583	11.4699	17.0871
db16	7	64	14.4539	11.3166	17.0727
db16	7	128	14.4991	10.554	16.8822

Tabela 4.8: Resultados finais para sinais de voz com wavelet sym9. L é o número de níveis de decomposição, e D é o fator de decimação.

<i>wavelet</i>	L	D	SIR (dB)	SAR (dB)	SDR (dB)
sym9	1	1	10.621	9.3807	13.8285
sym9	1	2	10.6328	9.3416	13.7806
sym9	2	1	10.7059	9.468	13.9858
sym9	2	2	10.7094	9.4619	13.9858
sym9	2	4	10.7321	9.3598	14.1333
sym9	3	1	10.7676	9.8612	13.9822
sym9	3	2	10.7656	9.8545	13.993
sym9	3	4	10.7554	9.8334	13.9889
sym9	3	8	10.776	9.3191	14.0316
sym9	4	1	11.1416	10.885	14.023
sym9	4	2	11.1374	10.8782	14.0207
sym9	4	4	11.1313	10.8693	14.0273
sym9	4	8	11.1404	10.7435	14.0445
sym9	4	16	11.1543	9.4297	13.824
sym9	5	1	11.9575	11.8512	14.9073
sym9	5	2	11.9573	11.8477	14.9085
sym9	5	4	11.9649	11.8461	14.8859
sym9	5	8	11.9704	11.7809	14.898
sym9	5	16	11.9567	11.2033	14.7912
sym9	5	32	11.6563	9.721	14.3222
sym9	6	1	12.6608	11.5503	15.3999
sym9	6	2	12.6604	11.548	15.3949
sym9	6	4	12.6635	11.5383	15.4008
sym9	6	8	12.6402	11.448	15.3152
sym9	6	16	12.5775	11.0599	15.2379
sym9	6	32	12.4986	10.3759	15.1029
sym9	6	64	12.4346	9.9404	14.9599
sym9	7	1	14.1004	12.1996	17.1699
sym9	7	2	14.1011	12.1967	17.1595
sym9	7	4	14.1024	12.1876	17.169
sym9	7	8	14.0709	12.0951	17.1009
sym9	7	16	13.9708	11.7921	17.0066
sym9	7	32	13.8821	11.2666	16.7658
sym9	7	64	13.9459	11.1625	16.7511
sym9	7	128	14.0138	10.267	16.3338

Tabela 4.9: Melhora obtida nas medidas SAR e SDR quando não se recorre à decimação. Foi utilizada uma decomposição de 7 níveis.

<i>wavelet</i>	melhora no SAR	melhora no SDR
bior6.8	1,4027	0,2624
coif5	1,7350	0,2724
db16	1,8534	0,5841
sym9	1,9326	0,8361

A Tabela 4.9 apresenta as melhoras no SAR e no SDR nas quatro *wavelets* escolhidas, com 7 níveis de decomposição. A melhora foi obtida subtraindo-se a medida obtida quando se utilizou fator de decimação unitário (sem decimação) da medida obtida quando a decimação foi crítica. Nesta tabela, verificamos melhoras no SAR variando de 1,4027 a 1,9326 dB e melhoras no SDR de 0,2624 a 0,8361 dB.

4.3.4.3 Variação do número de fontes nulas a cada instante

Os métodos que contemplam a separação no caso indeterminado supõem, via de regra, que o número de fontes ativas (ou não-nulas) a cada instante é fixo. Em geral, este número ou é 1 ou M . Assim, ou admite-se apenas uma fonte ativa a cada instante (sendo as restantes consideradas nulas, numa forma de *maskamento binário*) [61], ou iguala-se o número de fontes ativas em todos os instantes ao número de misturas M (abordagem classicamente efetuada pelo método do caminho mais curto, com melhores resultados que os do *maskamento binário*) [63],[64],[74].

Porém, [92] introduziu o conceito de “esparsidade generalizada” (*generalized sparseness*), o qual permite a variação do número estimado de fontes ativas a cada instante n (daqui em diante denominado $\hat{s}_{act}(n)$), no intervalo $[1, M]$. Inspirados nesta proposta, investigamos experimentalmente os benefícios de variar-se \hat{s}_{act} com n , não mantendo-o fixo em 1 ou em M .

Nossa abordagem (também denominada de “melhor caso”) mantém distinções importantes em relação à apresentada em [92], entre as quais destacamos: (a) avalia quantitativamente a maior melhora possível dentro do nosso universo de possibilidades, expondo uma espécie de *limite superior de desempenho*; (b) não estima as fontes a partir de uma mínima distância quadrática entre $\mathbf{H}\hat{\mathbf{s}}(n)$ e $\mathbf{x}(n)$, mas medi-

ante o método do caminho mais curto e (c) não testa todas as combinações de fontes sendo nulas, zerando apenas as fontes (estimadas) que produziriam um erro menor caso fossem iguais a zero, numa estirpe de *hard threshold*, o qual agiliza bastante a estimativa ao evitar um cálculo combinatorial.

Nossa proposta consiste das seguintes etapas, efetuadas para cada instante n :

- (I) estimar as fontes utilizando o método do caminho mais curto;
- (II) igualar a zero cada uma das fontes ativas estimadas, verificando qual o valor mais próximo do real (0 ou o estimado em (I))
- (III) no caso de a j -ésima fonte (estimada) $\hat{s}_j(n)$ tornada nula apresentar erro menor em relação à fonte real do que a fonte estimada ($|\hat{s}_j(n) - s_j(n)| > |s_j(n)|$), anular $\hat{s}_j(n)$.

Obviamente, esta proposta, necessitando do conhecimento das fontes reais, não se presta a propósitos de separação; apenas é útil para o estabelecimento de um limite de desempenho. Neste sentido, não cabe compará-la com o método apresentado em [92]. Foi contemplado nesta seção o caso $M \times M + 1$.

A Tabela 4.10 apresenta a redução em porcentagem do erro quadrático médio de reconstrução utilizando um número variável de fontes nulas (limite superior de desempenho). Os resultados que constam nesta tabela nos permitem concluir que variar o número de fontes ativas a cada instante é um procedimento que pode reduzir significativamente o erro na reconstrução (pelo menos 20%⁶, podendo chegar a 40%, no caso $N - 1 \times N$). O caso que apresentou maior redução percentual foi o 2×3 , o mais difícil.

Já a Figura 4.10 apresenta a medida logerr do método de reconstrução original (com $\hat{s}_{act}(n)$ fixo) e com o *melhor caso*, num caso onde foi utilizada outra matriz de misturas. Novamente, vemos uma redução significativa caso empreguemos o *melhor caso*, em geral da ordem de 1,7 dB de logerr.

A partir do momento em que avaliamos a magnitude da melhora obtida

⁶Sempre que utilizamos medidas de redução em porcentagem, não aplicamos o logaritmo na medida correspondente.

Tabela 4.10: Melhora via *melhor caso* no erro quadrático médio de reconstrução para o caso $N - 1 \times N$.

N	Redução no erro de reconstrução (%)
3	40,37
4	34,62
5	32,65
6	33,78
7	33,56
8	32,45
9	30,41
10	31,37
11	23,05
12	34,67
13	34,87
14	34,32

Tabela 4.11: Distribuição de $\hat{s}_{act}(n)$ para diferentes números de fontes.

N	0	1	2	3	4	5	6
3	0	77282	76574	-	-	-	-
4	0	9371	70331	74154	-	-	-
5	0	5243	29206	57147	62260	-	-
6	0	0	5341	36848	55662	56005	-
7	0	0	783	12229	36260	51266	53318

com o *melhor caso*, cumpre investigar a distribuição de $\hat{s}_{act}(n)$ nos casos analisados. Esta distribuição é importante, pois orienta possíveis algoritmos que recorram à mudança do número de fontes ativas a cada instante. A distribuição é mostrada na Tabela 4.11 e nas Figuras 4.11 e 4.12. O primeiro fato digno de nota é o de que nunca, em nossas simulações, encontramos uma solução ótima que consistia em zerar todas as estimativas. Outra tendência a ressaltar é a de que, na maior parte das vezes, $\hat{s}_{act}(n)$ deve estar próximo a M (número de misturas), embora na maioria das vezes seja interessante não fazer $\hat{s}_{act}(n) = M$. Em geral, o valor $\hat{s}_{act}(n) = M$ foi o mais escolhido. Porém, interessante, quando é grande o número de fontes ($N = 12, 13$ ou 14), a escolha mais numerosa foi $M - 1$ (possivelmente, com o aumento do número de fontes, outras escolhas possam ser mais vantajosas).

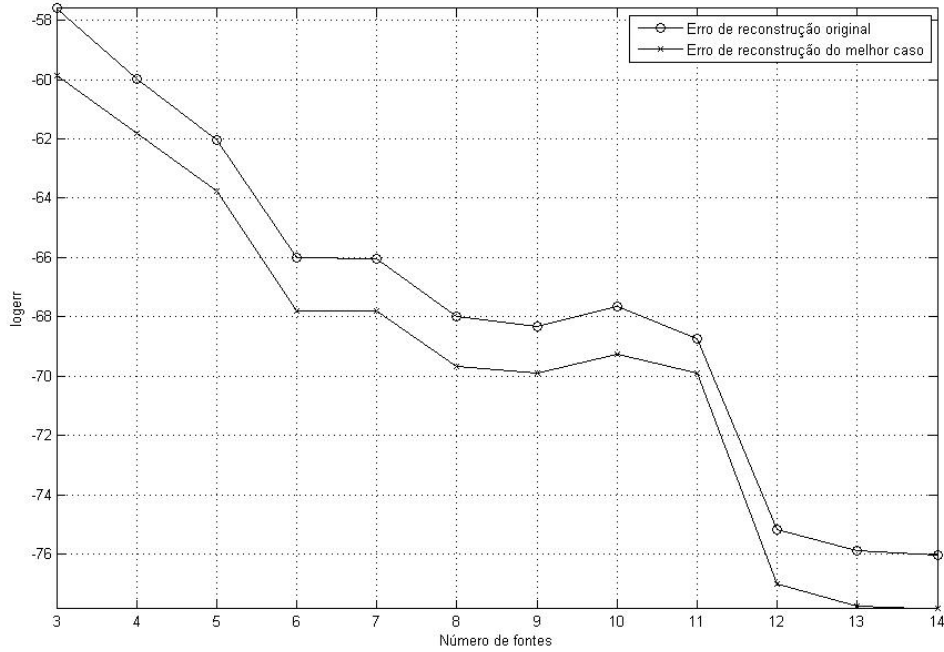


Figura 4.10: Comparação dos erros de reconstrução dos métodos caminho mais curto e *melhor caso*.

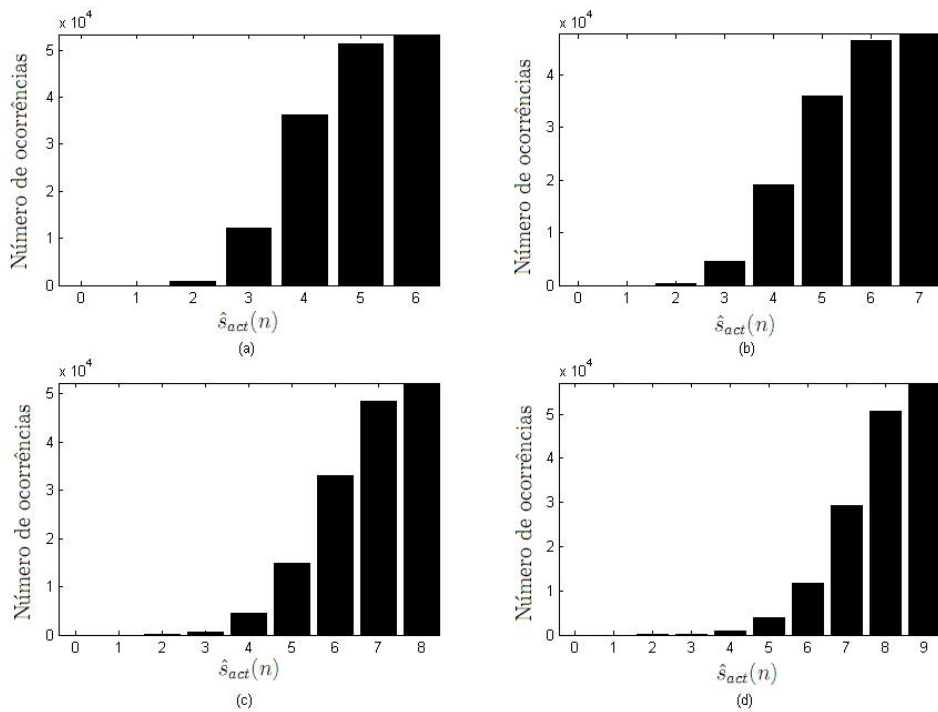


Figura 4.11: Distribuição de $\hat{s}_{act}(n)$ via *melhor caso* com diferentes números de fontes. (a) $N = 7$, (b) $N = 8$, (c) $N = 9$ e (d) $N = 10$.

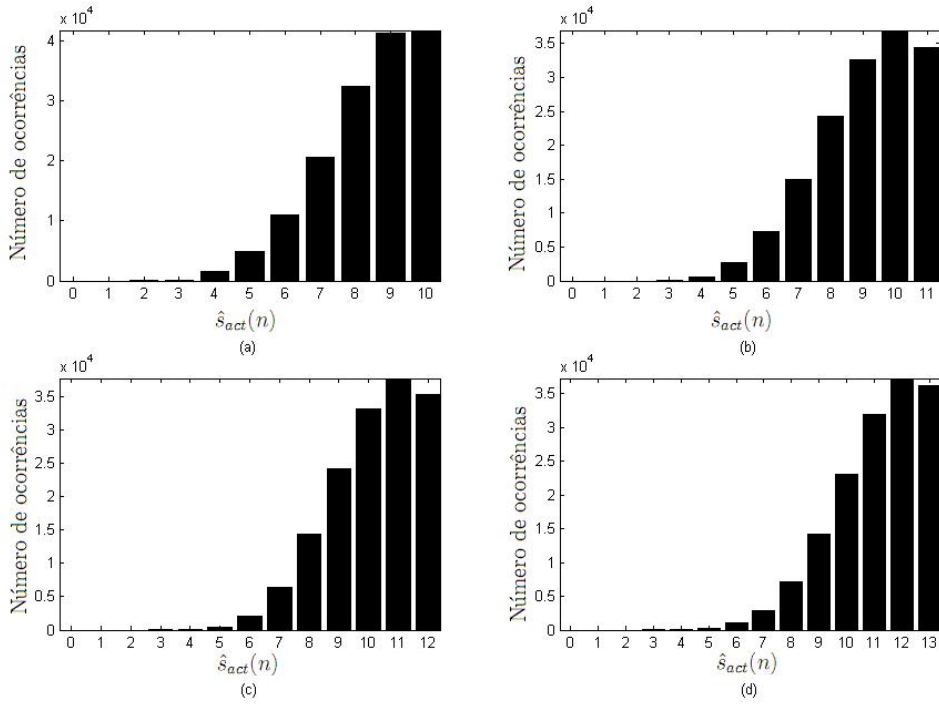


Figura 4.12: Distribuição de $\hat{s}_{act}(n)$ via *melhor caso* com diferentes números de fontes. (a) $N = 11$, (b) $N = 12$, (c) $N = 13$ e (d) $N = 14$.

4.3.4.4 Minimização da norma l_p ($p \neq 1$)

Para avaliar a influência do parâmetro p na separação de fontes, utilizamos três sinais de voz e geramos sinteticamente duas misturas instantânea. O parâmetro p foi variado de 0,7 a 1,3. A cada valor de p foi efetuado o cálculo da medida logerr correspondente. Este procedimento gerou os dados apresentados na Figura 4.13.

Nesta figura, observamos que o valor $p = 1$ não é o ótimo. Verifica-se também que valores de p superiores a 1 não resultam em benefícios explícitos para a separação. Em geral, há piora de separação nesta configuração; em alguns (poucos) valores, houve uma melhora muito pequena. Logo, não é interessante utilizar a condição $p > 1$.

Já quando empregamos $p < 1$, existem claramente valores de p que incrementam o desempenho do processo de estimativa das fontes. A Figura 4.13 mostra que a melhor situação reside em $p \approx 0,9$, onde a redução do erro de reconstrução foi de aproximadamente 1,5%, uma melhora não muito significativa. Em [86], verifica-se que este valor gera uma razão sinal-artefato ótima, embora, neste artigo, este valor não seja o melhor em termos de razão sinal-distorção e razão sinal-interferência, os

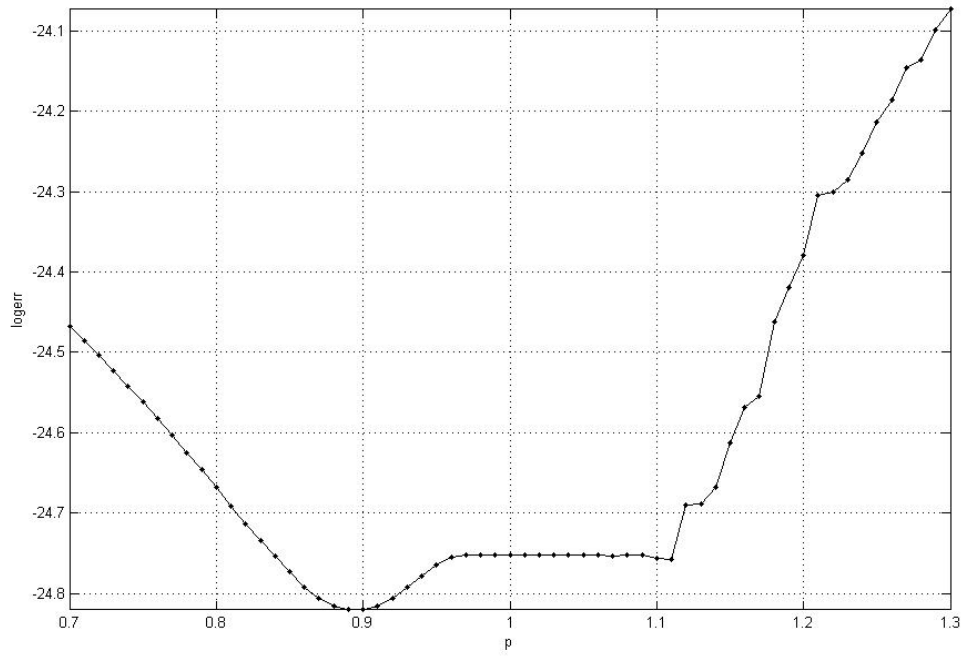


Figura 4.13: Erro de reconstrução em função da minimização da norma p .

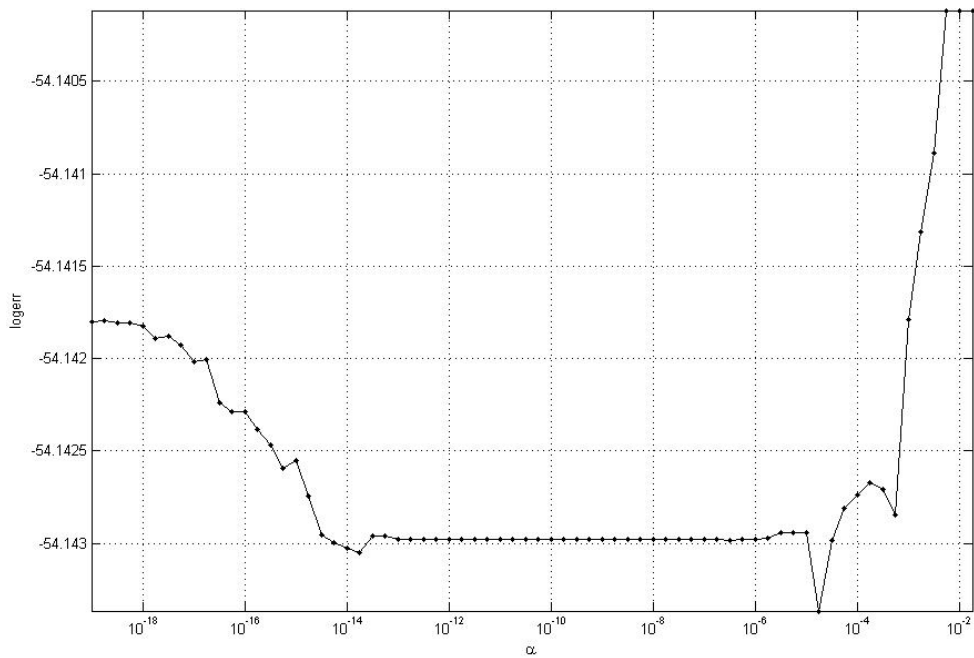


Figura 4.14: Erro de reconstrução em função do parâmetro α .

quais apresentam melhores resultados com valores de p ainda menores.

4.3.4.5 Adoção de um Critério de Continuidade

Utilizando novamente três sinais de voz e gerando duas misturas, variamos o parâmetro α e obtemos o logerr correspondente. Este procedimento gerou os dados apresentados na Figura 4.14. Valores de α não-negativos e maiores que o limite superior do intervalo considerado apresentam resultados praticamente constantes. O mesmo ocorre com valores α inferiores ao limite inferior do intervalo apresentado, com a diferença de o logerr da reconstrução ser menor.

Vemos que a introdução do critério de continuidade traz benefícios às estimativas das fontes. Porém, a melhora é insignificante, sendo da ordem de 0,036%. Em outras palavras, a inserção do critério de continuidade adotado não é interessante, à medida em que não se traduz em ganhos razoáveis de desempenho. Isto significa que a informação oriunda da diversidade espacial das fontes é muito mais confiável do que o critério de continuidade adotado.

4.4 Considerações Finais

Neste capítulo, abordamos o problema da separação de fontes em misturas indeterminadas lineares e instantâneas, no caso de existirem pelo menos duas misturas. Separar as fontes nesta configuração, como vimos, é um problema que costuma ser dividido em quatro passos.

O primeiro passo consiste em aplicar uma transformada linear esparsificadora às misturas. As transformadas mais utilizadas são a STFT e a *wavelet packet*, sendo esta última a adotada neste capítulo. Quando utilizamos um número de níveis de decomposição próximo de 7 (o mais comum na literatura), verificamos uma relativa independência da qualidade das estimativas das fontes em relação à *wavelet* escolhida.

O segundo passo se dedica a estimar a matriz de misturas. Neste capítulo, foram detalhadas as principais técnicas enfocadas neste problema: a varredura de

Zibulevsky, o modelo de misturas de exponenciais, os métodos de *clusterização* K-means, *fuzzy-C* e hierárquica, o método DUET, a extensão dos métodos DUET e TIFROM. Foi proposta uma nova técnica, a qual utiliza um modelo de misturas de gaussianas. Em nossos testes, o método proposto se revelou bastante competitivo em relação aos concorrentes.

O terceiro passo consiste em estimar a representação de cada uma das fontes no domínio da transformada. Foi avaliada a possibilidade de se minimizar normas das estimativas diferentes de l_p , com $p = 1$. Verificamos que valores de $p < 1$ realmente são os mais interessantes e que valores de $p > 1$ não resultam em melhoras significativas das estimativas das fontes. Adotamos um critério de continuidade já utilizado no caso de apenas uma única mistura, concluindo que este critério não é particularmente benéfico para a separação. Também verificamos que um número variável de fontes consideradas ativas a cada instante pode provocar ganhos significativos na separação, e que este número variável freqüentemente aproxima-se do número M de misturas (num contexto de otimalidade).

O quarto passo é a reconstrução das fontes. Primeiramente, verificamos que a redução do fator de decimação do banco de filtros utilizado incrementa significativamente as medidas de SAR e SDR das estimativas das fontes, o que indica a existência de efeitos danosos de *aliasing* devidos à decimação crítica, comumente empregada nestas técnicas de separação. Por fim, avaliamos a dificuldade de reconstrução quando variamos o número de fontes e de misturas, verificando que a reconstrução das fontes no caso indeterminado ainda está muito longe do caso determinado.

Capítulo 5

Separação de Fontes no Caso de Uma Única Mistura

Comumente, em problemas de separação de fontes, afirma-se que a relação entre o número M de misturas e o número N de fontes nos permite distinguir dois casos, os quais recorrem a técnicas bastante distintas. Estes casos seriam $M \geq N$ (determinado se $M = N$ ou sobredeterminado quando $M > N$) e o caso indeterminado (mais difícil, ocorrendo quando $M < N$). Porém, seria mais realista destacar-se da classificação “indeterminado” o caso extremo $M = 1$, pois mesmo as técnicas que tratam do caso indeterminado em geral não se aplicam à possibilidade - muito comum na prática - de termos acesso a apenas uma mistura. Isto se deve, em última análise, ao fato de que estas técnicas exploram a diversidade espacial das fontes entre os sensores, a qual inexistente quando há apenas uma mistura. As técnicas atualmente dedicadas à condição $M = 1$ são as menos bem-sucedidas dentre todas, devido à particular dificuldade desta configuração. Esta condição extrema é o tema deste capítulo.

Há duas estratégias muito comuns, relacionadas à decomposição das misturas, utilizadas por estas técnicas. As duas estratégias são:

- (I) a partir de uma decomposição fixa, estimar a contribuição de cada uma das fontes em cada um dos componentes;
- (II) encontrar uma decomposição cujos componentes sejam associados basicamente a apenas uma das fontes.

Na primeira estratégia, a decomposição é fixa, enquanto que, na segunda, é adaptativa. A estratégia (I) utiliza modelos das fontes (por exemplo, um modelo estatístico paramétrico devidamente treinado com trechos das fontes isoladas ou mesmo um recurso à estrutura harmônica dos sinais) para estimar a contribuição de cada fonte num dado componente [93]-[95]. Como é necessário um conhecimento das características das fontes, não podemos denominar as técnicas que utilizam a estratégia (I) de “cegas” (por esta razão as denominaremos, nesta dissertação, de “técnicas de separação supervisionada”).

Já a estratégia (II) não necessita de gravações isoladas das fontes, tampouco de um conhecimento *a priori* acerca de estruturas harmônicas, à medida em que pode recorrer a uma clusterização não supervisionada dos componentes de forma a associá-los, de forma cega (na realidade, com pressupostos bem gerais, como o da continuidade), a uma das fontes [90]. Porém, cabe ressaltar que os métodos para a clusterização supracitada [96],[97] costumam deteriorar bastante os resultados (vide [90]). Um outro exemplo promissor desta estratégia utiliza conhecimentos de psicoacústica e redes neurais de terceira geração (onde os neurônios são chamados de *spiking neurons*) [98].

O explicado acima nos permite concluir que os métodos de separação com apenas uma mistura baseiam-se principalmente na diversidade espectral das fontes (e não na diversidade espacial, como a maioria dos métodos para misturas indeterminadas com $N > M > 1$).

Este capítulo apresenta um novo método, fundamentado na estrutura harmônica dos instrumentos musicais, que se utiliza da estratégia (I). Para efeitos de comparação, apresentaremos o método proposto em [93], aqui denominado FGMM, o qual utiliza a mesma estratégia. Enfocaremos aqui o caso com duas fontes.

Da mesma forma que o método proposto em [93], o nosso método necessita de trechos isolados das fontes, para treinamento. Neste aspecto, cumpre lembrar que há métodos de separação de fontes harmônicas que não recorrem a trechos das fontes em separado para treinamento [99],[100]. Estes métodos, não avaliados aqui, são uma alternativa promissora, em especial quando não possuímos um banco de dados disponível.

5.1 O método FGMM

Adotamos a seguir a nomenclatura de [93]. Se as fontes s_1 e s_2 encontram-se misturadas de forma aditiva, podemos obviamente escrever a mistura x como:

$$x = s_1 + s_2 + n, \quad (5.1)$$

onde n é considerado ruído branco gaussiano de variância σ^2 . Dentro de um paradigma bayesiano, podemos escrever:

$$p(s_1, s_2|x) = \frac{p(s_1, s_2, x)}{p(x)}, \quad (5.2)$$

$$\frac{p(s_1, s_2, x)}{p(x)} = \frac{p(x|s_1, s_2)p(s_1, s_2)}{p(x)}, \quad (5.3)$$

$$p(s_1, s_2|x) \propto p(x|s_1, s_2)p_1(s_1)p_2(s_2). \quad (5.4)$$

A Eq. (5.4) foi obtida admitindo-se a independência das fontes e atentando para o fato de que $p(x)$ é fixo (a mistura é um dado constante). O modelo aditivo da mistura reflete-se na verossimilhança $p(x|s_1, s_2)$ e as informações disponíveis acerca das fontes são introduzidas em $p_1(s_1)$ e $p_2(s_2)$.

No caso de ausência de informações prévias acerca das fontes, uma estimativa de máxima verossimilhança poderia ser adotada. Porém, infelizmente, no contexto indeterminado a estimativa de máxima verossimilhança é inviável porque apresenta soluções múltiplas [93].

Logo, é necessário modelar as fontes no paradigma bayesiano adotado. O método FGMM admite que a distribuição das fontes (na realidade, como veremos mais adiante, do módulo das fontes no domínio da transformada de Fourier) é uma mistura de gaussianas de média zero (GMMs, detalhadas no Capítulo 4). Admitindo GMMs unidimensionais, obtemos as seguintes distribuições das fontes:

$$p(s_1) = \sum_{i=1}^{K_1} p_i^{(1)} \frac{1}{\sqrt{2\pi}\sigma_i^{(1)}} e^{-\frac{1}{2}\left(\frac{s_1}{\sigma_i^{(1)}}\right)^2}, \quad (5.5)$$

$$p(s_2) = \sum_{i=1}^{K_2} p_i^{(2)} \frac{1}{\sqrt{2\pi}\sigma_i^{(2)}} e^{-\frac{1}{2}\left(\frac{s_2}{\sigma_i^{(2)}}\right)^2}, \quad (5.6)$$

onde os parâmetros $(p_i^{(1)}, p_i^{(2)}, \sigma_i^{(1)}$ e $\sigma_i^{(2)})$ devem ser estimados a partir dos dados da mistura x . Após a estimativa destes parâmetros (efetuada comumente mediante o

algoritmo EM), cumpre estimar as fontes. Para tal devemos adotar uma medida do custo da substituição dos parâmetros reais por uma estimativa dos mesmos, medida que deve, num momento posterior, ser minimizada.

As duas medidas de custo mais comuns são a do erro quadrático e a da distribuição de Dirac (a qual apresenta valor nulo, exceto quando as estimativas coincidem com os parâmetros reais). É possível provar que a primeira medida implica uma estimativa que equivale à média posterior condicional, enquanto que a segunda implica a clássica estimativa MAP (*maximum a posteriori*).

Em termos matemáticos, supondo parâmetro θ , estimativa α e função custo $C(\alpha, \theta)$, a estimativa α_{opt} é calculada através de uma minimização do custo médio sobre todos os valores possíveis de θ :

$$\alpha_{opt} = \arg \min_{\alpha} \int_{\theta} C(\alpha, \theta) f(x|\theta) \pi(\theta) d\theta, \quad (5.7)$$

onde $\pi(\theta)$ representa o conhecimento que porventura tenhamos acerca do parâmetro θ , antes de observamos a mistura x . Admitindo $C(\alpha, \theta) = |\alpha - \theta|^2$, podemos deduzir [57] que $\alpha_{opt} = E(\theta|x)$ (estimativa de média posterior condicional - em inglês, abreviado por PM).

Quando $K_1 = K_2 = 1$, as GMMs associadas a cada fonte se degeneram em meras distribuições gaussianas; nesta condição, podemos facilmente demonstrar que tanto a estimativa MAP quanto a PM podem ser obtidas por:

$$\hat{s}_1 = \frac{(\sigma_1^{(1)})^2 x}{(\sigma_1^{(1)})^2 + (\sigma_1^{(2)})^2 + \sigma^2}, \quad (5.8)$$

$$\hat{s}_2 = \frac{(\sigma_1^{(2)})^2 x}{(\sigma_1^{(1)})^2 + (\sigma_1^{(2)})^2 + \sigma^2}. \quad (5.9)$$

Cumpre notar que o resultado acima pode ser compreendido como uma generalização bayesiana do filtro de Wiener, já que a condição $\sigma = 0$ resultaria neste filtro.

No caso de distribuições das fontes sendo GMMs não-degeneradas, a dedução não é direta, sendo fundamentada numa interpretação de uma mistura de gaussianas por meio de um modelo gerador, o qual admite que uma observação oriunda de uma GMM é obtida através de duas fases [68]:

- (1) seleção de uma das K gaussianas através de uma distribuição discreta que pode ser expressa por $\sum_{k=1}^K p_k \delta(q - k)$; q é associado ao componente (ou estado) ativo;
- (2) geração de uma amostra seguindo a distribuição gaussiana¹ com variância σ_q^2 .

Os estados ativos das fontes 1 e 2 serão denominados, respectivamente, de q_1 e q_2 . Estes estados são normalmente desconhecidos, devendo ser estimados. A probabilidade posterior de que $q_1 = i$ e $q_2 = j$ é dada por:

$$\begin{aligned} \gamma_{i,j}(x) &= p(i, j|x) \propto p(x|i, j)p(i)p(j) \\ &\propto p_i^{(1)} p_j^{(2)} g\left(x, \left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2\right), \end{aligned} \quad (5.10)$$

onde $g(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ (no Capítulo 4, a média da gaussiana era levada em consideração; aqui, supomos média nula). Admitindo os estados ativos $q_1 = i$ e $q_2 = j$, o estimador condicional de Wiener (o qual, num paradigma bayesiano, equivale tanto à estimativa MAP quanto à PM quando $\sigma = 0$) é dado por:

$$E(s_1|i, j) = \frac{\left(\sigma_i^{(1)}\right)^2 x}{\left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2}, \quad (5.11)$$

$$E(s_2|i, j) = \frac{\left(\sigma_j^{(2)}\right)^2 x}{\left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2}, \quad (5.12)$$

Caso modelemos as fontes com distribuições GMM não-degeneradas (com $K_1 > 1$ e $K_2 > 1$), as estimativas MAP e PM passam a ser distintas. A estimativa MAP estima o estado da GMM de cada fonte através do maior valor de $\gamma_{i,j}$, e então recorre às Eqs. (5.11) e (5.12) para estimar as fontes (numa espécie de filtragem de Wiener adaptativa [93]); já as estimativas PM são obtidas [101] por meio da ponderação a seguir mostrada:

$$E(s_1|x) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \gamma_{i,j}(x) \frac{\sigma_i^{(1)}}{\left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2} x, \quad (5.13)$$

¹Neste capítulo, supomos que as distribuições gaussianas possuem média zero; logo, elas são determinadas unicamente por sua variância.

$$E(s_2|x) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \gamma_{i,j}(x) \frac{\sigma_j^{(2)}}{\left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2} \cdot x, \quad (5.14)$$

onde $\gamma_{i,j} \propto p_i^{(1)} p_j^{(2)} g\left(x, \left(\sigma_i^{(1)}\right)^2 + \left(\sigma_j^{(2)}\right)^2 + \sigma^2\right)$.

Por fim, cabe ressaltar que todas as estimativas acima mostradas não se aplicam às amostras de sinais de áudio no domínio do tempo. Elas são aplicadas à cada raia (*bin*) da transformada discreta de Fourier (DFT, de *Discrete Fourier Transform*) dos sinais (na realidade, a transformada de Fourier de tempo curto, já que a DFT é aplicada em quadros, estes suavizados mediante a janela de Hanning), o que acaba por gerar observações unidimensionais. À cada raia são associadas duas GMMs: uma para a fonte 1 e outra para a fonte 2. Na etapa de treinamento e na Eq. (5.10), utiliza-se o módulo da DFT, o que implica uma modelagem que não atenta para a fase do sinal (no domínio de Fourier).

Para o treinamento dos parâmetros das GMMs, a partir de valores iniciais arbitrários, recorreremos ao clássico método iterativo EM [68],[80], já detalhado no Capítulo 4.

5.2 Método Proposto: FNH

Detalhado na seção anterior, o método FGMM almeja modelar (via GMMs) a distribuição dos módulos de cada raia da DFT dos sinais. Cada raia é modelada de forma independente. No entanto, é conhecida a alta dependência entre as raias quando um instrumento emite uma nota musical. Esta constatação inspirou nossa proposta (denominada, daqui em diante, FNH ²), a qual apresenta duas diferenças em relação ao método FGMM, a saber:

- (1) contempla a dependência entre harmônicos;
- (2) utiliza uma outra técnica de reconhecimento de padrões: redes neurais de múltiplas camadas.

O recurso a outras técnicas de reconhecimento é importante para evitar um grave problema existente na modelagem do método FGMM: o de um modelo de média

²No termo FNH, as letras F, N e H referem-se à, respectivamente, frequência, neural e harmônico.

zero de sinais não-negativos (módulos da DFT), o que é um tanto irreal.

5.2.1 Análise DFT para sinais harmônicos

A primeira dificuldade de recorrer à estrutura harmônica dos sinais reside na DFT. As frequências por ela analisadas estão uniformemente distribuídas entre 0 e a metade da frequência de amostragem. Isto provoca uma incompatibilidade com uma análise das frequências associadas aos harmônicos, dado que estes se distribuem uniformemente numa escala logarítmica. Para contornar este problema, é possível utilizar outras transformadas (como *wavelets*).

Neste ponto, optamos por uma estratégia já utilizada em modelagem senoidal: a partir dos módulos das três raias (da DFT) mais próximas à frequência desejada, interpolamos uma parábola. Nossa estimativa do módulo da DFT na frequência desejada será o valor apresentado pela parábola nesta frequência. Matematicamente, seja f a frequência a analisar; se as raias mais próximas a f apresentam frequências f_1 , f_2 e f_3 e módulos dados por, respectivamente, X_1 , X_2 e X_3 , o módulo da DFT em f é estimado por $af^2 + bf + c$, onde a , b e c são calculados por:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} f_1^2 & f_1 & 1 \\ f_2^2 & f_2 & 1 \\ f_3^2 & f_3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad (5.15)$$

5.2.2 Formação dos padrões

Para a tarefa de reconhecimento de padrões, são formados 12 diferentes conjuntos de padrões, cada qual associado a uma nota e aos seus respectivos harmônicos. O número exato de harmônicos depende da nota e da frequência de amostragem. Por exemplo, a nota A0 possui frequência de 27,5 Hz. Numa amostragem a 11 kHz (onde a frequência do sinal deve ser limitada a 5,5 kHz), ela apresenta harmônicos em 55, 110, 220, 440, ..., 3520 Hz; o padrão associado a esta nota possui 8 componentes (A0 mais seus sete harmônicos). Os padrões são formados apenas pelos módulos da DFT nas frequências de interesse. Os módulos são obtidos mediante a interpolação parabólica supracitada.

A idéia de se utilizar o módulo da DFT apenas de uma nota (e de seus

harmônicos) para cada rede neural se inspira no fato de que as relações entre o módulo da DFT na frequência da nota e de seus harmônicos costumam ser muito específicas para cada instrumento, possibilitando a distinção destes numa etapa de separação posterior.

5.2.3 Análise com Redes Neurais

Para treinamento, após a formação dos padrões, podemos utilizar redes neurais para classificar os padrões associados à fonte 1 ou à fonte 2. Para este propósito, recorreremos a redes neurais *feedforward* de duas camadas (a escondida e a de saída) com apenas um neurônio de saída, treinadas por meio do algoritmo rápido *resilient backpropagation* [102]. A função de ativação escolhida foi a tangente hiperbólica. No treinamento, associamos à saída do neurônio (da camada de saída) o valor 1 para a fonte 1 e o valor -1 para a outra fonte. Cabe ressaltar que há 12 redes neurais, uma para cada nota.

Na fase de teste, após o janelamento do sinal e posterior transformação para o domínio de Fourier, calculamos suas componentes nas frequências desejadas (novamente, por meio da interpolação parabólica). Então, para cada um dos 12 padrões, obtemos a resposta da rede neural, a qual varia entre +1 e -1.

Seja y_o a resposta de uma das rede neurais (obtida a partir do neurônio de saída). Este valor é associado a um nível de *mascaramento* M_k , dado pela fórmula $M_k = (y_o + 1)/2$. O valor de mascaramento M_k é associado a todas as frequências relacionadas à rede neural (a nota e seus harmônicos).

5.2.4 Mascaramento

As redes neurais geram, na fase de teste, um valor de mascaramento para as frequências associadas a cada uma das notas. Então, deparamo-nos novamente com o problema de compatibilizar estas frequências (uniformemente distribuídas por nota numa escala logarítmica) com as obtidas mediante uma DFT. Uma alternativa seria utilizar uma interpolação parabólica, como visto anteriormente. Porém, escolhemos utilizar uma interpolação *spline* cúbica [103] (notamos que os resultados são praticamente os mesmos que os oriundos de uma interpolação parabólica). Todos

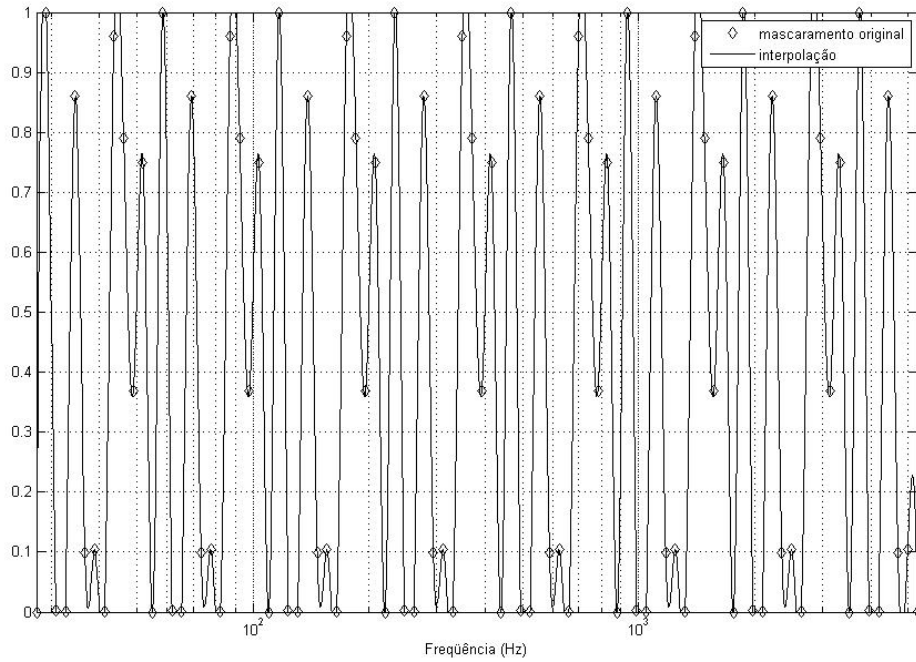


Figura 5.1: Exemplo de mascaramento via redes neurais e sua interpolação.

os valores negativos obtidos por esta interpolação são considerados nulos; já os superiores a 1 são igualados à unidade. De todo modo, estas retificações são pouco freqüentes.

A Figura 5.1 apresenta um exemplo de mascaramento oriundo de redes neurais e sua respectiva interpolação (o eixo das abcissas está na escala logarítmica, mais adequada para a visualização).

A estimativa da fonte 1 é obtida mediante o mero produto (no domínio de Fourier) do sinal de teste pela interpolação do mascaramento ($M_k^{int}(f)$, visto na Figura 5.1). Já a estimativa da fonte 2 é obtida pelo produto da DFT por $(1 - M_k^{int}(f))$. Dessa forma, o mascaramento é uma medida do grau em que determinada amostra (no espaço tempo \times freqüência) pertence à fonte 1 (e conseqüentemente o grau em que esta não pertence à fonte 2). Uma outra interpretação conveniente seria a da probabilidade de classe *a posteriori* [55] (no nosso caso, as “classes” seriam as fontes 1 e 2). Uma forma de mascaramento alternativa (e menos suave) seria a do mascaramento binário, a qual associa cada ponto (no espaço tempo \times freqüência) a apenas uma das fontes (o método DUET [61], detalhado no Capítulo 4).

5.3 Simulações

Auditivamente, os resultados dos métodos FGMM e FNH são muito ruins quando as fontes não estão sincronizadas. Cremos que, em [93], as fontes utilizadas faziam parte de um trecho coerente de música por esta razão. Eis uma limitação de ambos os métodos, a qual comumente não é muito grave. Em [93], os testes foram efetuados com dois instrumentos: bateria e piano. Nas simulações por nós efetuadas, optamos também por estes dois instrumentos, amostrados em 11kHz, devidamente janelados (janela de Hanning) e com sobreposição de 75%. Eis uma configuração interessante, a partir do momento em que temos um instrumento (piano) que obedece ao “paradigma harmônico” (objeto de atenção especial do método proposto FNH) e outro que não apresenta esta condição.

Duas medidas de qualidade de separação foram calculadas: SIR e SAR. Como vimos em capítulos anteriores, o SIR mede o resíduo da outra fonte na estimativa de cada fonte; já o SAR reflete a quantidade de distorção em cada sinal estimado. O cálculo dessas medidas segue o formulado em [27]. Estes resultados foram apresentados em [104].

Foi utilizado um trecho de cada fonte isolada com 1 minuto de duração para treinamento e validação. As redes neurais dedicaram 20% destes dados para validação. Foram inicializadas 10 diferentes redes neurais para cada nota, sendo escolhida para a fase de teste a que apresentar menor erro quadrático médio. Este procedimento, muito comum, reduz a probabilidade de uma rede neural se encontrar num mínimo local inadequado (da superfície de erro).

A Tabela 5.1 exhibe os resultados obtidos. O valor de P^3 (parâmetro) indica o número de gaussianas (para o método FGMM) ou o número de neurônios na camada escondida (para o método FNH) e N é o comprimento da janela de Hanning. Os subscritos “ p ” e “ b ” indicam piano e bateria, respectivamente.

Os resultados apresentados indicam que o SIR para a estimativa do piano obteve uma significativa melhora com o método proposto; já a da bateria sofre uma degradação muito grande, devido ao fato de esta não apresentar características harmônicas típicas. Interessante notar que o método proposto oferece um SAR su-

³Cumprer notar que neste capítulo a definição de P difere da utilizada no Capítulo 3, onde era o número de fontes.

Tabela 5.1: Medidas SIR e SAR obtidas pelos métodos FGMM e FNH.

Método	P	N	SIR _b	SIR _p	SAR _b	SAR _p
FGMM	4	512	13,42	3,56	-13,86	-15,06
FGMM	8	512	15,05	1,07	-14,12	-17,33
FGMM	16	512	17,56	0,23	-14,72	-18,27
FGMM	4	1024	27,73	-3,76	-13,54	-16,66
FGMM	8	1024	38,48	-7,02	-14,24	-17,72
FNH	3	512	-7,14	10,03	2,75	1,76
FNH	3	1024	-1,12	14,97	0,50	5,79
FNH	4	1024	-3,46	12,51	0,71	5,22
FNH	6	1024	-5,10	10,65	-0,09	4,86

perior, provavelmente devido ao fato de suas alterações no domínio da frequência tenderem a ser suaves (o que não ocorre com o método FGMM, pois este modela cada raia da DFT de forma independente das outras).

Outro padrão a ressaltar é a melhora em SIR_b no método FGMM com o aumento do número de gaussianas, acompanhado de uma degradação em SIR_p. No método FNH, notamos que um número de neurônios na camada escondida superior a 3 degrada os resultados (exceto SAR_b quando o número de neurônios é igual a 4; mas a diferença é muito pequena). Um tamanho de janela de 1024 se mostrou benéfico (em relação a um comprimento de 512) para o método proposto, enquanto que para o FGMM só o foi para a bateria. As Figuras 5.2-5.7 mostram as fontes originais (amostras no domínio do tempo, num trecho de duração de aproximadamente 19s), bem como suas estimativas mediante várias configurações dos métodos FGMM e a melhor estimativa do método proposto FNH. Embora a aparência da estimativa via FNH pareça muito superior, auditivamente a estimativa da bateria pelo método FGMM é significativamente melhor (o que é coerente com os valores de SIR_b apresentados). O método FGMM parece apresentar muito ruído (o que degrada o seu SAR). Provavelmente, tal fato se deve às discontinuidades no tratamento das raias (frequências), já que cada raia é modelada de forma completamente independente das outras.

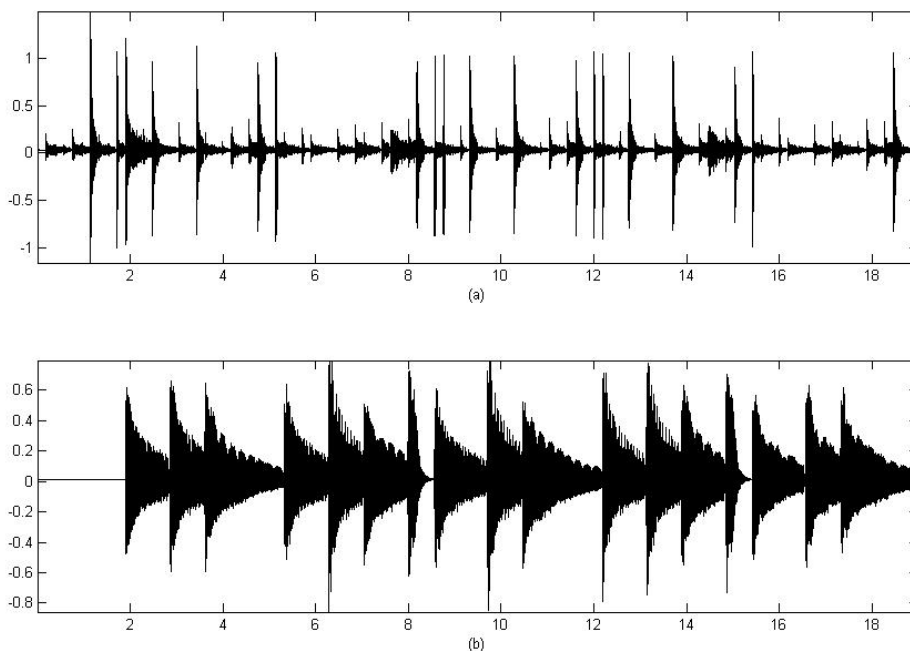


Figura 5.2: Fontes originais. (a) bateria, (b) piano.

A Figura 5.8 apresenta a evolução do mascaramento no espaço tempo \times frequência em um pequeno trecho do sinal; observamos que a superfície de mascaramento é relativamente suave. Foram suprimidas as frequências inferiores a 1000 Hz, para tornar o gráfico melhor visualizável (há uma grande densidade de picos nestas frequências, já que nesta figura o gráfico se apresenta numa escala linear).

5.4 Considerações Finais

Neste capítulo, apresentamos um método alternativo de separação de fontes mediante uma única mistura disponível. Este método adota o mesmo paradigma do método apresentado em [93]. Duas características que distinguem o método proposto são o seu recurso à estrutura harmônica de sinais oriundos de instrumentos musicais e o uso de redes neurais *feedforward* para reconhecimento de padrões.

A modelagem da estrutura harmônica nos permitiu obter melhoras significativas na separação, apresentando distorção menor e maior razão sinal-interferência (para instrumentos harmônicos). A especificidade da nossa proposta implica uma degradação de desempenho quando o instrumento a separar não possui uma frequência

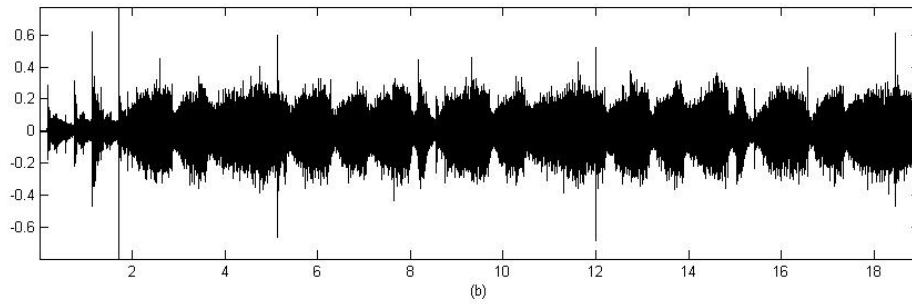
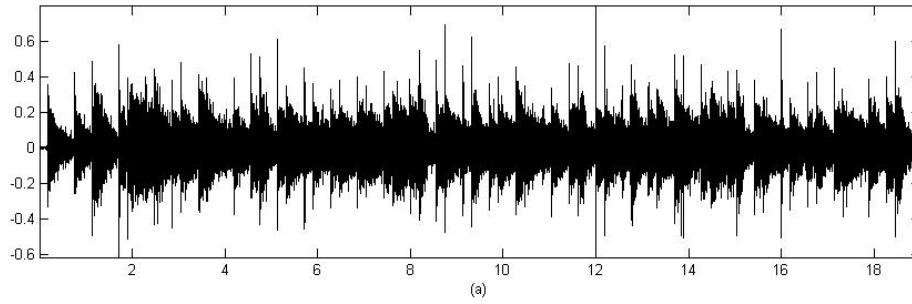


Figura 5.3: Fontes estimadas pelo método FGMM, com 4 gaussianas e janela de comprimento 512. (a) bateria, (b) piano.

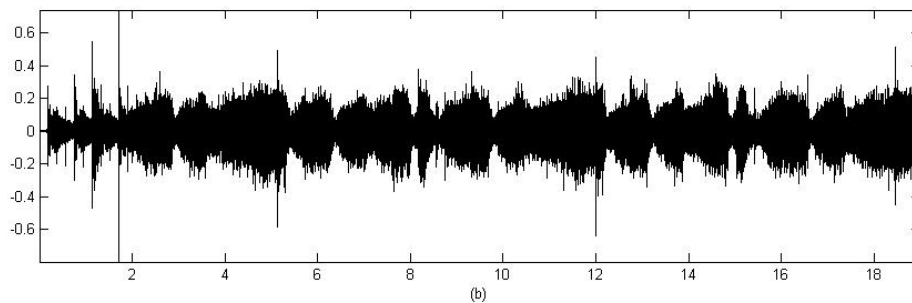
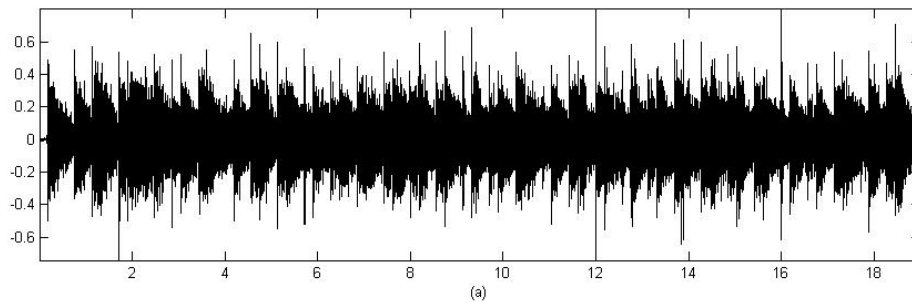


Figura 5.4: Fontes estimadas pelo método FGMM, com 16 gaussianas e janela de comprimento 512. (a) bateria, (b) piano.

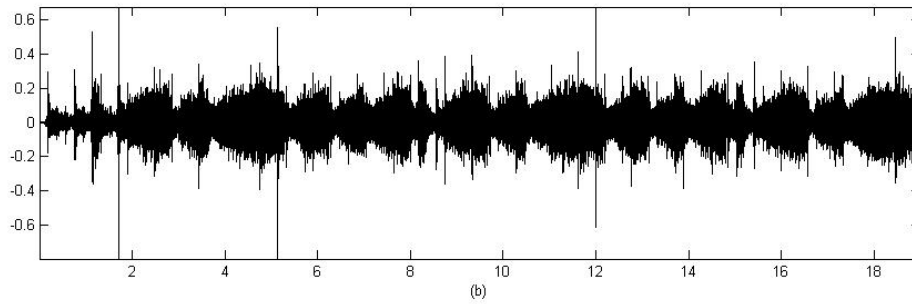
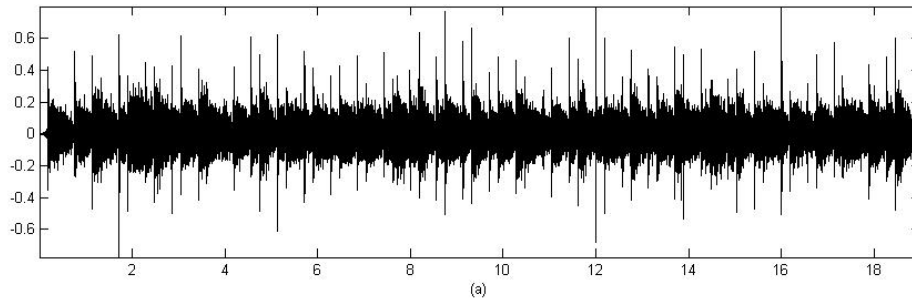


Figura 5.5: Fontes estimadas pelo método FGMM, com 4 gaussianas e janela de comprimento 1024. (a) bateria, (b) piano.

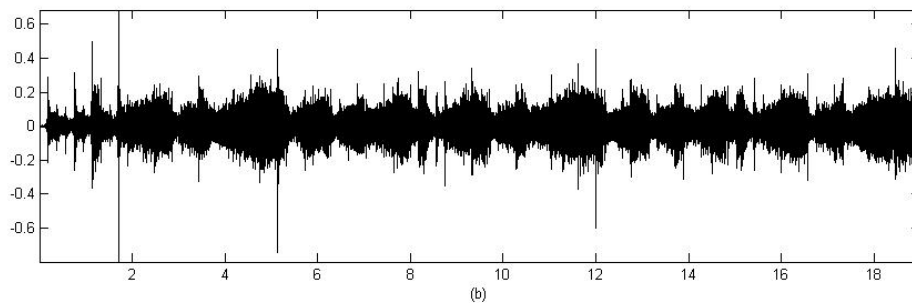
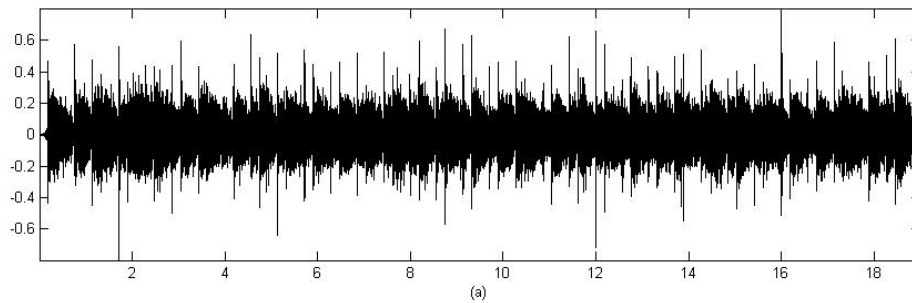


Figura 5.6: Fontes estimadas pelo método FGMM, com 8 gaussianas e janela de comprimento 1024. (a) bateria, (b) piano.

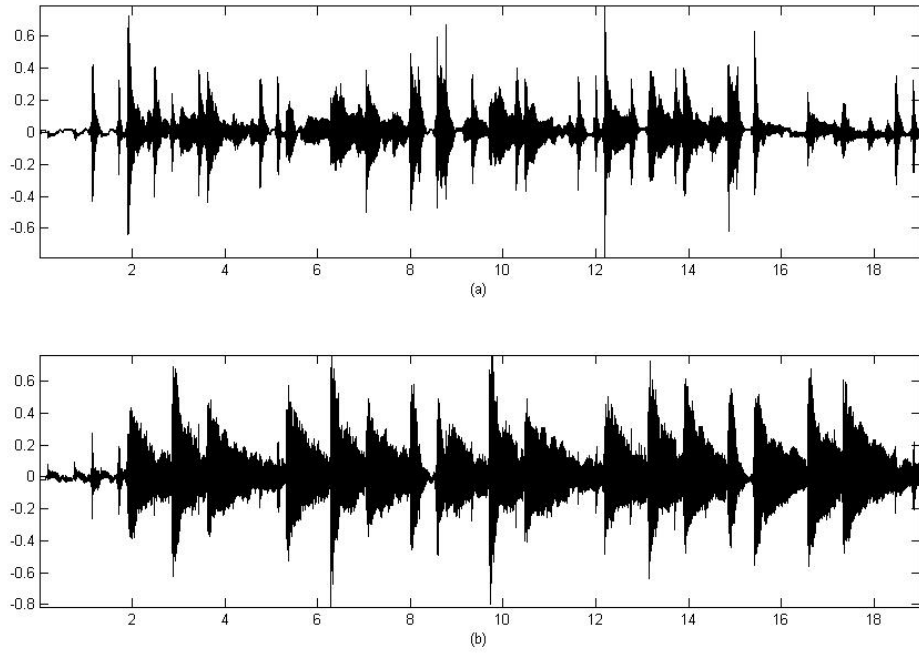


Figura 5.7: Fontes estimadas pelo método FNH, com 3 neurônios na camada escondida e janela de comprimento 1024. (a) bateria, (b) piano.

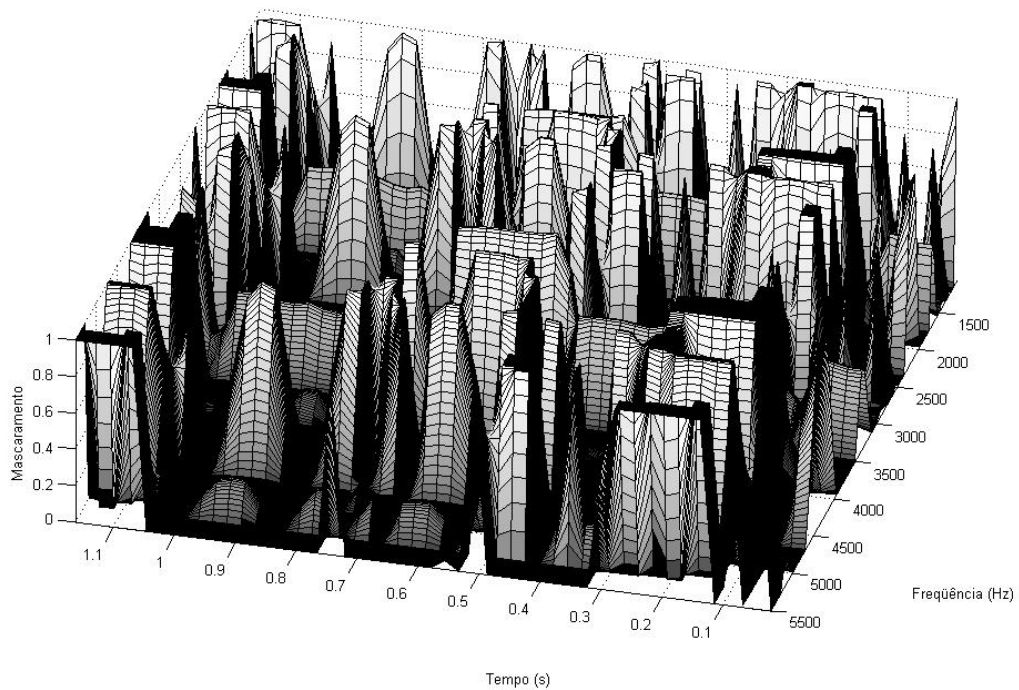


Figura 5.8: Padrão do mascaramento no espaço tempo \times frequência em um pequeno trecho de sinal.

fundamental bem definida (como é o caso da bateria). Este fato nos leva a concluir que uma estratégia interessante seria empregar o método proposto ou o método de [93] de forma dependente do instrumento. Uma alternativa para o mascaramento suave proposto seria utilizar um mascaramento binário ($M_k(f)$ podendo ser 0 ou 1 [61]).

Capítulo 6

Conclusão

Esta dissertação abordou o importante problema da separação de fontes em diferentes configurações. Todas as configurações de misturas contempladas pertencem ao caso linear, o qual se desmembra em dois: o caso instantâneo e o convolutivo.

No segundo capítulo, o caso instantâneo e determinado (número de fontes igual ao número de misturas) foi contemplado. O algoritmo InfoMax foi derivado a partir de uma abordagem bayesiana, a qual permitiu a explicitação de alguns de seus pressupostos. Um deles foi a distribuição uniforme dos coeficientes da matriz de misturas. Esta explicitação foi importante, à medida em que ensejou uma modificação no algoritmo, a qual supõe um conhecimento acerca dos coeficientes supracitados, caso tenhamos estes coeficientes amostrados segundo uma distribuição exponencial. Uma modificação do método Infomax, a qual leva em consideração informações acerca da matriz de misturas, apresentou vantagens na estimativa das fontes, particularmente para o caso onde o número de amostras é reduzido. Foi então testada a utilidade de um conhecimento estatístico inexato acerca destes coeficientes. Esta hipótese foi testada devido ao fato de, na prática, comumente o conhecimento estatístico que possuímos não ser exato. É comum, por exemplo, supor-se que uma distribuição a se estimar pertença ao grupo das distribuições paramétricas, o que muitas vezes não é verdade. Este teste nos permitiu concluir que mesmo um conhecimento bastante inexato acerca da distribuição dos coeficientes da matriz de mistura pode resultar em significativas melhoras dos métodos de separação.

As constatações acima inspiraram-nos a modificar um conhecido método de

deconvolução cega de fontes (também dedicado ao caso de misturas lineares e instantâneas), de modo a nele inserir um conhecimento estatístico acerca dos coeficientes da matriz de mistura. Verificamos que não raro esta modificação implica significativos ganhos de razão sinal-interferência mais ruído. Cabe notar que tanto neste caso quanto nos anteriores, a otimização é realizada via métodos de gradiente e almeja estimar a inversa da matriz de misturas, o que exige uma transformação do conhecimento acerca dos coeficientes da matriz de mistura para um conhecimento acerca dos coeficientes da inversa desta matriz.

A inserção deste tipo de conhecimento estatístico pode ser, obviamente, entendida para outros métodos de separação de fontes, e não necessariamente deve se restringir ao caso instantâneo. No caso de misturas anecóicas, onde o número de parâmetros a estimar também é relativamente pequeno, esta inserção pode se verificar vantajosa. No caso de misturas convolutivas, além deste conhecimento ser mais complicado de se possuir, é mais difícil tratá-lo matematicamente de forma a inseri-lo num método de separação. Uma proposta interessante para este caso seria utilizar características da continuidade da função de transferência do sistema, facilmente modelável, por exemplo, por cadeias de Markov de primeira ordem.

O conhecimento estatístico acima aludido não se revelou em geral proveitoso na possibilidade de termos acesso a um grande número de amostras das misturas. No entanto, particularmente em processamento de sinais de comunicações digitais, esta condição não costuma ser satisfeita, o que pode tornar este conhecimento bastante valioso em diversas situações práticas.

No terceiro capítulo, o problema da separação de fontes foi alçado a um maior grau de dificuldade, à medida em que foi abordado o problema da separação de fontes misturadas de forma convolutiva. Misturas convolutivas abarcam, como caso particular, o caso de misturas anecóicas (onde pode haver atraso entre sensores, embora não haja múltiplos percursos). Este tipo de mistura possui uma aplicabilidade bem maior, já que na prática as misturas convolutivas são as mais frequentes. Mas este é um problema bastante desafiador, pois é comum haver (em sinais de áudio, por exemplo) funções de transferência com comprimento da ordem de milhares.

Dois diferentes métodos de separação de fontes convolutivamente misturadas foram apresentados com maior detalhe, sendo denominados, nesta dissertação, por

GASFC e EDAO.

Foram feitas duas propostas para melhora de convergência do método GASFC, enfocadas na sua formulação *on-line*. A primeira consistiu em aplicar a este método algumas heurísticas de aceleração da convergência, as quais tornam variantes no tempo as taxas de aprendizagem, que passam a ser distintas para cada parâmetro ajustável. Após alguns melhoramentos, o método proposto apresentou uma taxa de convergência superior a do algoritmo original, sem perder desempenho após a estabilização dos parâmetros.

A segunda proposta para o algoritmo GASFC utilizou estimativas recursivas das matrizes de correlação. Esta proposta resultou numa grande aceleração da convergência do método, embora tenha havido uma perda de desempenho após a estabilização. Esta perda indesejável pode ser facilmente contornada, pois a recursão proposta depende de um fator de lembrança que pode ser reduzido à medida em que o algoritmo aproxima-se da estabilização.

Uma conjunção das duas propostas anteriores poderia agilizar ainda mais o processo de convergência do método GASFC.

Além das supracitadas modificações do método GASFC, duas avaliações relacionadas a este método foram efetuadas. A primeira consistiu na possibilidade de se utilizar filtros de separação de comprimento maior do que os filtros do sistema de mistura. Esta modificação, embora tenha aumentado o custo computacional, resultou em valores de SIR significativamente superiores, além de uma maior estabilidade de desempenho após a convergência do algoritmo. Claramente conclui-se, deste fato, que há um compromisso entre complexidade computacional e desempenho, o qual não havia sido devidamente enfatizado na literatura disponível.

A segunda avaliação no tocante ao método GASFC diz respeito à utilização de sua função custo para seleção de filtros IIR, rarissimamente utilizados em separação de fontes. Verificamos que, em geral, não há uma adequada seleção de filtros IIR caso escolhamos aqueles que resultem numa minimização da função custo. Comparações dos formatos das superfícies de SIR e das superfícies da função custo revelaram que muitas vezes os mínimos globais destas não coincidem com os máximos globais daquelas. Embora possamos não ter exaurido todas as possibilidades, identificamos duas configurações indesejáveis que podem ocorrer: o mínimo global da

função custo coincidir com um máximo local do SIR ou o máximo do SIR coincidir com um mínimo local da função custo. Estes inconvenientes descartaram a possibilidade de a função custo do método GASFC ser utilizada qual critério de seleção de filtros IIR de separação.

No tocante ao método EDAO, o qual pode ser interpretado como uma generalização multidimensional dos métodos-padrão de análise de componentes independentes, também foram efetuadas duas propostas: uma que conjuga suas estimativas com estimativas de direção de chegada das fontes (de modo a contornar o ainda presente problema da permutação) e outra que propicia ganhos de separação por meio de uma leve modificação na sua função custo.

Já no terceiro capítulo, o problema de mistura instantânea foi novamente contemplado. O nível de dificuldade acrescentado não mais diz respeito a uma maior quantidade de coeficientes dos filtros de misturas, mas a um reduzido número de misturas, problema cognominado “indeterminado”. A maioria das técnicas de análise de componentes independentes não mais se aplica neste caso, sendo em geral utilizadas técnicas de análise de componentes esparsos. Cabe ressaltar que estas técnicas ainda exploram a diversidade espacial das fontes, necessitando de pelo menos duas misturas para serem aplicáveis. Vimos que estas técnicas normalmente se valem de quatro passos para resolução do problema de separação de misturas indeterminadas.

Foi proposta uma nova técnica para a estimativa da matriz de misturas. Esta técnica recorre a modelos de misturas de gaussianas e apresentou desempenho superior a diversos outros métodos propostos na literatura. Dos três casos estudados, o método proposto foi superior a todos os outros em dois. Já no tocante a erro global, o algoritmo proposto apresentou desempenho (ligeiramente) inferior a apenas um dos algoritmos concorrentes.

Para a estimativa da representação das fontes no domínio da transformada, foram testados critérios de minimização de norma l_p , com $p \neq 1$, bem como introduzida uma função custo que penaliza soluções pouco contínuas. Verificamos que um valor de p superior a 1 não costuma provocar melhoras nas estimativas, embora valores inferiores à unidade o possam. O critério de continuidade testado, embora também utilizado no caso de uma única mistura disponível, não angariou ganhos significativos na separação, indicando que a informação oriunda da diversidade es-

pacial é muito mais informativa.

Comumente, os métodos que tratam do caso indeterminado supõem a cada instante que o número de fontes ativas (ou não-nulas) é igual à unidade ou ao número de misturas. Avaliamos experimentalmente os ganhos oriundos de um método que supõe um número de fontes ativas variante no tempo. Este método não se presta a propósitos de separação, pois o objetivo foi estabelecer, utilizando-se do conhecimento das fontes, um limite superior de desempenho. O desempenho deste método foi bem melhor do que nos métodos que supõem um número constante de fontes ativas. Interessantemente, um histograma do número ótimo de fontes ativas a cada instante tende a se concentrar em valores próximos de M . Observamos que este histograma cresce monotonamente, para um número de fontes inferior a 12, porém, para um número de fontes entre 12 e 14, verificamos a existência de picos no histograma para um número de fontes igual a $M - 1$. Não houve testes para um número de fontes maior do que 14, pois na prática é muito rara a situação de separar-se mais do que este número de fontes. A existência destes picos em valores menores do que M sugere que pode ter resultados interessantes uma estratégia que admita um número fixo de fontes ativas a cada instante, embora este número seja diferente de M , alternativa não encontrada por nós na literatura corrente. Uma técnica que utilize apenas as M misturas para estimar um número inferior (e maior do que 1) de fontes poderia ser desenvolvida para avaliar esta possibilidade.

No que diz respeito à reconstrução das fontes, verificamos que parte do erro de reconstrução deve-se a efeitos de *aliasing*, dado que os bancos de filtros de análise costumam ser criticamente decimados. Verificamos que um fator de decimação inferior ao crítico costuma melhorar sensivelmente as medidas de SAR e SDR das estimativas das fontes. Uma interessante sugestão seria analisar matematicamente o efeito do *aliasing* no erro de reconstrução, admitindo distribuições exponenciais dos coeficientes da decomposição das fontes. Outra avaliação quanto ao processo de reconstrução verificou que as atuais técnicas de análise de componentes esparsos, para um número de fontes menor do que 15, estimam as fontes com um erro muito superior ao que se pode obter no caso determinado (caso tenhamos, em ambas as situações, conhecimento pleno da matriz de misturas).

Por fim, no Capítulo 5, o problema de separação de fontes com apenas um

único sensor (ou mistura) foi contemplado. Esta é a configuração mais difícil para separação, pois é a que fornece menos informação possível. Não se pode explorar, neste tipo de problema, a diversidade espacial entre sensores. Uma forma de se contornar esta falta de conhecimento consiste em se recorrer a trechos das fontes isoladas, de forma a treinar um modelo de cada fonte. A partir do momento em que se utiliza um conhecimento (geralmente espectral) das fontes, não podemos mais denominar a separação de cega. Nesta dissertação, optamos por denominar este tipo de separação como “supervisionada”.

Foi detalhado um método de separação de fontes para o caso monosensor, o qual modela as raias da STFT com um modelo de misturas de gaussianas. Este método foi denominado como FGMM, apresentando algumas dificuldades, em especial a hipótese de independência entre raias, a qual não é atendida em, por exemplo, trechos de sustentação de instrumentos harmônicos.

Propomos então um método que utiliza redes neurais para separação, o qual contempla possíveis dependências estatísticas entre raias. Este método não modela a fase das raias das fontes, mas apenas o seu módulo, assim como o método FGMM. Ele efetua uma espécie de mascaramento na frequência do sinal de mistura, o qual diretamente gera as estimativas das fontes a partir do mesmo. A característica suave do mascaramento permitiu-nos obter um SAR bastante superior ao do método FGMM. Verificamos que as estimativas de instrumentos com características harmônicas foram visivelmente melhoradas com o método proposto, chamado de FNH.

Referências Bibliográficas

- [1] J. Anemuller, T. Sejnowski e S. Makeig, “*Complex independent component analysis of frequency domain eletroencephalographic data*”, *Neural Networks*, pp. 16:1311-1323, 2003.
- [2] D. B. Haddad, M. R. Petraglia e P. B. Batalheiro, “Sensibilidade da Separação Cega de Fontes à Variação do Comprimento dos Filtros de Separação e da Ordem da Estatística Utilizada na Função Custo”, XXV Simpósio Brasileiro de Telecomunicações - SBrT 2007.
- [3] H. Sawada, R. Mukai, S. Araki e S. Makino, “*A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation*”, *IEEE Transactions on Speech And Audio Processing*, vol. 12, no. 5, pp. 530-538, 2004.
- [4] M. S. Naceur, M. A. Loghmari e M. R. Boussema, “*The Contribution of the Sources Separation Method in the Decomposition of Mixed Pixels*”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2642-2653, no. 11, 2004.
- [5] F. Acernese, A. Ciaramella, S. De Martino, R. De Rosa, M. Falanga e R. Tagliaferri, “*Neural Networks for Blind-Source Separation of Stromboli Explosion Quakes*”, *IEEE Transactions on Neural Networks*, vol. 14, pp. 167-175, no. 1, 2003.
- [6] M. M. Ichir e A. Mohammad-Djafari, “*Hidden markov models for wavelet-based blind source separation*”, *IEEE Transactions on Image Processing*, vol. 15, pp. 1887-1899, no. 7, 2006.

- [7] A. T. Erdogan, “*Globally Convergent Deflationary Instantaneous Blind Source Separation Algorithm for Digital Communication Signals*”, *IEEE Transactions on Signal Processing*, vol. 55, pp. 2182-2192, no. 5, 2007.
- [8] N. Gadhok e W. Kinsner, “*An Implementation of β -Divergence for Blind Source Separation*”, *IEEE CCECE/CCGEL*, pp. 1446-1449, 2006.
- [9] C. C. Cheung e L. Xu, “*Some global and local convergence analysis on the information-theoretic independent component analysis approach*”, *Neurocomputing*, vol. 30, pp. 79-102, 2000.
- [10] X. Cao e R. Liu, “*General Approach to Blind Source Separation*”, *IEEE Transactions on Signal Processing*, vol. 44, pp. 562-571, no. 3, 1996.
- [11] H. Buchner, R. Aichner e W. Kellerman, “*A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics*”, *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 120-134, no. 1, 2005.
- [12] T. Kim, H. T. Attias, S. Lee e T. Lee, “*Blind Source Separation Exploiting Higher-Order Frequency Dependencies*”, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 70-78, 2007.
- [13] F. Gamboab e E. Gassiat, “*Source Separation when the Input Sources Are Discrete or Have Constant Modulus*”, *IEEE Transactions on Signal Processing*, vol. 45, pp. 3062-3071, no. 12, 1997.
- [14] P. Comon, “*Independent Component Analysis: A new concept?*”, *Signal Processing*, vol. 36, no. 3, pp. 287-314, 1994.
- [15] J. Cardoso, “*Blind Signal Separation: Statistical Principles*”, *Proceeding of the IEEE*, vol. 86, no. 10, pp. 2009-2025, 1998.
- [16] C. Févotte and S. J. Godsill, “*A Bayesian Approach for Blind Separation of Sparse Sources*”, *IEEE Transactions on Audio and Speech Processing*, pp. 2174-2188, 2006.

- [17] A. J. Bell e T. J. Sejnowski, “*An information-maximization approach to blind separation and blind deconvolution*”, *Neural Computation*, vol. 7, no. 6, pp. 1004-1034, 1995.
- [18] A. Hyvärinen, “*Fast and robust fixed-point algorithms for independent component analysis*”, *IEEE Transactions on Neural Networks*, pp. 626-634, 1999.
- [19] K. H. Knuth, “*A Bayesian Approach to source separation*”, *Proc First Int Workshop on ICA and Signal Separation*, pp. 283-288, 1999.
- [20] K. H. Knuth, “*Informed Source Separation: A Bayesian Tutorial*”, *Proc. 13th EUSIPCO 2005*, 2005.
- [21] S. Amari, A. Cichocki e H. H. Yang, “*A new learning algorithm for blind signal separation*”, *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT, vol. 8, pp.757-763, 1996.
- [22] A. Cichocki e S. Amari, *Adaptive Blind Signal and Image Processing*, New York: Wiley, 2002.
- [23] T. M. Cover, “*Elements of Information Theory*”, *John Wiley & Sons*, 1991.
- [24] S. Amari e H. Nagaoka, “*Methods of Information Geometry*”, *AMS e Oxford University Press*, 2000.
- [25] A. Rényi, “*On measures of entropy and information*”, *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, *University of California Press*, vol. 1, pp. 547-561, 1961.
- [26] L. Castedo, C. J. Escudero e A. Dapena, “*A Blind Signal Separation Method for Multiuser Communications*”, *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1343-1348, 1997.
- [27] R. Gribonval, E. Vincent e C. Févotte, “*Proposals for performance measurement in source separation*”, *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pp. 763-768, 2003.

- [28] E. Vincent, R. Gribonval e C. Févotte, “*Performance Measurement in Blind Audio Source Separation*”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462-1469, 2006.
- [29] D. B. Haddad, M. R. Petraglia e P. B. Batalheiro, “*Performance Evaluation of Two Semi-Blind Source Separation Methods*”, SPAWC08, 2008. Aceito para publicação.
- [30] L. Tong, G. Xu e T. Kailath, “*Blind identification and equalization based on second-order statistics: A time domain approach*”, *IEEE Transaction on Information Theory*, vol. 40, no. 2, pp. 340-349, 1994.
- [31] M. I. Gurelli e C. L. Nikias, “*EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals*”, *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 134-149, 1995.
- [32] G. Xu, H. Liu, L. Tong e T. Kailath, “*A least-squares approach to blind channel identification*”, *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982-2993, 1995.
- [33] Y. Hua, “*Fast maximum likelihood for blind identification of multiple FIR channels*”, *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661-672, 1996.
- [34] A. Gorokhov, P. Loubaton e E. Moulines, “*Second Order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach*”, *Proc. ICASSP*, vol. 5, pp. 2415-2418, 1996.
- [35] P. Smaragdis, “*Blind separation of convolved mixtures in the frequency domain*”, *Neurocomputing*, vol. 22, pp. 21-34, 1998.
- [36] S. Amari, S. C. Douglas, A. Cichocki e H. Yang, “*Multichannel blind deconvolution and equalization using the natural gradient*”, *Proc. 1st IEEE Workshop Signal Processing Advanced Wireless Commun.*, pp. 101-104, 1997.
- [37] K. Matsuoka e S. Nakashima, “*Minimal distortion principle for blind source separation*”, *Proc. ICA*, pp. 722-727, 2001.

- [38] L. Parra e C. Spence, “*Convolutional Blind Separation of Non-Stationary Sources*”, *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 320-327, 2000.
- [39] J. Anemüller e B. Kollmeier, “*Amplitude Modulation decorrelation for convolutional blind source separation*”, *Proc. ICA*, pp. 215-220, 2000.
- [40] N. Murata, S. Ikeda e A. Ziehe, “*An approach to blind source separation based on temporal structure of speech signals*”, *Neurocomputing*, vol. 41, pp. 1-24, 2001.
- [41] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda e F. Itakura, “*Evaluation of blind signal separation method using directivity pattern under reverberant conditions*”, *Proc. ICASSP*, pp. 3140-3143, 2002.
- [42] M. Z. Ikram e D. R. Morgan, “*A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation*”, *Proc. ICASSP*, pp. 881-884, 2002.
- [43] K. E. Hild II, “*Blind separation of convolutional mixtures using Renyi’s divergence*”, tese de doutorado, Universidade da Flórida, 2003.
- [44] K. E. Hild II, David Pinto, Deniz Erdogmus e Jose C. Principe, “*Convolutional Blind Source Separation by Minimizing Mutual Information Between Segments of Signals*”, *IEEE Transactions on Circuits and Systems - I: Regular Papers*, vol. 52, no. 10, pp. 2188-2196, 2005.
- [45] Y. Huang e J. Benesty (eds.), *Audio Signal Processing*, cap. 10, *Kluwer Academic Publishers*, 2004.
- [46] L. Molgedey e H. G. Schuster, “*Separation of a mixture of independent signals using time delayed correlations*”, *Physical Review Letters*, vol. 72, pp. 3634-3636, 1994.
- [47] E. Weinstein, M. Feder e A. Oppenheim, “*Multi-channel signal separation by decorrelation*”, *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 405-413, 1993.

- [48] S. Van Gerven e D. Van Compermolle, “*Signal Separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness*”, *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1602-1612, 1995.
- [49] H. Buchner, R. Aichner e W. Kellerman, “*Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity*”, *Proc. Int. Workshop Acoustic Echo Noise Control (IWAENC)*, pp. 275-278, 2003.
- [50] K. Matsuoka, M. Ohya e M. Kawamoto, “*A neural net for blind separation of nonstationary signals*”, *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.
- [51] H. Buchner, R. Aichner e W. Kellerman, “*A generalization of a class of blind source separation algorithms for convolutive mixtures*”, *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation (ICA)*, pp. 945-950, 2003.
- [52] A. Oppenheim, “*Inequalities connected with definite hermitian forms*”, *J. London Math. Soc.*, vol. 5, pp. 114-119, 1930.
- [53] R. Aichner, H. Buchner e W. Kellerman, “*Exploiting Narrowband Efficiency for Broadband Convolutive Blind Source Separation*”, *EURASIP Journal on Advances in Signal Processing*, ID 16381, 2007.
- [54] R. A. Jacobs, “*Increased rates of convergence through learning rate adaptation*”, *Neural Networks*, vol. 1, pp. 295-307, 1988.
- [55] S. Haykin, *Neural Networks*, Prentice Hall, 2a. edição, 1999.
- [56] R. Aichner, H. Buchner, S. Araki e S. Makino, “*On-line time-domain blind source separation of nonstationary convolved signals*”, *Proc. Int. Symp. Independent Component Analysis Blind Signal Separation (ICA)*, pp. 833-840, 2003.
- [57] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, terceira edição, 1996.
- [58] S. Araki, R. Mukai, S. Makino, T. Nishikawa e H. Saruwatari, “*The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutive Mixtures of Speech*”, *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 2, pp. 109-116, 2003.

- [59] M. Miyoshi e Y. Kaneda, “*Inverse filtering of room acoustics*”, *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 36, pp. 145-152, 1988.
- [60] P. Kisilev, M. Zibulevsky e Y. Y. Zeevi, “*Multiscale Framework For Blind Separation of Linearly Mixed Signals*”, *Journal of Machine Learning Research*, 2003.
- [61] Ö. Yilmaz e S. Rickard, “*Blind Separation of speech mixtures via time-frequency masking*”, *IEEE Transaction on Signal Processing*, vol. 52, no. 7, pp. 1830-1847, 2004.
- [62] R. Gribonval e S. Lesage, “*A survey of Sparse Component Analysis for Blind Source Separation: principles, perspectives and new challenges*”, *ESANN proceedings*, pp. 323-330, 2006.
- [63] P. Bofill e M. Zibulevsky, “*Underdetermined blind source separation using sparse representations*”, *Signal Processing*, vol. 81, no. 11, pp. 2353-2362, 2001.
- [64] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, S. Xie, “*Underdetermined Blind Source Separation Based on Sparse Representation*”, *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 423-437, 2006.
- [65] S. Lesage, S. Krstulovic e R. Gribonval, “*Under-Determined Source Separation: Comparison of Two Approaches Based on Sparse Decompositions*”, *ICA 2006*, pp. 633-640, 2006.
- [66] N. Mitianoudis e T. Stathaki, “*Overcomplete Source Separation Using Laplacian Mixture Models*”, *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 277-280, 2005.
- [67] N. Mitianoudis e T. Stathaki, “*Batch and Online underdetermined source separation using Laplacian Mixture Models*”, *IEEE Transactions on Audio, Speech & Language Processing*, vol. 15, no. 6, pp. 1818-1832, 2007.
- [68] C. Tomasi, “*Estimating Gaussian Mixture Densities with EM - A Tutorial*”, *Duke University*, data desconhecida.
- [69] J. R. Deller Jr., J. H. L. Hansen e J. G. Proakis, “*Discrete-Time Processing of Speech Signals*”, *IEEE Press*, 2000.

- [70] Y. Li, A. Cichocki e S. Amari, “*Analysis of sparse representation and blind source separation*”, *Neural Comput.*, vol. 16, pp. 1193-1234, 2004.
- [71] J. C. Dunn, “*A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters*”, *Journal of Cybernetics*, vol. 3, pp. 32-57, 1973.
- [72] J. C. Bezdek, “*Pattern Recognition with Fuzzy Objective Function Algorithms*”, *Plenum Press*, 1981.
- [73] M. Zibulevsky e B. A. Pearlmutter, “*Blind Source Separation by Sparse Decomposition*”, *Neural Computing*, vol. 13, no. 4, pp. 863-882, 2001.
- [74] S. Winter, W. Kellerman, H. Sawada e S. Makino, “*MAP-Based Underdetermined Blind Source Separation of Convolutional Mixtures by Hierarchical Clustering and l_1 -Norm Minimization*”, *EURASIP Journal on Advances in Signal Processing*, vol. 2007.
- [75] P. P. Vaidyanathan, “*Multirate systems and filter banks*”, *Prentice Hall*, 1993.
- [76] A. Jourjine, S. Rickard e O. Yilmaz, “*Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures*”, *Proc. 2000 IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 5, pp. 2985-2988, 2000.
- [77] F. Abrard, Y. Deville e P. White, “*From blind source separation to blind source cancellation in the underdetermined case: A new approach based on time-frequency analysis*”, *Proc. 3rd. Int. Conf. Independent Component Analysis*, pp.734-739 , 2001.
- [78] F. Abrard e Y. Deville, “*Blind separation of dependent sources using the time-frequency ratio of mixtures approach*”, *7th Int. Symp. Signal Processing Applications (ISSPA)* (apresentação), 2003.
- [79] F. Abrard e Y. Deville, “*A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources*”, *Signal Process.*, vol. 85, no. 7, pp. 1389-1403, 2005.

- [80] J. A. Bilmes, “*A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Mixture Models*”, Tech. Rep., U. C. Berkeley, 1998.
- [81] F. J. Theis, C. G. Puntonet e E. W. Lang, “*Median-based clustering for underdetermined blind signal processing*”, *IEEE Signal Processing Letters*, vol. 13, no. 2, pp. 96-99, 2003.
- [82] F. J. Theis, “*Mathematics in independent component analysis*”, Tese de Ph.D., Universidade de Regensburg, 2002.
- [83] I. Takigawa, M. Kudo e J. Toyama, “*Performance analysis of minimum l_1 -norm solutions for underdetermined source separation*”, *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pp. 582-591, 2004.
- [84] A. Hyvärinen, J. Karhunen e E. Oja, “*Independent Component Analysis*”, *New York: Wiley Interscience*, 2001.
- [85] M. Zibulevsky, B. A. pearlmutter, “*Blind source separation by sparse decompositions*”, *Technical Report* no. CS99-1, Universidade de Nova México, Albuquerque, 1999.
- [86] R. Saab, Ö. Yilmaz, M. J. McKeown e R. Abugharbieh, “*Underdetermined sparse blind source separation with delays*”, *Proc. First Workshop on Signal Processing with Sparse/Structured Representation (SPARS'05)*, Rennes, França, pp. 67-70, 2005.
- [87] R. Gribonval, S. Lesage, “*A survey of Sparse Component Analysis for Blind Source Separation: principles, perspectives, and new challenges*”, *European Symposium on Artificial Neural Networks (ESANN'2006)*, Bruges, Bélgica, pp. 323-330, 2006.
- [88] M. M. Ichir e A. Mohammad-Djafari, “*Hidden Markov Models for Wavelet-Based Blind Source Separation*”, *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1887-1899, 2006.

- [89] A. Tonazzini, L. Bedini e E. Salerno, “A Markov Model for Blind Image Separation by a Mean-Field EM Algorithm”, *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 473-482, 2006.
- [90] T. Virtanen, “Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066-1074, 2007.
- [91] J. C. Lagarias, J. A. Reeds, M. H. Wright e P. E. Wright, “Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions”, *SIAM Journal of Optimization*, Vol. 9 Number 1, pp. 112-147, 1998.
- [92] M. Togami, T. Sumiyoshi e A. Amano, “Sound Source Separation Of Overcomplete Convolutive Mixtures Using Generalized Sparseness”, *IWAENC*, 2006.
- [93] L. Benaroya, F. Bimbot e R. Gribonval, “Audio Source Separation With a Single Sensor”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 191-199, 2006.
- [94] S. T. Roweis, “One microphone source separation”, *Proc. NIPS*, 2000.
- [95] H. Sameti, H. Sheikhzadeh, L. Deng e R. L. Brennan, “HMM strategies for enhancement of speech signals embedded in nonstationary noise”, *IEEE Transactions Speech Audio Processing*, vol. 6, no. 5, pp. 445-455, 1998.
- [96] M. A. Casey e A. Westner, “Separation of mixed audio sources by independent subspace analysis”, *Proc. Int. Comp. Music Conf*, 2000.
- [97] S. Dubnov, “Extracting sound objects by independent subspace analysis”, *Proc. 22nd Int. Audio Eng. Soc. Conf.*, 2002.
- [98] J. Rouat, “Source Separation with One Ear: Proposition for an Anthropomorphic Approach”, *EURASIP Journal on Applied Signal Processing*, pp. 1365-1373, 2005.
- [99] D. L. Wang e G. J. Brown, “Separation of speech from interfering sounds based on oscillatory correlation”, *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp.684-697, 1999.

- [100] D. L. Wang e G. J. Brown, “*Separation of stop consonants*”, *ICASSP*, vol. 2, pp. 749-752, 2003.
- [101] Y. Ephraim e N. Merhav, “*Hidden Markov Processes*”, *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1518-1569, 1995.
- [102] M. Riedmiller e H. Braun, “*A direct adaptive method for faster backpropagation learning: The RPROP Algorithm*”, *Proceedings of the IEEE International Conference On Neural Networks*, 1993.
- [103] C. de Boor, *A Practical Guide to Splines, Applied Math. Sciences, Springer-Verlag*, 1978.
- [104] D. B. Haddad, M. R. Petraglia, P. B. Batalheiro, “*Separação de instrumentos musicais com uma única mistura*”, 6^o Congresso da AES Brasil, 2008. Aceito para publicação.