

ESTRUTURAS EM SUBBANDAS PARA FILTRAGEM ADAPTATIVA E
SEPARAÇÃO CEGA E SEMI-CEGA DE SINAIS DE VOZ

Diego Barreto Haddad

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientadores: Mariane Rembold Petraglia
Paulo Bulkool Batalheiro

Rio de Janeiro
Junho de 2013

ESTRUTURAS EM SUBBANDAS PARA FILTRAGEM ADAPTATIVA E
SEPARAÇÃO CEGA E SEMI-CEGA DE SINAIS DE VOZ

Diego Barreto Haddad

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. Paulo Bulkool Batalheiro, D.Sc.

Prof. João Marcos Travassos Romano, D.Sc.

Prof. Marcello Luiz Rodrigues de Campos, Ph.D.

Prof. Lisandro Lovisolo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2013

Haddad, Diego Barreto

Estruturas em Subbandas para Filtragem Adaptativa e Separação Cega e Semi-Cega de Sinais de Voz/Diego Barreto Haddad. – Rio de Janeiro: UFRJ/COPPE, 2013.

XXXI, 235 p.: il.; 29, 7cm.

Orientadores: Mariane Rembold Petraglia

Paulo Bulkool Batalheiro

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2013.

Referências Bibliográficas: p. 215 – 235.

1. Separação Cega de Fontes. 2. Filtragem Adaptativa.
3. Estruturas em Subbandas. I. Petraglia, Mariane Rembold *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

À minha família e amigos.

Agradecimentos

Esta tese encerra um importante ciclo de minha vida acadêmica. O aprendizado foi decerto proporcional ao dispêndio de energia, o qual, graças à minha orientadora (Mariane), jamais sobrepujou minhas forças. Por isso a ela reservo uma gratidão ilimitada. Desde o mestrado, o convívio com sua sabedoria e amplo conhecimento tem sido uma experiência muito rica. Também ao meu coorientador Paulo explico minha alegria por seu auxílio incessante, por suas contribuições e por sua amizade.

Minha dedicação à tese foi possibilitada, reconheço, por abnegações inenarráveis por parte de minha família e de meus amigos. Agradeço, portanto, à minha esposa Angélica, aos meus pais, irmão, primos, tios, avós e amigos, por pacientemente se privarem da minha atenção enquanto estive imerso em leituras de artigos, codificação de programas, na escrita de capítulos e na submissão de artigos.

Aos meus companheiros de trabalho no CEFET-RJ, que sempre apoiaram minha dedicação à pesquisa, registro meu profundo agradecimento. Ressalto que o trabalho com profissionais deste quilate é um privilégio que jamais valorizarei o suficiente.

Gostaria de listar nominalmente muitas pessoas que tanto contribuíram para minha formação pessoal, profissional e acadêmica, porém não poderia fazê-lo sem cometer, por omissão, profundas injustiças. A essas pessoas, que decerto se reconhecem implicitamente citadas aqui, meu muito obrigado!

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

ESTRUTURAS EM SUBBANDAS PARA FILTRAGEM ADAPTATIVA E SEPARAÇÃO CEGA E SEMI-CEGA DE SINAIS DE VOZ

Diego Barreto Haddad

Junho/2013

Orientadores: Mariane Rembold Petraglia
Paulo Bulkool Batalheiro

Programa: Engenharia Elétrica

Este trabalho contempla algoritmos de filtragem adaptativa em versões supervisionadas e autodidatas, dando ênfase a características típicas de sinais de voz, como não-estacionariedade, espectro colorido e não-gaussianidade. Apresenta-se uma análise de transiente e de desempenho em regime permanente do algoritmo IMPNLMS, proposto para identificação de respostas ao impulso esparsas, bem como de uma extensão deste algoritmo em subbandas. Fatores de aprendizagem variantes no tempo são aplicados ao IMPNLMS, bem como à técnica GBSS-SOS de separação no domínio do tempo, a qual foi estendida para uma estrutura em subbandas. Uma nova forma de inicialização das técnicas de separação no domínio da frequência é detalhada, assim como a inserção de coeficientes adaptativos interfrequenciais e o emprego da informação lateral do tempo de reverberação para determinação cega do tamanho do quadro da transformada. O emprego de decomposições atômicas para a identificação do sistema de mistura é avaliado, com os parâmetros da decomposição mais importantes sendo detalhados. O emprego de dicionários adaptados se revela promissor nesta identificação. Por fim, o algoritmo de equalização DCGN foi objeto de análises de transiente e de obtenção das soluções em regime permanente.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

SUBBAND STRUCTURES FOR ADAPTIVE FILTERING AND (SEMI-)BLIND
SEPARATION OF SPEECH SIGNALS

Diego Barreto Haddad

June/2013

Advisors: Mariane Rembold Petraglia

Paulo Bulkool Batalheiro

Department: Electrical Engineering

This work encompasses adaptive filtering algorithms in supervised and unsupervised versions, stressing typical characteristics of speech signals, as non-stationarity, coloring and non-Gaussianity. A transient and steady-state analysis of the IMPNLMS algorithm, proposed for sparse impulse responses identification, is presented, as well as a subband extension of this algorithm. The use of variant learning factor is proposed for the IMPNLMS, as well as for the GBSS-SOS technique for time-domain blind source separation, which is also extended to a subband structure. A new initialization method for frequency-domain blind source separation techniques is detailed, as well as the insertion of interfrequency adaptive coefficients. Reverberation time information, provided by blind estimation techniques, are also used for the selection of the transform frame length in these techniques. Atomic decompositions are used for identification of the mixture system, and the most important decomposition parameters are analyzed. The adaptation of dictionaries is implemented, with advantages in performance. At last, transient and steady-state properties of the DCGN equalization algorithm were analyzed.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xxii
Lista de Símbolos	xxiv
Lista de Abreviaturas	xxx
1 Introdução	1
1.1 Escopo da Tese	1
1.2 Relevância das Técnicas de Filtragem Adaptativa	2
1.3 Características dos Sinais de Voz	3
1.3.1 Não-estacionariedade	4
1.3.2 Esparsidade	5
1.3.3 Espectro Colorido	6
1.3.4 Não-gaussianidade	7
1.3.5 Variabilidades Extrínsecas dos Registros de Sinais de Voz: Funções de Transferência Acústicas	9
1.4 Bancos de Filtros e Sistemas Multitaxas	11
1.4.1 Bancos de Filtros	12
1.5 Estrutura da Tese	13
2 Algoritmos Proporcionais de Filtragem Adaptativa	16
2.1 Introdução	16
2.2 Algoritmos de Filtragem Adaptativa Proporcionais	19
2.3 Revisão Bibliográfica	20
2.4 Interpretação geométrica da família de algoritmos PNLMS	23
2.5 Considerações Finais	25
3 Análises e Extensões da Família de Algoritmos PNLMS	26
3.1 Introdução	26
3.2 Análise de Transiente do Algoritmo IMPNLMS	27

3.2.1	Simulações	40
3.3	Estimativa do MSE Final de algoritmos de IRE	43
3.4	Análise do MSE em Regime Permanente de uma Estrutura em Sub- bandas Aplicada a Algoritmos de IRE	49
3.4.1	Simulações	55
3.5	Estabilidade do Algoritmo	56
3.5.1	Simulações	58
3.6	Fatores de Aprendizagem Variantes no Tempo	59
3.6.1	Simulações	60
3.7	Considerações Finais	63
4	O Problema da Separação Cega de Fontes	65
4.1	O Problema da Separação Cega de Fontes	65
4.2	Estrutura Básica e Ambiguidades dos Algoritmos de Separação Cega de Fontes	67
4.3	Modelagem do Processo de Mistura	68
4.4	Análise de Componentes Principais	70
4.5	Princípios de Separação	73
4.6	Estatísticas de Ordem Superior	74
4.7	Análise de Componentes Independentes via Estimativa de Máxima Verossimilhança	75
4.8	Gradiente Natural como Ferramenta de Otimização	78
4.9	Esparsidade	80
4.10	Estimativa da Matriz de Mistura	82
4.10.1	Varredura de Zibulevsky	84
4.11	Métricas de Avaliação das Técnicas Não-Supervisionadas de Separa- ção e Desconvolução Cegas de Fontes	86
4.11.1	Razão Fonte-Interferências (SIR) e Razão Fonte-Distorção (SDR)	86
4.11.2	Interferência Interssimbólica	89
4.11.3	Avaliação Subjetiva	89
4.12	Considerações Finais	90
5	Separação Cega de Fontes no Domínio da Frequência	91
5.1	Introdução	91
5.2	Separação Cega de Fontes no Domínio da Frequência	92
5.2.1	Inicialização	93
5.2.2	Separação	95
5.2.3	Correção da Permutação	99
5.2.4	Resolução de Escalamento	108

5.3	Proposta de Procedimento de Inicialização Empregando Estimativas de Direção de Chegada	108
5.3.1	Resultados	111
5.4	Proposta de Inserção de Coeficientes Adaptativos Interfrecuenciais . .	114
5.4.1	Estrutura Proposta com Coeficientes Adaptativos Cruzados .	114
5.4.2	Restrição do Número de Coeficientes Adaptativos Cruzados .	116
5.4.3	Determinação dos Coeficientes Adaptativos da Estrutura Proposta	116
5.4.4	Simulações	118
5.5	Proposta de Emprego de Estimativas Cegas do Tempo de Reverberação	122
5.5.1	Seleção de F via Limiares de Decisão	124
5.5.2	Resultados	125
5.6	Considerações Finais	128
6	Estruturas em Subbandas para Técnicas de Separação Cega de Fontes no Domínio do Tempo	129
6.1	Introdução	129
6.2	Estrutura Multitaxas Aplicada ao Algoritmo GBSS-SOS	130
6.2.1	Algoritmo GBSS-SOS em Subbandas	135
6.3	Proposta de Adoção de Fatores de Aprendizagem Variantes	135
6.4	Estratégia de Pós-Processamento	138
6.5	Resultados	139
6.6	Considerações Finais	143
7	Identificação do Sistema de Mistura por meio de Técnicas de Matching Pursuit	145
7.1	Introdução	145
7.2	Motivação para o Emprego de Representações Redundantes	146
7.3	Decomposição via Dicionários Redundantes: Algoritmo MP	148
7.4	O Processo Anecoico de Mistura	154
7.5	Emprego do MP para BSS	155
7.5.1	Nova Proposta de um Algoritmo MP Multicanal para BSS . .	155
7.5.2	Estimativa Conjunta dos Coeficientes de Atenuação e dos Atrasos	157
7.5.3	Avaliação de Consistência das Estimativas	159
7.5.4	Comparação com o DUET	168
7.6	Aplicação da Aprendizagem de Dicionários a Técnicas Semi-cegas de Separação	169
7.6.1	Algoritmo K -SVD	171
7.6.2	Resultados	172

7.7	Considerações Finais	181
8	Análises de um Algoritmo de Desconvolução Cega	182
8.1	Algoritmo DCGN	184
8.1.1	Versão monocanal do algoritmo DCGN	184
8.2	Análises da Versão Monocanal do algoritmo DCGN	185
8.2.1	Equivariância	185
8.2.2	Estimativa de Primeira Ordem dos Parâmetros Finais do Algoritmo DCGN	188
8.2.3	Análise de Transiente do Algoritmo DCGN	197
8.2.4	Método EDO	200
8.2.5	Estimativa Analítica Precisa em Regime Permanente	201
8.2.6	Critério de Seleção das Soluções	205
8.3	Considerações Finais	207
9	Conclusões	208
	Referências Bibliográficas	215

Lista de Figuras

1.1	Ilustração do processo de cálculo da STFT. A despeito de sua natureza discreta, os sinais são apresentados como contínuos para melhor visualização.	5
1.2	Exemplos de distribuições subgaussiana (em azul), gaussiana (em vermelho) e supergaussiana (em verde). As caudas das distribuições foram ampliadas para que suas magnitudes relativas possam ser observadas.	8
1.3	Distribuições de probabilidade de sinais de voz.	9
1.4	Função de transferência medida.	11
1.5	Bancos de filtros digitais.	13
1.6	Principais características dos capítulos centrais da tese (excluindo-se os capítulos de introdução e de conclusões). Os capítulos que contém contribuições estão assinalados em <i>itálico</i>	15
2.1	Estrutura de um algoritmo adaptativo de identificação supervisionado.	17
2.2	Interpretação geométrica de um algoritmo da família PNLMS para $L = 3$	24
2.3	Interpretação geométrica da atualização dos algoritmos de IRE para $L = 2$	25
3.1	Análise experimental da Hipótese III.	40
3.2	Análise experimental da Hipótese IV.	41
3.3	Análise experimental da Hipótese V.	41
3.4	Análise da Hipótese VI.	42
3.5	Análise da Hipótese VII.	42
3.6	Evolução do MSE (em dB) da primeira configuração (caso 1). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).	44

3.7	Evolução do MSE (em dB) da segunda configuração (caso 2). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).	44
3.8	Evolução do MSE (em dB) da terceira configuração (caso 3). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).	45
3.9	Evolução do MSE (em dB) da quarta configuração (caso 4). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em ciano: evolução teórica empregando a técnica proposta (modelo 2).	45
3.10	Evolução do MSE (em dB) da quinta configuração (caso 5). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em ciano: evolução teórica empregando a técnica proposta (modelo 2).	46
3.11	Estrutura WT-SF com banco de filtros não-uniforme e filtros adaptativos esparsos.	50
3.12	Evolução do MSE de diversos algoritmos de identificação de respostas esparsas.	56
3.13	(a) MSE em regime permanente do algoritmo WTIMPNLMS-SF: linhas sólidas representam resultados experimentais e as tracejadas refletem os resultados teóricos; (b) Probabilidade de divergência.	57
3.14	Evolução do MSE dos algoritmos NLMS, VSS-NLMS, IMPNLMS, VSS-IMPNLMS, WTIMPNLMS-SF e VSS-WTIMPNLMS-SF, aplicados na mesma configuração de identificação, com sinal de entrada colorido.	62
3.15	Evolução do MSE dos algoritmos NLMS, VSS-NLMS, IMPNLMS, VSS-IMPNLMS, WTIMPNLMS-SF e VSS-WTIMPNLMS-SF, aplicados na mesma configuração de identificação.	63

4.1	Problema e objetivo ideal da separação cega de fontes (BSS).	66
4.2	Diagrama de blocos funcional que ilustra o processo de separação cega de fontes.	67
4.3	Ilustração de um processo de separação de fontes considerado bem-sucedido. As fontes estão denotadas por $s_i(t)$, as misturas por $x_i(t)$ e as estimativas das fontes por $y_i(t)$. Os sinais são contínuos para melhor visualização.	68
4.4	Geração da i -ésima mistura a partir das N fontes.	69
4.5	Exemplo de componente principal (vetor \mathbf{v}_1) para uma determinada distribuição bidimensional de \mathbf{x}_k	71
4.6	Exemplo de compressão de uma imagem via PCA. (a) Imagem original; (b) 5 componentes principais; (c) 10 componentes principais; (d) 15 componentes principais; (e) 20 componentes principais e (f) 25 componentes principais.	72
4.7	Exemplo de transformações num sinal \mathbf{s}_k bidimensional. (a) sinal \mathbf{s}_k ; (b) $\mathbf{x}_k = \mathbf{H}\mathbf{s}_k$, onde \mathbf{H} é uma matriz gerada aleatoriamente e (c) versão esferizada \mathbf{z}_k de \mathbf{x}_k	73
4.8	Logaritmo da verossimilhança das misturas em função dos parâmetros u e v . Tanto a distribuição real das fontes quanto seu modelo são exponenciais.	76
4.9	Logaritmo da verossimilhança das misturas em função dos parâmetros u e v . A despeito da distribuição real das fontes (exponencial e, portanto, super-gaussiana), foi empregado um modelo de distribuição subgaussiano para o cálculo da verossimilhança (uma mistura de gaussianas).	77
4.10	O vetor gradiente em duas diferentes parametrizações da mesma função \mathcal{F} . (a) $\mathcal{F}(\mathbf{w})$; (b) $\mathcal{F}(\boldsymbol{\theta})$	79
4.11	Exemplo dos vetores de mistura e da matriz de mistura numa configuração determinada; os vetores em preto são as colunas da matriz de mistura; o vetor \mathbf{x}_k aparece em azul.	82
4.12	Pontos dos vetores \mathbf{x}_k . (a) configuração 2×3 (domínio do tempo); (b) configuração 2×3 (domínio da transformada); (c) configuração 3×5 (domínio do tempo) e (d) configuração 3×5 (domínio da transformada).	83
4.13	Histograma típico obtido pela varredura de Zibulesvsky.	85
4.14	A decomposição, num contexto de misturas instantâneas, de um mesmo vetor. (a) supondo que seja uma estimativa da fonte 1; (b) supondo que seja uma estimativa da fonte 2. Os vetores das fontes, do ruído e da estimativa foram mantidos constantes.	88

4.15	Estrutura de desconvolução que será empregada nesta tese. O filtro $H(z)$ (desconhecido) modela a transformação empreendida pelo canal e $\check{W}(z)$ é o filtro equalizador.	89
5.1	Diagrama de blocos das técnicas de separação de fontes no domínio da frequência.	93
5.2	Ilustração da importância da resolução de permutação. Em azul, componentes da fonte 1; em verde, componentes da fonte 2. Embora em todas as raias a separação tenha sido bem-sucedida, a falta de consistência na associação das estimativas às fontes corretas degrada de forma intolerável o desempenho do processo de separação. Por simplicidade, supõe-se que o problema de escalamento foi perfeitamente resolvido.	97
5.3	Ilustração dos módulos das funções <i>score</i> complexas derivadas da tangente hiperbólica real; (a) versão “cartesiana”; (b) versão “polar”.	98
5.4	Vetores empregados na modelagem de campo próximo.	101
5.5	Modelo de campo distante.	103
5.6	Evolução do envelope de um sinal de voz de 8 segundos de duração, com $F = 1024$, $S = 256$ e janela de Hanning. (a) frequência de 773,43 Hz; (b) frequência (adjacente) de 781,25 Hz; (c) frequência harmônica de 1554,68 Hz e (d) frequência não harmônica de 1929,68 Hz.	106
5.7	Configuração circular empregada para a gravação de misturas de sinais de voz com um arranjo de microfones.	111
5.8	Média da SIR (em dB) e das avaliações subjetivas para o ambiente ruidoso. Em azul, B-EDAOR; em vermelho, M-EDAOR.	113
5.9	Estruturas de sistemas de separação no domínio da STFT. (a) Estrutura padrão; (b) estrutura proposta com coeficientes adaptativos interfrequenciais entre raias adjacentes.	115
5.10	Diagrama de blocos do método proposto para a determinação dos coeficientes adaptativos, incluindo os cruzados, os quais são aplicados apenas às subbandas adjacentes.	119
5.11	Razão fonte-interferência em função do ângulo da fonte 2. Círculos: método proposto; asteriscos: método padrão. (a) e (b): configuração 1; (c) e (d): configuração 2.	121
5.12	T_{60} em função de $\sigma_{x^-}^2$ em misturas de sinais de voz, com janelamento de Hanning com sobreposição de 50% e $N_f = 20$. (a) $f_s = 8$ kHz e $F_T = 32$; (b) $f_s = 8$ kHz e $F_T = 64$; (c) $f_s = 8$ kHz e $F_T = 128$; (d) $f_s = 16$ kHz e $F_T = 32$; (e) $f_s = 16$ kHz e $F_T = 64$ e (f) $f_s = 16$ kHz e $F_T = 128$	124

5.13	Escolha ótima de F para os casos que constam no conjunto de teste. Em amarelo: $F = 512$, em vermelho: $F = 1024$, em verde: $F = 2048$ e em azul, $F = 4096$. As linhas tracejadas assinalam os limiares obtidos pela maximização da SDR média no conjunto de treinamento. Os limiares assinalados determinam quatro intervalos, para os quais a técnica proposta escolheria, da esquerda para a direita, os seguintes valores de F , respectivamente: 4096, 2048, 1024 e 512. (a) $f_s = 8$ kHz e $F_T = 32$; (b) $f_s = 16$ kHz e $F_T = 128$	125
5.14	Estimativa da função densidade de probabilidade de $S(n)$ para 8 kHz. Em azul: estratégia proposta; em vermelho: escolha ótima. (a) escolha fixa $F = 512$; (b) escolha fixa $F = 1024$; (c) escolha fixa $F = 2048$ e (d) escolha fixa $F = 4096$	126
5.15	Estimativa da função densidade de probabilidade de $S(n)$ para 16 kHz. Em azul: estratégia proposta; em vermelho: escolha ótima. (a) escolha fixa $F = 512$; (b) escolha fixa $F = 1024$; (c) escolha fixa $F = 2048$ e (d) escolha fixa $F = 4096$	126
6.1	Configuração MIMO linear para a solução do problema de BSS determinado e superdeterminado.	131
6.2	Estrutura empregada para a separação em subbandas.	132
6.3	SIR média (em dB) obtida pelos algoritmos na versão original (linha sólida) e na versão com variação do fator de aprendizagem (linha tracejada) para controle de instabilidade, no caso 2. (a) algoritmo padrão, com $L = 512$; (b) algoritmo padrão, com $L = 1024$; (c) algoritmo padrão, com $L = 2048$; (d) algoritmo em subbandas.	141
6.4	SIR média (em dB) obtida pelos algoritmos sem processamento multitaxas (linha sólida) e na versão em subbandas. (a) caso 1 ($L = 512$); (b) caso 2 ($L = 1024$). Foram escolhidos, para o algoritmo padrão, os valores de L que apresentaram melhor desempenho.	142
6.5	SIR média (em dB) obtida pelos algoritmos sem processamento multitaxas (linha sólida) e na versão em subbandas. Em ambas as versões, o pós-processamento proposto foi inserido, com $\gamma_p = 0,8$ e $N_p = 10$. (a) caso 1 ($L = 512$); (b) caso 2 ($L = 1024$). Foram escolhidos, para o algoritmo padrão, os valores de L que apresentaram melhor desempenho.	142

6.6	SIR média (em dB) obtida pelos algoritmos, para o caso 3. Em vermelho: algoritmo padrão; em verde: algoritmo com controle de instabilidade; em azul tracejado: algoritmo com controle de instabilidade e pós-processamento. Quando utilizado, o pós-processamento utilizou como parâmetros $\gamma_p = 0,8$ e $N_p = 10$. Nesta configuração, o baixo desempenho não permitiu que o pós-processamento ensejasse ganhos. (a) algoritmo em banda cheia ($L = 512$); (b) algoritmo em banda cheia ($L = 1024$); (c) algoritmo em banda cheia ($L = 2048$) e (d) algoritmo em subbandas.	143
7.1	Exemplo de projeção ortogonal de um sinal $\bar{\mathbf{x}}_i$ no átomo $\phi^{(m)}$	149
7.2	Exemplo de decomposição via MP de um sinal tridimensional ($L = 3$), com um dicionário composto por 5 átomos (vetores em azul), cada qual com componentes ϕ_i , para $i \in \{1, 2, 3\}$, os quais não devem ser confundidos com o i -ésimo átomo $\phi^{(i)}$. Em verde, os resíduos de cada iteração. Os vetores tracejados em azul indicam as projeções do sinal (na primeira iteração) ou dos resíduos (nas iterações posteriores) no átomo que apresenta o maior produto interno. (a) Primeira iteração com sinal original em preto; (b) segunda iteração, com o resíduo $\mathcal{R}^{(1)}$ em preto; (c) terceira iteração, com o resíduo $\mathcal{R}^{(2)}$ em preto e (d) aproximação (em cinza tracejado) do sinal original (em preto), com o resíduo final $\mathcal{R}^{(3)}$ da representação assinalado em vermelho. Ao contrário deste exemplo ilustrativo, em geral o número de iterações é bem inferior a L	151
7.3	Exemplo de decomposição via MP de um sinal de dimensão 512, no qual a maior parte da energia do sinal é extraída em apenas duas iterações.	152
7.4	Exemplos de átomos de um dicionário com $L = 512$ e $\bar{\phi}(k) = e^{-\pi k^2}$. (a) $\vartheta_3 = 0,0101$ e $\vartheta_2 = 64$; (b) $\vartheta_3 = 0,0101$ e $\vartheta_2 = 256$; (c) $\vartheta_3 = 0,0477$ e $\vartheta_2 = 256$ e (d) $\vartheta_3 = 0,0226$ e $\vartheta_2 = 128$	153
7.5	Exemplo gráfico do conceito de interferência. Ao vetor preto adicionamos dois vetores (o azul e o vermelho) de normas equivalentes ao raio do círculo verde. Se o vetor resultante for externo ao círculo cinza (caso do vetor azul), a soma é construtiva; sendo interno, a soma é destrutiva (caso do vetor vermelho).	160

7.6	Exemplo real de decomposição MP com o surgimento de componentes que não modelam o sinal original, cujos surgimentos se devem apenas à redução de erros progressos de modelagem. (a) sinal original (em azul) e a aproximação obtida com os 6 primeiros componentes (em vermelho); (b) sinal original (em azul) e o sétimo componente (em verde) extraído via MP, o qual, por não ter relação com o sinal original, apresenta interferência destrutiva; (c) ampliação do trecho assinalado em (a); (d) ampliação do trecho assinalado em (b).	160
7.7	Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = 0,30798z^9$ e $\tilde{H}_{22}(z) = 2,7917z^{-8}$	161
7.8	Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = -4,0355z^9$ e $\tilde{H}_{22}(z) = -3,6803z^{10}$	162
7.9	Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = 2,8023z^9$ e $\tilde{H}_{22}(z) = -4,1887z^6$	162
7.10	Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = -4,6226z^9$ e $\tilde{H}_{22}(z) = 3,8517z^6$	163
7.11	Exemplos de átomos extraídos do dicionário adaptado para o idioma fr , com $\mathcal{Q} = 3$ e $\mathcal{N} = 8$	174
7.12	Exemplos de átomos extraídos do dicionário adaptado para o idioma us , com $\mathcal{Q} = 3$ e $\mathcal{N} = 8$	174
7.13	Exemplos de átomos extraídos do dicionário adaptado para o idioma uk , com $\mathcal{Q} = 3$ e $\mathcal{N} = 8$	175
7.14	MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 1$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$	176
7.15	MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 2$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$	177

- 7.16 MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 3$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$ 177
- 7.17 MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para o caso subdeterminado e com 500 blocos decompostos via MP. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$ 179
- 8.1 Avaliação da estimativa da Eq. (8.14), com inicialização $\check{\mathbf{w}}_0 = [1 \ 0]$, $\beta = 10^{-5}$ e 5×10^5 iterações. O sinal $s(k)$ foi gerado a partir da distribuição $f(s) = \frac{1}{2}e^{-|s|}$. A reta identidade está em preto e o termo $\frac{1}{\mathbb{E}[f(c(0)s(k))s(k)]}$ (calculado via médias de Monte Carlo) em azul. Em vermelho, assinala-se o valor real para o qual o algoritmo converge (calculado por meio da média das últimas 3×10^5 iterações). (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0 \ 3]$; (b) ampliação de parte do gráfico em (a); (c) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$ (d) ampliação de parte do gráfico de (c). 189
- 8.2 Evolução da ISI nos testes efetuados na Fig. 8.1, com inicialização $\check{\mathbf{w}}_0 = [1 \ 0]$ e $\beta = 10^{-5}$. (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0 \ 3]$; (b) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$ 189
- 8.3 Evolução dos parâmetros $\check{w}_k(0)$, $\check{w}_k(1)$ e $\check{w}_k(2)$ para os quatro casos do experimento com $\mathbf{M} = \mathbf{L} = 3$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 1 \ 0]$. As linhas horizontais refletem o valor teórico. (a) filtro $\mathbf{h}^{(1)}$; (b) filtro $\mathbf{h}^{(2)}$; (c) filtro $\mathbf{h}^{(3)}$ e (d) filtro $\mathbf{h}^{(4)}$. 194
- 8.4 Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(1)}$ do experimento com $\mathbf{M} = \mathbf{L} = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB). 195

8.5	Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(2)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).	195
8.6	Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(3)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).	196
8.7	Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(4)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).	196
8.8	Aplicação da análise de transiente para o algoritmo DCGN.	199
8.9	Aplicação da análise de transiente para o algoritmo DCGN para um canal.	199
8.10	Evolução dos coeficientes de uma simulação (linha tracejada) do algoritmo DCGN.	200
8.11	Comparação das funções $\check{w}_{k,i}(t)$ (em tracejado) com o valor final dos parâmetros $\check{w}_k(0)$ e $\check{w}_k(1)$ (em linha sólida; todos foram obtidos com uma simulação com 25000 iterações, obtendo-se a média das últimas 5000 iterações e $\beta = 3 \times 10^{-3}$, $f(y(k)) = y^3(k)$ e $s(k)$ com amostras iid e pdf uniforme no intervalo $[-1,1]$). (a) filtro do canal $\mathbf{h} = [1 \ 0, 15]$; (b) filtro do canal $\mathbf{h} = [-1 \ 0, 2]$; (c) filtro do canal $\mathbf{h} = [2 \ -0, 4]$ e (d) filtro do canal $\mathbf{h} = [0, 4 \ 1, 5 \ 1]$. Para todas as simulações arbitrou-se $L = 2$, com a inicialização $\check{\mathbf{w}}_0 = [0 \ 1]$	202
8.12	Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função <i>score</i> parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $\check{w}_k(0)$; em azul: valores de $\check{w}_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 0, 5$ e (b) $\alpha = 1$	204

8.13 Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função *score* parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $\check{w}_k(0)$; em azul: valores de $\check{w}_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 1,5$ e (b) $\alpha = 2$ 204

Lista de Tabelas

1.1	Características dos sinais de voz empregados em cada capítulo de inovação da tese.	15
2.1	Algoritmo IMPNLMS	22
3.1	Principais equações recursivas da análise de convergência do algoritmo IMPNLNS.	39
3.2	Configurações testadas nas simulações. Em todos os casos, os valores de δ , ϵ , λ_ξ e $\xi(-1)$ foram, respectivamente, 10^{-2} , 10^{-3} , 10^{-1} e 0,96.	43
3.3	Configurações do sinal de entrada $x(k)$ e do ruído de medição $\nu(k)$. Em todos os casos, o sinal $x(k)$ foi obtido através do colorimento de um sinal branco e gaussiano de variância $\sigma_x^2 = 1$	43
3.4	Algoritmo WTPNLMS-SF	52
3.5	Algoritmo VSS-IMPNLMS	61
3.6	MSE (em dB) em regime permanente dos algoritmos testados.	63
5.1	SIR no ambiente ruidoso e em silêncio.	112
5.2	Número de iterações (N_{it}) para a convergência dos algoritmos e tempo total de execução (T).	112
5.3	Resultados subjetivos no ambiente ruidoso.	112
5.4	Especificações adicionais das configurações testadas. \mathcal{N} é o número de raias empregadas para a separação e \mathcal{N}_{prop} é o número de raias (a partir da segunda) nas quais os CACs foram inseridos. O limiar de correlação é relativo ao maior módulo do coeficiente de correlação de cada agrupamento.	120
5.5	Arquivos da base de dados referentes às fontes utilizadas em cada caso.	120
5.6	Resultados adicionais das configurações testadas. #sup indica o número de vezes (de um total de 91, pois o ângulo da segunda fonte varia de 0° a 90°) nas quais o novo método supera o método padrão.	121
5.7	SIR no método padrão e no método com coeficientes interfrequenciais aplicados a todas as raias ($\mathcal{N} = 1024$).	122
5.8	Intervalo de variação dos parâmetros de cada simulação.	123

5.9	SDR média (em dB) em função de diferentes escolhas de K , no conjunto de teste.	126
5.10	Probabilidade P_b dos casos em que a escolha automática (proposta) de F obtém um desempenho igual ou superior à escolha de um F fixo, para $f_s = 8$ kHz e $f_s = 16$ kHz, avaliada no conjunto de teste.	127
5.11	Probabilidade P_p da estratégia proposta escolher um valor específico de F , para $f_s = 8$ kHz e $f_s = 16$ kHz, avaliada no conjunto de teste.	127
6.1	Características dos casos testados.	140
7.1	Valores de σ_F^2 para diferentes combinações de p e L (média das 10 configurações testadas).	164
7.2	Percentual de componentes selecionados em função de L , p , número da iteração e de $\log_2(\vartheta_2)$	165
7.3	Percentual de componentes selecionados em função de L , p , número da iteração e de intervalos de ϑ_3	166
7.4	Valor de σ_F^2 em função da escala $\log_2(\vartheta_2)$ (média das configurações testadas).	167
7.5	Valor de σ_F^2 em função da frequência ϑ_3 (média das configurações testadas).	167
7.6	Valor de σ_F^2 em função do número da iteração que extraiu o componente (média das configurações testadas).	168
7.7	Taxa de acerto dos atrasos ao longo das 40 misturas efetuadas.	169
7.8	MSE das estimativas dos coeficientes de atenuação, ao longo das 40 misturas.	169
7.9	MSE (em dB) obtido com emprego de diferentes dicionários, para o caso determinado, para diferentes fatores de redundância \mathcal{N} e número de quadros \mathcal{Q} , com 500 blocos decompostos via MP.	179
7.10	MSE (em dB) obtido com emprego de diferentes dicionários, para o caso subdeterminado, para diferentes fatores de redundância \mathcal{N} e número de quadros \mathcal{Q} , com 500 blocos decompostos via MP.	180

Lista de Símbolos

$C(z)$	resposta global correspondente ao processo de filtragem do canal e ao equalizador, p. 89
D_n	número de correlações com diferentes atrasos relativos na n -ésima subbanda, p. 133
$F(\cdot)$	função linear, definida por partes, empregada pelo algoritmo IMPNLMS, p. 21
L	número de coeficientes de um filtro adaptativo ou de elementos num dicionário, p. 16
L'_n	comprimento dos subfiltros de separação na n -ésima subbanda, p. 133
M	número de misturas, p. 65
M_{F_n}	comprimento do n -ésimo filtro de síntese, p. 50
M_{H_n}	comprimento do n -ésimo filtro de análise, p. 50
N	número de fontes, p. 65
N_{it}	número de iterações, p. 112
$N_{b,n}$	total de blocos considerados na otimização da n -ésima subbanda, p. 135
$\#\mathcal{D}$	cardinalidade do dicionário \mathcal{D} , p. 150
Δ_{ji}	atraso existente entre a i -ésima fonte e o j -ésimo sensor, num contexto de mistura anecoica, p. 154
$\Phi(\theta_t, \theta_k, \mathbf{x}_k)$	função potencial empregada na varredura de Zibulevsky, p. 85
$\Phi_M^{(i)}(\mathbf{p})$	norma l_p das projeções do i -ésimo átomo nas misturas, p. 156

Π_k	hiperplano dos coeficientes do filtro adaptativo onde o erro <i>a posteriori</i> se anula, p. 23
$\alpha(k)$	Parâmetro utilizado pelo algoritmo IMPNLMS para a distribuição da energia de atualização, p. 21
Γ_k	matriz diagonal responsável pela distribuição da energia de atualização dos algoritmos proporcionais, p. 22
β	fator de aprendizagem do algoritmo adaptativo, p. 18
$\beta_{n,p}$	fator de aprendizagem referente à p -ésima estimativa na n -ésima subbanda, p. 136
$\check{W}(z)$	filtro correspondente ao equalizador, p. 89
$\check{\mathbf{w}}_k$	filtro adaptativo de equalização, na k -ésima iteração, p. 184
δ	fator de regularização empregado para evitar divisões por zero, p. 19
ϵ	fator de regularização empregado para evitar divisões por zero, p. 22
$\gamma_{\Delta}^{(q)}$	diferença de deslocamento (circular) entre a q -ésima e a primeira mistura do átomo selecionado, p. 158
γ_{β}	fator de aumento ou redução do fator de aprendizagem do algoritmo GBSS-SOS, p. 136
λ	fator de suavidade da função potencial $\Phi(\theta_t, \theta_k, \mathbf{x}_k)$, p. 85
λ_{ξ}	fator de lembrança da medida de esparsidade do algoritmo IMPNLMS., p. 22
$\lambda_h(\mathbf{f}_0, k_0)$	decaimento associado à raia \mathbf{f}_0 e ao quadro k_0 , p. 123
λ_i	i -ésimo autovalor de $E[\Gamma_k] \mathbf{R}_{\mathbf{x}}$, p. 36
Λ	matriz diagonal que contém os autovalores de $\mathbf{R}_{\mathbf{x}}$, p. 73
Λ_k	matriz diagonal que contém os autovalores de $E[\Gamma_k] \mathbf{R}_{\mathbf{x}}$, p. 36
\mathcal{D}	dicionário utilizado pela técnica <i>Matching Pursuit.</i> , p. 149
$\mathcal{F}(\cdot)$	função-custo, p. 78

$\mathcal{I}(n)$	interferência do n -ésimo átomo selecionado numa representação de J átomos, p. 159
$\mathcal{J}(\mathbf{X}, \mathbf{D}, \mathbf{A})$	função-custo associada ao algoritmo K -SVD, p. 171
$\mathcal{M}_j(\mathbf{f}, k)$	máscara binária aplicada no domínio da STFT, p. 6
\mathcal{Q}	número de quadros adjacentes utilizados para formar um átomo, p. 173
$\mathcal{R}^{(j)}$	resíduo obtido pelo MP na j -ésima iteração, p. 149
$\mathcal{X}_i(\mathbf{f}, k)$	coeficiente da transformada STFT do sinal $x_i(k)$, p. 4
$\phi^{(m)}$	m -ésimo átomo de um dicionário, p. 149
$\mathcal{P}_{\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q\}} \{\mathbf{y}_j\}$	projeção ortogonal de \mathbf{y}_j no subespaço gerado pelos vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q$, p. 87
$\mu_k(i)$	média da distribuição de $\hat{h}_k(i)$, p. 29
$\nu(k)$	ruído de medição aditivo, p. 17
ν_i^τ	i -ésimo sinal de ruído atrasado de τ unidades, p. 87
ω_l	l -ésima raiz de $\det(\mathbf{I} + 2\omega \mathbf{E}[\mathbf{\Gamma}_k] \mathbf{R}_x)$, p. 37
$\overline{\mathbf{G}}(\mathbf{w})$	tensor da métrica de Riemann, p. 79
$\overline{\mathbf{x}}_i$	sinal a ser decomposto pela técnica MP, p. 149
$\phi_i(y_i)$	função <i>score</i> associada à i -ésima estimativa, p. 80
σ_ν^2	variância do ruído de medição $\nu(k)$, p. 17
$\sigma_k^2(i)$	variância da distribuição de $\hat{h}_k(i)$, p. 29
\mathbf{H}	matriz que modela um processo de mistura instantânea, p. 69
\mathbf{Q}_k	matriz que contém os autovetores de $\mathbf{E}[\mathbf{\Gamma}_k] \mathbf{R}_x$, p. 36
\mathbf{R}_x	matriz de autocorrelação de \mathbf{x}_k , p. 30
$\mathbf{R}_{ij}^n(l)$	matriz de dimensões $D_n \times D_n$ que contém as estimativas da correlação dos sinais $\underline{y}_i(m)$ e $\underline{y}_j(m)$ para D_n blocos, p. 135
\mathbf{c}_k	vetor que contém as respostas combinadas do canal e do filtro equalizador, p. 186

$\mathbf{f}_s(\mathbf{s}_k)$	função densidade de probabilidade das fontes, p. 75
$\mathbf{f}_x(\mathbf{x}_k)$	função densidade de probabilidade das misturas, p. 75
\mathbf{h}	vetor-linha que contém os coeficientes da resposta impulsiva de um sistema linear e invariante no tempo, p. 17
\mathbf{h}^j	j -ésima coluna da matriz de mistura \mathbf{H} , p. 82
\mathbf{v}_i	vetor linha que contém a i -ésima linha da matriz de branqueamento, p. 70
\mathbf{x}_k	vetor coluna que contém as amostras de um filtro adaptativo na k -ésima iteração, p. 16
M	comprimento da resposta impulsiva, p. 16
W_a	janela de análise, p. 4
θ_k	ângulo da reta que passa pelo ponto $(x_1(k), x_2(k))$ e pela origem, p. 84
$\tilde{f}_i(y_i)$	função densidade de probabilidade aproximada das fontes, p. 80
$\underline{F}_i(z)$	i -ésimo filtro do banco de filtros de síntese, p. 12
$\underline{H}_i(z)$	i -ésimo filtro do banco de filtros de análise, p. 12
\underline{L}_m	fator de decimação da m -ésima subbanda, p. 12
\underline{M}	número de subbandas, p. 12
\underline{N}_n	comprimento dos blocos extraídos na n -ésima subbanda, pelo algoritmo GBSS-SOS, das misturas na n -ésima subbanda, p. 133
$\underline{\nu}_j$	vetor que contém algumas amostras consecutivas do ruído na j -ésima mistura, p. 87
$\underline{\mathbf{s}}_j$	vetor que contém algumas amostras consecutivas da j -ésima fonte, gerado para fins de avaliação, p. 87
$\underline{\mathbf{y}}_j$	vetor que contém algumas amostras consecutivas da j -ésima estimativa, gerado para fins de avaliação, p. 87
$\underline{\mathbf{y}}_j^n(l)$	j -ésimo vetor de saída do bloco de índice l da n -ésima subbanda, p. 134

$\underline{y}_j^n(m)$	i -ésimo sinal de saída na n -ésima subbanda, p. 133
$\hat{\sigma}_v^2$	estimativa da variância do ruído, p. 60
$\hat{\sigma}_e^2$	estimativa da variância do erro, p. 60
$\hat{\mathbf{h}}_k$	vetor-linha que contém os coeficientes do filtro adaptativo na k -ésima iteração, p. 16
$\hat{\mathbf{z}}_k$	vetor que contém os desvios de todos os coeficientes do filtro adaptativo na k -ésima iteração, p. 28
$\hat{d}(k)$	resposta de referência a ser fornecida por um algoritmo adaptativo supervisionado, composta pela resposta ideal degradada por um ruído de medição, p. 17
$\hat{h}_k(i)$	i -ésimo coeficiente do filtro adaptativo na k -ésima iteração, p. 16
$\hat{x}(k)$	versão reconstruída de $x(k)$ após a passagem por bancos de filtros de análise e de síntese, p. 12
$\hat{z}_j(k)$	desvio do j -ésimo elemento de $\hat{\mathbf{h}}_k$, p. 28
$\xi(k)$	medida de esparsidade no instante k , estimada adaptativamente., p. 22
$\xi_{\hat{\mathbf{h}}_k}$	medida de esparsidade de $\hat{\mathbf{h}}_k$, p. 21
c	velocidade do som, p. 103
$d(k)$	resposta ideal a ser fornecida pelo algoritmo adaptativo supervisionado, p. 16
$e(k)$	diferença entre a resposta de referência e a saída de um filtro adaptativo na k -ésima iteração (erro), p. 17
$e_a(k)$	erro <i>a priori</i> , p. 47
$e_p(k)$	erro <i>a posteriori</i> , p. 47
e_{artef}	componente da estimativa associado a artefatos, p. 87
e_{interf}	componente da estimativa associado a interferências, p. 87
$e_{\text{ruído}}$	componente da estimativa associado ao ruído, p. 87
f_s	frequência de amostragem, p. 5

$g_k(i)$	i -ésimo elemento da diagonal principal de $\mathbf{\Gamma}_k$, p. 22
$s_i(k)$	i -ésima fonte, p. 66
s_j^τ	j -ésima fonte atrasada de τ unidades, p. 87
s_{alvo}	componente possivelmente distorcido da estimativa, correspondente à fonte a ela associada, p. 87
$x(k)$	sinal de entrada de um filtro adaptativo, p. 16
$x_i(k)$	sinal obtido pelo i -ésimo sensor, p. 4
$x_q^n(m)$	sinal da q -ésima mistura observado na n -ésima subbanda, p. 132
$y(k)$	saída de um filtro adaptativo no instante k , p. 17
$y_i(k)$	i -ésima estimativa de um método de separação cega de fontes, p. 67
$z_i(k)$	i -ésima componente principal, p. 70
T_{60}	tempo de reverberação, p. 10
Q	matriz que contém os autovetores de \mathbf{R}_x , p. 73
V	matriz de branqueamento ou esferização, p. 73
F	comprimento da janela, p. 4
L	atraso máximo permitido dos componentes da fonte na estimativa correspondente, para efeito de avaliação, p. 87
S	deslocamento entre quadros adjacentes, p. 5
f	frequência discreta, p. 4
p	parâmetro da norma l_p , p. 156
$\det(\mathbf{A})$	determinante da matriz \mathbf{A} , p. 75
$\text{diag}(\mathbf{v})$	matriz diagonal que dispõe os elementos de \mathbf{v} ao longo de sua diagonal principal, p. 80
$\text{diag}^0(\mathbf{X})$	operador que anula todos os elementos que não estão situados na diagonal principal da matriz \mathbf{X} , p. 108
$\text{curt}(s)$	custose da variável aleatória s , p. 7
$\text{round}(x)$	aproximação de x pelo inteiro mais próximo, p. 107

Lista de Abreviaturas

AGL	Algoritmo Generalizado de Lloyd, p. 170
BP	<i>Basis Pursuit</i> , p. 148
BSS	<i>Blind Source Separation</i> , p. 65
CMA	<i>Constant Modulus Algorithm</i> , p. 183
DCGN	Desconvolução Cega por meio de Gradiente Natural, p. 184
DDT	<i>Deterministic Discrete-Time</i> , p. 205
DOA	<i>Difference of Arrival</i> , p. 103
DSP	<i>Digital Signal Processor</i> , p. 11
DUET	<i>Degenerate Unmixing Estimation Technique</i> , p. 168
EAO	Estimativa de Alta Ordem, p. 197
EDAOR	Exploração de Dependências de Alta Ordem entre Raias, p. 99
EDO	Equações Diferenciais Ordinárias, p. 200
EMSE	MSE em excesso, p. 62
EPO	Estimativa de Primeira Ordem, p. 190
FFT	<i>Fast Fourier Transform</i> , p. 152
GBSS-SOS	<i>Generalization of Blind Source Separation methods based on Second-Order Statistics</i> , p. 129
ICA	Independent Component Analysis, p. 70
IMPNLMS	<i>Improved Mu-Law PNLMS</i> , p. 19
IRE	Identificação de Respostas Esparsas, p. 19
ISI	<i>Intersymbol Interference</i> , p. 89

KLT	<i>Karhunen-Loève Transform</i> , p. 9
LMS	<i>Least Mean Square</i> , p. 16
MIMO	<i>Multiple-Input and Multiple-Output</i> , p. 68
MOD	<i>Method of Optimal Directions</i> , p. 172
MP	<i>Matching Pursuit</i> , p. 148
MSE	<i>Mean Square Error</i> , p. 17
MTF	<i>Multiplicative Transfer Function</i> , p. 114
NLMS	<i>Normalized Least Mean Square</i> , p. 16
NP	<i>Nondeterministic Polynomial time</i> , p. 148
NS	<i>negative-side</i> , p. 123
OMP	<i>Orthogonal Matching Pursuit</i> , p. 148
OOMP	<i>Optimized Orthogonal Matching Pursuit</i> , p. 148
PCA	<i>Principal Component Analysis</i> , p. 70
PNLMS	<i>Proportionate Normalized Least-Mean-Square</i> , p. 19
QMF	<i>Quadrature Mirror Filter Bank</i> , p. 12
RI	<i>Resposta Impulsiva</i> , p. 19
SCA	<i>Sparse Component Analysis</i> , p. 80
SDR	<i>Source-to-Distortion Ratio</i> , p. 86
SIR	<i>Source-to-Interference Ratio</i> , p. 86
STFT	<i>Short-Time Fourier Transform</i> , p. 4
SVD	<i>Singular Value Decomposition</i> , p. 171
TDOA	<i>Time Difference of Arrival</i> , p. 103
TITO	<i>Two-Input Two-Output</i> , p. 132
VQ	<i>Vector Quantization</i> , p. 170
WDO _M	<i>W-disjoint orthogonality</i> , p. 6
WT-SF	<i>Estrutura Wavelet com Filtros Esparsos</i> , p. 49

Capítulo 1

Introdução

1.1 Escopo da Tese

Esta tese contemplará tanto algoritmos de filtragem adaptativa supervisionada (no contexto de identificação de sistemas em tempo real) quanto algoritmos cegos (aplicados aos problemas de separação e desconvolução), voltados para o processamento de sinais de voz. A especialização para este tipo de sinal permite-nos levar em consideração algumas características típicas dos registros de fala no projeto e na análise desses algoritmos. Uma das principais ferramentas que utilizaremos, ao longo dos capítulos, para um aprimoramento desses algoritmos é o processamento em subbandas.

A identificação de sistemas permite, entre outras possibilidades, cancelamento de eco em sistemas telefônicos, redução de interferência interssimbólica oriunda de múltiplos percursos em comunicações digitais, modelagem de camadas geológicas em sismologia e projeto de mecanismos de controle de plantas. Em cada uma destas aplicações, é interessante obter uma taxa de convergência alta, para garantir efetividade do filtro adaptativo. Para angariar um aumento na taxa de convergência quando os sinais de entrada são coloridos (caso dos sinais de voz), propomos o recurso à estruturas em subbandas e o emprego de algoritmos proporcionais, mais adequados para a identificação de respostas que sejam possivelmente esparsas, uma característica comum em sistemas telefônicos. Para contornar uma possível degradação de desempenho em regime permanente do processo de identificação, introduzimos uma variação temporal nos fatores de aprendizagem envolvidos na adaptação. Uma análise de transiente de um algoritmo proporcional também é efetuada, o que permite prever a evolução temporal do desempenho do algoritmo proporcional.

Técnicas de separação cega de fontes são úteis em diversas tarefas, como edição de áudio, transcrição musical, realce de gravações e reconhecimento de fala. Mesmo as técnicas situadas no estado da arte se deparam com diversos desafios, como de-

sempenho insuficiente, falta de robustez ao problema da permutação, falta de regras para ajuste de alguns parâmetros (como comprimento do quadro), instabilidade e grande número de iterações. Estes desafios serão o foco de algumas propostas desta tese, seja no que tange a técnicas de separação no domínio do tempo, seja a técnicas no domínio da frequência. No contexto de registros anecoicos, o impacto das características de decomposições atômicas é avaliado no processo de identificação do sistema de mistura, bem como o apelo a formas de onda mais consentâneas com os sinais de voz envolvidos.

Técnicas de desreverberação de sinais de voz podem envolver o aprendizado cego através de algoritmos de desconvolução, cujas equações de atualização, sendo não-lineares, apresentam desafios técnicos para a obtenção de resultados analíticos concernentes a seu desempenho em regime permanente, assim como os relativos ao transiente. Nesta tese, um algoritmo de desconvolução terá algumas de suas propriedades elucidadas por meio de fórmulas analíticas.

Este capítulo introdutório está estruturado da seguinte forma: a Seção 1.2 enfatiza a relevância das técnicas de filtragem adaptativa; na Seção 1.3 as características dos registros de sinais de fala são contempladas, enquanto que uma abordagem concisa das técnicas de processamento multitaxas é apresentada na Seção 1.4. Por fim, a Seção 1.5 disserta a respeito da estrutura da tese.

1.2 Relevância das Técnicas de Filtragem Adaptativa

No contexto de processamento digital de sinais, sendo-nos disponíveis amostras confiáveis de um determinado sinal, importa-nos selecionar o algoritmo mais apropriado aos nossos propósitos de extração da informação desejada. Todavia, não raro os sensores têm acesso a uma versão deturpada do sinal original, de sorte que seu registro encontra-se comprometido por ruídos, interferências, propagação via múltiplos percursos e distorções. A menos que assumamos um controle sobre essas modificações às quais o sinal foi submetido (o que é amiúde impossível), observamos uma distorção de desafiadora reversibilidade em suas propriedades estatísticas, a qual costuma se refletir em alterações na estrutura temporal do mesmo. Neste contexto, cumpre efetuar um processamento prévio do sinal registrado, de modo a garantir-lhe um nível de integridade condizente com as etapas posteriores de processamento de sinais, as quais não necessariamente se restringem a sistemas artificiais, já que podem incluir, por exemplo, a aquisição, mediada por órgãos sensoriais, de informações pelo cérebro.

Dada a significativa imprevisibilidade das transformações efetuadas no sinal ori-

ginal, torna-se impraticável, no objetivo de reduzi-las, o recurso a um algoritmo de processamento dotado de parâmetros fixos que permita a extração da informação desejada. Tal impossibilidade motiva o emprego de técnicas de filtragem adaptativa, cuja flexibilidade as torna capazes de contemplar uma ampla variedade de contextos. Essas técnicas podem ser compreendidas quais refinadores sucessivos dos parâmetros de um sistema (comumente composto por um ou mais filtros), cujos ajustes permitem um contínuo e (espera-se) adequado pré-processamento dos sinais de entrada. Uma estimativa inicial costuma ser iterativamente refinada por meio de uma regra geral:

$$\begin{pmatrix} \text{nova estimativa} \\ \text{dos parâmetros} \end{pmatrix} = \begin{pmatrix} \text{antiga estimativa} \\ \text{dos parâmetros} \end{pmatrix} + \begin{pmatrix} \text{termo de} \\ \text{correção} \end{pmatrix},$$

onde o termo de correção costuma depender dos antigos parâmetros do sistema e do sinal de entrada. Caso o cálculo desse termo de correção exija, adicionalmente, um sinal de referência, cuja discrepância em relação à saída da estrutura adaptativa deve ser por esta reduzida, classificamos o algoritmo como supervisionado. Algoritmos que devem atuar na ausência deste sinal de referência são denominados cegos, autodidatas ou não-supervisionados. Ainda que se possa justificar uma diferenciação entre métodos cegos e não-supervisionados [1], nesta tese esses termos serão empregados indistintamente. À custa de apresentarem em média uma complexidade computacional mais intensa, a maioria dos algoritmos cegos apresentam, comparados às estratégias supervisionadas, uma menor complexidade de informação, porquanto requerem menos conhecimento estatístico acerca do sinal [2].

As diversas técnicas de filtragem adaptativa não-supervisionada são aplicáveis a vários problemas relacionados entre si, como extração cega de sinais, análise de componentes independentes, separação cega de fontes, análise de componentes esparsos, desconvolução cega e equalização cega.

1.3 Características dos Sinais de Voz

Como observação inicial, importa notar que alguns pesquisadores diferenciam o termo fala (relativo à produção sonora de informação inteligível) da voz (característica individual da emissão da fala). Tais distinções, sendo alheias ao escopo da tese, serão doravante ignoradas.

A comunicação por voz se refere a todo o processo de transmissão de informação de um indivíduo (o locutor) para outro (o ouvinte), por meio da fala. Tal pro-

cesso é frequentemente descrito como uma cadeia de eventos [3], na qual o meio de propagação do sinal acústico surge como uma oportunidade para que tecnologias eletrônicas atuem como ferramentas aptas para promover uma melhor qualidade da comunicação [4]. Sendo a fala a modalidade mais natural de comunicação entre seres humanos, há uma grande motivação para o estudo de técnicas de processamento de sinais que reduzam as imperfeições capazes de obstruir a recuperação do fluxo de informações contidas na voz [4]. A confecção de métodos de processamento mais adequados a sinais de voz requer um exame de suas particularidades, que será a seguir efetuado.

1.3.1 Não-estacionariedade

A intensa não-estacionariedade dos registros de sinais de voz se deve a uma plethora de causas promotoras de variabilidade. Por clareza, classificaremos tais causas como intrínsecas à produção da fala (associadas ao locutor) ou como exógenas (ou extrínsecas), quando dele independem. No momento, abordaremos apenas as causas intrínsecas.

Os diversos fonemas apresentam características temporais e espectrais diferentes e mesmo cada som elementar é modificado por seu contexto próximo, como as unidades fonéticas que lhe precedem e sucedem [5], devido à coarticulação entre fonemas consecutivos provocada pela economia de esforço na produção contínua da fala [6]. A variabilidade *intra*-locutor corresponde ao modo da fala, o qual depende tanto de características controláveis e conscientes (como no canto), quanto de características não controláveis (quando o locutor está com frio ou estressado, por exemplo).

A não-estacionariedade dos sinais de voz é de menor intensidade a curto prazo. Esse fato motiva o emprego de segmentação em quadros destes sinais em análises no espaço tempo \times frequência. O comprimento destes quadros deve refletir uma duração média na qual a informação frequencial pode ser considerada inalterada, duração esta normalmente estimada entre 10 e 30 ms [7].

Um mapeamento tempo-frequência popular para este acompanhamento temporal é a STFT (transformada de Fourier de tempo curto, do inglês *Short-Time Fourier Transform*), a qual, a partir do sinal $x_i(k)$ gravado no i -ésimo sensor, gera $\mathcal{X}_i(\mathbf{f}, k)$ (no domínio da transformada) segundo a fórmula:

$$\mathcal{X}_i(\mathbf{f}, k) = \sum_{k'} x_i(k') \mathbf{W}_a(k' - k\mathbf{S}) e^{-j2\pi \frac{\mathbf{f}}{\mathbf{F}} k'}, \quad (1.1)$$

onde \mathbf{W}_a é a janela de análise cujo suporte temporal, centralizado na origem, apresenta comprimento \mathbf{F} , j é a raiz de -1 , \mathbf{f} pertence ao conjunto $\{0, 1, \dots, \mathbf{F} - 1\}$ de

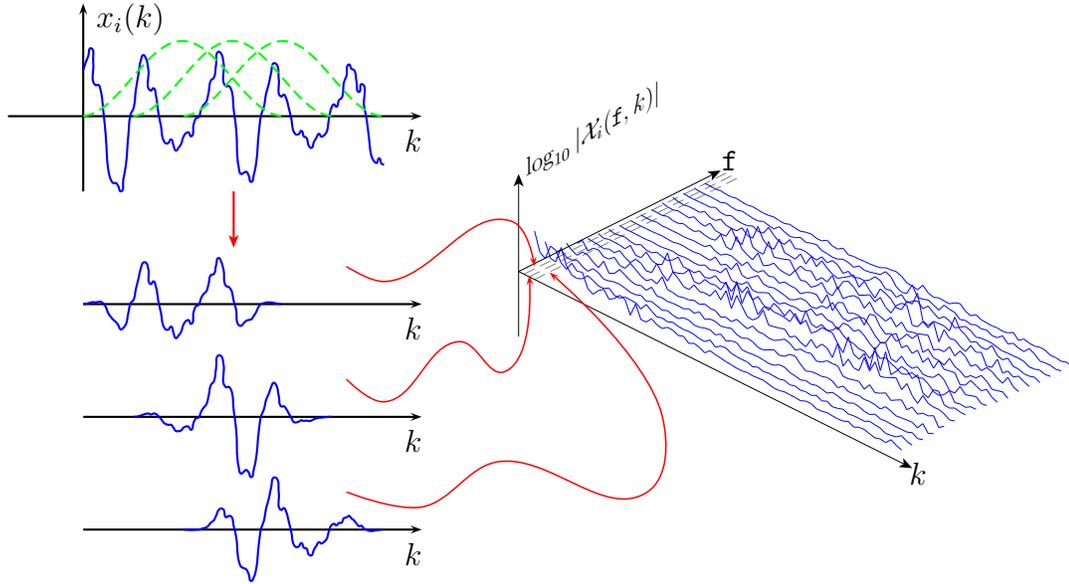


Figura 1.1: Ilustração do processo de cálculo da STFT. Apesar de sua natureza discreta, os sinais são apresentados como contínuos para melhor visualização.

frequências discretas¹, \mathbf{S} representa o deslocamento entre quadros adjacentes e k representa o índice de cada quadro, cujo centro de localização temporal se encontra em torno da amostra $(k + 0,5)\mathbf{S}$. Normalmente a janela W_a apresenta um máximo no entorno da origem, decaindo suavemente para zero nas suas extremidades.

A Fig. 1.1 ilustra o processo de cálculo da STFT, a qual, após segmentação do sinal em quadros (possivelmente sobrepostos no tempo) de duração curta, os suaviza por meio de uma janela que atenua os efeitos das segmentações abruptas no tempo.

1.3.2 Esparsidade

A variabilidade entre locutores refere-se, entre outras, a diferenças de timbre, de idade, de prosódia e de gênero [5]. Experimentos demonstram que tal variabilidade enseja uma significativa disjunção no espaço tempo \times frequência entre diferentes sinais de voz. Num contexto de misturas de sinais nos quais pretendemos recuperar os sinais em separado, tal disjunção pode ser explorada mediante a construção de um mascaramento binário [8],[9] que, aplicado numa mistura de um sensor de referência (em geral arbitrariamente escolhido), permita uma estimativa dos sinais originais numa representação esparsa (como a STFT), na qual espera-se que o montante de sobreposição dos sinais seja reduzido. Desse modo, a estimativa $\mathcal{Y}_j(\mathbf{f}, k)$ (no domínio

¹A frequência contínua associada à \mathbf{f} é dada por $\frac{\mathbf{f} \cdot f_s}{F}$, com f_s denotando a frequência de amostragem.

da STFT) do j -ésimo sinal seria expressa por [9]:

$$\mathcal{Y}_j(\mathbf{f}, k) = \mathcal{M}_j(\mathbf{f}, k)\mathcal{X}_i(\mathbf{f}, k), \quad (1.2)$$

onde $\mathcal{M}_j(\mathbf{f}, k) \in \{0, 1\}$ e j é escolhido arbitrariamente, pois o procedimento de mascaramento é aplicado a apenas uma das misturas. Podemos interpretar $\mathcal{M}_j(\mathbf{f}, k)$ como uma função indicador, a qual associa cada ponto na grade tempo \times frequência a apenas uma das fontes ou como uma partição do suporte da representação, o que é uma estratégia razoável quando este suporte é aproximadamente disjunto entre cada par de fontes [8] (ou seja: $\mathcal{S}_1(\mathbf{f}, k)\mathcal{S}_2(\mathbf{f}, k) \approx 0, \forall \mathbf{f}, k$, no caso de duas fontes $s_1(k)$ e $s_2(k)$).

Para aferir o grau de disjunção no espaço tempo \times frequência entre sinais de voz, [8] cria o conceito de WDO_M (do inglês *W-disjoint orthogonality*), associado à máscara $\mathcal{M}_i(\mathbf{f}, k)$:

$$\text{WDO}_M = \frac{\|\mathcal{M}_i(\mathbf{f}, k)\mathcal{S}_i(\mathbf{f}, k)\|^2 - \|\mathcal{M}_i(\mathbf{f}, k)\overline{\mathcal{Y}}_i(\mathbf{f}, k)\|^2}{\|\mathcal{S}_i(\mathbf{f}, k)\|^2}, \quad (1.3)$$

onde $\overline{\mathcal{Y}}_i(\mathbf{f}, k)$ está associado à transformada de $\overline{y}_i(k) = \sum_{n=1, n \neq i}^N s_i(k)$ e $\mathcal{S}_i(\mathbf{f}, k)$ é o i -ésimo sinal a recuperar. O valor de WDO_M se restringe ao intervalo $[0, 1]$, com o limite superior sendo atingido quando as fontes são perfeitamente disjuntas. A condição $\text{WDO}_M = 0$ implica que o mascaramento anula completamente a energia da fonte de interesse ou que a energia desta (na estimativa) iguala-se à energia da interferência. Diversos testes constataram que o grau de disjunção, no tocante à medida WDO_M para uma máscara ótima, entre sinais de voz (amostrados a 16 kHz) é elevado no domínio da STFT (com janelamento de Hamming e sobreposição de 50%), sendo superior a 93% para 2 sinais de voz e aproximadamente igual a 64% para 10 sinais de voz [8].

1.3.3 Espectro Colorido

O espectro de um sinal de voz apresenta consistentemente um caráter colorido, o que pode ser inferido a partir da modelagem física do fenômeno de geração da fala. Na produção de uma vogal, por exemplo, a passagem do fluxo de ar nas cordas vocais as faz vibrar periodicamente, numa frequência que fornece o *pitch* (ou frequência fundamental) da vogal emitida. Uma modelagem para a produção de voz, plausível na emissão de vogais em regime permanente, consiste em entendê-la como oriunda da saída de um sistema linear invariante no tempo (correspondente ao trato vocal), cuja entrada é um trem de impulsos. Esta entrada, de característica harmônica, tem sua estrutura frequencial conformada (ou colorida) pelas características do trato vocal.

É claro que o fenômeno da produção da fala é complexo e as fontes (entradas do sistema supracitado) que a originam podem não aderir ao modelo acima, pois, além de periódicas, podem ser ruidosas ou impulsivas [10]. Além disso, não aludimos ao fato de que o sistema correspondente ao trato vocal é influenciado, por exemplo, pela posição da língua e dos lábios. Porém, o espectro colorido do sinal de voz é uma propriedade ubíqua, não se restringindo às vogais. Como outro exemplo, o espectro dos sons nasais é predominantemente influenciado pelas características da cavidade nasal, cujo elevado volume gera formantes de baixa frequência [10]. Essas considerações permitem-nos concluir que um modelo de sinal branco para sinais de voz é inadequado.

1.3.4 Não-gaussianidade

Um exame da distribuição unidimensional das amostras de um sinal de voz (descartando as dependências entre amostras adjacentes) permite-nos concluir que tal distribuição difere de uma gaussiana. Uma popular medida estatística para a avaliação da gaussianidade de um sinal é a curtose, a qual pode ser compreendida tanto como uma versão normalizada do momento de quarta ordem quanto o mais simples indicativo de não-gaussianidade de uma variável aleatória [11]. A curtose da variável aleatória s é definida por:

$$\text{curt}(s) = \frac{\text{E}[|s - \text{E}[s]|^4]}{\sigma^4} - 3, \quad (1.4)$$

onde σ é o desvio-padrão de s . Variáveis gaussianas apresentam $\frac{\text{E}[|s - \text{E}[s]|^4]}{\sigma^4} = 3$, tendo portanto curtose nula. Em geral, variáveis aleatórias não-gaussianas apresentam curtose não-nula. As variáveis aleatórias que apresentam curtose negativa, nula e positiva são regidas por distribuições chamadas de, respectivamente, subgaussiana, mesocurtótica e supergaussiana. Na literatura de estatística, outros nomes para distribuições subgaussianas e supergaussianas são, respectivamente, platicurtótica e leptocurtótica.

Distribuições supergaussianas costumam apresentar um pico mais concentrado do que suas congêneres gaussianas, assim como uma cauda mais pronunciada. Análises de séries econômicas ressaltam que o retorno de ativos financeiros costuma se adequar a um modelo supergaussiano, pois costuma variar aleatoriamente, na maior parte do tempo, em torno de sua tendência (o que gera um pico em torno da origem mais pronunciado do que o de uma variável aleatória gaussiana de idêntica variância), porém, quando há um desvio deste padrão aleatório (seja na direção de aumento, seja na de decréscimo), a variação costuma ser mais rápida do que o esperado (fenômeno responsável pela cauda mais pronunciada da distribuição) [12]. Já

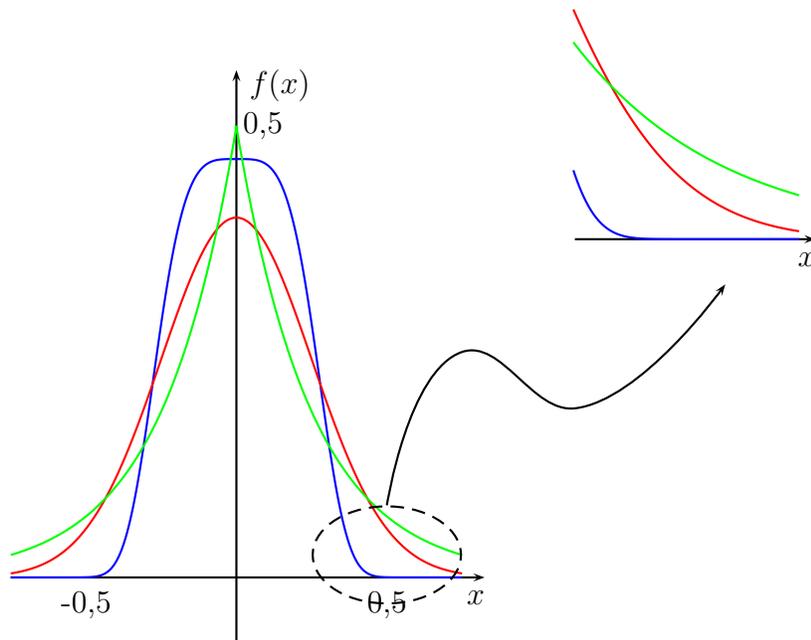


Figura 1.2: Exemplos de distribuições subgaussiana (em azul), gaussiana (em vermelho) e supergaussiana (em verde). As caudas das distribuições foram ampliadas para que suas magnitudes relativas possam ser observadas.

as distribuições subgaussianas (como por exemplo a uniforme) tendem a apresentar modos mais “espalhados” do que as gaussianas, com caudas tendendo mais rapidamente a zero. A Fig. 1.2 permite uma visualização destas propriedades.

A Fig. 1.3 apresenta as distribuições de probabilidade (estimadas por meio de *kernels* gaussianos) das amostras no domínio do tempo de alguns registros de sinais de voz, normalizados para variância unitária e comparados com uma distribuição gaussiana de idêntica variância. Os gráficos ilustram uma consistente propriedade das amostras de sinais de voz: a supergaussianidade, explicada, dentre outros fatores, pelos trechos de silêncio que aumentam a probabilidade de as amostras apresentarem magnitudes próximas de zero. Mesmo em contextos reverberantes, que têm como efeito a redução dos trechos de silêncio e a gaussianização dos sinais (relacionada ao fato de a soma de muitas variáveis aleatórias tender a uma distribuição gaussiana), a supergaussianidade ainda é bastante nítida (vide Fig. 1.3.(b) e 1.3.(d)). Tal supergaussianidade também se manifesta no domínio da STFT [13]. O artigo [14] fornece evidências de que em trechos em que a voz está ativa a distribuição marginal de suas amostras segue uma distribuição laplaciana e que, sendo contemplados os trechos de silêncio, a caracterização de primeira ordem da distribuição experimental se aproxima de uma distribuição gamma, a qual é capaz de aproximar melhor a distribuição de sinais de voz na região próxima à origem. Atentando para uma descrição estatística de ordem superior, [14] também argumenta que os coeficien-

tes da KLT (transformada de Karhunen-Loève) de um sinal de voz constituem um vetor aleatório com elementos decorrelacionados distribuídos segundo uma distribuição laplaciana, sendo cada um deles gerado por meio de parâmetros distintos e lentamente variantes no tempo desta distribuição. Tais características distanciam as amostras de um sinal de voz da gaussianidade.

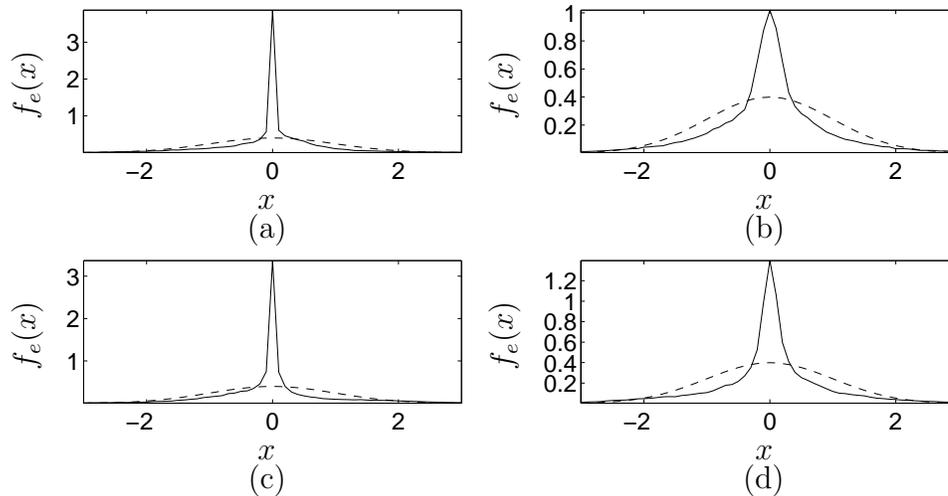


Figura 1.3: Linha sólida: distribuições de probabilidade de sinais de voz com 10s de duração (extraídos de <http://www.irisa.fr/metiss/SASSECO7/?show=test>), amostrados em 16 kHz e normalizados para variância unitária. Em linha tracejada, uma distribuição gaussiana de mesma variância. (a) fala em japonês emitida por um locutor masculino, numa câmara anecoica; (b) fala em japonês emitida por um locutor masculino, numa sala com tempo de reverberação de aproximadamente 250ms; (c) fala em inglês emitida por uma locutora feminina, numa câmara anecoica; (d) fala em inglês emitida por uma locutora feminina, numa sala com tempo de reverberação de aproximadamente 250ms.

1.3.5 Variabilidades Extrínsecas dos Registros de Sinais de Voz: Funções de Transferência Acústicas

Até o momento, foram citadas apenas as variabilidades intrínsecas dos sinais de voz. Entre as exógenas, verificamos que a mera escolha do dispositivo de gravação utilizado é causa de modificações no sinal registrado e que mesmo o ambiente costuma contribuir com perturbações adicionais, como ruídos, ecos e interferências. Tais variabilidades extrínsecas não são necessariamente indesejáveis pois, por exemplo, a diversidade espacial pode contribuir para uma audição mais realista ou mesmo para propósitos de localização e conformação de feixe.

Técnicas que atenuem as variabilidades extrínsecas indesejáveis (perturbações) da fala são importantes, por exemplo, para que um ouvinte experimente uma melhora na qualidade da conversação, ou para que um sistema computacional possa extrair

informações da mesma (como a frequência fundamental num determinado trecho). Mesmo perturbações localizadas em determinadas frequências afetam significativamente sistemas automáticos de reconhecimento de voz, por exemplo, tornando-os sensíveis a ruídos, reverberação e interferências [15].

Uma importante influência no registro de sinais sonoros reside nas respostas impulsivas associadas às propriedades acústicas de cada ambiente. Após a excitação por um som impulsivo, verificam-se as seguintes etapas:

Primeiramente, há o decaimento inicial do som nos primeiros 10 ou 20 ms de queda de energia após a explosão (*burst*) inicial. O tempo de reverberação baseado nesta região é chamado de tempo de decaimento inicial. Após o primeiro impulso há uma sequência de impulsos, oriundos por reflexões das superfícies próximas ao receptor e à fonte. Então segue-se um complicado trem de impulsos, oriundos das primeiras reflexões das superfícies do ambiente. Nesta região os defeitos acústicos presentes no ambiente começam a aparecer. Reflexões muito atrasadas aparecem quais pulsos isolados. Ecos flutuantes aparecem como reflexões repetidas que não se esvanecem tão rapidamente quanto a cauda normal de reverberação [16].

Um parâmetro importante de uma função de transferência acústica é o tempo de reverberação (mais conhecido pelo termo T_{60}), o qual, grosso modo, corresponde ao tempo necessário para que, finda a excitação sonora, o nível de pressão sonora decaia 60 dB. O tempo de reverberação, correlacionado com a inteligibilidade da fala, é altamente relacionado à banda mais baixa de frequência da resposta ao impulso [17]. Embora, em termos perceptuais, o tempo inicial da razão de decaimento do som reverberante tenda a ser mais importante do que o tempo total de reverberação (já que um rápido decaimento inicial é interpretado pelo ouvido como associado a um tempo curto de reverberação), em aplicações como o cancelamento de eco deve-se atenuar também as reflexões mais tardias, pois do contrário o eco se manterá nítido [16]. Para maior clareza da audição, o nível do som inicial somado ao componente direto deve superar o nível do som reverberante [16].

Além da geometria do ambiente, as características dos materiais nele presentes (como as referentes à difusão, reflexão ou absorção) são muito importantes para a caracterização da resposta impulsiva. Um exemplo de resposta impulsiva (medida) de um ambiente é apresentado na Fig. 1.4. Verificamos neste exemplo que a energia da função de transferência é concentrada na sua parte inicial, o que dá origem a uma esparsidade que será explorada nesta tese por meio de algoritmos proporcionais de adaptação.

Experimentos em ambientes de diferentes dimensões demonstraram que o tempo

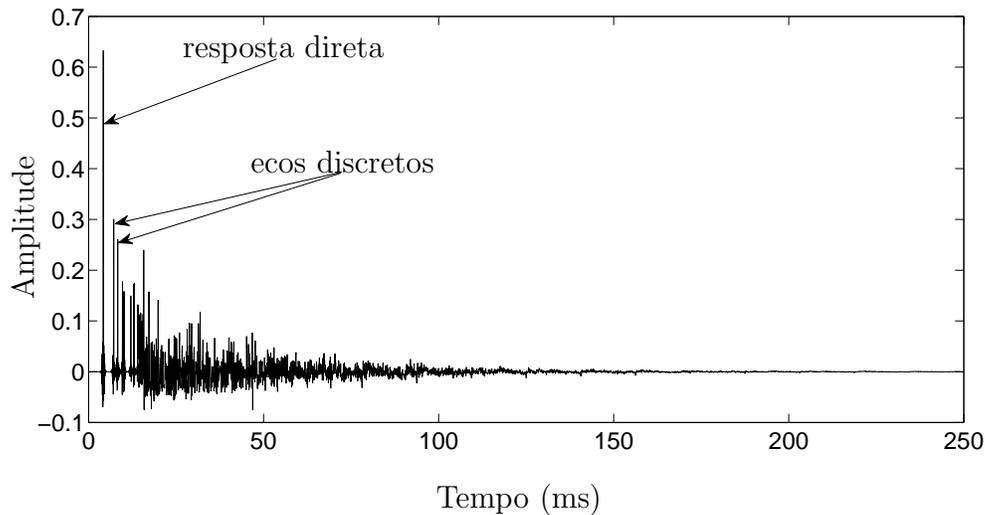


Figura 1.4: Função de transferência medida num ambiente, disponibilizada em <http://www.irisa.fr/metiss/SASSECO7>.

médio de reverberação se reduz à medida que a frequência dos sinais se eleva; a principal causa deste efeito reside nas perdas de propagação provocadas pela viscosidade do ar em frequências superiores a 1 kHz [18]. Relacionado a este fato, as respostas impulsivas costumam apresentar uma característica passa-baixas porque as ondas acústicas de alta frequência costumam ser absorvidas mais fortemente pelas paredes e por outras superfícies [19].

Uma forma versátil de se estudar o efeito das funções de transferência acústica consiste em efetuar simulações da acústica de um determinado ambiente, as quais estão, dia a dia, se tornando mais sofisticadas, obtendo resultados razoavelmente próximos aos de uma sala real. Assim, é possível a modificação de diversos parâmetros de um ambiente (como as dimensões e o tempo de reverberação), prescindindo de custosas intervenções físicas [20].

Em todas as simulações de funções de transferências acústicas desta tese, foi empregado o simulador proposto em [21], [22], cuja implementação em MATLAB foi disponibilizada em <http://www.eric-lehmann.com>. Este simulador emprega uma versão aprimorada da técnica imagem-fonte, muito popular na síntese de respostas impulsivas.

1.4 Bancos de Filtros e Sistemas Multitaxas

O termo *filtragem multitaxas* refere-se a situações nas quais o sistema de filtragem envolve mais do que uma taxa de amostragem [23]. Sistemas multitaxas são reconhecidamente uma das mais poderosas abordagens para a exploração de alguns benefícios angariados por DSPs e circuitos programáveis [24]. O processamento

multitaxas pode ser empregado quando uma filtragem mais sofisticada é necessária, como por exemplo na redução do número de operações por amostra de entrada, sendo amplamente utilizado em comunicações, remoção de ruído, processamento de imagens, áudio digital e multimídia, entre outros [24]. Sua eficiência computacional fundamenta-se em sua habilidade de emprego simultâneo de diferentes taxas de amostragem em partes distintas do sistema.

Já na década de 1970, a estrutura QMF (do inglês *quadrature mirror filter bank*) de dois canais consistiu numa quebra de paradigma na tecnologia de compressão de sinais de voz, efetuada por meio da quantização dos sinais nas subbandas [25]. Desde então, o interesse pelo processamento multitaxas tem se mantido constante.

Nesta tese investigaremos algumas aplicações de estruturas em subbandas para filtragem adaptativa, tendo em vista, por exemplo, um incremento na taxa de convergência destes algoritmos, passível de ser obtido devido ao caráter colorido dos sinais de voz. Conquanto não pretendamos detalhar toda a teoria de sistemas multitaxas, iremos apresentar concisamente os principais conceitos de bancos de filtros que serão utilizados.

1.4.1 Bancos de Filtros

Bancos de filtros são uma coleção de filtros digitais, cuja estrutura padrão encontra-se na Fig. 1.5. O conjunto de filtros $\{\underline{H}_0(z), \underline{H}_1(z), \dots, \underline{H}_{M-1}(z)\}$ corresponde ao denominado “banco de filtros de análise” o qual, aplicado em um sinal de entrada $x(k)$, gera versões do mesmo nas M subbandas, nas quais costumam atuar decimadores que tornam a taxa de amostragem de cada um destes sinais consentânea com sua largura de banda. Após o processo de decimação, tem lugar a etapa de processamento (como por exemplo a codificação), em geral mais flexível no domínio das subbandas. Após este processamento, a recuperação da taxa original do sinal é efetuada pelos expansores, os quais precedem os filtros de síntese, pertencentes ao conjunto $\{\underline{F}_0(z), \underline{F}_1(z), \dots, \underline{F}_{M-1}(z)\}$. Na ausência da etapa de processamento, em geral espera-se ser a versão reconstruída $\hat{x}(k)$ uma cópia fiel de $x(k)$, a menos de fatores de atraso e escalamento.

Uma das palavras-chave em processamento multitaxas é o *aliasing*, que advém do fato de que os filtros de análise não são idealmente limitados em banda. Não raro o efeito do *aliasing* é perceptível, cabendo-nos portanto atenuá-lo o quanto for possível [23].

Apesar de exigirem maior atenção no tocante a distorções de amplitude, de fase e de *aliasing*, os bancos de filtros mais interessantes para a redução de custo computacional são os denominados de maximamente (ou criticamente) decimados, nos quais os fatores de decimação devem satisfazer o critério $\sum_{m=0}^{M-1} \frac{1}{L_m} = 1$.

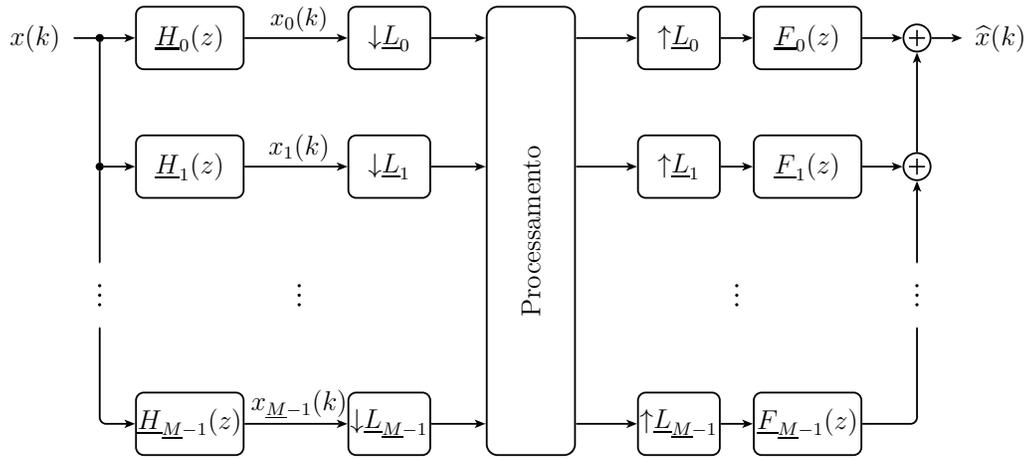


Figura 1.5: Bancos de filtros digitais.

Sendo constante a largura de faixa dos filtros de análise e de síntese, o banco de filtros é dito uniforme; caso contrário, é não-uniforme. A depender das características do sinal de entrada (em especial se este é branco ou colorido) ou das funções de transferência envolvidas ou mesmo de um processo fundamentado em considerações perceptuais, bancos de filtros não-uniformes apresentam uma flexibilidade adicional que pode se revelar útil.

1.5 Estrutura da Tese

Esta tese tem como foco o emprego de técnicas de filtragem adaptativa (em versões supervisionada, cega ou semi-cega) aplicadas a sinais de voz. A especificação do tipo de sinal a processar permite a inserção de características ou modificações nas diferentes técnicas de modo a contemplar suas idiossincrasias. Embora as técnicas adaptadas para sinais de voz não sejam necessariamente ótimas para outros tipos de sinais, sua especialização para um sinal particular não raro engendra um desempenho superior quando aplicadas a sinais cujas propriedades estatísticas sejam consentâneas com aquelas para as quais foram projetadas. A seguir, descrevemos sucintamente os conteúdos dos capítulos centrais da tese.

De caráter basicamente introdutório, o Cap. 2 apresenta, no contexto de filtragem adaptativa supervisionada, os algoritmos LMS, NLMS e a família de algoritmos proporcionais PNLMS, apresentando uma interpretação geométrica dos dois últimos. Os algoritmos proporcionais são úteis para configurações nas quais as funções de transferência acústica são esparsas, o que é comum em aplicações como cancelamento de eco.

As contribuições relativas aos algoritmos proporcionais são efetuadas no Cap. 3.

Esse capítulo, enfocando o algoritmo IMPNLMS, efetua uma análise de transiente do algoritmo, estima teoricamente o erro quadrático médio em regime permanente e explora questões de estabilidade. Tais análises são estendidas para uma versão em subbandas deste algoritmo, capaz de explorar a característica colorida dos sinais de voz. Algumas dessas análises são facilmente aplicáveis a diversos algoritmos da família PNLMS. Por fim, encontra-se um razoável compromisso entre taxa de convergência e desempenho em regime permanente do algoritmo IMPNLMS (ou de sua versão em subbandas) com a utilização de fatores de aprendizagem variantes no tempo.

O Cap. 4 introduz o problema da separação cega de fontes, bem como princípios comumente utilizados para a resolução deste problema, como o da maximização da independência ou o da esparsidade. No tocante às técnicas de separação cega de fontes que atuam no domínio da frequência, o Cap. 5 apresenta um resumo do estado da arte no tema e apresenta inovações desta tese, como um novo procedimento de inicialização, a inserção de coeficientes adaptativos interfrequenciais e o emprego da informação lateral obtida via estimativa cega do T_{60} .

No Cap. 6, estruturas em subbandas são empregadas para melhorar o desempenho de um algoritmo de separação cega de fontes que atua no domínio do tempo. A inserção de fatores de aprendizagem variantes no tempo e de pós-processamento se mostra adequada à aceleração da taxa de convergência.

Por sua vez, o Cap. 7 utiliza decomposições com dicionários redundantes para identificação do sistema de mistura em contextos instantâneos e anecoicos. Os atrasos e atenuações do sistema de misturas são identificados por meio de parâmetros das decomposições das misturas, sendo avaliadas as características dos componentes que resultam em estimativas mais confiáveis. O emprego de dicionários adaptados a locutores (ou idiomas) específicos é avaliado como promissor para uma maior rapidez ou acurácia do processo de identificação.

O Cap. 8 empreende análises de um algoritmo de desconvolução cega, capaz de separar sinais de voz. São estudadas, além do comportamento do transiente, características de equivariância e estimados os coeficientes em regime permanente. Neste último caso, uma técnica de seleção das soluções corretas dos coeficientes finais é apresentada.

O Cap. 9 apresenta as conclusões da tese. A Fig. 1.6 resume as características principais de seus capítulos centrais. Entre outras classificações, esta figura ressalta a distinção entre capítulos que apelam a estruturas redundantes daqueles que não as empregam. O conceito utilizado de estrutura redundante é amplo. No contexto de transformações, como a STFT, redundância significa que o número de coeficientes da transformação, quando esta apela a sobreposição entre quadros adjacentes, é superior ao número de amostras do sinal original. Bancos de filtros que não apre-

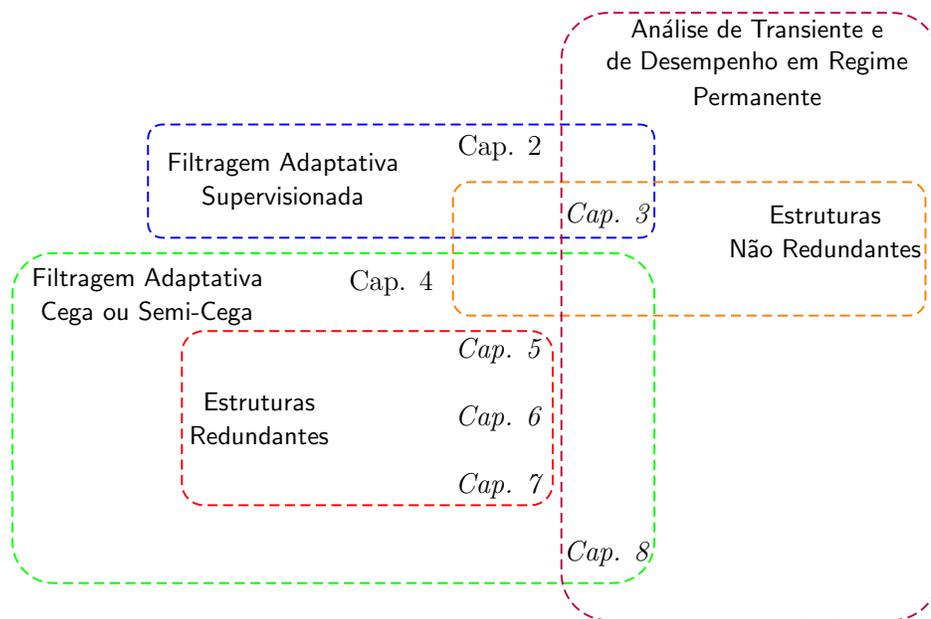


Figura 1.6: Principais características dos capítulos centrais da tese (excluindo-se os capítulos de introdução e de conclusões). Os capítulos que contém contribuições estão assinalados em itálico.

sentem decimação crítica são também classificados como redundantes, assim como as estruturas adaptativas (particularmente as cegas, nesta tese) acopladas ao processamento multitaxas com decimação não crítica. No que tange às decomposições atômicas, objeto de estudo do Cap. 7, o fato de o dicionário conter mais elementos do que os necessários para gerar o espaço associado ao número de amostras destes elementos também nos permite classificá-las como redundantes.

A Tab. 1.1 apresenta as propriedades dos sinais de voz levadas em consideração nos capítulos que contém inovações.

Propriedade	Cap. 3	Cap. 5	Cap. 6	Cap. 7	Cap. 8
Não-estacionariedade			×	×	
Esparsidade		×			
Espectro colorido	×	×	×	×	
Não-gaussianidade		×		×	×

Tabela 1.1: Características dos sinais de voz empregados em cada capítulo de inovação da tese.

Capítulo 2

Algoritmos Proporcionais de Filtragem Adaptativa

2.1 Introdução

A identificação de sistemas é uma das mais importantes aplicações de filtragem adaptativa supervisionada, na qual um filtro adaptativo gradualmente se desloca, a partir de um estado desadaptado arbitrário, em direção a um determinado sistema de referência [26], o qual deseja emular. Esta identificação permite que se estime o canal de um sistema de comunicação ou a função de transferência acústica entre um sinal indesejado e o sensor, o que possibilita, entre outras opções, o cancelamento de eco e de ruído. Outra aplicação consiste na estimativa dos parâmetros de uma predição linear de um determinado sinal.

Todos os sistemas de identificação contemplados nesta tese admitem uma estrutura de adaptação em comum. Se o filtro adaptativo apresenta L parâmetros adaptativos, podemos definir o vetor de entrada \mathbf{x}_k em função do sinal de entrada $x(k)$:

$$\mathbf{x}_k = \begin{bmatrix} x(k) & x(k-1) & \cdots & x(k-L+1) \end{bmatrix}^T. \quad (2.1)$$

Na k -ésima iteração, o vetor-linha $\hat{\mathbf{h}}_k$ que contém os coeficientes do filtro adaptativo é expresso por:

$$\hat{\mathbf{h}}_k = \begin{bmatrix} \hat{h}_k(0) & \hat{h}_k(1) & \cdots & \hat{h}_k(L-1) \end{bmatrix}. \quad (2.2)$$

Suporemos, nesta tese, que o comprimento M da resposta impulsiva \mathbf{h} (definida de forma semelhante a $\hat{\mathbf{h}}_k$) a identificar é igual ou menor do que L .

A cada instante k , existe uma resposta ideal $d(k)$ a ser fornecida pelo filtro adaptativo, a qual é dada por:

$$d(k) = \mathbf{h}\mathbf{x}_k, \quad (2.3)$$

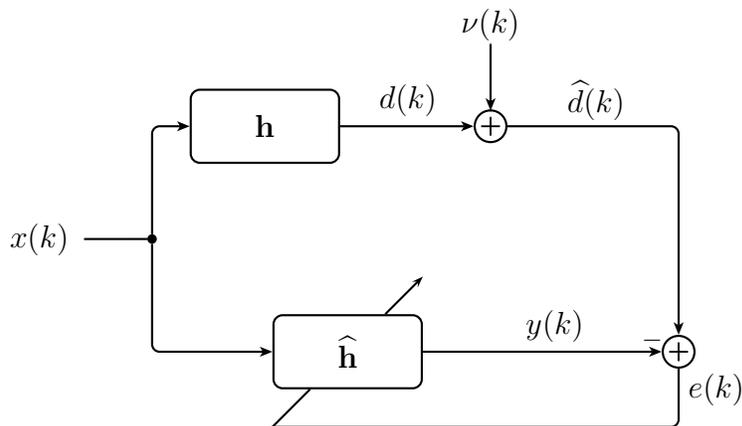


Figura 2.1: Estrutura de um algoritmo adaptativo de identificação supervisionado.

onde a ausência de transposição no produto interno se dá porque \mathbf{h} é um vetor linha.

Para o aprendizado supervisionado, modelaremos a incerteza do processo de medição da resposta ideal pela adição de um ruído (de medição) branco e gaussiano $\nu(k)$, o qual independe do sinal de entrada e apresenta variância σ_ν^2 , cuja perturbação é responsável por gerar uma variância não-nula nas estimativas dos parâmetros do sistema desconhecido [27]. A imprevisibilidade deste ruído impede sua remoção; portanto, o sistema adaptativo tem acesso a $\hat{d}(k) = d(k) + \nu(k)$. Cumpra a ele alterar seus parâmetros de forma a minimizar alguma função custo dependente do erro medido $e(k)$:

$$e(k) = \hat{d}(k) - y(k) = [\mathbf{h} - \hat{\mathbf{h}}_k] \mathbf{x}_k + \nu(k), \quad (2.4)$$

onde $y(k) = \hat{\mathbf{h}}_k \mathbf{x}_k$ é a saída do filtro adaptativo no instante k .

O processo de minimização da função custo, em conjunto com a escolha da mesma, especifica as características de aprendizado do sistema adaptativo. A Fig. 2.1 ilustra a estrutura de um típico algoritmo adaptativo de identificação supervisionado. É desejado que o sinal de entrada $\mathbf{x}(k)$ seja um sinal de banda larga, de modo a garantir a convergência do filtro adaptativo para um bom modelo do sistema desconhecido a identificar [27].

Entre as diversas funções custo encontradas na literatura, a mais popular é o erro quadrático instantâneo, o qual pode ser interpretado como uma estimativa estocástica do erro quadrático médio (MSE, do inglês *Mean Square Error*). A proposta de utilizar essa estimativa instantânea ocorreu a Bernard Widrow quando explicava a Marcian E. Hoff sua técnica de filtragem adaptativa fundamentada em minimização do MSE (que Widrow chamava de *performance feedback*) e que se popularizou

pela sigla LMS (*Least Mean Square*), cunhada por James S. Koford [28].

O algoritmo LMS é o mais simples algoritmo de aprendizagem. A ele cabe minimizar $e^2(k)$. Para utilizar o método de otimização orientado pela direção do gradiente descendente, cumpre calcular o gradiente em função de $\hat{\mathbf{h}}_k$:

$$\frac{\partial e^2(k)}{\partial \hat{\mathbf{h}}_k} = 2e(k) \frac{\partial e(k)}{\partial \hat{\mathbf{h}}_k} = -2e(k) \mathbf{x}_k^T. \quad (2.5)$$

Então podemos atualizar o filtro $\hat{\mathbf{h}}_k$ por meio de:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k - \beta' \frac{\partial e^2(k)}{\partial \hat{\mathbf{h}}_k} = \hat{\mathbf{h}}_k + \underbrace{2\beta'}_{\beta} e(k) \mathbf{x}_k^T = \hat{\mathbf{h}}_k + \beta e(k) \mathbf{x}_k^T, \quad (2.6)$$

onde β é o fator (ou passo) de aprendizagem do algoritmo, e a condição mais comum $0 < \beta \ll 1$ implica que a alteração no filtro adaptativo pode ser expressa como uma fração do gradiente, projetada sobre o hiperplano dos coeficientes a otimizar. O passo de aprendizagem é fixo ao longo das iterações, sendo idêntico para todos os elementos de $\hat{\mathbf{h}}_k$.

Embora comumente a atualização expressa na Eq. (2.6) seja derivada de um gradiente estocástico, o algoritmo LMS também pode ser articulado num contexto determinístico, no qual apresenta otimalidade no tocante à razão entre a energia dos erros de estimação e a energia do erro de medição ou pode ser interpretado como uma otimização do tipo min-max, relacionada à teoria de estimação H_∞ [29]. Essa vertente teórica não será por nós explorada.

Para filtros adaptativos longos (como os empregados em cancelamento de eco), o algoritmo LMS apresenta uma convergência deveras lenta, o que comumente o torna preterido por sua versão normalizada (NLMS, do inglês *Normalized Least Mean Square*). O NLMS acelera o processo de convergência variando o fator de aprendizagem ao longo das iterações, evitando, por meio de uma normalização, que eventualmente vetores \mathbf{x}_k com módulo elevado amplifiquem o ruído no processo adaptativo. Diversos princípios distintos podem ser empregados para a derivação da equação de atualização do NLMS; uma delas é a minimização do ganho máximo de energia dos distúrbios com relação aos erros filtrados [30]. Outro problema de otimização associado a esse algoritmo será visto mais à frente neste capítulo.

A seguir, apresentamos a equação de atualização do NLMS, na qual costuma-se inserir um fator β responsável por controlar o desajuste (razão entre o MSE em excesso e o MSE mínimo) [27]:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta \frac{e(k) \mathbf{x}_k^T}{\mathbf{x}_k^T \mathbf{x}_k + \delta}, \quad (2.7)$$

onde δ é um número ligeiramente maior que zero, cuja introdução evita uma possível divisão por um valor muito pequeno. Esse parâmetro de regularização pode se tornar dependente da iteração de modo a acelerar a taxa de convergência [31],[32].

Neste capítulo, pretendemos apresentar uma introdução à família de algoritmos proporcionais. O próximo capítulo contempla as inovações da tese concernentes a estes algoritmos.

2.2 Algoritmos de Filtragem Adaptativa Proporcional

Esta tese enfoca uma classe do conjunto dos algoritmos supervisionados, a dos que almejam identificar uma resposta ao impulso esparsa, comum em diversos processos acústicos, químicos e sísmicos, bem como em canais *wireless* [33],[34]. Neste contexto, uma sequência é considerada esparsa caso a maior parte de seus elementos esteja próxima de zero, o que implica um conceito de esparsidade mais fraco que o usualmente empregado em análise numérica [35].

Não há nenhuma dificuldade matemática na aplicação dos algoritmos supervisionados de identificação para as configurações nas quais parte significativa da resposta ao impulso a identificar apresenta magnitudes próximas de zero. Porém alguns problemas podem surgir, sendo o principal deles uma convergência lenta. Neste capítulo, estamos interessados em explorar a interpretação geométrica da família de algoritmos PNLMS (do inglês *Proportionate Normalized Least-Mean-Square*, os quais seguem o paradigma inaugurado em [36]) proposta para acelerar a convergência da identificação e apresentar um dos algoritmos proporcionais (denominado IMPNLMS [37], do inglês *Improved Mu-law PNLMS*) mais bem-sucedidos. O algoritmo IMPNLMS será alvo de análises teóricas no Cap. 3.

Basicamente, os algoritmos proporcionais inserem explicitamente a esparsidade da RI (resposta impulsiva) no algoritmo adaptativo, promovendo um aumento na taxa de convergência do processo de identificação, o que é particularmente útil quando a RI apresenta comprimento elevado. O algoritmo NLMS é o algoritmo-padrão com o qual a maior parte das propostas para aceleração de convergência da identificação de respostas esparsas (IRE) é comparada. Elas almejam contornar a lentidão na convergência do NLMS em configurações esparsas mediante a distribuição dos fatores de aprendizagem ao longo dos coeficientes de $\hat{\mathbf{h}}_k$, com passos de aprendizagem mais elevados para coeficientes de maior magnitude. Podemos encarar esta estratégia como decorrente de uma cooperação estabelecida por um administrador central de recursos, o qual concede atualizações mais pronunciadas para os coeficientes de maior magnitude. Importa notar a existência de outras famílias de

métodos que contemplam o problema de IRE, as quais, por recorrerem a outras estratégias, encontram-se fora do escopo desta tese; por essa razão, a despeito da imprecisão conceitual, doravante as expressões “família de algoritmos PNLMS” e “algoritmos de IRE” serão empregadas indistintamente.

A principal diferença entre os algoritmos de IRE reside na forma como esta distribuição é efetuada. Assim, podemos entender que o algoritmo NLMS apresenta melhor comportamento de convergência do que o LMS porque recorre a um fator de aprendizagem que varia ao longo das iterações e que os algoritmos de IRE almejam incrementar a convergência em relação ao NLMS alterando o fator de aprendizagem também ao longo de um novo eixo: o dos coeficientes do filtro adaptativo, já que supõe-se que estes apresentam grandes diferenças de magnitude. Na Seção 2.4 compararemos o algoritmo NLMS e a família de algoritmos PNLMS sob uma perspectiva geométrica.

2.3 Revisão Bibliográfica

Para situar melhor o estado da arte dos algoritmos de IRE, apresentamos resumidamente as principais contribuições da literatura pertinente.

É comum o uso do NLMS para a adaptação de um filtro com um número elevado de coeficientes [36]. Uma das primeiras alternativas propostas consistiu no emprego de filtros com menor número de coeficientes adaptativos do que os necessários para obter todas as RIs de um dado comprimento, atualizando subconjuntos de coeficientes que correspondem às regiões dispersivas [38], [39]. Uma das grandes vantagens deste tipo de estratégia encontra-se na substancial redução do custo computacional [36].

Outra possibilidade para acelerar a convergência num contexto de RI esparsa (quando o NLMS se comporta de forma sub-ótima [33],[40]) consiste numa distribuição do fator de aprendizagem β . Originalmente, ele é idêntico para todos os coeficientes do filtro adaptativo. Como já explicado, uma estratégia que se revelou bem-sucedida consiste em incrementar o passo de aprendizagem relativo aos coeficientes de maior magnitude, concomitantemente à sua redução para os coeficientes próximos de zero, o que redundava em significativa aceleração da taxa de convergência para respostas impulsivas esparsas.

A primeira proposta nesse sentido foi denominada PNLMS [36], da qual todos os algoritmos estudados neste capítulo são derivados. O aumento do custo computacional deste algoritmo (comparativamente ao NLMS) pode ser contornado, por exemplo, mediante o emprego de técnicas de *Set-Membership* [41].

Almejando distribuir de forma desigual a “energia” de atualização disponível, o algoritmo, em resumo, aumenta o fator de aprendizagem correspondente a $\hat{h}_k(n)$ de

forma proporcional ao seu módulo. Os elementos de $\hat{\mathbf{h}}_k$ que forem mais distantes de zero terão uma atualização mais pronunciada do que aqueles de menor magnitude. O algoritmo também efetua uma regularização de pequenos sinais [36].

O algoritmo PNLMS apresenta uma vantagem em relação ao NLMS em configurações esparsas: uma convergência inicial rápida [36]. Porém, a taxa de convergência reduz-se dramaticamente após um período inicial e se exige grande esparsidade da RI a identificar; o algoritmo é mais lento que o NLMS em configurações pouco esparsas ou não-esparsas [37]. Por essa razão, o algoritmo PNLMS++ [42] sugere uma alternância entre os algoritmos NLMS e PNLMS, de forma a reduzir essa degradação em configurações pouco esparsas.

Em [43], propõe-se uma aproximação do fator de aprendizagem ótimo, de modo a contornar esta desvantagem do algoritmo PNLMS. Uma versão ligeiramente modificada deste algoritmo é apresentada em [44], cuja contribuição principal consiste em não utilizar a magnitude dos coeficientes diretamente, mas o logaritmo desta magnitude, seja por motivos teóricos, seja por razões experimentais. A função logarítmica é aproximada por uma função definida por partes, com cada parte sendo linear, de modo a reduzir o custo computacional, similarmente à lei μ utilizada para compressão não-uniforme em aplicações de telecomunicações [45].

Todas as abordagens citadas apresentam o inconveniente de exigirem, para que se verifique o aumento da taxa de convergência, a esparsidade da resposta impulsiva a identificar, o que nem sempre ocorre. O artigo [37] contorna este problema, apresentando o algoritmo IMPNLMS, o qual recorre a uma função que estima a esparsidade do sistema a partir dos coeficientes do filtro adaptativo. Esta medida de esparsidade é então empregada de modo a não permitir uma degradação da taxa de convergência mesmo em contextos em que a RI a identificar não é esparsa. Devido às suas desejáveis propriedades, o algoritmo IMPNLMS é o alvo principal das análises teóricas propostas nesta tese.

No IMPNLMS, a medida de esparsidade $\xi_{\hat{\mathbf{h}}_k}$ é definida como [46]:

$$\xi_{\hat{\mathbf{h}}_k} = \frac{L}{L - \sqrt{L}} \left(1 - \frac{\|\hat{\mathbf{h}}_k\|_1}{\sqrt{L}\|\hat{\mathbf{h}}_k\|_2} \right), \quad (2.8)$$

onde $\|\hat{\mathbf{h}}_k\|_p = \left(\sum_{n=0}^{L-1} |\hat{\mathbf{h}}_k(n)|^p \right)^{(1/p)}$. A função de esparsidade assume valores no intervalo $[0,1]$, aproximando-se de 1 quando a RI é esparsa e de zero quando a RI é dispersiva [46]. O algoritmo IMPNLMS é apresentado na Tabela 2.1. A conversão de $\xi_{\hat{\mathbf{h}}}$ para o domínio do parâmetro $\alpha(k)$ foi arbitrada por meio de simulações. A função $F(\cdot)$ empregada, a qual reduz a complexidade do algoritmo por ser linear definida por partes, ao invés da função logarítmica (medida inspirada em [44]), foi

Tabela 2.1: Algoritmo IMPNLMS

Inicialização (valores típicos)

$$\delta = 0,01, \quad \epsilon = 0,001, \quad \beta = 0,25, \quad \lambda_\xi = 0,1, \quad \xi(-1) = 0,96$$

$$\hat{\mathbf{h}}_0 = \left[\hat{\mathbf{h}}_0(0) \quad \hat{\mathbf{h}}_0(1) \quad \cdots \quad \hat{\mathbf{h}}_0(L-1) \right] = \mathbf{0}$$

Processamento e Adaptação

Para $k = 0, 1, 2, \dots$

$$\mathbf{x}_k = \left[x(k) \quad x(k-1) \quad \cdots \quad x(k-L+1) \right]^T$$

$$y(k) = \hat{\mathbf{h}}_k \mathbf{x}_k$$

$$e(k) = \hat{d}(k) - y(k)$$

$$\xi_{\hat{\mathbf{h}}_k} = \frac{L}{L - \sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} |\hat{h}_k(j)|}{\sqrt{L \sum_{j=0}^{L-1} \hat{h}_k^2(j)}} \right)$$

$$\xi(k) = (1 - \lambda_\xi)\xi(k-1) + \lambda_\xi \xi_{\hat{\mathbf{h}}_k}$$

$$\alpha(k) = 2\xi(k) - 1$$

Para $i = 0, 1, \dots, L-1$

$$g_k(i) = \frac{1 - \alpha(k)}{2L} + \frac{(1 + \alpha(k))F(|\hat{h}_k(i)|)}{2 \sum_{j=0}^{L-1} F(|\hat{h}_k(j)|) + \epsilon}$$

Fim

$$\mathbf{\Gamma}_k = \text{diag}\{g_k(0), \dots, g_k(L-1)\}$$

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k + \delta}$$

Fim

escolhida como:

$$F(|\hat{h}_k(n)|) = \begin{cases} 400|\hat{h}_k(n)|, & |\hat{h}_k(n)| < 0,005 \\ 8,51|\hat{h}_k(n)| + 1,96, & \text{caso contrário} \end{cases}, \quad (2.9)$$

Os outros algoritmos da família PNLMS não serão apresentados porque não estamos interessados em demonstrar a superioridade do IMPNLMS (isto já foi efetuado em [37]); a diferença básica entre eles reside na dependência da matriz diagonal $\mathbf{\Gamma}_k$ com relação aos coeficientes do filtro adaptativo $\hat{\mathbf{h}}_k$. Uma outra pequena distinção reside no denominador do termo $\beta \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k + \delta}$; enquanto que o artigo original [36] propõe $\mathbf{x}_k^T \mathbf{x}_k + \delta$, a maior parte de suas variantes mais recentes prefere empregar $\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k + \delta$ (vide, por exemplo, [33]).

Entre outras, algumas estratégias alternativas (não exploradas neste trabalho) para a identificação de respostas esparsas consistem no emprego de uma aproximação da norma l_0 para obter uma medida mais acurada de esparsidade [47] e no recurso a subespaços de Krylov [48].

2.4 Interpretação geométrica da família de algoritmos PNLMS

Para $\beta = 1$, podemos associar o algoritmo NLMS ao seguinte problema de otimização:

$$\min_{\hat{\mathbf{h}}_{k+1}} \frac{1}{2} \|\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k\|^2 \quad (2.10)$$

$$\text{sujeito a } \hat{d}(k) - \hat{\mathbf{h}}_{k+1} \mathbf{x}_k = 0. \quad (2.11)$$

Já a família de algoritmos do tipo PNLMS tem seu problema de minimização com restrição correspondente:

$$\min_{\hat{\mathbf{h}}_{k+1}} \frac{1}{2} (\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k) \mathbf{\Gamma}_k^{-1} (\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k)^T \quad (2.12)$$

$$\text{sujeito a } \hat{d}(k) - \hat{\mathbf{h}}_{k+1} \mathbf{x}_k = 0$$

A demonstração é efetuada por meio da técnica de multiplicadores de Lagrange [41]. Já que $(\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k) \mathbf{\Gamma}_k^{-1} (\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k)^T = \sum_{l=0}^{L-1} \frac{(\hat{h}_{k+1}(l) - \hat{h}_k(l))^2}{g_k(l)}$, a minimização associada a um membro da família de algoritmos PNLMS não é igualmente sensível às distâncias $(\hat{h}_{k+1}(l) - \hat{h}_k(l))^2$ para valores distintos de l . À medida que o coeficiente $\hat{h}_k(l)$ aumenta, o termo $g_k(l)$ a ele associado adquire maior amplitude (pela Eq. (2.9)), tornando a minimização menos sensível às suas alterações. Isto permite uma atualização mais pronunciada dos coeficientes de maior magnitude, o que é equivalente a empregarmos um fator de aprendizagem maior para eles. A distorção da medida de distância empregada (que passa a ser oriunda de uma métrica euclidiana ponderada [49]) motiva uma interpretação geométrica do família de algoritmos PNLMS.

Definamos por Π_k o hiperplano dos coeficientes do filtro adaptativo onde o erro *a posteriori* (medido após a atualização utilizando-se os sinais desejado \mathbf{x}_k e de referência $\hat{d}(k)$ atuais) se anula:

$$\Pi_k = \left\{ \hat{\mathbf{h}}_{k+1} : \hat{d}(k) - \hat{\mathbf{h}}_{k+1} \mathbf{x}_k = 0 \right\}. \quad (2.13)$$

Atualizar $\hat{\mathbf{h}}_k$ por meio do algoritmo NLMS com $\beta = 1$ e $\delta = 0$ equivale a efetuar a projeção de $\hat{\mathbf{h}}_k$ em Π_k . Como a direção de atualização é dada por \mathbf{x}_k , temos que $\mathbf{x}_k \perp \Pi_k$. Já os algoritmos derivados do PNLMS não efetuam uma projeção ortogonal de $\hat{\mathbf{h}}_k$ em Π_k , pois a direção de atualização passa a ser alterada por meio de $\mathbf{\Gamma}_k$, conforme ilustra, para o caso $L = 3$, a Figura 2.2. É possível efetuar interpreta-

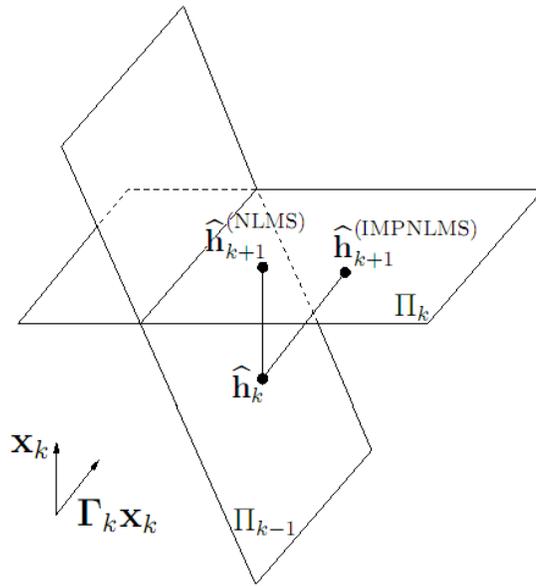


Figura 2.2: Interpretação geométrica de um algoritmo da família PNLMS para $L = 3$.

ções geométricas correspondentes a outros valores de β , porém elas são basicamente idênticas às já efetuadas na literatura. Para uma revisão detalhada destas interpretações, vide [50].

O algoritmo NLMS almeja minimizar $\|\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k\|^2$. O lugar geométrico dos pontos onde esta distância euclidiana entre $\hat{\mathbf{h}}_{k+1}$ e $\hat{\mathbf{h}}_k$ é constante é uma hiperesfera, cujo centro está localizado em $\hat{\mathbf{h}}_k$. Isto explica a ortogonalidade entre \mathbf{x}_k e Π_k , pois um hiperplano tangente a uma esfera é ortogonal a um vetor cuja origem reside no centro da esfera e com a extremidade situada no ponto de tangência. Já no caso dos algoritmos da família PNLMS, a Eq. (2.12) permite-nos observar que as curvas de nível associadas ao problema de minimização são hiperelipses. A Fig. 2.3 ilustra a diferença de atualização entre o algoritmo NLMS e os derivados do PNLMS para o caso $L = 2$. A reta que contém os pontos B, C e E é o lugar geométrico dos pontos nos quais o erro *a posteriori* se anula. Os pontos A e D representam possíveis localizações de $\hat{\mathbf{h}}_k$. No ponto D, $\hat{\mathbf{h}}_k = [0, 9 \ 0, 9]$. Neste caso, a taxa de aprendizado será igual para cada um dos coeficientes, pois estes são idênticos. Logo, tanto o NLMS quanto os algoritmos de IRE atualizam $\hat{\mathbf{h}}_k$ na mesma direção (deslocando o filtro adaptativo do ponto D para o E), tornando-se indistinguíveis. Já no ponto A temos $\hat{\mathbf{h}}_k = [0, 1 \ 1]$; o segundo coeficiente do filtro adaptativo apresenta magnitude significativamente superior à do primeiro, o que resulta, para um algoritmo derivado do PNLMS, numa distribuição desigual da “energia” de atualização. A direção de atualização do algoritmo de IRE (A \rightarrow B) provoca em relação ao NLMS (A \rightarrow C) uma alteração maior em $\hat{h}_{k+1}(1)$ do que em $\hat{h}_{k+1}(0)$, devido à menor magnitude deste.

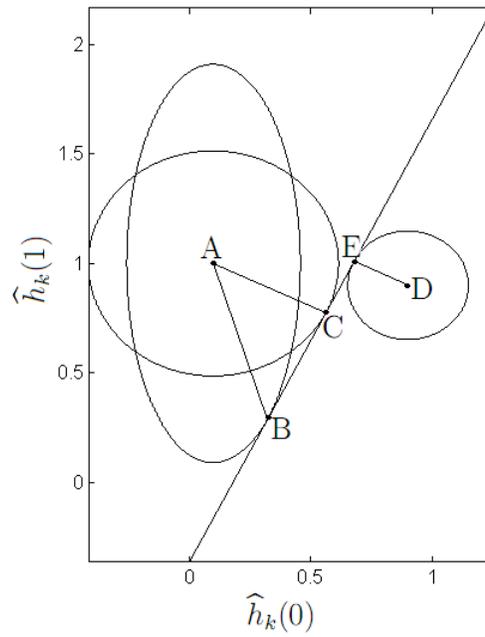


Figura 2.3: Interpretação geométrica da atualização dos algoritmos de IRE para $L = 2$.

No contexto particular em que $L = 2$, o termo $(\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k)\mathbf{\Gamma}_k^{-1}(\hat{\mathbf{h}}_{k+1} - \hat{\mathbf{h}}_k)^T$ apresenta elipses como lugares geométricos das curvas de nível. Isto implica que a direção de atualização de um algoritmo de IRE difere da do NLMS, a menos que a elipse seja degenerada, possibilidade que ocorre quando $\hat{h}_k(0) = \hat{h}_k(1)$, o que é ilustrado no ponto D da Fig. 2.3.

2.5 Considerações Finais

Este capítulo teve por objetivo detalhar o paradigma de identificação adaptativa e supervisionada de sistemas, enfatizando a família de algoritmos proporcionais e sua interpretação geométrica. Tal família de algoritmos (e em especial o algoritmo IMPNLMS) será alvo de análises teóricas e de extensões no próximo capítulo.

Capítulo 3

Análises e Extensões da Família de Algoritmos PNLMS

3.1 Introdução

O capítulo anterior elaborou um panorama da família de algoritmos PNLMS, os quais têm sido alvo de grande interesse por parte da comunidade acadêmica. Como a análise do comportamento de um algoritmo adaptativo no transiente é importante para a caracterização de suas propriedades (permitindo, entre outras vantagens, a comparação entre algoritmos distintos), este capítulo, na Seção 3.2, explora teoricamente a evolução do algoritmo IMPNLMS na fase de transiente. O algoritmo IMPNLMS foi escolhido por ser um dos mais bem-sucedidos algoritmos proporcionais.

Embora o sinal de entrada dos filtros adaptativos possa ser colorido, a grande maioria das análises de transiente encontradas na literatura de algoritmos de IRE supõe ser este branco. A violação desta hipótese resulta em substanciais diferenças na taxa de convergência dos algoritmos. Como esta tese tem em vista aplicações em sinais de voz (que, como visto na Seção 1.3.3, se caracterizam por apresentarem espectro colorido), nossa análise de transiente contemplará explicitamente a correlação entre amostras do sinal de entrada, o que acarreta um significativo aumento de complexidade da análise de transiente.

O MSE em regime permanente é uma outra métrica importante para se avaliar o desempenho de um algoritmo adaptativo, já que reflete, em termos estatísticos, o grau de acurácia do algoritmo após a etapa de aprendizado. Por essa razão, desenvolvemos na Seção 3.3 estimativas teóricas para o MSE dos algoritmos da família PNLMS.

Estruturas em subbandas têm contribuído para acelerar a convergência de algoritmos adaptativos, em particular quando o sinal de entrada não é branco. No con-

texto de aplicações com sinais de voz, utilizamos um processamento em subbandas em conjunção com o algoritmo IMPNLMS, estendendo, na Seção 3.4, a estimativa do MSE da Seção 3.3 para a configuração em subbandas. As análises do MSE permitem uma aproximação heurística dos limites superiores dos fatores de aprendizagem, o que é efetuado na Seção 3.5. Por fim, técnicas de variação do fator de aprendizagem ao longo das iterações são avaliadas, na Seção 3.6, como promissoras para evitar um aumento do MSE em regime permanente à medida que se aumenta o número de subbandas.

3.2 Análise de Transiente do Algoritmo IMPNLMS

Estimar teoricamente a convergência do erro quadrático médio de um algoritmo adaptativo elimina a necessidade de efetuarmos médias de Monte Carlo, entre outras vantagens já devidamente reconhecidas pela literatura. Nesta seção, almejamos encontrar equações recursivas que descrevam de modo razoavelmente preciso a evolução do MSE do algoritmo IMPNLMS ao longo das iterações.

Dentre os algoritmos pertencentes à família PNLMS, o mais bem-sucedido é o IMPNLMS [37], cuja evolução de transiente encontra-se falta de análise matemática na literatura. Por isso, ele foi escolhido como alvo desta análise. Esta portanto é a única seção do capítulo que não se aplica a todos os algoritmos da família PNLMS, mas apenas a um exemplar deles. Os artigos [51] e [52] analisam a convergência de outros algoritmos desta família.

As equações do algoritmo que nos interessam no momento são (ver Tab. 2.1):

$$\xi_{\hat{\mathbf{h}}_k} = \frac{L}{L - \sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} |\hat{h}_k(j)|}{\sqrt{L} \sum_{j=0}^{L-1} \hat{h}_k^2(j)} \right) \quad (3.1)$$

$$\xi(k) = (1 - \lambda_\xi)\xi(k-1) + \lambda_\xi \xi_{\hat{\mathbf{h}}_k} \quad (3.2)$$

$$\alpha(k) = 2\xi(k) - 1 \quad (3.3)$$

$$g_k(i) = \frac{1 - \alpha(k)}{2L} + \frac{(1 + \alpha(k))F(|\hat{h}_k(i)|)}{2 \sum_{j=0}^{L-1} F(|\hat{h}_k(j)|) + \epsilon} \quad (3.4)$$

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k + \delta} \quad (3.5)$$

Para simplificar os cálculos dos valores esperados, definamos o desvio $\widehat{z}_j(k)$ do j -ésimo elemento de $\widehat{\mathbf{h}}_k$ como:

$$\widehat{z}_j(k) = \widehat{h}_k(j) - h(j), \quad (3.6)$$

e $\widehat{\mathbf{z}}_k = [\widehat{z}_0(k) \ \widehat{z}_1(k) \ \cdots \ \widehat{z}_{L-1}(k)]$, como o vetor que contém os desvios de todos os coeficientes do filtro adaptativo na k -ésima iteração. Tais desvios são variáveis aleatórias, já que dependem de um sinal de entrada de natureza estocástica.

Desejamos calcular os seguintes termos: *i*) $E[\xi_{\widehat{\mathbf{h}}(k)}]$; *ii*) $E[\xi(k)]$, *iii*) $E[\alpha(k)]$, *iv*) $E[g_k(i)]$, *v*) $E[\widehat{\mathbf{z}}_{k+1}]$ e *vi*) $E[\widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l)]$. Os termos *v*) e *vi*) serão os responsáveis pela estimativa do MSE em cada iteração¹; podemos considerar o cálculo dos demais termos como uma etapa necessária para a obtenção destes. Na obtenção dos termos *i*)-*vi*), torna-se necessário o cálculo de alguns termos intermediários, que são: *vii*) $E[|\widehat{h}_k(j)|]$, *viii*) $E[F(|\widehat{h}_k(j)|)]$ e *ix*) $E\left[\frac{1}{[\sum_{j=0}^{L-1} g_k(j)x^{2(k-j)} + \delta]^n}\right]$.

Alguns dos valores esperados não apresentam respostas analíticas, o que obriga-nos a lançar mão de algumas hipóteses simplificadoras (avaliadas experimentalmente na Seção 3.2.1) de modo a encontrar as equações recursivas que permitam-nos estimar o erro quadrático médio do algoritmo em cada iteração, semelhante ao que foi feito, para outros algoritmos da família PNLMS, em [51] e [52]. Em alguns termos, empregamos hipóteses mais adequadas, de modo a não comprometer em demasia a acurácia da análise.

As principais hipóteses empregadas na obtenção das fórmulas finais são:

- Hipótese I: O sinal de entrada \mathbf{x}_k é gaussiano, estacionário e apresenta média zero, sendo decorrelacionado do ruído de medição $\nu(k)$, o qual também tem média zero. Esta hipótese justifica-se pelo uso do modelo linear (ainda que desconhecido) e pelo fato de que a ordem da solução ótima de Wiener do filtro adaptativo coincide com a do sistema desconhecido [51].
- Hipótese II-a: O valor esperado da razão de variáveis aleatórias que dependem dos coeficientes do filtro adaptativo pode ser aproximado pela razão entre os valores esperados dessas variáveis.
- Hipótese II-b: O valor esperado da razão de variáveis aleatórias que dependem dos coeficientes do filtro adaptativo pode ser aproximado pelo valor esperado do numerador multiplicado pelo valor esperado do inverso do denominador².

¹Na literatura relacionada, a suposição de sinais de entrada brancos só torna necessária a estimativa de segunda ordem dos termos $E[\widehat{z}_{k+1}^2(i)]$; como nossa análise não utiliza esta restrição importa-nos também calcular os termos cruzados $E[\widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l)]$, para $i \neq l$.

²Um denominador importante em nossa análise é $(\sum_{j=0}^{L-1} g_k(j)x^{2(k-j)} + \delta)^n$, o qual, sendo constante (ou se tivermos a condição $L \gg \sqrt{2 \sum_{i=1}^L E[g_{k+1}^2(i)]}$ [53]), torna válida a aproximação

Se X e Y são variáveis aleatórias, a Hipótese II-a implica que $E\left[\frac{X}{Y}\right] \approx \frac{E[X]}{E[Y]}$. Já a Hipótese II-b (utilizada sempre que possível) empregaria a aproximação alternativa $E\left[\frac{X}{Y}\right] \approx E[X] E\left[\frac{1}{Y}\right]$.

- Hipótese III: $E\left[\sqrt{\left(L \sum_{j=0}^{L-1} \widehat{h}_k^2(j)\right)}\right] \approx \sqrt{\left(L \sum_{j=0}^{L-1} E\left[\widehat{h}_k^2(j)\right]\right)}$.
- Hipótese IV: $E\left[\alpha(k)F(|\widehat{h}_k(i)|)\right] \approx E[\alpha(k)] E\left[F(|\widehat{h}_k(i)|)\right]$, onde F é a função apresentada na Eq. (2.9).
- Hipótese V: O estimador dos coeficientes atua como um filtro passa-baixa; daí supomos que $\widehat{h}_k(i)$ e $g_k(i)$ variam lentamente em relação a $x(k-j)$. Logo, podemos efetuar aproximações como:

$$E[g_k(i)x^n(k-j)] \approx E[g_k(i)] E[x^n(k-j)], \quad (3.7)$$

$$E[g_k(i)^n \widehat{\mathbf{z}}_k(j)^m] \approx E[g_k(i)]^n E[\widehat{\mathbf{z}}_k(j)^m], \quad (3.8)$$

e

$$E\left[\widehat{h}_k(i)x^n(k-j)\right] \approx E\left[\widehat{h}_k(i)\right] E\left[x^n(k-j)\right]. \quad (3.9)$$

- Hipótese VI: os coeficientes $\widehat{h}_k(i)$ apresentam distribuição gaussiana³, de média $\mu_k(i)$ e variância $\sigma_k^2(i)$.
- Hipótese VII: $E[g_k(i)g_k(l)] \approx E[g_k(i)] E[g_k(l)]$.

A seguir, explicitamos os cálculos feitos para a resolução do valor esperado de cada termo:

i) termo $E\left[\xi_{\widehat{\mathbf{h}}_k}\right]$.

A partir da Eq. (3.1), temos:

$$E\left[\xi_{\widehat{\mathbf{h}}_k}\right] = \frac{L}{L - \sqrt{L}} \left(1 - E\left[\frac{\sum_{j=0}^{L-1} |\widehat{h}_k(j)|}{\sqrt{L \sum_{j=0}^{L-1} \widehat{h}_k^2(j)}}\right] \right). \quad (3.10)$$

Simplificando mediante a Hipótese II(a)

$$E\left[\xi_{\widehat{\mathbf{h}}_k}\right] \approx \frac{L}{L - \sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} E\left[|\widehat{h}_k(j)|\right]}{E\left[\sqrt{L \sum_{j=0}^{L-1} \widehat{h}_k^2(j)}\right]} \right). \quad (3.11)$$

da Hipótese II-a. Exploraremos o cálculo deste termo por meio da Hipótese II-b, mais acurada.

³Devido à recursividade do algoritmo, podemos evocar o Teorema Central do Limite para justificar esta hipótese, ao menos a partir de um certo número de iterações.

e recorrendo à Hipótese III, temos:

$$\mathbb{E} \left[\xi_{\widehat{\mathbf{h}}_k} \right] \approx \frac{L}{L - \sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} \mathbb{E} \left[|\widehat{h}_k(j)| \right]}{\sqrt{L} \sum_{j=0}^{L-1} \mathbb{E} \left[\widehat{h}_k^2(j) \right]} \right). \quad (3.12)$$

Verificamos o surgimento de um termo $\mathbb{E} \left[\left| \widehat{h}_k(j) \right| \right]$, o qual será posteriormente obtido mediante a Hipótese VI; a expressão $\mathbb{E} \left[\widehat{h}_k^2(j) \right]$ pode ser obtida mediante uma mera manipulação de $\mathbb{E} \left[\widehat{z}_k(j) \right]$, $h(j)$ e do termo (vi), impondo-se, neste último, as identidades $i = l = j$.

ii) termo $\mathbb{E} \left[\xi(k) \right]$.

A partir da Eq. (3.2), encontramos:

$$\mathbb{E} \left[\xi(k) \right] = \mathbb{E} \left[(1 - \lambda_\xi) \xi(k-1) + \lambda_\xi \xi_{\widehat{\mathbf{h}}_k} \right] = (1 - \lambda_\xi) \mathbb{E} \left[\xi(k-1) \right] + \lambda_\xi \mathbb{E} \left[\xi_{\widehat{\mathbf{h}}_k} \right], \quad (3.13)$$

cujas modelagem, em termos estatísticos, é exata (sem aproximações), devido ao fato de λ_ξ ser determinístico.

iii) termo $\mathbb{E} \left[\alpha(k) \right]$:

Recorrendo à Eq. (3.3), encontramos o termo (iii) sem efetuar nenhuma aproximação:

$$\mathbb{E} \left[\alpha(k) \right] = 2\mathbb{E} \left[\xi(k) \right] - 1.$$

iv) termo $\mathbb{E} \left[g_k(i) \right]$.

Partindo da Eq. (3.4), encontramos:

$$\mathbb{E} \left[g_k(i) \right] = \frac{1 - \mathbb{E} \left[\alpha(k) \right]}{2L} + \mathbb{E} \left[\frac{(1 + \alpha(k)) F(|\widehat{h}_k(i)|)}{2 \sum_{j=0}^{L-1} F(|\widehat{h}_k(j)|) + \epsilon} \right]. \quad (3.14)$$

Pelas Hipóteses II-a e IV, obtemos:

$$\mathbb{E} \left[g_k(i) \right] \approx \frac{1 - \mathbb{E} \left[\alpha(k) \right]}{2L} + \frac{(1 + \mathbb{E} \left[\alpha(k) \right]) \mathbb{E} \left[F(|\widehat{h}_k(i)|) \right]}{2 \sum_{j=0}^{L-1} \mathbb{E} \left[F(|\widehat{h}_k(j)|) \right] + \epsilon}. \quad (3.15)$$

O termo $\mathbb{E} \left[F(|\widehat{h}_k(j)|) \right]$ será derivado posteriormente, mediante a Hipótese VI.

v) termo $\mathbb{E} \left[\widehat{\mathbf{z}}_{k+1} \right]$.

Seja a matriz de autocorrelação de \mathbf{x}_k denotada por $\mathbf{R}_\mathbf{x} = \mathbb{E} \left[\mathbf{x}_k \mathbf{x}_k^T \right]$. Utilizando as Hipóteses II-b e V, obtemos:

$$\mathbb{E} \left[\widehat{\mathbf{z}}_{k+1} \right] \approx \mathbb{E} \left[\widehat{\mathbf{z}}_k \right] - \beta \mathbb{E}_{g,1} \mathbf{\Gamma}_k \mathbf{R}_\mathbf{x} \mathbb{E} \left[\widehat{\mathbf{z}}_k \right], \quad (3.16)$$

onde $E_{g,n} = E \left[\frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \epsilon \right)^n} \right]$. Uma das contribuições deste trabalho reside num cálculo mais preciso de $E_{g,n}$ por meio da Hipótese II-b, o que será feito na derivação do termo (*ix*).

vi) termo $E[\widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l)]$.

Expressando as atualizações do i -ésimo coeficiente de \widehat{h}_k , temos:

$$\begin{aligned} \widehat{h}_{k+1}(i) &= \widehat{h}_k(i) + \frac{\beta g_k(i)x(k-i)\nu(k)}{\sum_{j=0}^{L-1} g_k(j)x^2(n-j) + \delta} \\ &\quad - \frac{\beta g_k(i)x(n-i) \sum_{j=0}^{L-1} \overbrace{\left[\widehat{h}_k(j) - h(j) \right]}^{\widehat{z}_k(j)} x(n-i)x(n-j)}{\sum_{j=0}^{L-1} g_k(j)x^2(n-j) + \delta} \end{aligned} \quad (3.17)$$

Em relação a $\widehat{z}_k(i)$, temos:

$$\widehat{z}_{k+1}(i) = \widehat{z}_k(i) + \frac{\beta g_k(i)x(k-i)\nu(k)}{\sum_{j=0}^{L-1} g_k(j)x^2(n-j) + \delta} - \frac{\beta g_k(i)x(n-i) \sum_{j=0}^{L-1} \widehat{z}_k(j)x(n-i)x(n-j)}{\sum_{j=0}^{L-1} g_k(j)x^2(n-j) + \delta} \quad (3.18)$$

A atualização de $\widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l)$ segue abaixo:

$$\begin{aligned} \widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l) &= \widehat{z}_k(i)\widehat{z}_k(l) + \frac{\beta g_k(l)x(k-l)\widehat{z}_k(i)\nu(k)}{\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta} \\ &\quad - \frac{\beta g_k(l) \sum_{j=0}^{L-1} \widehat{z}_k(j)\widehat{z}_k(i)x(k-l)x(k-j)}{\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta} \\ &\quad + \frac{\beta g_k(i)x(k-i)\widehat{z}_k(l)\nu(k)}{\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta} + \frac{\beta^2 g_k(i)g_k(l)x(k-i)x(k-l)\nu^2(k)}{\left(\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta \right)^2} \quad (3.19) \\ &\quad - \frac{\beta g_k(i) \sum_{n=0}^{L-1} \widehat{z}_k(n)\widehat{z}_k(l)x(k-i)x(k-n)}{\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta} \\ &\quad - \frac{\beta^2 g_k(i)g_k(l) \sum_{n=0}^{L-1} \widehat{z}_k(n)x(k-i)x(k-l)x(k-n)\nu(k)}{\left(\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta \right)^2} \\ &\quad + \frac{\beta^2 g_k(i)g_k(l) \sum_{j=0}^{L-1} \sum_{n=0}^{L-1} \widehat{z}_k(j)\widehat{z}_k(n)x(k-i)x(k-j)x(k-l)x(k-n)}{\left(\sum_{n=0}^{L-1} g_k(n)x^2(k-n) + \delta \right)^2} \end{aligned}$$

Seja $r(n) = E[x(k)x(k-n)]$. Por meio das Hipóteses II, V e VII, podemos simplificar a expressão do valor esperado:

$$\begin{aligned}
E[\widehat{z}_{k+1}(i)\widehat{z}_{k+1}(l)] &= E[\widehat{z}_k(i)\widehat{z}_k(l)] - \beta E_{g,1} g_k(l) \sum_{j=0}^{L-1} E[\widehat{z}_k(j)\widehat{z}_k(i)] r(l-j) \\
&\quad + \beta^2 E_{g,2} E[g_k(i)] E[g_k(l)] r(i-l) \sigma_\nu^2 \\
&\quad - \beta E_{g,1} E[g_k(i)] \sum_{n=0}^{L-1} E[\widehat{z}_k(n)\widehat{z}_k(l)] r(i-n) \\
&\quad + \beta^2 E_{g,2} E[g_k(i)] E[g_k(l)] \sum_{j=0}^{L-1} \sum_{n=0}^{L-1} E[\widehat{z}_k(j)\widehat{z}_k(n)] r_{i,j,l,n},
\end{aligned} \tag{3.20}$$

onde $r_{i,j,l,n} = E[x(k-i)x(k-j)x(k-l)x(k-n)]$. Utilizando a Hipótese I, podemos calcular $r_{i,j,l,n}$ por meio das fórmulas de momentos de variáveis gaussianas.

Nos cálculos dos valores esperados acima, surgiram os termos intermediários (vii) - (ix) cujo cálculo será agora detalhado.

$vii)$ termo $E\left[\left|\widehat{h}_k(j)\right|\right]$

Por meio da Hipótese VI, $\widehat{h}_k(i)$ é gaussiano, sendo portanto completamente determinado por meio de sua média $\mu_k(i) = E\left[\widehat{h}_k(i)\right]$ e de sua variância $\sigma_k^2(i) = E\left[\widehat{h}_k^2(i)\right] - \left(E\left[\widehat{h}_k(i)\right]\right)^2$, ambas já calculadas no termos $v)$ e $vi)$ ⁴.

Logo, a pdf (função densidade de probabilidade, do inglês *probability density function*) de $\left|\widehat{h}_k(i)\right|$ é dada por:

$$f\left(\left|\widehat{h}_k(i)\right|\right) = \frac{1}{\sqrt{2\pi}\sigma_k(i)} \left[e^{-\frac{(\left|\widehat{h}_k(i)\right| - \mu_k(i))^2}{2\sigma_k^2(i)}} + e^{-\frac{(\left|\widehat{h}_k(i)\right| + \mu_k(i))^2}{2\sigma_k^2(i)}} \right] U\left(\widehat{h}_k(i)\right), \tag{3.21}$$

onde $U(\cdot)$ corresponde à função degrau.

Após o cálculo de $E\left[\widehat{h}_k(i)\right]$ e $E\left[\widehat{h}_k^2(i)\right]$, podemos calcular a média $\mu_k(i)$ e o desvio-padrão $\sigma_k(i)$, determinando a pdf de cada $\widehat{h}_k(i)$.

Sejam as seguintes integrais definidas:

$$\begin{aligned}
\Phi_0(a, b, \mu_k(i), \sigma_k(i)) &= \int_a^b e^{-\frac{(\widehat{h}_k(i) - \mu_k(i))^2}{2\sigma_k^2(i)}} d\widehat{h}_k(i) \\
&= \sigma_k(i) \sqrt{\frac{\pi}{2}} \left[\operatorname{erfc}\left(\frac{a - \mu_k(i)}{\sqrt{2}\sigma_k(i)}\right) - \operatorname{erfc}\left(\frac{b - \mu_k(i)}{\sqrt{2}\sigma_k(i)}\right) \right], \tag{3.22}
\end{aligned}$$

⁴No termo $vi)$, $E\left[\widehat{h}_k^2(i)\right]$ pode ser encontrado impondo-se $l = i$; a conversão de $\widehat{z}_k(i)$ para $\widehat{h}_k(i)$ é regida pela Eq. (3.6).

$$\begin{aligned}
\Phi_1(a, b, \mu_k(i), \sigma_k(i)) &= \int_a^b \widehat{h}_k(i) e^{-\frac{(\widehat{h}_k(i) - \mu_k(i))^2}{2\sigma_k^2(i)}} d\widehat{h}_k(i) \\
&= \sigma_k^2(i) \left[e^{-\frac{(a - \mu_k(i))^2}{2\sigma_k^2(i)}} - e^{-\frac{(b - \mu_k(i))^2}{2\sigma_k^2(i)}} \right] \\
&\quad + \mu_k(i) \sigma_k(i) \sqrt{\frac{\pi}{2}} \left[\operatorname{erfc} \left(\frac{a - \mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) - \operatorname{erfc} \left(\frac{b - \mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) \right],
\end{aligned} \tag{3.23}$$

onde $\operatorname{erfc}(x)$ é função erro complementar, definida por $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-x^2} dx$.

Com a pdf de $|\widehat{h}_k(i)|$ e a integral Φ_1 acima definidas⁵, a expressão $E \left[\left| \widehat{h}_k(i) \right| \right]$ pode por fim ser determinada:

$$\begin{aligned}
E \left[\left| \widehat{h}_k(i) \right| \right] &= \int_0^\infty \left| \widehat{h}_k(i) \right| f \left(\left| \widehat{h}_k(i) \right| \right) d \left| \widehat{h}_k(i) \right| \\
&= \frac{1}{\sqrt{2\pi}\sigma_k(i)} \int_0^\infty \left| \widehat{h}_k(i) \right| e^{-\frac{(\widehat{h}_k(i) - \mu_k(i))^2}{2\sigma_k^2(i)}} d \left| \widehat{h}_k(i) \right| \\
&\quad + \frac{1}{\sqrt{2\pi}\sigma_k(i)} \int_0^\infty \left| \widehat{h}_k(i) \right| e^{-\frac{(\widehat{h}_k(i) + \mu_k(i))^2}{2\sigma_k^2(i)}} d \left| \widehat{h}_k(i) \right| \\
&= \frac{\Phi_1(0, \infty, \mu_k(i), \sigma_k(i))}{\sqrt{2\pi}\sigma_k(i)} + \frac{\Phi_1(0, \infty, -\mu_k(i), \sigma_k(i))}{\sqrt{2\pi}\sigma_k(i)}.
\end{aligned} \tag{3.24}$$

viii) termo $E \left[F \left(\left| \widehat{h}_k(j) \right| \right) \right]$

Utilizando a média estatística, temos:

$$\begin{aligned}
E \left[F \left(\left| \widehat{h}_k(i) \right| \right) \right] &= \int_0^{0,005} \frac{400\widehat{h}_k(i)}{\sqrt{2\pi}\sigma_k(i)} \left[e^{-\frac{(\widehat{h}_k(i) - \mu_k(i))^2}{2\sigma_k^2(i)}} + e^{-\frac{(\widehat{h}_k(i) + \mu_k(i))^2}{2\sigma_k^2(i)}} \right] d\widehat{h}_k(i) \\
&\quad + \int_{0,005}^\infty \frac{8,51|\widehat{h}_k(i)| + 1,96}{\sqrt{2\pi}\sigma_k(i)} \left[e^{-\frac{(\widehat{h}_k(i) - \mu_k(i))^2}{2\sigma_k^2(i)}} + e^{-\frac{(\widehat{h}_k(i) + \mu_k(i))^2}{2\sigma_k^2(i)}} \right] d\widehat{h}_k(i)
\end{aligned} \tag{3.25}$$

Podemos então calcular $E \left[F \left(\left| \widehat{h}_k(i) \right| \right) \right]$ explicitamente:

$$\begin{aligned}
E \left[F \left(\left| \widehat{h}_k(i) \right| \right) \right] &= \frac{1}{\sqrt{2\pi}\sigma} \left[400\Phi_1(5 \times 10^{-3}, \mu, \sigma) + 400\Phi_1(5 \times 10^{-3}, -\mu, \sigma) \right. \\
&\quad + 8,51\Phi_1(5 \times 10^{-3}, \infty, \mu, \sigma) + 8,51\Phi_1(5 \times 10^{-3}, \infty, -\mu, \sigma) \\
&\quad \left. + 1,96\Phi_0(5 \times 10^{-3}, \infty, \mu, \sigma) + 1,96\Phi_0(5 \times 10^{-3}, \infty, -\mu, \sigma) \right].
\end{aligned} \tag{3.26}$$

⁵No obtenção do termo (viii), a integral definida Φ_0 também será utilizada.

ix) termo $E_{g,n} = E \left[\frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \delta \right)^n} \right]$

O termo $E_{g,n}$ costuma ser estimado [54], [55], [56], [57] por meio da Hipótese II-a⁶:

$$\begin{aligned} E \left[\frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \delta \right)^n} \right] &\approx \frac{1}{E \left[\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \delta \right)^n \right]} \\ &\approx \frac{1}{\left(\sum_{j=0}^{L-1} E[g_k(j)] E[x^2(k-j)] + \delta \right)^n} = \frac{1}{\left(\sigma_x^2 \sum_{j=0}^{L-1} E[g_k(j)] + \delta \right)^n} \approx \frac{1}{(\sigma_x^2 + \delta)^n}, \end{aligned} \quad (3.27)$$

onde σ_x^2 é a variância de $x(k)$. Identificamos a aproximação acima como a mais crítica para a exatidão da análise, especialmente quando os filtros adaptativos apresentam comprimentos curtos, pois nestes casos a variância do denominador é elevada. Segundo a definição de $\xi_{\hat{\mathbf{h}}_k}$, mesmo uma resposta ao impulso curta pode ter grau máximo de esparsidade (quando apenas um de seus elementos é não-nulo), o que poderia motivar o emprego do IMPNLMS nestas configurações. Em aplicações que utilizam processamento em subbandas, os filtros adaptativos atuantes em cada subbanda também podem ser curtos o suficiente para degradar a acurácia da aproximação acima. Assim, adaptando as técnicas de [58], [59], [55], [60] e [61], desenvolvemos uma estimativa mais acurada de $E_{g,n}$, para $n = 1, 2$, a qual será detalhada a seguir.

Sendo δ um parâmetro de regularização próximo de zero, podemos, truncando a expansão de $E_{g,n}$ numa série de Taylor, efetuar as aproximações⁷:

$$\begin{aligned} E_{g,1} &= E \left[\frac{1}{\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \delta} \right] \\ &\approx \frac{1}{\sum_{j=0}^{L-1} g_k(j)x^2(k-j)} - \delta \frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) \right)^2}, \end{aligned} \quad (3.28)$$

$$E_{g,2} = E \left[\frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) + \delta \right)^2} \right] \quad (3.29)$$

$$\approx \frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) \right)^2} - 2\delta \frac{1}{\left(\sum_{j=0}^{L-1} g_k(j)x^2(k-j) \right)^3}. \quad (3.30)$$

Se $\bar{E}_{g,n} = \frac{1}{(\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k)^n}$, as relações (aproximadas) entre $E_{g,n}$ e $\bar{E}_{g,n}$ são, portanto, dadas por:

$$E_{g,1} \approx \bar{E}_{g,1} - \delta \bar{E}_{g,2}, \quad (3.31)$$

⁶No caso do algoritmo NLMS, $g_k(j) = 1, \forall j$, o que modifica levemente o resultado.

⁷O artigo [62], num contexto similar, propõe essa abordagem para o cálculo de $E_{g,1}$.

$$E_{g,2} \approx \bar{E}_{g,2} - 2\delta\bar{E}_{g,3}. \quad (3.32)$$

Estas relações permitem-nos concentrar na estimativa de $\bar{E}_{g,n}$, para $n = 1, 2$ e 3 , os quais dependem das variáveis aleatórias $g_k(j)$ e $x(k-j)$. Já que $g_k(j)$ varia muito mais lentamente que $x(k-j)$, a maior parte da inconstância de $\bar{E}_{g,n}$ deriva deste último. Por isso, aproximaremos o valor esperado efetuando integrações nas variáveis $x(k-j)$, supondo que $g_k(j)$ pode ser substituído por $E[g_k(j)]$. Matematicamente, isso implica:

$$E[\bar{E}_{g,n}] \approx E\left[\frac{1}{(\mathbf{x}_k^T E[\mathbf{\Gamma}_k] \mathbf{x}_k)^n}\right]. \quad (3.33)$$

Deve-se notar que o valor esperado $E[\mathbf{\Gamma}_k]$ já foi calculado, de forma recursiva, no termo *iv*).

Assim, podemos escrever:

$$E\left[\frac{1}{(\mathbf{x}_k^T E[\mathbf{\Gamma}_k] \mathbf{x}_k)^n}\right] = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{L \times} \frac{1}{(\mathbf{x}_k^T E[\mathbf{\Gamma}_k] \mathbf{x}_k)^n} f(\mathbf{x}_k) d\mathbf{x}_k. \quad (3.34)$$

Utilizando a Hipótese I, a pdf $f(\mathbf{x}_k)$ é conhecida⁸, resultando em:

$$E\left[\frac{1}{(\mathbf{x}_k^T E[\mathbf{\Gamma}_k] \mathbf{x}_k)^n}\right] = \frac{1}{(2\pi)^{L/2} \sqrt{\det(\mathbf{R}_x)}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{L \times} \frac{1}{(\mathbf{x}^T E[\mathbf{\Gamma}_k] \mathbf{x})^n} e^{-\frac{\mathbf{x}^T \mathbf{R}_x^{-1} \mathbf{x}}{2}} d\mathbf{x}. \quad (3.35)$$

O objetivo então é encontrar uma forma de resolver, com mais acurácia, a integral múltipla na Eq. (3.35). Começemos por definir uma função auxiliar⁹ $\Psi_n(\omega)$:

$$\Psi_n(\omega) = \frac{1}{(2\pi)^{L/2} \sqrt{\det(\mathbf{R}_x)}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{L \times} \frac{1}{(\mathbf{x}^T E[\mathbf{\Gamma}_k] \mathbf{x})^n} e^{-\omega \mathbf{x}^T E[\mathbf{\Gamma}_k] \mathbf{x}} e^{-\frac{\mathbf{x}^T \mathbf{R}_x^{-1} \mathbf{x}}{2}} d\mathbf{x}. \quad (3.36)$$

Obviamente, temos que:

$$E\left[\frac{1}{(\mathbf{x}_k^T E[\mathbf{\Gamma}_k] \mathbf{x}_k)^n}\right] = \Psi_n(\omega)|_{\omega=0} = \Psi_n(0) = \underbrace{\int \cdots \int}_{n \times} \frac{\partial^n \Psi_n(\omega)}{\partial^n \omega} \underbrace{d\omega \cdots d\omega}_{n \times} \bigg|_{\omega=0}, \quad (3.37)$$

⁸Observe que, pela Hipótese I, o sinal \mathbf{x}_k é estacionário, o que permite-nos elidir o índice k .

⁹Ao propor esta função auxiliar, não tínhamos conhecimento que uma versão muito parecida já havia aparecido em [61], num contexto similar. Importa realçar que [61] analisa a convergência de primeira ordem dos coeficientes adaptativos e não contempla a evolução do MSE. Além disso, o algoritmo analisado difere do aqui focado.

onde a constante de integração foi omitida, já que $\lim_{\omega \rightarrow \infty} \Psi(\omega) = 0$. O objetivo de derivar n vezes a função $\Psi(\omega)$ reside no fato de podermos efetuar algumas simplificações oriundas de propriedades de pdfs gaussianas. Assim:

$$\begin{aligned}
\frac{\partial^n \Psi_n(\omega)}{\partial^n \omega} &= \frac{(-1)^n}{(2\pi)^{L/2} \sqrt{\det(\mathbf{R}_x)}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{L \times} e^{-\omega \mathbf{x}^T \mathbf{E}[\Gamma_k] \mathbf{x}} e^{-\frac{\mathbf{x}^T \mathbf{R}_x^{-1} \mathbf{x}}{2}} d\mathbf{x} \\
&= \frac{(-1)^n}{(2\pi)^{L/2} \sqrt{\det(\mathbf{R}_x)}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{\mathbf{x}^T (\mathbf{R}_x^{-1} + 2\omega \mathbf{E}[\Gamma_k]) \mathbf{x}}{2}} d\mathbf{x} \\
&= \frac{(-1)^n}{(2\pi)^{L/2} \sqrt{\det(\mathbf{R}_x)}} (2\pi)^{L/2} \sqrt{\det[(\mathbf{R}_x^{-1} + 2\omega \mathbf{E}[\Gamma_k])^{-1}]} \\
&= \frac{(-1)^n}{\sqrt{\det(\mathbf{I} + 2\omega \mathbf{E}[\Gamma_k] \mathbf{R}_x)}} \tag{3.38}
\end{aligned}$$

Pelas Equações (3.37) e (3.38), temos:

$$\bar{E}_{g,n} = \underbrace{\int \cdots \int}_{n \times} \frac{(-1)^n}{\sqrt{\det(\mathbf{I} + 2\omega \mathbf{E}[\Gamma_k] \mathbf{R}_x)}} \underbrace{d\omega \cdots d\omega}_{n \times} \Bigg|_{\omega=0}. \tag{3.39}$$

A expressão acima exige o cálculo de integrais hiper-elípticas, as quais (exceto em casos em que L é pequeno) não apresentam solução analítica. Por isso, teremos que recorrer a aproximações. Sejam \mathbf{Q}_k e $\mathbf{\Lambda}_k$ as matrizes que contém, respectivamente, os autovetores e os autovalores de $\mathbf{E}[\Gamma_k] \mathbf{R}_x$; $\mathbf{\Lambda}_k$ é uma matriz diagonal. Logo, $\mathbf{E}[\Gamma_k] \mathbf{R}_x = \mathbf{Q}_k \mathbf{\Lambda}_k \mathbf{Q}_k^T$. A ortogonalidade de \mathbf{Q}_k implica $\mathbf{E}[\Gamma_k] \mathbf{R}_x \mathbf{Q}_k = \mathbf{Q}_k \mathbf{\Lambda}_k$. Como:

$$(\mathbf{I} + 2\omega \mathbf{E}[\Gamma_k] \mathbf{R}_x) \mathbf{Q}_k = \mathbf{Q}_k + 2\omega \overbrace{\mathbf{E}[\Gamma_k] \mathbf{R}_x \mathbf{Q}_k}^{\mathbf{Q}_k \mathbf{\Lambda}_k} = \mathbf{Q}_k (\mathbf{I} + 2\omega \mathbf{E}[\Gamma_k] \mathbf{R}_x), \tag{3.40}$$

os autovetores de $\mathbf{I} + 2\omega \mathbf{E}[\Gamma_k] \mathbf{R}_x$ são os mesmos que os de $\mathbf{E}[\Gamma_k] \mathbf{R}_x$, sendo seus autovalores¹⁰ dados por $1 + 2\omega \lambda_i$, onde λ_i é o i -ésimo autovalor de $\mathbf{E}[\Gamma_k] \mathbf{R}_x$.

Já que o determinante de uma matriz equivale ao produto de seus autovalores, podemos reescrever a Eq. (3.39) da forma a seguir:

$$\bar{E}_{g,n} = \underbrace{\int \cdots \int}_{n \times} \frac{(-1)^n}{\sqrt{\prod_{l=1}^L (1 + 2\omega \lambda_l)}} \underbrace{d\omega \cdots d\omega}_{n \times} \Bigg|_{\omega=0}. \tag{3.41}$$

¹⁰A nossa abordagem para a análise do IMPNLMS acaba por ser bem mais custosa computacionalmente do que a similar já efetuada para o NLMS porque os autovalores de $\mathbf{E}[\Gamma_k] \mathbf{R}_x$ passam a diferir em cada iteração, devido à matriz $\mathbf{\Gamma}_k$.

As raízes de $\det(\mathbf{I} + 2\omega\mathbf{E}[\mathbf{\Gamma}_k]\mathbf{R}_x)$ são $\omega_l = -\frac{1}{2\lambda_l}$. Uma mera expansão do determinante mostra que o coeficiente do termo ω^L é dado por $2^L \prod_{l=1}^L \lambda_l$, o que nos permite obter:

$$\bar{E}_{g,n} = \int \dots \underbrace{\int}_{n \times} \frac{(-1)^n}{\sqrt{\left(2^L \prod_{l=1}^L \lambda_l\right) (\omega - \omega_1)(\omega - \omega_2) \dots (\omega - \omega_L)}} \underbrace{d\omega \dots d\omega}_{n \times} \Bigg|_{\omega=0}. \quad (3.42)$$

O fato de, como já explicado, a integral acima não apresentar solução explícita compele-nos a empregar alguma aproximação. Optamos por utilizar a estratégia de [55], substituindo raízes adjacentes do polinômio por uma raiz única (ω'_q , com multiplicidade 2), obtida pela média geométrica das raízes adjacentes¹¹:

$$\omega'_q = -\sqrt{\omega_{2q-1}\omega_{2q}} \quad (3.43)$$

Com esta aproximação, chegamos a:

$$\bar{E}_{g,n} \approx \frac{(-1)^n}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \int \dots \int \underbrace{\int}_{n \times} \frac{\overbrace{1}^{\wp(\omega)}}{(\omega - \omega'_1)(\omega - \omega'_2) \dots (\omega - \omega'_{L/2})} \underbrace{d\omega \dots d\omega}_{n \times} \Bigg|_{\omega=0}. \quad (3.44)$$

Seja o termo $\wp(\omega)$. Façamos sua expansão em frações parciais:

$$\wp(\omega) = \frac{1}{(\omega - \omega'_1)(\omega - \omega'_2) \dots (\omega - \omega'_{L/2})} = \frac{A_1}{\omega - \omega'_1} + \frac{A_2}{\omega - \omega'_2} + \dots + \frac{A_{L/2}}{\omega - \omega'_{L/2}} \quad (3.45)$$

Logo:

$$\bar{E}_{g,n} \approx \frac{(-1)^n}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \int \dots \int \underbrace{\int}_{n \times} \left(\frac{A_1}{\omega - \omega'_1} + \frac{A_2}{\omega - \omega'_2} + \dots + \frac{A_{L/2}}{\omega - \omega'_{L/2}} \right) \underbrace{d\omega \dots d\omega}_{n \times} \Bigg|_{\omega=0}, \quad (3.46)$$

onde

$$A_q = \frac{1}{\prod_{j=1, j \neq q}^{L/2} (\omega'_q - \omega'_j)}. \quad (3.47)$$

¹¹Sendo ω_q sempre negativo, ω'_q também deve sê-lo; daí a razão do sinal negativo.

Para $n = 1, 2$ e 3 , obtemos¹²:

$$\begin{aligned}\bar{E}_{g,1} &\approx \frac{-1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \int \left(\frac{A_1}{\omega - \omega'_1} + \frac{A_2}{\omega - \omega'_2} + \dots + \frac{A_{\frac{L}{2}}}{\omega - \omega'_{\frac{L}{2}}} \right) d\omega \Big|_{\omega=0} \\ &= \frac{-1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \ln(\omega - \omega'_l) \Big|_{\omega=0} = \frac{-1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \ln(-\omega'_l)\end{aligned}\quad (3.48)$$

$$\begin{aligned}\bar{E}_{g,2} &\approx \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \iint \left(\frac{A_1}{\omega - \omega'_1} + \frac{A_2}{\omega - \omega'_2} + \dots + \frac{A_{\frac{L}{2}}}{\omega - \omega'_{\frac{L}{2}}} \right) d\omega d\omega \Big|_{\omega=0} \\ &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l [\omega'_l - \omega - \omega'_l \ln(\omega - \omega'_l) + \omega \ln(\omega - \omega'_l)] \Big|_{\omega=0} \\ &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l [\omega'_l - \omega'_l \ln(-\omega'_l)]\end{aligned}\quad (3.49)$$

$$\begin{aligned}\bar{E}_{g,3} &\approx \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \iiint \left(\frac{A_1}{\omega - \omega'_1} + \frac{A_2}{\omega - \omega'_2} + \dots + \frac{A_{\frac{L}{2}}}{\omega - \omega'_{\frac{L}{2}}} \right) d\omega d\omega d\omega \Big|_{\omega=0} \\ &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \left[\frac{\omega'^2_l \ln(\omega - \omega'_l)}{2} + \frac{3\omega'_l \omega}{2} + \frac{\omega^2 \ln(\omega - \omega'_l)}{2} - \frac{3\omega^2}{4} - \omega'_l \omega \ln(\omega - \omega'_l) \right] \Big|_{\omega=0} \\ &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \left[\frac{\omega'^2_l \ln(-\omega'_l)}{2} \right]\end{aligned}\quad (3.50)$$

As Equações (3.48)-(3.50), em conjunção com (3.31) e (3.32), apresentam as fórmulas necessárias para o cálculo mais refinado dos termos $E_{g,n}$.

Os termos acima calculados foram expressos em função de valores esperados das iterações predecentes. Para a inicialização do cálculo recursivo, arbitramos $E[\hat{h}_0(i)] = \mu_0(i) = 0$ e $E[\hat{h}_0^2(i)] = \delta_1$, onde δ_1 é um valor muito pequeno e diferente de zero, empregado por razões de precisão numérica.

As principais equações recursivas propostas para a análise de convergência do algoritmo IMPNLMS estão listadas na Tab. 3.1.

¹²Não encontramos na literatura nenhum exemplo de emprego analítico do caso $n = 3$ para a função $\Psi(\omega)$.

Tabela 3.1: Principais equações recursivas da análise de convergência do algoritmo IMPNLNS.

$$\begin{aligned}
\mathbb{E} \left[\xi_{\hat{\mathbf{h}}_k} \right] &= \frac{L}{L-\sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} \mathbb{E}[\left| \hat{h}_k(j) \right|]}{\sqrt{L \sum_{j=0}^{L-1} \mathbb{E}[\hat{h}_k^2(j)]}} \right) \\
\mathbb{E} [\xi(k)] &= (1 - \lambda_\xi) \mathbb{E} [\xi(k-1)] + \lambda_\xi \mathbb{E} \left[\xi_{\hat{\mathbf{h}}_k} \right] \\
\mathbb{E} [\alpha(k)] &= 2\mathbb{E} [\xi(k)] - 1 \\
\mathbb{E} [g_k(i)] &= \frac{1-\mathbb{E}[\alpha(k)]}{2L} + \frac{(1+\mathbb{E}[\alpha(k)])\mathbb{E}[F(|\hat{h}_k(i)|)]}{2 \sum_{j=0}^{L-1} \mathbb{E}[F(|\hat{h}_k(j)|)] + \epsilon} \\
\mathbb{E} [\hat{\mathbf{z}}_{k+1}] &= \mathbb{E} [\hat{\mathbf{z}}_k] - \beta \mathbb{E}_{g,1} \mathbf{\Gamma}_k \mathbf{R}_x \mathbb{E} [\hat{\mathbf{z}}_k] \\
\mathbb{E} [\hat{z}_{k+1}(i) \hat{z}_{k+1}(l)] &= \mathbb{E} [\hat{z}_k(i) \hat{z}_k(l)] - \beta \mathbb{E}_{g,1} g_k(l) \sum_{j=0}^{L-1} \mathbb{E} [\hat{z}_k(j) \hat{z}_k(i)] r(l-j) \\
&\quad + \beta^2 \mathbb{E}_{g,2} \mathbb{E} [g_k(i)] \mathbb{E} [g_k(l)] r(i-l) \sigma_\nu^2 \\
&\quad - \beta \mathbb{E}_{g,1} \mathbb{E} [g_k(i)] \sum_{n=0}^{L-1} \mathbb{E} [\hat{z}_k(n) \hat{z}_k(l)] r(i-n) \\
&\quad + \beta^2 \mathbb{E}_{g,2} \mathbb{E} [g_k(i)] \mathbb{E} [g_k(l)] \sum_{j=0}^{L-1} \sum_{n=0}^{L-1} \mathbb{E} [\hat{z}_k(j) \hat{z}_k(n)] r_{i,j,l,n} \\
\Phi_0(a, b, \mu_k(i), \sigma_k(i)) &= \sigma_k(i) \sqrt{\frac{\pi}{2}} \left[\operatorname{erfc} \left(\frac{a-\mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) - \operatorname{erfc} \left(\frac{b-\mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) \right] \\
\Phi_1(a, b, \mu_k(i), \sigma_k(i)) &= \sigma_k^2(i) \left[e^{-\frac{(a-\mu_k(i))^2}{2\sigma_k^2(i)}} - e^{-\frac{(b-\mu_k(i))^2}{2\sigma_k^2(i)}} \right] \\
&\quad + \mu_k(i) \sigma_k(i) \sqrt{\frac{\pi}{2}} \left[\operatorname{erfc} \left(\frac{a-\mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) - \operatorname{erfc} \left(\frac{b-\mu_k(i)}{\sqrt{2}\sigma_k(i)} \right) \right] \\
\mathbb{E} \left[\left| \hat{h}_k(i) \right| \right] &= \frac{\Phi_1(0, \infty, \mu_k(i), \sigma_k(i))}{\sqrt{2\pi}\sigma_k(i)} + \frac{\Phi_1(0, \infty, -\mu_k(i), \sigma_k(i))}{\sqrt{2\pi}\sigma_k(i)} \\
\mathbb{E} \left[F \left(\left| \hat{h}_k(i) \right| \right) \right] &= \frac{1}{\sqrt{2\pi}\sigma} [400\Phi_1(5 \times 10^{-3}, \mu, \sigma) + 400\Phi_1(5 \times 10^{-3}, -\mu, \sigma) \\
&\quad + 8, 51\Phi_1(5 \times 10^{-3}, \infty, \mu, \sigma) + 8, 51\Phi_1(5 \times 10^{-3}, \infty, -\mu, \sigma) \\
&\quad + 1, 96\Phi_0(5 \times 10^{-3}, \infty, \mu, \sigma) + 1, 96\Phi_0(5 \times 10^{-3}, \infty, -\mu, \sigma)] \\
E_{g,1} &= \bar{E}_{g,1} - \delta \bar{E}_{g,2} \\
E_{g,2} &= \bar{E}_{g,2} - 2\delta \bar{E}_{g,3} \\
\bar{E}_{g,1} &= \frac{-1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \ln(-\omega'_l) \\
\bar{E}_{g,2} &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l [\omega'_l - \omega'_l \ln(-\omega'_l)] \\
\bar{E}_{g,3} &= \frac{1}{\sqrt{2^L \prod_{l=1}^L \lambda_l}} \sum_{l=1}^{L/2} A_l \left[\frac{\omega_l'^2 \ln(-\omega'_l)}{2} \right]
\end{aligned}$$

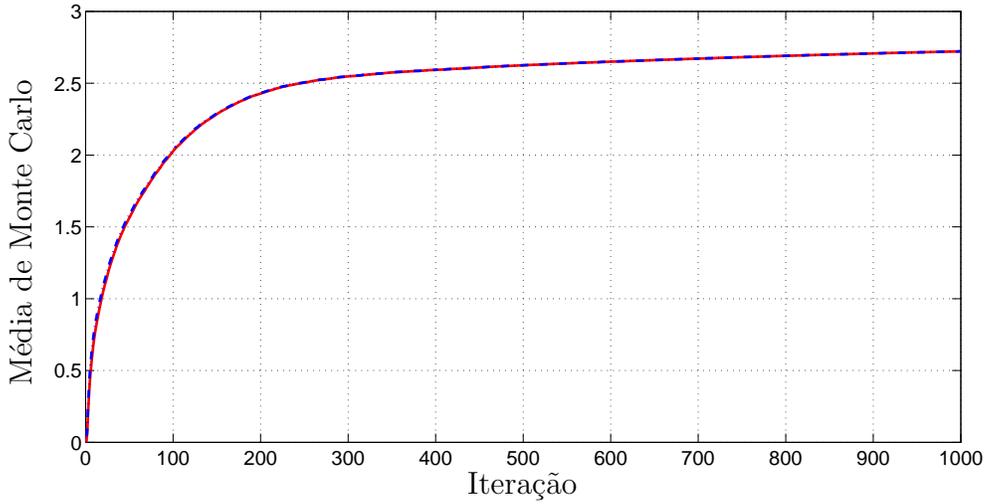


Figura 3.1: Comparação experimental dos termos $E \left[\sqrt{\left(L \sum_{j=0}^{L-1} \widehat{h}_k^2(j) \right)} \right]$ (em vermelho) e $\sqrt{\left(L \sum_{j=0}^{L-1} E \left[\widehat{h}_k^2(j) \right] \right)}$ (aproximação oriunda da Hipótese III, em azul pontilhado e tracejado). Os resultados são oriundos de 1000 médias de Monte Carlo.

3.2.1 Simulações

Para avaliar experimentalmente a acurácia das hipóteses empregadas na análise de transiente do algoritmo IMPNLMS, recorreremos a uma configuração na qual um filtro adaptativo almeja identificar a 5ª função de transferência da recomendação da ITU-T G.168 [63], que apresenta 96 coeficientes. O filtro adaptativo apresenta idêntico número de coeficientes e o sinal de entrada é obtido pela passagem de um sinal branco e gaussiano pelo filtro $\frac{0,25}{1-1,5z^{-1}+z^{-2}-0,25z^{-3}}$. Os demais parâmetros da simulação do algoritmo IMPNLMS são: 1000 iterações, $\beta = 0,5$, $\delta = 0,01$, $\epsilon = 0,001$, $\lambda_\xi = 0,1$ e $\xi(-1) = 0,96$. As Figs. 3.1-3.5 comparam as médias dos termos originais e as suas aproximações obtidas por recurso a algumas das hipóteses utilizadas. Estas figuras revelam uma razoável concordância destas hipóteses com os valores experimentais.

Para ilustrar a maior exatidão da nossa análise do comportamento do transiente do MSE do algoritmo IMPNLMS, 5 diferentes funções de transferência a identificar foram testadas. Obtivemo-las a partir de trechos dos modelos de função de transferência da recomendação da ITU-T G.168 para cancelamento de eco [63]. As Tabs. 3.2 e 3.3 apresentam as configurações de cada caso simulado.

A análise proposta estima o termo $E_{g,2}$ (Eq. (3.32)), sem recorrer à aproximação $E_{g,2} \approx [E_{g,1}]^2$. Para verificar o efeito desta aproximação, dois modelos foram testados: proposta sem a aproximação (modelo 1) e proposta com a aproximação (modelo 2). As Figs. 3.6-3.10 comparam as evoluções do MSE experimental com os teóricos. Nota-se que o modelo 1 é ligeiramente melhor do que o modelo 2, sendo a

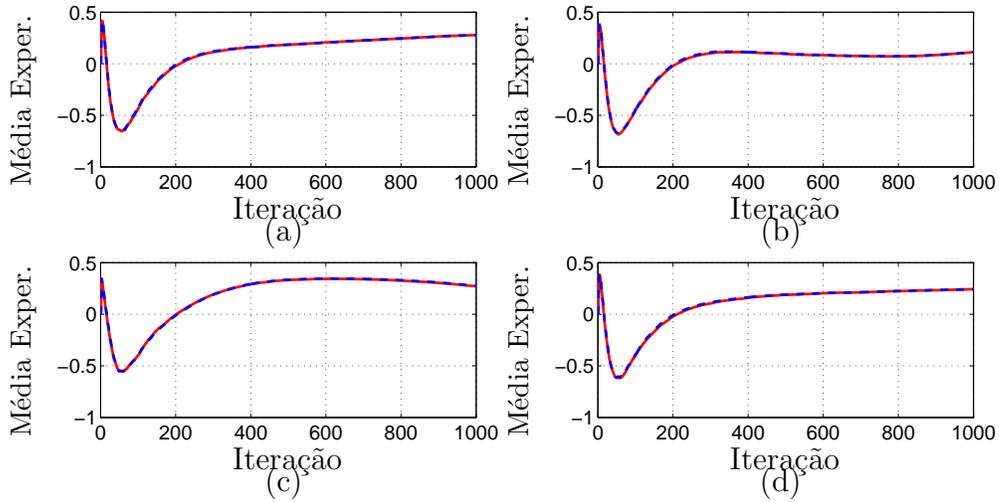


Figura 3.2: Comparação do termo $E[\alpha(k)F(|\hat{h}_k(i)|)]$ (em vermelho) com a aproximação (oriunda da Hipótese IV) $E[\alpha(k)]E[F(|\hat{h}_k(i)|)]$ (em azul tracejado), ao longo das iterações. (a) $i = 10$; (b) $i = 20$; (c) $i = 30$; (d) $i = 40$. Os resultados são oriundos de 1000 médias de Monte Carlo.

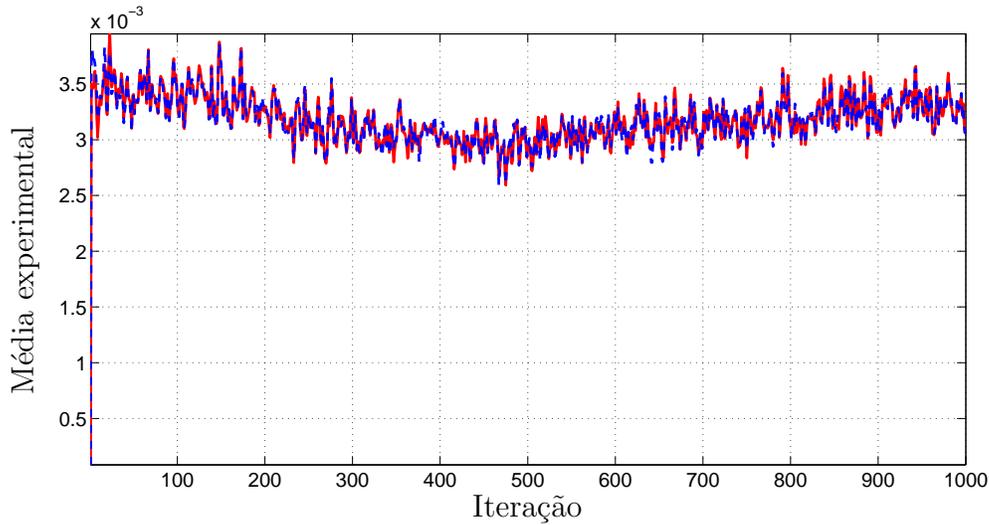


Figura 3.3: Comparação do termo $E[g_k(10)x^2(k-30)]$ (em vermelho) com sua aproximação (obtida mediante a Hipótese V) $E[g_k(10)]E[x^2(k-30)]$ (em azul tracejado). Os resultados são oriundos de 1000 médias de Monte Carlo.

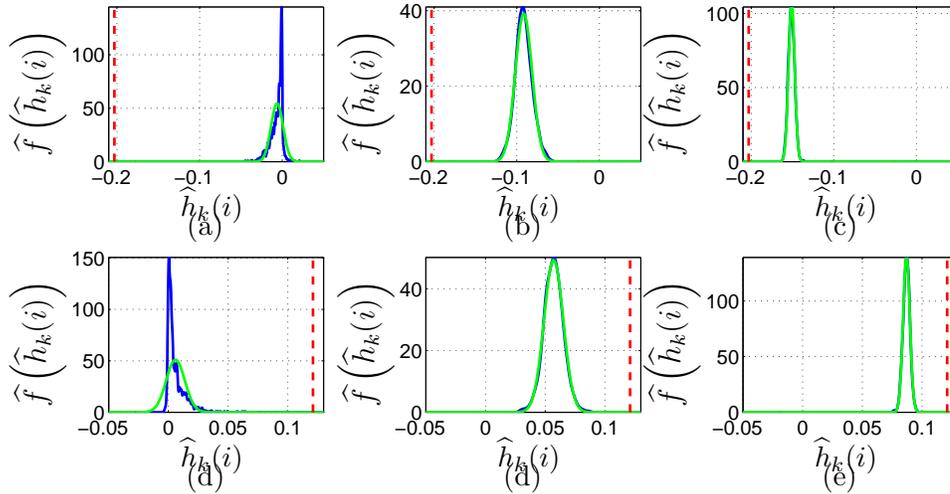


Figura 3.4: Função densidade de probabilidade, estimada via *kernels* gaussianos, do coeficiente adaptativo $\hat{h}_k(i)$ (em azul). (a) $k = 10$ e $i = 29$; (b) $k = 200$ e $i = 29$; (c) $k = 1000$ e $i = 29$; (d) $k = 10$ e $i = 31$; (e) $k = 200$ e $i = 31$; (f) $k = 1000$ e $i = 31$. Os resultados são oriundos de 1000 simulações. Os valores ideais destes coeficientes encontram-se assinalados pela linha tracejada vermelha. Em verde, apresenta-se uma distribuição gaussiana com média e desvio padrão idênticos aos das amostras empregadas para estimativa da função densidade de probabilidade. A Hipótese VI supõe que a distribuição destes coeficientes é gaussiana.

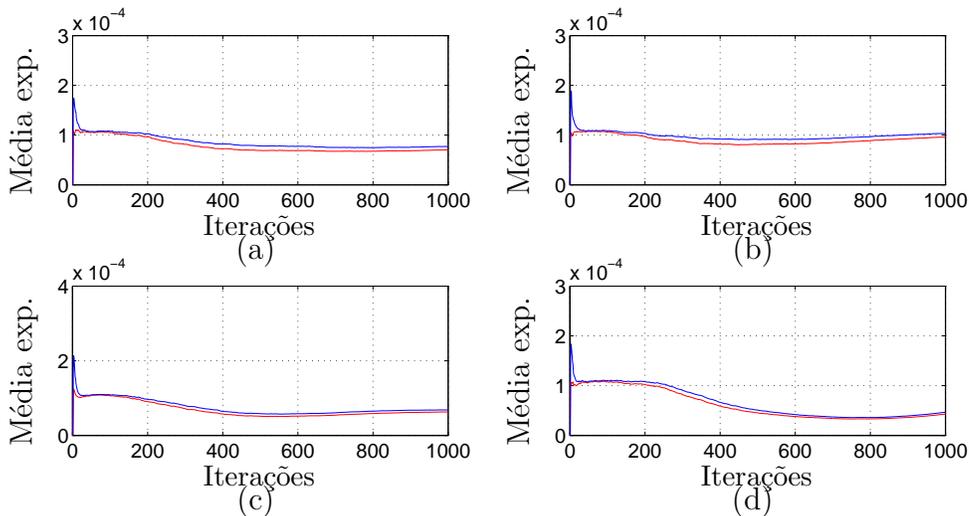


Figura 3.5: Comparação da média dos termos $E[g_k(i)^2]$ (em azul) e sua aproximação (oriunda da Hipótese VII) $E[g_k(i)]E[g_k(i)]$. (a) $i = 5$; (b) $i = 10$; (c) $i = 15$ e (d) $i = 20$. Os resultados são oriundos de 1000 médias de Monte Carlo.

Tabela 3.2: Configurações testadas nas simulações. Em todos os casos, os valores de δ , ϵ , λ_ξ e $\xi(-1)$ foram, respectivamente, 10^{-2} , 10^{-3} , 10^{-1} e 0,96.

Caso	Nº Modelo	Amostra inicial	Amostra final	L	β
1	5	20	35	16	0,25
2	1	1	20	20	0,25
3	2	1	10	10	0,25
4	3	1	25	25	0,2
5	4	1	30	30	0,2

Tabela 3.3: Configurações do sinal de entrada $x(k)$ e do ruído de medição $\nu(k)$. Em todos os casos, o sinal $x(k)$ foi obtido através do colorimento de um sinal branco e gaussiano de variância $\sigma_x^2 = 1$.

Caso	σ_ν^2	Filtro de colorimento (comando equivalente do MATLAB)
1	10^{-6}	<code>fir2(5,[0 .25 .5 .75 1],[1 1 0.5 0 0])</code>
2	10^{-5}	<code>fir2(10,[0 .25 .5 .75 1],[1 1 0.5 0 0])</code>
3	10^{-6}	<code>fir2(24,[0 .25 .5 .75 1],[1 0.75 0.5 0 1])</code>
4	10^{-6}	<code>fir2(36,[0 .25 .5 .75 1],[1 0.5 0.25 0.8 .2])</code>
5	10^{-6}	<code>fir2(36,[0 .25 .5 .75 1],[1 0.5 0.25 0.8 .2])</code>

diferença pouco significativa. Como o modelo 1 é mais pesado computacionalmente que o modelo 2, a melhor relação custo \times exatidão parece portanto se localizar no modelo 2. O fato de o comprimento do filtro adaptativo ser baixo significa que a aproximação padrão (que emprega a hipótese $E\left[\frac{A}{B}\right] \approx \frac{E[A]}{E[B]}$) é pouco precisa para a análise do estado transiente do MSE. Embora em tese a análise de transiente possa servir para uma estimativa do MSE em regime permanente para um número alto de iterações, a acumulação dos erros ao longo das iterações torna esta abordagem pouco precisa¹³ (vide, por exemplo, a Fig. 3.8). O fato de o sinal de entrada ser colorido implica uma convergência bem mais lenta do MSE. Na literatura, são escassos os exemplos de análise de transiente de algoritmos de IRE com entradas coloridas. A análise elaborada neste capítulo não compartilha esta restrição.

3.3 Estimativa do MSE Final de algoritmos de IRE

O valor de MSE para o qual o algoritmo tende após a convergência é uma importante medida de desempenho. A seguir, utilizando a técnica do balanceamento de energia

¹³Como veremos, a proposta de estimativa do MSE em regime permanente via balanceamento de energia não está sujeita a esta imprecisão.

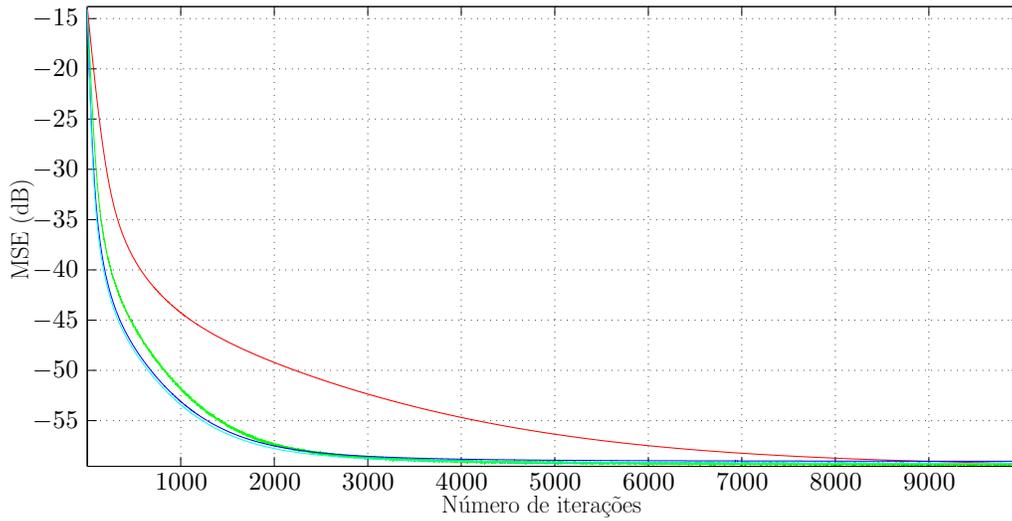


Figura 3.6: Evolução do MSE (em dB) da primeira configuração (caso 1). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).

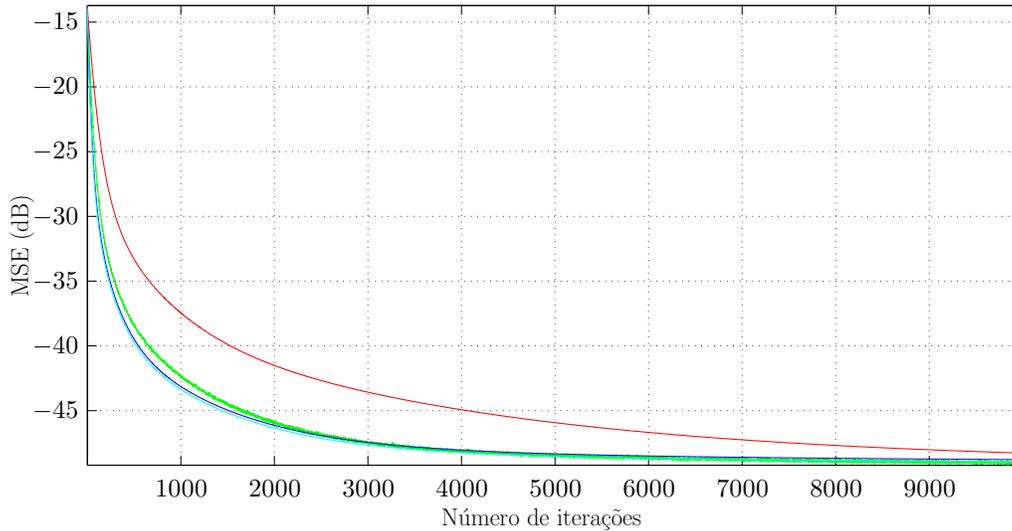


Figura 3.7: Evolução do MSE (em dB) da segunda configuração (caso 2). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).

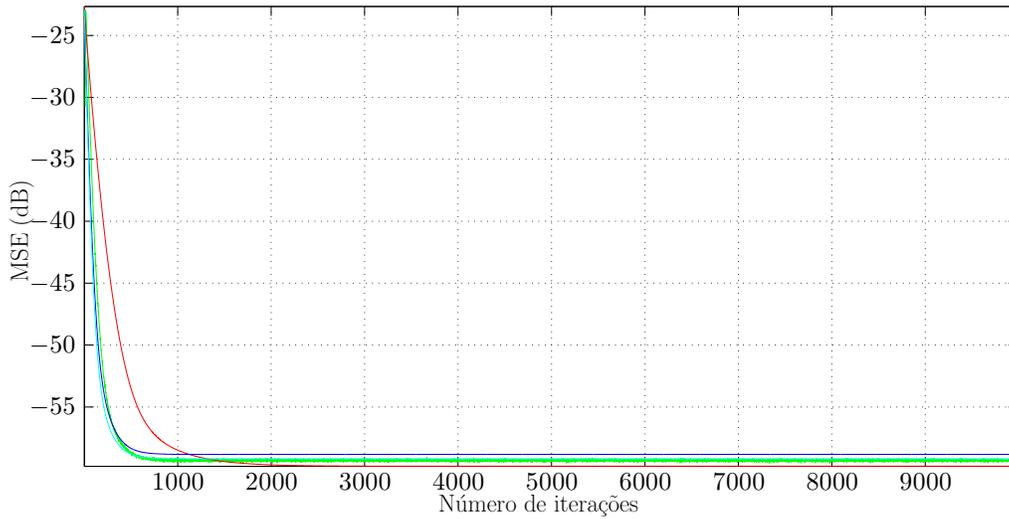


Figura 3.8: Evolução do MSE (em dB) da terceira configuração (caso 3). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em azul claro: evolução teórica empregando a técnica proposta (modelo 2).

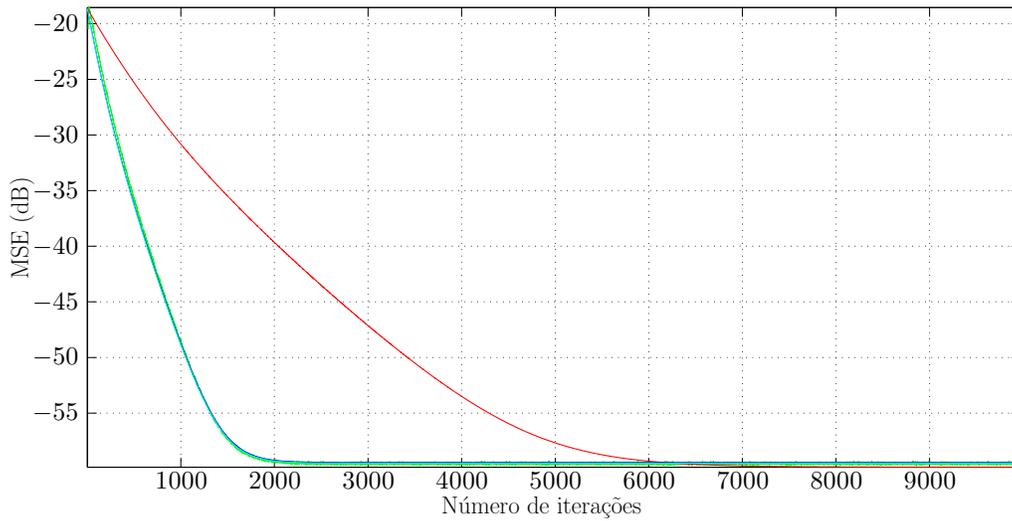


Figura 3.9: Evolução do MSE (em dB) da quarta configuração (caso 4). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em ciano: evolução teórica empregando a técnica proposta (modelo 2).

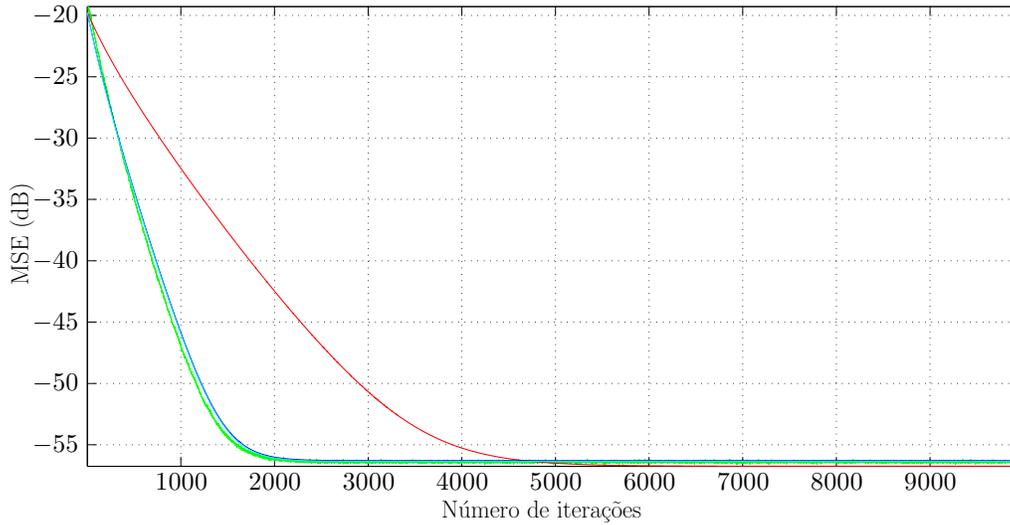


Figura 3.10: Evolução do MSE (em dB) da quinta configuração (caso 5). Em vermelho: evolução teórica (empregando a aproximação padrão) do MSE; em verde: MSE experimental (oriundo de 10000 médias de Monte Carlo); em azul: evolução teórica empregando a técnica proposta (modelo 1); em ciano: evolução teórica empregando a técnica proposta (modelo 2).

[64], obteremos uma fórmula teórica para o MSE final de todos os algoritmos da família PNLMS.

As hipóteses utilizadas na obtenção do MSE em regime permanente não são restritivas:

- Hipótese I: O filtro adaptativo não diverge, estabilizando-se quando $k \rightarrow \infty$.
- Hipótese II: o ruído $\nu(k)$ apresenta média nula e independe de $\mathbf{x}_j, \forall j$.
- Hipótese III: No $\lim_{k \rightarrow \infty}, \frac{1}{\|\mathbf{x}_k\|^2}$ e $e_a(k)$ são variáveis aleatórias decorrelacionadas¹⁴.
- Hipótese IV: A ordem do filtro adaptativo supera ou iguala a ordem do filtro a identificar e durante o transiente o fator de aprendizagem associado a cada coeficiente não é nulo.

A equação de atualização de qualquer algoritmo da família PNLMS é:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \frac{\beta \mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}. \quad (3.51)$$

¹⁴O erro *a priori* $e_a(k)$ será definido logo a seguir

Expressando-a com relação aos desvios $\widehat{\mathbf{z}}_k$ e $\widehat{\mathbf{z}}_{k+1}$, obtemos:

$$\widehat{\mathbf{z}}_{k+1} = \widehat{\mathbf{z}}_k + \frac{\beta \mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}. \quad (3.52)$$

Definamos os erros *a priori* $e_a(k)$ e *a posteriori* $e_p(k)$ como¹⁵:

$$e_a(k) = -\widehat{\mathbf{z}}_k \mathbf{x}_k, \quad (3.53)$$

$$e_p(k) = -\widehat{\mathbf{z}}_{k+1} \mathbf{x}_k. \quad (3.54)$$

Após multiplicar os termos da Eq. (3.52) à direita por \mathbf{x}_k , encontramos:

$$e_p(k) = e_a(k) - \beta e(k), \quad (3.55)$$

De onde deriva:

$$e(k) = \frac{e_a(k) - e_p(k)}{\beta}. \quad (3.56)$$

Lançando o resultado de (3.56) em (3.52), temos:

$$\widehat{\mathbf{z}}_{k+1} + \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e_p(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} = \widehat{\mathbf{z}}_k + \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e_a(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}. \quad (3.57)$$

Calculando a energia do lado esquerdo (E_e) da Eq. (3.57), obtemos:

$$\begin{aligned} E_e &= \left(\widehat{\mathbf{z}}_{k+1} + \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e_p(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} \right) \left(\widehat{\mathbf{z}}_{k+1} + \frac{\mathbf{\Gamma}_k \mathbf{x}_k e_p(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} \right) \\ &= \underbrace{\widehat{\mathbf{z}}_{k+1} \widehat{\mathbf{z}}_{k+1}^T}_{\|\widehat{\mathbf{z}}_{k+1}\|^2} + \frac{\widehat{\mathbf{z}}_{k+1} \mathbf{\Gamma}_k \mathbf{x}_k e_p(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} + \frac{e_p(k) \mathbf{x}_k^T \mathbf{\Gamma}_k \widehat{\mathbf{z}}_{k+1}}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} + \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k^2 \mathbf{x}_k e_p^2(k)}{\underbrace{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}_{\|\mathbf{x}_k\|^2}} \quad (3.58) \\ &= \|\widehat{\mathbf{z}}_{k+1}\|^2 - \frac{\widehat{\mathbf{z}}_{k+1} \mathbf{\Gamma}_k \mathbf{x}_k \mathbf{x}_k^T \widehat{\mathbf{z}}_{k+1}}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} - \frac{\widehat{\mathbf{z}}_{k+1} \mathbf{x}_k \mathbf{x}_k^T \mathbf{\Gamma}_k \widehat{\mathbf{z}}_{k+1}}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k} + \frac{e_p^2(k)}{\|\mathbf{x}_k\|^2}. \end{aligned}$$

Fazendo $k \rightarrow \infty$ e aplicando o valor esperado, chegamos a:

$$\mathbb{E}[E_e] = \mathbb{E}[\|\widehat{\mathbf{z}}_\infty\|^2] - \mathbb{E}\left[\frac{\widehat{\mathbf{z}}_\infty \mathbf{\Gamma}_\infty \mathbf{x}_k \mathbf{x}_k^T \widehat{\mathbf{z}}_\infty}{\mathbf{x}_k^T \mathbf{\Gamma}_\infty \mathbf{x}_k}\right] - \mathbb{E}\left[\frac{\widehat{\mathbf{z}}_\infty \mathbf{x}_k \mathbf{x}_k^T \mathbf{\Gamma}_\infty \widehat{\mathbf{z}}_\infty}{\mathbf{x}_k^T \mathbf{\Gamma}_\infty \mathbf{x}_k}\right] + \mathbb{E}\left[\frac{e_p^2(\infty)}{\|\mathbf{x}_k\|^2}\right]. \quad (3.59)$$

¹⁵Para manter compatibilidade com a nossa notação, nosso desenvolvimento apresenta algumas diferenças em relação à análise por fluxo de energia proposta em [64], o qual de todo modo não analisou os algoritmos da família PNLMS.

Efetuando um procedimento semelhante no lado direito da Eq. (3.57), pode-se encontrar:

$$E[E_d] = E[|\widehat{\mathbf{z}}_\infty|^2] - E\left[\frac{\widehat{\mathbf{z}}_\infty \mathbf{\Gamma}_\infty \mathbf{x}_k \mathbf{x}_k^T \widehat{\mathbf{z}}_\infty^T}{\mathbf{x}_k^T \mathbf{\Gamma}_\infty \mathbf{x}_k}\right] - E\left[\frac{\widehat{\mathbf{z}}_\infty \mathbf{x}_k \mathbf{x}_k^T \mathbf{\Gamma}_\infty \widehat{\mathbf{z}}_\infty^T}{\mathbf{x}_k^T \mathbf{\Gamma}_\infty \mathbf{x}_k}\right] + E\left[\frac{e_a^2(\infty)}{|\mathbf{x}_k|^2}\right]. \quad (3.60)$$

Até o momento, os resultados são exatos, não sendo empregadas quaisquer aproximações. Por meio da Hipótese I, ao igualar as energias dos lados direito e esquerdo de (3.57) obtemos:

$$\lim_{k \rightarrow \infty} E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] = \lim_{k \rightarrow \infty} E\left[\frac{e_p^2(k)}{|\mathbf{x}_k|^2}\right]. \quad (3.61)$$

A partir da relação $e(k) = e_a(k) + \nu(k)$ e da Eq. (3.56), pode-se escrever:

$$e_p(k) = (1 - \beta)e_a(k) - \beta\nu(k). \quad (3.62)$$

De onde obtemos o termo $e_p^2(k)$:

$$e_p^2(k) = (1 - \beta)^2 e_a^2(k) - 2\beta(1 - \beta)e_a(k)\nu(k) + \beta^2 \nu^2(k), \quad (3.63)$$

cujo resultado, empregado na Eq. (3.61), nos fornece:

$$\lim_{k \rightarrow \infty} E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] = \lim_{k \rightarrow \infty} \left\{ (1 - \beta)^2 E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] - 2\beta(1 - \beta) E\left[\frac{e_a(k)\nu(k)}{|\mathbf{x}_k|^2}\right] + \beta^2 E\left[\frac{\nu^2(k)}{|\mathbf{x}_k|^2}\right] \right\}. \quad (3.64)$$

Utilizando a Hipótese II, temos:

$$\lim_{k \rightarrow \infty} E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] = \lim_{k \rightarrow \infty} \left\{ (1 - \beta)^2 E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] + \beta^2 \sigma_\nu^2 E\left[\frac{1}{|\mathbf{x}_k|^2}\right] \right\}, \quad (3.65)$$

onde σ_ν^2 é a variância de $\nu(k)$.

Empregando a Hipótese III, temos que $E\left[\frac{e_a^2(k)}{|\mathbf{x}_k|^2}\right] = E[e_a^2(k)] E\left[\frac{1}{|\mathbf{x}_k|^2}\right]$, o que simplifica a equação acima para:

$$\lim_{k \rightarrow \infty} E[e_a^2(k)] = \frac{\beta \sigma_\nu^2}{2 - \beta}. \quad (3.66)$$

Utilizando a Hipótese IV, o termo $\lim_{k \rightarrow \infty} E[e_a^2(k)]$ é o erro quadrático médio em excesso (ou seja, acima da variância do ruído) do algoritmo em regime permanente [64] e portanto:

$$\lim_{k \rightarrow \infty} E[e^2(k)] = \sigma_\nu^2 + \frac{\beta \sigma_\nu^2}{2 - \beta} = \frac{\sigma_\nu^2}{1 - \frac{\beta}{2}}. \quad (3.67)$$

O resultado da Eq. (3.66) coincide com o apresentado em [36], por meio de outra técnica de análise. Porém, nossa derivação apresenta duas vantagens: é mais simples

e emprega menos hipóteses. Em [36], duas hipóteses adicionais foram utilizadas: a de que \mathbf{x}_k é gaussiano e que $\beta \ll 1$ ou $g_k(j) \ll 1, \forall j$. A nossa demonstração revela que tais hipóteses não são necessárias. As restrições que as Hipóteses I e IV exigem quanto aos fatores de aprendizagem e $g_k(j)$ são triviais ou mais fracas do que as exigidas em [36].

A Eq. (3.67) significa que o MSE final de qualquer algoritmo da família PNLMS praticamente independe da distribuição da “energia” de atualização empreendida pela matriz $\mathbf{\Gamma}$ e que as diferentes escolhas desta matriz influenciam basicamente a taxa de convergência do MSE.

3.4 Análise do MSE em Regime Permanente de uma Estrutura em Subbandas Aplicada a Algoritmos de IRE

Sinais de voz, como muitos outros sinais de interesse, são coloridos. Tal fato pode implicar certas vantagens das estruturas em subbandas, em particular quando empregamos um algoritmo de gradiente descendente normalizado na atualização dos coeficientes em cada subfiltro [27]. Este trabalho pretende explorar a flexibilidade advinda de uma estrutura não uniforme, a qual, por ter se revelado promissora, será a seguir descrita e analisada.

O artigo [65] propõe uma estrutura não uniforme em subbandas para os algoritmos de IRE tratados por este capítulo (ressaltamos que o leque de aplicações desta estrutura não se restringe a técnicas adaptativas de identificação). Esta estrutura conjuga transformadas *wavelet* com filtros adaptativos esparsos em cada subbanda, adequadamente escolhidos de modo a identificar qualquer RI de comprimento M . A estrutura será aqui denominada de WT-SF (do inglês *Wavelet Transform with Sparse Filters*). A estrutura não utiliza decimação, o que evita quaisquer problemas oriundos de *aliasing*, consistindo numa versão não-uniforme da estrutura preconizada por [66], com algumas modificações; a principal delas é a inserção de atrasos nos filtros de análise, o que permite abandonar a restrição de que todos devem ter a mesma ordem.

A Fig. 3.11 ilustra a estrutura WT-SF, a qual emprega um banco de filtros não-uniforme com filtros de análise¹⁶ $\underline{H}_n(z)$ e filtros adaptativos esparsos $\hat{H}_{n,k}(z^{L_n})$. Para uma *wavelet* cujas bandas estão divididas em oitavas, os filtros de análise

¹⁶A variável n refere-se à subbanda e a variável k ao número da iteração; observe que os filtros de análise são fixos e portanto independem de k .

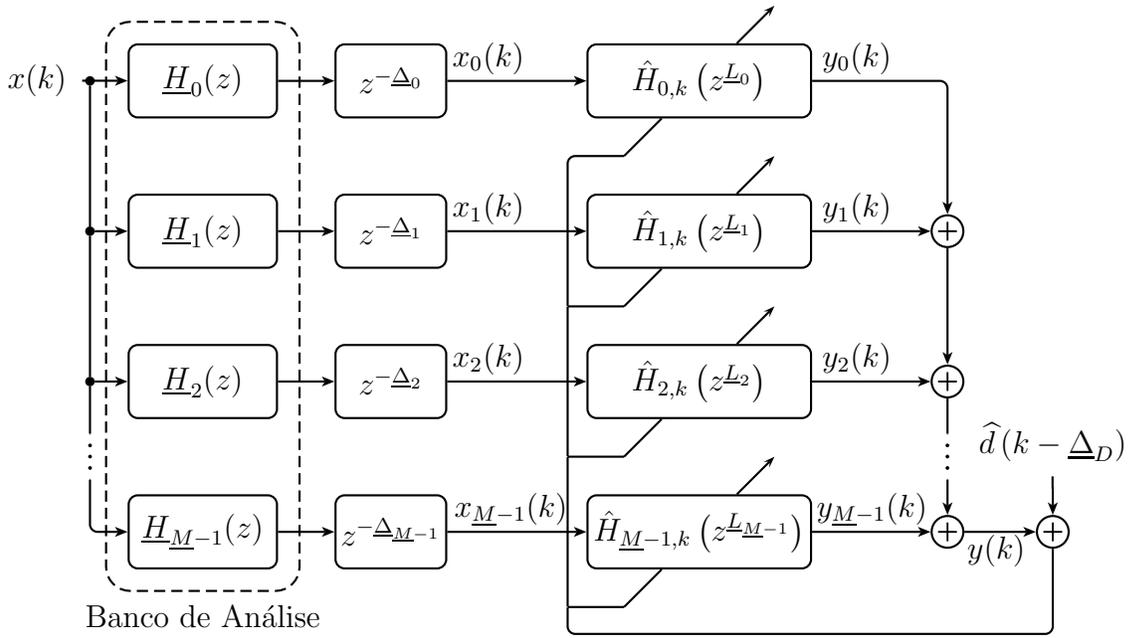


Figura 3.11: Estrutura WT-SF com banco de filtros não-uniforme e filtros adaptativos esparsos.

equivalentes do banco de filtros com \underline{M} subbandas são [67]:

$$\underline{H}_0(z) = \prod_{j=0}^{\underline{M}-2} \underline{H}^0(z^{2^j}),$$

$$\underline{H}_n(z) = \underline{H}^1(z^{2^{\underline{M}-1-k}}) \prod_{j=0}^{\underline{M}-k-2} \underline{H}^0(z^{2^j}), n = 1, \dots, \underline{M} - 1, \quad (3.68)$$

onde $\underline{H}^0(z)$ e $\underline{H}^1(z)$ são, respectivamente, os filtros passa-baixa e passa-alta associados às funções *wavelet* [67]. Os fatores de esparsidade são:

$$\underline{L}_0 = 2^{\underline{M}-1}, \quad \underline{L}_n = 2^{\underline{M}-n}, n = 1, \dots, \underline{M} - 1. \quad (3.69)$$

Os atrasos $\underline{\Delta}_n$ que constam na Fig. 3.11 são dados por $\underline{\Delta}_n = M_{\underline{H}_0} - M_{\underline{H}_n}$, onde $M_{\underline{H}_n}$ é o comprimento do n -ésimo filtro de análise. Para a modelagem de um filtro FIR de comprimento M , o número de coeficientes adaptativos (não-nulos) de $\hat{H}_{n,k}(z^{\underline{L}_n})$ deve ser pelo menos igual a $M_n = \left\lceil \frac{M+M_{F_n}}{\underline{L}_n} \right\rceil + 1$, onde M_{F_n} é o comprimento do n -ésimo filtro de síntese [65]. A estrutura WT-SF não inclui os filtros de síntese; os filtros adaptativos encarregam-se também da reconstrução.

Mantendo fixo o fator de aprendizagem, o emprego conjugado da estrutura WT-SF com um algoritmo da família PNLMS gera uma convergência do MSE mais rápida do que quando utilizamos o algoritmo correspondente na sua formulação pa-

drão, conforme demonstram as simulações presentes em [68].

Uma vantagem adicional de se utilizar a estrutura em subbandas advém do fato de que a contribuição de cada um dos subfiltros $\widehat{H}_{n,k}(z^{L_n})$ está basicamente restrita à faixa de frequências de uma subbanda, o que permite, em algumas aplicações, a eliminação de subbandas que não contribuam significativamente para a resposta em frequência global [66]. Uma desvantagem da estrutura WT-SF é a inserção do atraso $\underline{\Delta}_D = M_{\underline{H}_0}$, que é nulo nos algoritmos padrão da família PNLMS.

Para uma compreensão mais ampla dos efeitos do emprego da estrutura WT-SF (em comparação com a implementação correspondente em banda cheia), analisaremos, por meio da técnica de balanceamento de energia [64], o MSE final do algoritmo WTPNLMS-SF (algoritmo PNLMS que emprega a estrutura WT-SF), apresentado na Tab. 3.4. As diferentes possibilidades de escolha das matrizes $\mathbf{\Gamma}_{n,k}$ dependem do algoritmo de IRE utilizado com a estrutura WT-SF (os resultados se aplicam a qualquer algoritmo da família PNLMS e mesmo para o NLMS, desde que, neste último caso, arbitremos as matrizes $\mathbf{\Gamma}_{n,k}$ como sendo matrizes identidades). Como a análise contempla todos os algoritmos da família PNLMS, não é necessário empregar uma definição específica de $\mathbf{\Gamma}_{n,k}$.

Para obtermos o balanceamento de energia do algoritmo WTPNLMS-SF, cumpre escrever todas as suas atualizações numa única equação:

$$\widehat{\mathbf{h}}_{a_{k+1}} = \widehat{\mathbf{h}}_{a_k} + \beta \mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (||\mathbf{x}_{a_k}||_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e(k), \quad (3.70)$$

onde:

$$\widehat{\mathbf{h}}_{a_j} = \begin{bmatrix} \widehat{\mathbf{h}}_{0,j} & \widehat{\mathbf{h}}_{1,j} & \cdots & \widehat{\mathbf{h}}_{\underline{M}-1,j} \end{bmatrix}, \quad (3.71)$$

$$\mathbf{\Gamma}_{a_k} = \begin{bmatrix} \mathbf{\Gamma}_{0,k} & \mathbf{0}_{M_0,M_1} & \cdots & \mathbf{0}_{M_0,M_{\underline{M}-1}} \\ \mathbf{0}_{M_1,M_0} & \mathbf{\Gamma}_{1,k} & \cdots & \mathbf{0}_{M_1,M_{\underline{M}-1}} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{M_{\underline{M}-1},M_0} & \mathbf{0}_{M_{\underline{M}-1},M_1} & \cdots & \mathbf{\Gamma}_{\underline{M}-1,k} \end{bmatrix}, \quad (3.72)$$

$$\mathcal{H} = \begin{bmatrix} \mathcal{H}_0 & \mathcal{H}_1 & \cdots & \mathcal{H}_{\underline{M}-1} \end{bmatrix}, \quad (3.73)$$

$$\mathcal{H}_i = \begin{bmatrix} \mathbf{h}_{i,0} & \mathbf{h}_{i,1} & \cdots & \mathbf{h}_{i,\underline{M}-1} \end{bmatrix}, \quad (3.74)$$

$$\mathbf{h}_{i,j} = \left[\mathbf{0}_{1 \times \underline{\Delta}_i + (j-1)\underline{L}_i} \quad h_i(0) \quad h_i(1) \quad \cdots \quad h_i(M_{H_i}) \quad \mathbf{0}_{1 \times (M_i-j)\underline{L}_i + M_{H_i}} \right]^T, \quad (3.75)$$

Tabela 3.4: Algoritmo WTPNLMS-SF

Inicialização (valores típicos)

$$\delta = 0,01$$

$$\hat{\mathbf{h}}_{n,0} = \left[\hat{\mathbf{h}}_n(0) \quad \hat{\mathbf{h}}_n(1) \quad \cdots \quad \hat{\mathbf{h}}_n(M_n - 1) \right] = \mathbf{0}$$

Processamento e Adaptação

Para $k = 0, 1, 2, \dots$

Para $n = 0, 1, \dots, \underline{M} - 1$

$$x_n(k) = \sum_{i=0}^{M_{H_n}-1} h_n(i)x(k-i)$$

$$\mathbf{x}_{n,k} = \left[x_n(k - \underline{\Delta}_n) \quad x_n(k - \underline{\Delta}_n - \underline{L}_n) \quad \cdots \quad x_n(k - \underline{\Delta}_n - (M_n - 1)\underline{L}_n) \right]^T$$

$$y_n(k) = \hat{\mathbf{h}}_{n,k} \mathbf{x}_{n,k}$$

Fim

$$y(k) = \sum_{n=0}^{\underline{M}-1} y_n(k)$$

$$e(k) = \hat{d}(k - \underline{\Delta}_D) - y(k)$$

Para $n = 0, 1, \dots, \underline{M} - 1$

Cálculo de $\mathbf{\Gamma}_{n,k}$

$$\hat{\mathbf{h}}_{n,k+1} = \hat{\mathbf{h}}_{n,k} + \beta \frac{\mathbf{x}_{n,k}^T \mathbf{\Gamma}_{n,k} e(k)}{\mathbf{x}_{n,k}^T \mathbf{\Gamma}_{n,k} \mathbf{x}_{n,k} + \delta}$$

Fim

Fim

$$\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 =$$

$$\begin{bmatrix} \mathbf{x}_{a_k}^T \mathcal{H}_0 \mathbf{\Gamma}_{0,k} \mathcal{H}_0^T \mathbf{x}_{a_k} \mathbf{I}_{M_0} & \mathbf{0}_{M_0, M_0} & \cdots & \mathbf{0}_{M_0, M_{M-1}} \\ \mathbf{0}_{M_1, M_0} & \mathbf{x}_{a_k}^T \mathcal{H}_1 \mathbf{\Gamma}_{1,k} \mathcal{H}_1^T \mathbf{x}_{a_k} \mathbf{I}_{M_1} & \cdots & \mathbf{0}_{M_1, M_{M-1}} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{M_{M-1}, M_0} & \mathbf{0}_{M_{M-1}, M_1} & \cdots & \mathbf{x}_{a_k}^T \mathcal{H}_{M_{M-1}} \mathbf{\Gamma}_{M-1,k} \mathcal{H}_{M_{M-1}}^T \mathbf{x}_{a_k} \mathbf{I}_{M_{M-1}, M-1} \end{bmatrix}, \quad (3.76)$$

$$\mathbf{x}_{a_k} = \left[x(k) \quad x(k-1) \quad x(k-2) \quad \cdots \quad x(k - \max_n [\underline{\Delta}_n + (M_n - 1)\underline{L}_n + M_{\underline{H}_n}]) \right]^T, \quad (3.77)$$

$$e(k) = \hat{d}(k) - y(k) = d(k) - \hat{\mathbf{h}}_{a_k}^T \mathcal{H}^T \mathbf{x}_{a_k} + \nu(k), \quad (3.78)$$

onde \mathbf{I}_n , $\mathbf{0}_{n \times m}$ e $\mathbf{0}_n$ são, respectivamente, a matriz identidade de ordem n , a matriz nula de dimensões $n \times m$ e a matriz nula de dimensões $n \times n$.

Seja \mathbf{h}_{a_k} o vetor ótimo. Se $\tilde{\mathbf{h}}_{a_k} = \mathbf{h}_{a_k} - \hat{\mathbf{h}}_{a_k}$, a atualização do WTPNLMS-SF pode ser expressa em função dos desvios $\tilde{\mathbf{h}}_{a_k}$:

$$\tilde{\mathbf{h}}_{a_{k+1}} = \tilde{\mathbf{h}}_{a_k} - \beta \mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e(k). \quad (3.79)$$

Multiplicando à direita por $\mathcal{H}^T \mathbf{x}_{a_k}$, temos:

$$\overbrace{\tilde{\mathbf{h}}_{a_{k+1}}^T \mathcal{H}^T \mathbf{x}_{a_k}}^{e_p(k)} = \overbrace{\tilde{\mathbf{h}}_{a_k}^T \mathcal{H}^T \mathbf{x}_{a_k}}^{e_a(k)} - \beta \mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k} e(k), \quad (3.80)$$

$$\Rightarrow e_p(k) = e_a(k) - \beta \mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k} e(k), \quad (3.81)$$

$$\Rightarrow e(k) = \frac{e_a(k) - e_p(k)}{\beta \mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k}}, \quad (3.82)$$

um resultado que, aplicado à Eq. (3.79), nos permite obter:

$$\tilde{\mathbf{h}}_{a_{k+1}} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e_a(k)}{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k}} = \tilde{\mathbf{h}}_{a_k} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e_p(k)}{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k}} \quad (3.83)$$

Já que $\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k} = \sum_{i=0}^{M-1} \frac{\mathbf{x}_{a_k}^T \mathcal{H}_i \mathbf{\Gamma}_{i,k} \mathcal{H}_i^T \mathbf{x}_{a_k}}{\mathbf{x}_{a_k}^T \mathcal{H}_i \mathbf{\Gamma}_{i,k} \mathcal{H}_i^T \mathbf{x}_{a_k}} = \underline{M}$, podemos simplificar a equação acima para:

$$\tilde{\mathbf{h}}_{a_{k+1}} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e_a(k)}{\underline{M}} = \tilde{\mathbf{h}}_{a_k} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} e_p(k)}{\underline{M}} \quad (3.84)$$

A energia do lado esquerdo da Eq. (3.84) é:

$$\begin{aligned}
E_e &= \tilde{\mathbf{h}}_{a_{k+1}} \tilde{\mathbf{h}}_{a_{k+1}}^T \\
&+ \frac{\tilde{\mathbf{h}}_{a_{k+1}} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a(k)}{\underline{M}} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \tilde{\mathbf{h}}_{a_{k+1}}^T e_a(k)}{\underline{M}} \\
&+ \frac{\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a^2(k)}{\underline{M}^2} \\
&= \|\tilde{\mathbf{h}}_{a_{k+1}}\|^2 + \frac{\tilde{\mathbf{h}}_{a_{k+1}} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} \mathbf{x}_{a_k}^T \mathcal{H} \tilde{\mathbf{h}}_{a_{k+1}}^T}{\underline{M}} \\
&+ \frac{\tilde{\mathbf{h}}_{a_k}^T \mathcal{H}^T \mathbf{x}_{a_k} \mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \tilde{\mathbf{h}}_{a_{k+1}}}{\underline{M}} + \frac{\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a^2(k)}{\underline{M}^2}
\end{aligned} \tag{3.85}$$

Por meio de uma sequência semelhante de cálculos, obtemos a energia do lado direito da Eq. (3.84):

$$\begin{aligned}
E_d &= \|\tilde{\mathbf{h}}_{a_k}\|^2 + \frac{\tilde{\mathbf{h}}_{a_k} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} \mathbf{x}_{a_k}^T \mathcal{H} \tilde{\mathbf{h}}_{a_{k+1}}^T}{\underline{M}} \\
&+ \frac{\tilde{\mathbf{h}}_{a_{k+1}} \mathcal{H}^T \mathbf{x}_{a_k} \mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \tilde{\mathbf{h}}_{a_k}^T}{\underline{M}} \\
&+ \frac{\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_p^2(k)}{\underline{M}^2}
\end{aligned} \tag{3.86}$$

Para $k \rightarrow \infty$ e supondo que houve convergência (de forma similar ao que fizemos ao analisar o MSE do algoritmo PNLMS em banda cheia), encontramos:

$$\begin{aligned}
&\lim_{k \rightarrow \infty} \mathbb{E} \left[\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a^2(k) \right] \\
&= \lim_{k \rightarrow \infty} \mathbb{E} \left[\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_p^2(k) \right].
\end{aligned} \tag{3.87}$$

Já que $e_p(k) = (1 - \beta \underline{M})e_a(k) - \beta \underline{M}\nu(k)$, sabemos que:

$$e_p^2(k) = e_a^2(k) - 2\beta \underline{M} e_a^2(k) + \beta^2 \underline{M}^2 e_a^2(k) - 2\beta(1 - \beta \underline{M})e_a(k)\nu(k) + \beta^2 \underline{M}^2 \nu^2(k), \tag{3.88}$$

cujo resultado, lançado na Eq. (3.87) (e utilizando a Hipótese de $\nu(k)$ de média zero e independente de \mathbf{x}_{a_k} e $e_a(k)$), acaba por gerar:

$$\begin{aligned}
&\lim_{k \rightarrow \infty} \beta^2 \underline{M}^2 \sigma_\nu^2 \mathbb{E} \left[\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} \right] \\
&= \lim_{k \rightarrow \infty} 2\beta \underline{M} \mathbb{E} \left[\mathbf{x}_{a_k}^T \mathcal{H} \Gamma_{a_k} \left[\left(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2 \right)^{-1} \right]^2 \Gamma_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a^2(k) \right]
\end{aligned}$$

$$- \beta^2 \underline{M}^2 \mathbb{E} \left[\mathbf{x}_{a_k}^T \mathcal{H} \mathbf{\Gamma}_{a_k} \left[(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \right]^2 \mathbf{\Gamma}_{a_k} \mathcal{H}^T \mathbf{x}_{a_k} e_a^2(k) \right] \quad (3.89)$$

Supondo que \mathbf{x}_{a_k} e $e_a(k)$ sejam independentes, encontramos:

$$(2\beta \underline{M} - \beta^2 \underline{M}^2) \mathbb{E} [e_a^2(k)] = \beta^2 \underline{M} \sigma_\nu^2, \quad (3.90)$$

o que, colocando em evidência $e_a^2(k)$, fornece¹⁷:

$$\mathbb{E} [e_a^2(k)] = \frac{\beta \underline{M} \sigma_\nu^2}{2 - \beta \underline{M}}. \quad (3.91)$$

Este resultado, como vimos, é uma estimativa teórica do erro quadrático médio em excesso (EMSE, do inglês *Excess Mean-Squared Error*). Cabe ressaltar a independência do EMSE em relação à escolha dos filtros de análise, bem como (da mesma forma que os algoritmos da família PNLMS em banda cheia) em relação aos elementos da diagonal da matriz $\mathbf{\Gamma}_{a_k}$.

3.4.1 Simulações

Para ilustrar o ganho de convergência quando empregamos a estrutura em subbandas preconizada por [65], apresentamos resultados de uma simulação. O sinal de entrada gerado é gaussiano e branco, de variância unitária, que é colorido pela função de transferência $H_e(z)$:

$$H_e(z) = \frac{0,25}{1 - 1,5z^{-1} + z^{-2} - 0,25z^{-3}}. \quad (3.92)$$

Como algoritmo padrão para comparação, escolhemos o NLMS regularizado. Para acelerar a convergência da identificação de respostas esparsas, optamos pelo algoritmo IMPNLMS [37], utilizado tanto na sua versão original quanto na estrutura WT-SF com 2 e 3 subbandas. A função de transferência a identificar foi obtida a partir do quarto modelo da recomendação ITU [63], adicionando-se 128 zeros ao seu final, de modo que seu comprimento total é igual a 256. Utilizando na estrutura WT-SF a wavelet biortogonal 4.4, a Fig. 3.12 apresenta a evolução do MSE¹⁸, revelando o efeito benéfico do emprego da estrutura WT-SF no tocante à convergência.

¹⁷As hipóteses empregadas nesta análise são exatamente as mesmas da Seção 3.3, desde que substituamos \mathbf{x}_k por \mathbf{x}_{a_k} ; devido à semelhança das derivações não explicitamos nesta seção as hipóteses empregadas.

¹⁸Suavizamos o MSE substituindo cada 400 valores consecutivos por sua média. Foi empregado um ruído de medição aditivo, gaussiano e de variância $\sigma_\nu^2 = 10^{-6}$. O algoritmo NLMS teve parâmetro de regularização arbitrado como $\delta = 0,01$. Os parâmetros do algoritmo IMPNLMS utilizados foram $\delta = 0,01$, $\epsilon = 0,001$, $\lambda_\xi = 0,1$ e $\xi = 0,96$. Para os algoritmos NLMS e IMPNLMS, o fator de aprendizagem utilizado foi $\beta = 0,9$. Para o algoritmo WTIMPNLMS-SF foi empregado $\beta = 0,45$ para $\underline{M} = 2$ e $\beta = 0,3$ para $\underline{M} = 3$.

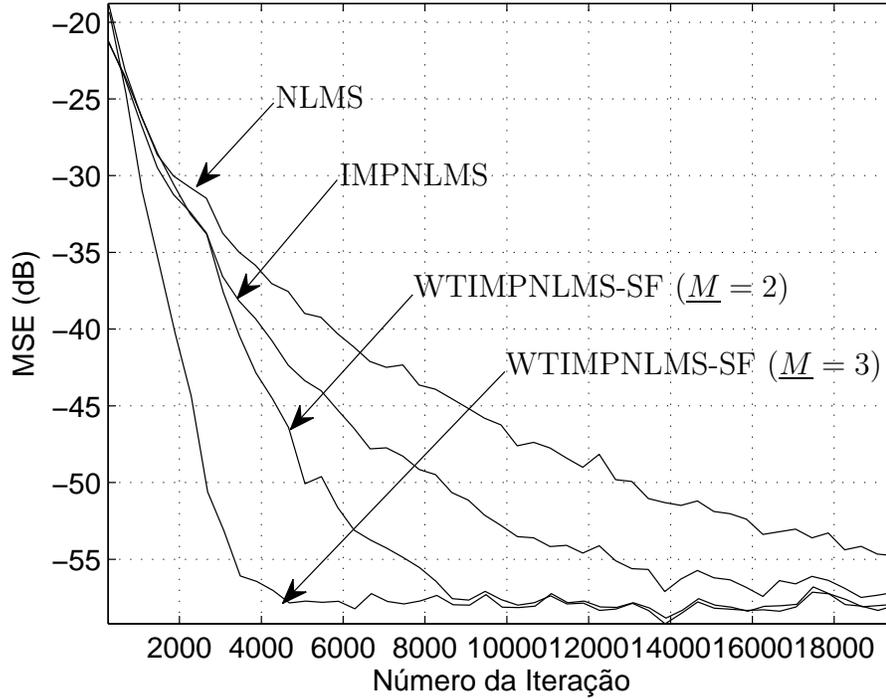


Figura 3.12: Evolução do MSE de diversos algoritmos de identificação de respostas esparsas.

Utilizando como resposta ao impulso a identificar a do segundo modelo da recomendação ITU [63] para cancelamento de eco, que apresenta comprimento 96, empregando como *wavelet* a Daubechie 8 e alterando o número de subbandas \underline{M} e o fator de aprendizagem, comparamos o MSE experimental com o teórico na Fig. 3.13-(a). Observamos grande concordância entre a teoria e o resultado experimental, especialmente para $\underline{M} \leq 3$, situação mais comum, pois em nossas simulações o emprego de mais do que 3 subbandas não resulta em ganhos substanciais de convergência.

3.5 Estabilidade do Algoritmo

No caso do algoritmo NLMS padrão, o EMSE é $\frac{\beta\sigma_v^2}{2-\beta}$. O fato de que o fator de aprendizagem apareça multiplicado por \underline{M} na Eq. (3.91) apresenta duas implicações importantes: a primeira é que o MSE da estrutura WT-SF é maior do que o da versão em banda cheia, já que $\left| \frac{\beta M \sigma_v^2}{2-\beta M} \right| > \left| \frac{\beta \sigma_v^2}{2-\beta} \right|$ para $\underline{M} > 1$. Quanto maior o número de subbandas \underline{M} , maior é a perda do MSE no regime permanente. Este resultado tem uma interpretação simples: os filtros adaptativos podem ser compreendidos como \underline{M} estimadores em paralelo, e a estimativa resultante tem um erro quadrático médio amplificado pelo número de estimadores. Em última análise, este resultado deriva

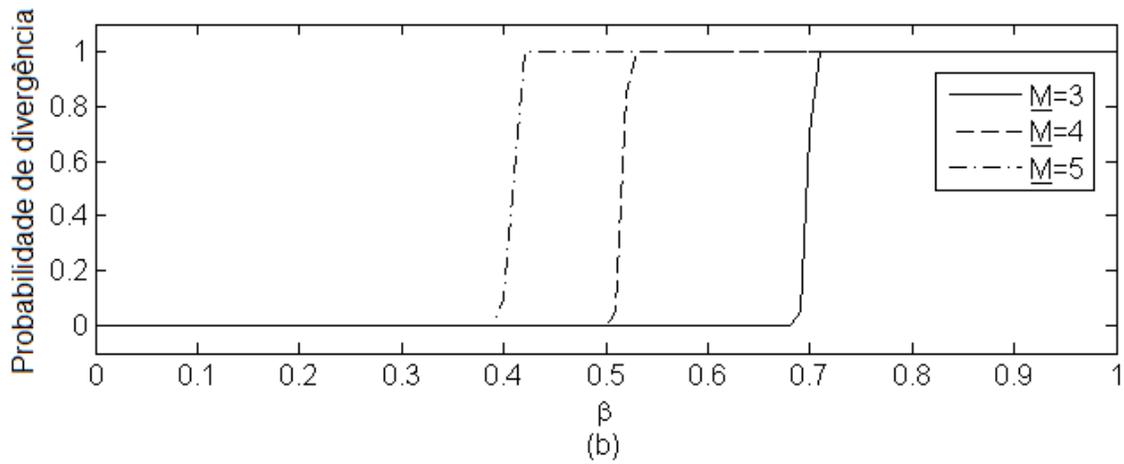
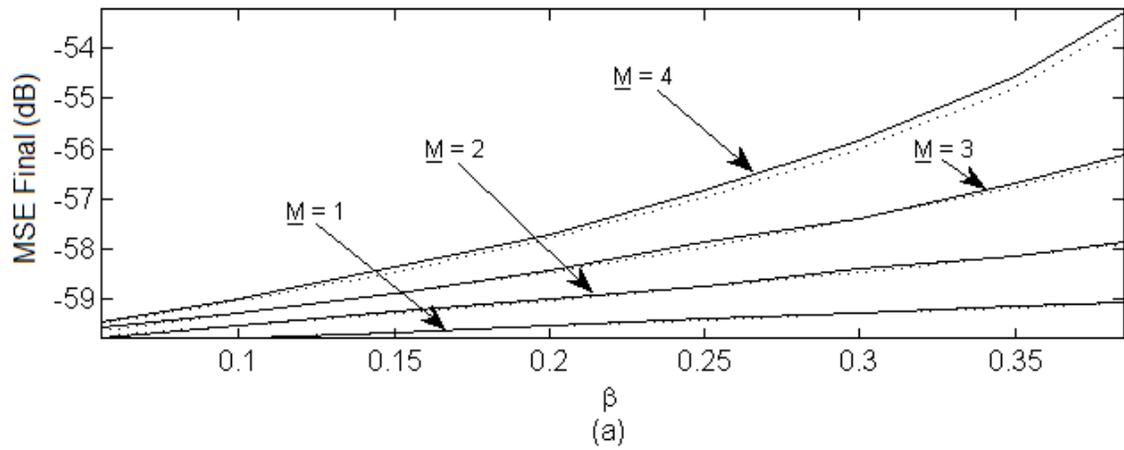


Figura 3.13: (a) MSE em regime permanente do algoritmo WTIMPNLMS-SF: linhas sólidas representam resultados experimentais e as tracejadas refletem os resultados teóricos; (b) Probabilidade de divergência.

do fato de que $\mathbf{x}_{a_k}^T \mathcal{H} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k} = \underline{M}$. Com efeito, se alterarmos a definição de $(\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1}$ de modo a que tenhamos $\mathbf{x}_{a_k}^T \mathcal{H} (\|\mathbf{x}_{a_k}\|_{\mathcal{H}\mathcal{H}^T}^2)^{-1} \mathcal{H}^T \mathbf{x}_{a_k} = 1$, o MSE em estado permanente passa a ser muito semelhante (empiricamente) ao do algoritmo equivalente em banda cheia. Mesmo nesse caso, a convergência da estrutura WT-SF é superior ao do algoritmo em subbanda.

A análise de estabilidade do algoritmo (limite superior de β que evita a divergência) não pode ser efetuada a partir da análise proposta na Seção 3.4, já que o EMSE teórico (Eq. (3.91)) foi obtido a partir da suposição de que houve convergência. Porém, a Eq. (3.91) pode servir como indicativo de que a faixa de β que evita a divergência se restringe à medida que aumentamos \underline{M} . Isto pode ser visto por meio da constatação de que $E[e_a^2(k)]$ é não-negativo, o que imporia $2 - \beta \underline{M} > 0$ e $\beta < \frac{2}{\underline{M}}$. Ressalte-se que esta conclusão não foi obtida de modo rigoroso, devendo portanto ser posta à prova por meio de simulações. Independentemente da precisão do limitante superior $\frac{2}{\underline{M}}$, realmente se verifica que a faixa permitida de valores de β se restringe com o aumento de \underline{M} . De todo modo, este limitante superior menor não é um grave problema, pois mesmo a divisão de β por \underline{M} na estrutura em subbandas ainda a mantém vantajosa (no contexto de sinais coloridos) com relação ao algoritmo em banda cheia.

3.5.1 Simulações

Na Fig. 3.13-(b), medimos a probabilidade de divergência do algoritmo WTIMPNLMS-SF em função de \underline{M} e β , utilizando-se o primeiro modelo da recomendação [63] e a *wavelet* Haar. Apenas 20 médias de Monte Carlo foram necessárias para este experimento. Os valores testados de β variam de 0,01 a 1 com incrementos de 0,01; a Fig. 3.13-(b) revela¹⁹ que o limite do superior para o fator de aprendizagem que evita a divergência é razoavelmente bem definido, num valor próximo ao obtido teoricamente ($2/\underline{M}$), confirmando a acurácia da análise teórica. Note que mesmo para $\underline{M} = 4$ o limite superior para o fator de aprendizagem (0,5) é razoavelmente elevado ($\beta = 0,25$ é um valor típico), significando que a redução provocada pela estrutura WTPNLMSF-SF no intervalo de valores de β admissíveis não é demasiado restritiva para o emprego desta. Na realidade, verifica-se que dividir o valor de β do algoritmo padrão por \underline{M} implica um algoritmo em subbandas com MSE em regime permanente idêntico, com a vantagem de possuir uma maior taxa de convergência.

¹⁹Nesta figura, o ruído de medição foi modelado como aditivo, gaussiano e de variância $\sigma_v^2 = 10^{-6}$. Em (b) foram calculadas as probabilidades de divergência para $\underline{M} = 3$, $\underline{M} = 4$ e $\underline{M} = 5$.

3.6 Fatores de Aprendizagem Variantes no Tempo

Os algoritmos da família PNLMS proporcionam fatores de aprendizagem diferentes para coeficientes adaptativos distintos; a manutenção dos valores dos coeficientes adaptativos implica a constância dos fatores de aprendizagem. Porém é sabido que a variação dos fatores de aprendizagem ao longo das iterações é uma estratégia que pode manter convergência rápida sem comprometer o desajuste em regime permanente, desde que eles assumam valores relativamente altos durante o transiente, reduzindo-se à medida que o filtro aproxima-se do regime permanente.

A aplicação deste tipo de técnica ao algoritmo WTIMPNLMS-SF permitiria manter o acréscimo da taxa de convergência e, concomitantemente, reduzir o desajuste. Por isso a inserção de variações temporais de β torna-se uma estratégia relevante, a qual tem sido proposta por diversos artigos, entre os quais [32] angariou interesse por apresentar uma forma fácil de controle (sem a necessidade de constantes adicionais serem arbitradas), intitulada VSS (do inglês *Variable Step Size*). [32] considera esta técnica como uma generalização de uma estratégia de filtragem *Set-Membership* [69].

Para que a equação de atualização resultante adira ao formato da maioria das técnicas proporcionais, iremos adaptar a derivação de [32] para o nosso contexto. A seguir, obteremos a versão VSS dos algoritmos PNLMS. A extensão para os algoritmos WTPNLMS-SF é trivial.

A equação de atualização dos algoritmos da família PNLMS é, por motivos de clareza, rerepresentada a seguir:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta_k \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}, \quad (3.93)$$

onde o índice k em β_k indica que o fator de aprendizagem é dependente da iteração.

Subtraindo \mathbf{h} de ambos os lados e multiplicando à direita por \mathbf{x}_k , obtemos:

$$\left(\hat{\mathbf{h}}_{k+1} - \mathbf{h}\right) \mathbf{x}_k = \left(\hat{\mathbf{h}}_k - \mathbf{h}\right) \mathbf{x}_k + \beta_k e(k). \quad (3.94)$$

Subtraindo $\nu(k)$ de ambos os lados, temos:

$$\hat{e}_p(k) = e(k) - \beta_k e(k) = (1 - \beta_k)e(k), \quad (3.95)$$

onde $\hat{e}_p(k) = \overbrace{(\mathbf{h} - \hat{\mathbf{h}}_{k+1}) \mathbf{x}_k}^{e_p(k)} + \nu(k)$ é o erro *a posteriori* medido. Elevando a equação acima ao quadrado, aplicando o valor esperado e escolhendo β_k de modo que

E $[\hat{e}_p^2(k)] = \sigma_\nu^2$, chegamos à seguinte equação de segundo grau:

$$\beta_k^2 - 2\beta_k + \left(1 - \frac{\hat{\sigma}_\nu^2}{\hat{\sigma}_e^2}\right) = 0, \quad (3.96)$$

onde $\hat{\sigma}_e^2$ e $\hat{\sigma}_\nu^2$ são estimativas atualizáveis das variâncias do erro e do ruído de medição. A solução fisicamente plausível da equação acima é:

$$\beta_k = 1 - \frac{\hat{\sigma}_\nu}{\hat{\sigma}_e}, \quad (3.97)$$

na qual costuma-se inserir um fator constante β de controle da taxa de convergência²⁰, de modo que os algoritmos VSS-PNLMS têm sua equação de atualização descrita por:

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta \left(1 - \frac{\hat{\sigma}_\nu}{\hat{\sigma}_e}\right) \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k}. \quad (3.98)$$

A Tab. 3.5 apresenta os detalhes de nossa implementação do algoritmo VSS-IMP NLMS, incluindo a parte de estimativa de $\hat{\sigma}_e$. Em nossas simulações, a estimativa do desvio padrão do ruído $\hat{\sigma}_\nu$ será igual ao desvio padrão real σ_ν ; na prática, recorre-se a estimativas deste desvio padrão (como por exemplo a proposta de [70]), as quais podem ser obtidas durante trechos de silêncio [32].

A técnica VSS pode ser facilmente generalizada para o algoritmo WTIMP NLMS-SF, desde que multipliquemos β pelo fator $\left(1 - \frac{\hat{\sigma}_\nu}{\hat{\sigma}_e}\right)$ (vide Eq. (3.98)). Este fator, ainda que inferior, é praticamente igual a 1 quando $\hat{\sigma}_e$ é muito maior do que $\hat{\sigma}_\nu$, o que ocorre durante a maior parte do transiente. Em regime permanente, temos que $\hat{\sigma}_e \approx \hat{\sigma}_\nu$, de sorte que este fator é praticamente nulo, gerando um desajuste muito reduzido e um MSE em regime permanente muito próximo do menor valor possível (σ_ν^2). Efetuaremos a seguir avaliações empíricas dos algoritmos VSS-IMP NLMS e VSS-WTIMP NLMS-SF.

3.6.1 Simulações

Almejando contornar o aumento do desajuste em regime permanente do algoritmo WTIMP NLMS-SF, implementamos as versões VSS (com fatores de aprendizagem variantes ao longo das iterações) deste algoritmo. Utilizando sinal de entrada um sinal branco e gaussiano de variância unitária que passa por um filtro cuja função de transferência é $\frac{0,25}{1-1,5z^{-1}+z^{-2}-0,25z^{-3}}$ e como RI a identificar o modelo 6 da recomendação da ITU [63], a qual apresenta 120 coeficientes, testamos diversos algoritmos apresentados neste capítulo. Para os algoritmos derivados do IMP NLMS, arbitrou-se $\sigma_\nu^2 = 10^{-6}$, $\delta = 0,01$, $\epsilon = 0,001$, $\lambda_\xi = 0,1$, $\xi = 0,96$ e $L = 120$. Para o NLMS,

²⁰[32] propõe $\beta = 1$.

Tabela 3.5: Algoritmo VSS-IMPNLMS

Inicialização (valores típicos)

$$\delta = 0,01, \quad \epsilon = 0,001, \quad \beta = 0,25, \quad \lambda_\xi = 0,1, \quad \xi(-1) = 0,96$$

$$\hat{\mathbf{h}}_0 = \left[\hat{\mathbf{h}}_0(0) \quad \hat{\mathbf{h}}_0(1) \quad \cdots \quad \hat{\mathbf{h}}_0(L-1) \right] = \mathbf{0}$$

$$\hat{\sigma}_e^2(-1) = 0, \quad \text{variância estimada do erro}$$

$$\hat{\sigma}_\nu^2, \quad \text{variância estimada do ruído}$$

Processamento e Adaptação

Para $k = 0, 1, 2, \dots$

$$\mathbf{x}_k = \left[x(k) \quad x(k-1) \quad \cdots \quad x(k-L+1) \right]^T$$

$$y(k) = \hat{\mathbf{h}}_k \mathbf{x}_k$$

$$e(k) = \hat{d}(k) - y(k)$$

$$\hat{\sigma}_e^2(k) = \left(1 - \frac{1}{2L}\right) \hat{\sigma}_e^2(k-1) + \frac{1}{2L} e^2(k)$$

$$\xi_{\hat{\mathbf{h}}_k} = \frac{L}{L - \sqrt{L}} \left(1 - \frac{\sum_{j=0}^{L-1} |\hat{h}_k(j)|}{\sqrt{L \sum_{j=0}^{L-1} \hat{h}_k^2(j)}}\right)$$

$$\xi(k) = (1 - \lambda_\xi) \xi(k-1) + \lambda_\xi \xi_{\hat{\mathbf{h}}_k}$$

$$\alpha(k) = 2\xi(k) - 1$$

Para $i = 0, 1, \dots, L-1$

$$g_k(i) = \frac{1 - \alpha(k)}{2L} + \frac{(1 + \alpha(k)) F(|\hat{h}_k(i)|)}{2 \sum_{j=0}^{L-1} F(|\hat{h}_k(j)|) + \epsilon}$$

Fim

$$\mathbf{\Gamma}_k = \text{diag}\{g_k(0), \dots, g_k(L-1)\}$$

$$\text{Se } \left(1 - \frac{\hat{\sigma}_\nu}{\hat{\sigma}_e}\right) > 0$$

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \beta \left(1 - \frac{\hat{\sigma}_\nu}{\hat{\sigma}_e}\right) \frac{\mathbf{x}_k^T \mathbf{\Gamma}_k e(k)}{\mathbf{x}_k^T \mathbf{\Gamma}_k \mathbf{x}_k + \delta}$$

Fim

Fim

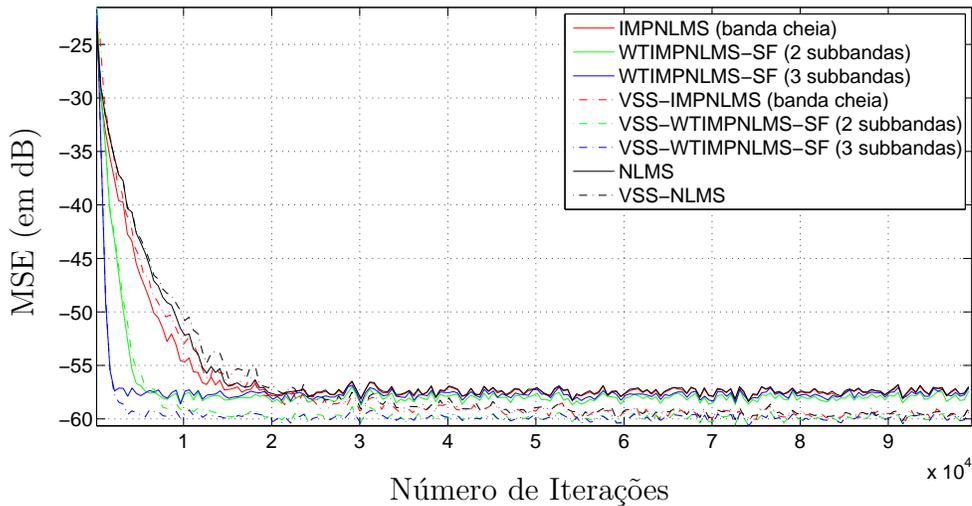


Figura 3.14: Evolução do MSE dos algoritmos NLMS, VSS-NLMS, IMPNLMS, VSS-IMPNLMS, WTIMPNLMS-SF e VSS-WTIMPNLMS-SF, aplicados na mesma configuração de identificação. O MSE foi suavizado por meio da substituição de 500 amostras consecutivas por sua média.

o VSS-NLMS, o IMPNLMS e o VSS-IMPNLMS, β foi arbitrado em 0,9 e para os algoritmos que empregam a estrutura WT-SF β foi fixado em $\frac{0,9}{M}$, com a *wavelet* bior4.4. Os resultados constam na Fig. 3.14, os quais demonstram que o emprego da técnica VSS reduz levemente a taxa de convergência, porém implica uma significativa redução do desajuste, com o MSE em regime permanente se aproximando do seu valor mínimo (-60 dB). A estrutura WT-SF com $M = 3$ e técnica VSS apresenta o melhor resultado entre todas as outras opções, pois a redução da taxa de convergência derivada do emprego do VSS é imperceptível e seu MSE em regime permanente é compatível com os demais. Dessa forma, revela-se interessante o emprego da estrutura WT-SF aliada ao algoritmo VSS, de modo a manter o aumento da taxa de convergência sem sacrifício do desajuste em regime permanente. O MSE em regime permanente (obtido por meio da média das últimas 40000 iterações) é apresentado na Tab. 3.6, a qual revela que o EMSE (MSE em excesso) obtido pela técnica VSS é muito reduzido (o menor valor possível do MSE seria -60 dB, devido à variância do ruído).

Para enfatizar que a vantagem da estrutura WT-SF somente se manifesta quando o sinal de entrada é colorido, a Fig. 3.15 apresenta a evolução do MSE para os mesmos algoritmos da Fig. 3.14, quando o sinal de entrada é branco, gaussiano e de variância unitária. Neste caso, o emprego de subbandas não se revela interessante.

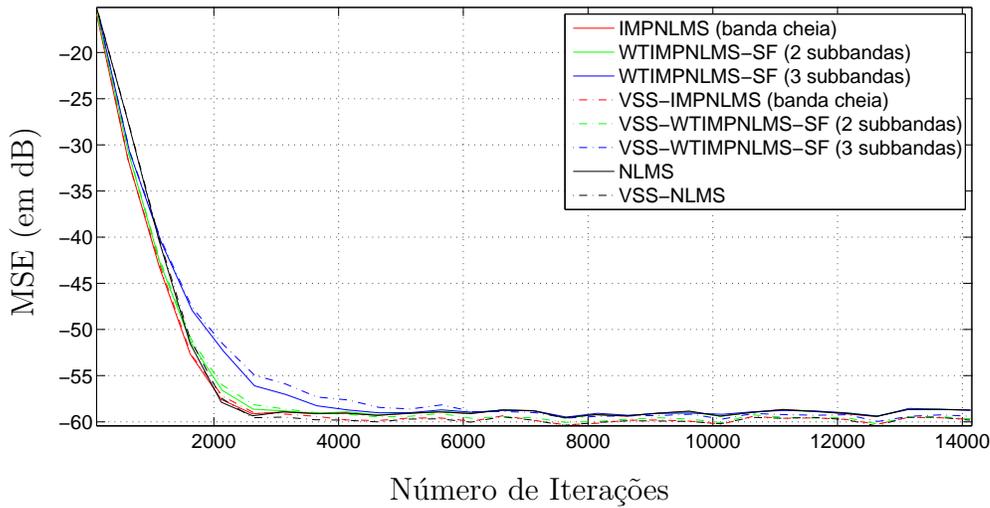


Figura 3.15: Evolução do MSE dos algoritmos NLMS, VSS-NLMS, IMPNLMS, VSS-IMPNLMS, WTIMPNLMS-SF e VSS-WTIMPNLMS-SF, aplicados na mesma configuração de identificação, com sinal de entrada branco. O MSE foi suavizado por meio da substituição de 500 amostras consecutivas por sua média.

3.7 Considerações Finais

Este capítulo concentrou-se na análise do algoritmo de identificação de respostas impulsivas possivelmente esparsas denominado IMPNLMS.

Na Seção 3.2, uma análise do transiente do MSE do algoritmo IMPNLMS foi efetuada. Esta análise guarda semelhanças com as efetuadas em [51],[52], porém concomitantemente apresenta significativas distinções. Em primeiro lugar, [51],[52] supõem que o sinal de entrada é branco. Esta hipótese é violada por diversos sinais na prática, entre eles sinais de voz. Quando o sinal de entrada é colorido, as características de convergência tornam-se bastante diferentes e portanto não cabe restringir a análise para este caso. O algoritmo sob análise também difere; [51] contempla um algoritmo PNLMS simplificado, enquanto que [52] se concentra sobre o algoritmo IPNLMS [71]. Já nossa análise é voltada para o algoritmo IMPNLMS [37], mais sofisticado que o IPNLMS no tocante ao uso de uma medida de esparsi-

Tabela 3.6: MSE (em dB) em regime permanente dos algoritmos testados na configuração da Fig. 3.14.

Algoritmo	MSE (versão padrão)	MSE (versão VSS)
NLMS	-57,34	-59,53
IMPNLMS	-57,41	-59,59
WTIMPNLMS-SF ($\underline{M} = 2$)	-57,90	-59,83
WTIMPNLMS-SF ($\underline{M} = 3$)	-57,59	-59,85

dade, na estimativa automática do parâmetro $\alpha(k)$ e no emprego da função $F(\cdot)$. Outra característica inovadora da nossa análise é a introdução de técnicas (adaptadas) oriundas de [58], [59], [55] e [60], as quais obtêm precisão significativamente superior para filtros adaptativos curtos (pois não utilizam a aproximação da Eq. (3.27)), permitindo que configurações nas quais L seja pequeno possam ser devidamente contempladas com acuradas análises teóricas.

Na Seção 3.3, a análise por balanceamento de energia preconizada por [64] foi empregada para estimar o MSE em regime permanente dos algoritmos da família PNLMS. O resultado coincide com o apresentado em [36], porém com duas vantagens: nossa análise é mais simples e emprega menos hipóteses. Por exemplo, a suposição acerca da gaussianidade de $x(k)$ se mostrou desnecessária.

Na Seção 3.4 uma estrutura em subbandas não uniforme proposta para os algoritmos de IRE teve alguns aspectos teóricos elucidados, particularmente a respeito do MSE em regime permanente e da influência do número de subbandas no intervalo admissível (no tocante à convergência) para o fator de aprendizagem β .

Por fim, a Seção 3.6 detalha um procedimento de inserção de fatores de aprendizagens variantes no tempo, de modo a permitir um desajuste em regime permanente próximo do menor possível, sem prejuízo da taxa de convergência da estrutura WT-SF.

Capítulo 4

O Problema da Separação Cega de Fontes

4.1 O Problema da Separação Cega de Fontes

Comumente utilizamos sensores ou transdutores para recolher dados de uma determinada fonte (seja ela, por exemplo, de natureza sonora, cerebral ou sísmica) a qual coexiste, no cenário de gravação, com outras fontes. Tal configuração implica termos gravações de misturas das fontes. A existência dessas misturas pode ser intencional, como se dá, por exemplo, em gravações musicais.

O problema da separação cega de fontes consiste em recuperar N sinais originais (também denominados de fontes), apenas a partir do conhecimento de M misturas destes. Trechos isolados de cada fonte são, nessa configuração, inacessíveis; daí o emprego do adjetivo “cega”, justificado por tratar-se de uma estratégia não-supervisionada. Num ambiente onde coexistem diversas fontes sonoras, o ser humano se desincumbe facilmente dessa tarefa, ao concentrar sua atenção em apenas uma das fontes, descartando as demais como interferências não passíveis de processamento de alto nível. O fato de que indivíduos com desordens no sistema auditivo central possam apresentar dificuldades de compreensão de uma fala num ambiente ruidoso [72] ilustra a complexidade da resolução desse problema. Dessa nossa habilidade advém o outro nome - *cocktail-party* [11] - atribuído ao problema de BSS (separação cega de fontes, do inglês *Blind Source Separation*). Embora a separação cega de fontes em contextos genéricos, dada a pouca informação disponível, possa parecer à primeira vista um problema impossível para um sistema artificial [73], atualmente consiste num ramo maduro de pesquisa, sendo passível de inúmeras aplicações. Dentre estas, podemos elencar exemplos nas áreas de sensoriamento remoto [74], análise de sinais biomédicos [75] e sísmicos [76], bem como em comunicações digitais [77].

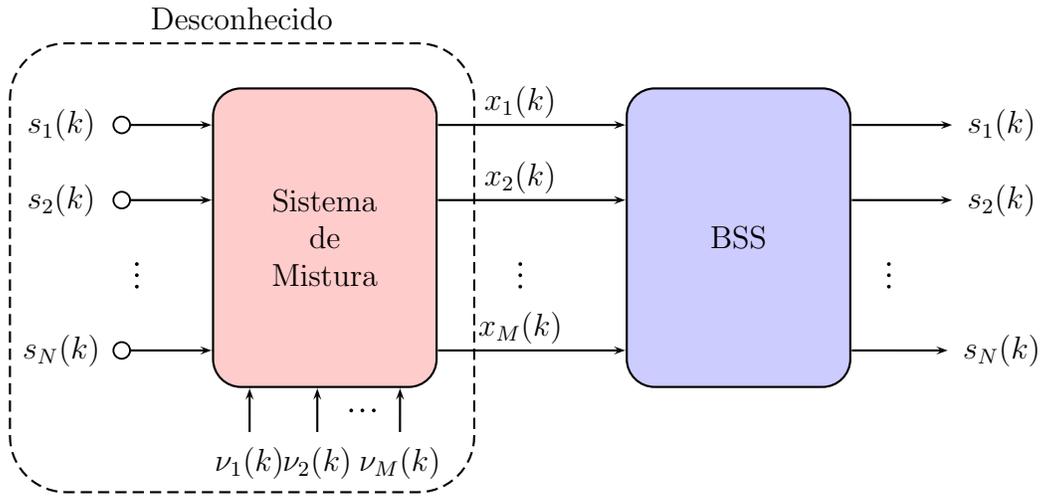


Figura 4.1: Problema e objetivo ideal da separação cega de fontes (BSS).

A estrutura do problema de BSS é apresentada na Fig. 4.1. Essa figura apresenta as N fontes $s_1(k)$, $s_2(k)$, \dots , $s_N(k)$, que, após passarem em um processo que pode conter sinais de perturbação $\nu_i(k)$, dão origem a M misturas $x_i(k)$, a partir das quais almeja-se reconstituir as fontes originais.

Diversas características do sistema de mistura impactam a escolha do algoritmo de separação, bem como seu custo computacional. Uma classificação importante dos sistemas de misturas os subdivide em sistemas lineares e não-lineares. Enquanto estes são em teoria menos restritivos nas hipóteses, aqueles aliam à facilidade na modelagem a ampla aplicabilidade [78]. Se a transformação efetuada pelo sistema de mistura mantém-se constante ao longo do tempo, dizemos ser invariante no tempo o processo de mistura. Doravante consideraremos que os sistemas de mistura são lineares e invariantes no tempo.

A relação entre o número N de fontes e o número M de misturas é outra variável relevante no processo de seleção do algoritmo de BSS. Sistemas nos quais $N < M$, $N = M$ e $N > M$ são denominados, respectivamente, de superdeterminados, determinados e subdeterminados. As misturas subdeterminadas, por disponibilizarem menor quantidade de informação, são as mais desafiadoras. Cumpre destacar, dessas configurações, o caso $M = 1$ (monossensor) por ser extremamente difícil e por demandar técnicas específicas [79].

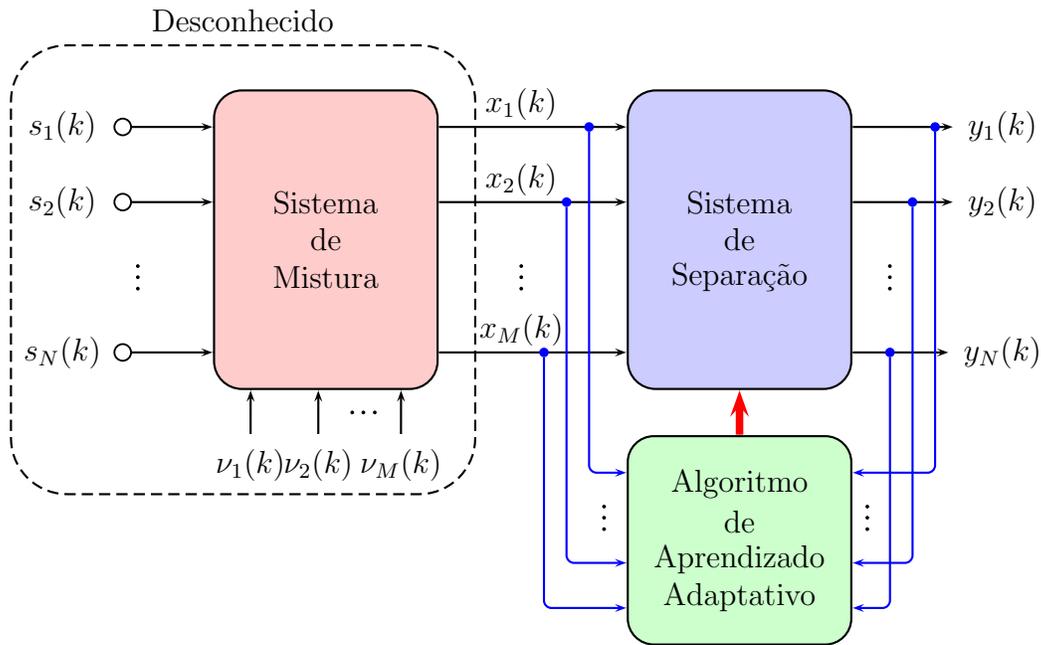


Figura 4.2: Diagrama de blocos funcional que ilustra o processo de separação cega de fontes.

4.2 Estrutura Básica e Ambiguidades dos Algoritmos de Separação Cega de Fontes

O objetivo de uma técnica de BSS é determinar um sistema adaptativo inverso (ou de reconstrução) que, atuando sobre as informações disponíveis (misturas), obtenha versões aproximadas $y_i(k)$ das fontes. A Fig. 4.2 apresenta a estrutura geral dos algoritmos de separação cega de fontes. A técnica de BSS pode efetuar sua busca no espaço de mistura, quando almeja identificar os parâmetros do sistema de mistura para então empreender a reconstrução das fontes, ou no espaço de separação, caso em que os parâmetros de um sistema direto de separação são determinados. Quando o sistema de mistura não é inversível (o que ocorre em configurações subdeterminadas), em geral recorre-se à busca no espaço de mistura; caso contrário, é comum o emprego de métodos de busca no espaço de separação.

A ausência de qualquer conhecimento exato a respeito das fontes acaba por ter profundas implicações. A primeira delas é o incontornável problema de permutação, o qual implica que a reconstituição das fontes, ainda que bem-sucedida, pode apresentar estimativas “embaralhadas” com relação à ordem das fontes, de modo que a i -ésima fonte seja associada à j -ésima estimativa sem que necessariamente tenhamos $i = j$. A ordenação das fontes consiste em mero produto de convenção; veremos posteriormente que nas técnicas de separação cega de fontes no domínio da

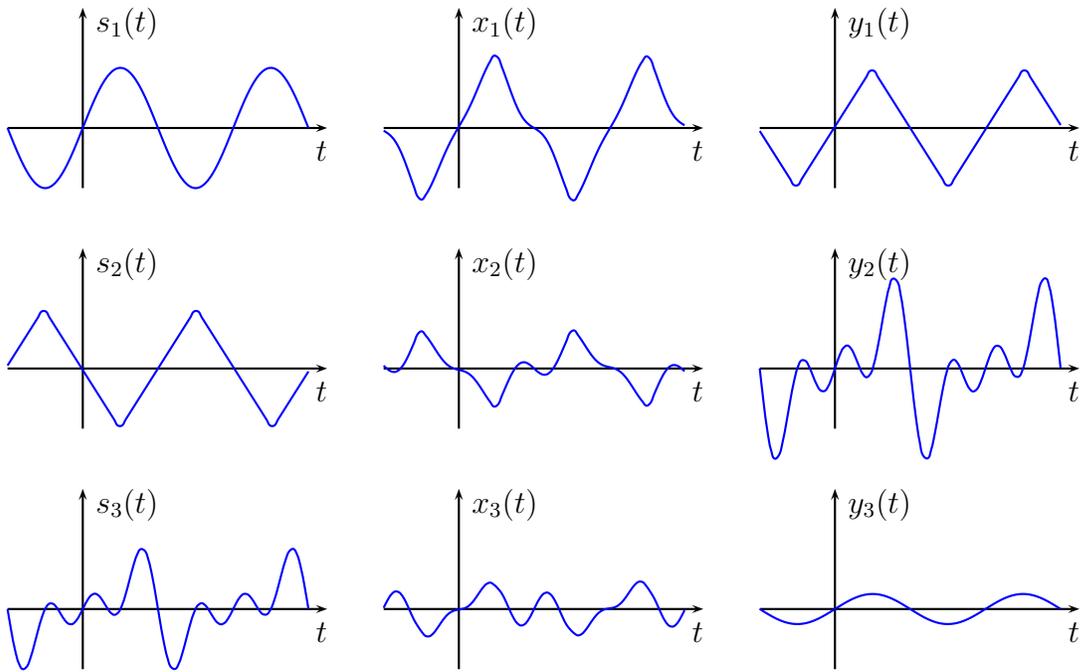


Figura 4.3: Ilustração de um processo de separação de fontes considerado bem-sucedido. As fontes estão denotadas por $s_i(t)$, as misturas por $x_i(t)$ e as estimativas das fontes por $y_i(t)$. Os sinais são contínuos para melhor visualização.

frequência a permutação torna-se não trivial.

Outra característica inerente às técnicas de BSS é a ambiguidade de filtragem, a qual implica que as estimativas sejam versões filtradas das fontes. Em sua versão mais simples (no contexto de misturas instantâneas), essa filtragem pode se manifestar numa ambiguidade de escalamento pois, sendo-nos desconhecida a energia das fontes a recuperar, não nos é possível recuperar a escala dinâmica das mesmas. A permutação e a ambiguidade de escalamento são decorrentes da ausência de unicidade da estimativa das fontes [80].

A Fig. 4.3 ilustra, num contexto de misturas instantâneas, os problemas de permutação e de escalamento. Cabe ressaltar nessa figura a mudança de ordenação das estimativas (a primeira estimativa, por exemplo, está associada à segunda fonte), bem como o escalamento arbitrário efetuado nas fontes (compare por exemplo $s_1(t)$ com $y_3(t)$).

4.3 Modelagem do Processo de Mistura

Já que consideraremos processos de mistura lineares e invariantes no tempo, o sistema MIMO (do inglês *Multiple-Input and Multiple-Output*) pode ser modelado por filtros $H_{ij}(z)$, os quais correspondem, num contexto de fontes sonoras, à função de

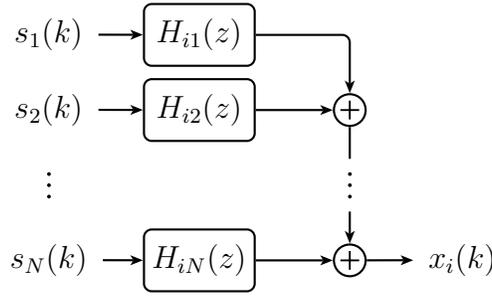


Figura 4.4: Geração da i -ésima mistura a partir das N fontes.

transferência acústica existente entre a j -ésima fonte e o i -ésimo sensor. A Fig. 4.4 apresenta o diagrama de blocos correspondente ao processo gerador da i -ésima mistura.

A depender do comprimento M dos filtros de mistura $H_{ij}(z)$, temos algumas configurações passíveis de atenção especial na literatura. Caso tenhamos $M = 1$, dizemos ser instantânea a mistura. Esse modelo considera infinita a velocidade de propagação das fontes no espaço. Nesse caso particular, a geração das k -ésimas amostras do vetor-coluna das misturas é dada por:

$$\mathbf{x}_k = \mathbf{H}\mathbf{s}_k, \quad (4.1)$$

onde \mathbf{H} é uma matriz $M \times N$ que modela o processo de mistura, com seu elemento h_{ij} denotado como “coeficiente de atenuação”, o qual é a razão, invariante ao longo das amostras, entre a contribuição da j -ésima fonte à i -ésima mistura e a j -ésima fonte. Fisicamente, a redução da distância entre a j -ésima fonte e o i -ésimo sensor costuma se traduzir num acréscimo da magnitude do coeficiente de atenuação.

Caso todos os filtros de mistura apresentem apenas um coeficiente não-nulo (de modo que $H_{ij}(z) = h_{ij}z^{-\Delta_{ij}}$), o processo ilustrado pela Figura 4.4 é chamado de anecoico, pois implica a ausência de reverberações ou propagação via múltiplos percursos (ecos), consistindo num modelo cujo nome se deve ao fato de aderir ao processo de gravação numa câmara anecoica.

Se o modelo de mistura não é instantâneo, tampouco anecoico, dizemos que é convolutivo. Um modelo de mistura convolutiva é poderoso o suficiente para representar parte significativa dos processos de mistura que ocorrem na prática.

A pesquisa em separação cega de fontes concentrou seus esforços iniciais no problema de misturas instantâneas. Embora esses modelos sejam às vezes irrealistas (em especial no contexto de sinais de voz e de áudio), as técnicas desenvolvidas para resolvê-los geraram ferramentas e conceitos úteis para a separação de misturas con-

volutivas. Por essa razão, este capítulo irá apresentar alguns conceitos importantes das técnicas popularmente empregadas para a resolução de misturas instantâneas, os quais serão utilizados em capítulos posteriores, no contexto de misturas anecoicas e convolutivas.

4.4 Análise de Componentes Principais

Útil como etapa auxiliar para a separação cega de fontes (especialmente em casos superdeterminados e determinados), a análise de componentes principais (PCA, do inglês *Principal Component Analysis* [81], também conhecida análise de vetor latente ou, na área de meteorologia, como “funções ortogonais empíricas”) almeja reduzir o número de variáveis a partir de um conjunto maior delas (supostamente interrelacionadas), ao mesmo tempo em que almeja manter uma boa representação. É estreitamente relacionada à transformada de Karhunen-Loève em teoria de comunicações, consistindo numa ferramenta clássica de análise estatística, extração de características, compressão de espaços e até de detecção de aberrações (em inglês, *outliers*). Embora seu objetivo mantenha semelhanças com o da ICA (do inglês *Independent Component Analysis*), cabe ressaltar que para a PCA a redundância é medida via correlações entre os dados, enquanto que a ICA emprega o conceito estatisticamente mais rico da independência, sem dar relevo à redução do número de variáveis [11]. Devido ao grande apelo geométrico, exploramos nesta seção a PCA (a qual, a despeito das semelhanças, descabe confundir com a técnica de “análise de fatores” [81]) de variáveis reais. A extensão para variáveis complexas é simples, sendo sua apresentação efetuada posteriormente.

Dentre os diversos princípios, não raro equivalentes, que permitem a derivação da componente principal, destaca-se o da maximização da variância. Seja um vetor coluna \mathbf{x}_k definido por:

$$\mathbf{x}_k = \begin{bmatrix} x_1(k) & x_2(k) & \dots & x_M(k) \end{bmatrix}^T, \quad (4.2)$$

cujos elementos $x_i(k)$ são variáveis aleatórias. Dizemos que um determinado vetor-linha de norma unitária $\mathbf{v}_1 \in \mathbb{R}^M$ gera a componente principal $z_1(k)$ de \mathbf{x}_k caso $z_1(k) = \mathbf{v}_1 \mathbf{x}_k$ tenha variância máxima. Cabe ressaltar que supomos, ao longo desta discussão, que os dados possuem média nula.

A Fig. 4.5 apresenta o vetor \mathbf{v}_1 para uma determinada distribuição bidimensional do vetor aleatório \mathbf{x}_k . O vetor \mathbf{v}_1 é paralelo ao eixo de maior variação de \mathbf{x}_k , de modo que $z_1(k) = \mathbf{v}_1 \mathbf{x}_k$ apresente a maior variância possível.

Sendo a projeção de um determinado vetor \mathbf{x}_k no vetor \mathbf{v}_1 dada por $(\mathbf{v}_1 \mathbf{x}_k) \mathbf{v}_1$, deparamo-nos com uma interpretação popular do vetor \mathbf{v}_1 , cuja direção garante a

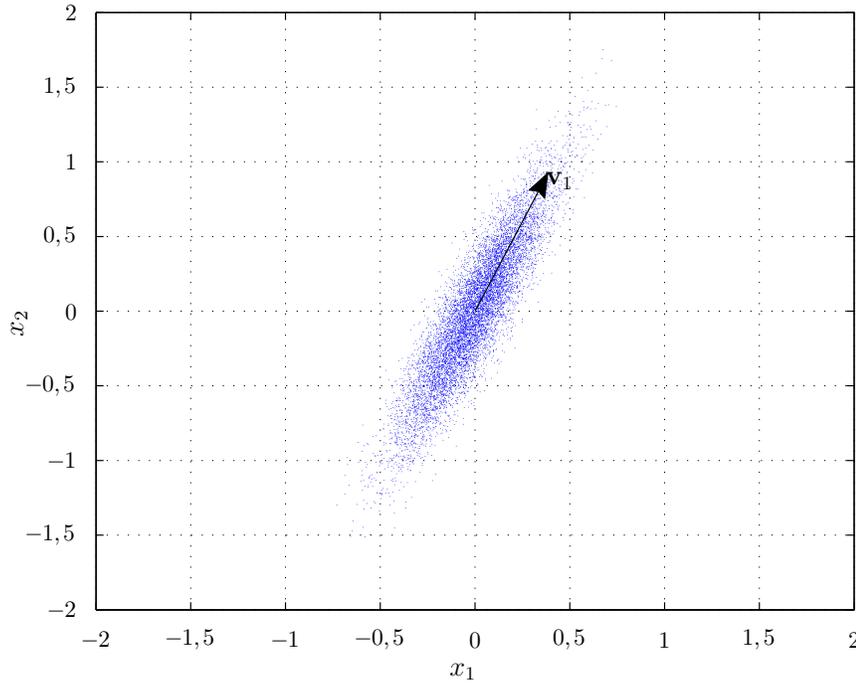


Figura 4.5: Exemplo de componente principal (vetor \mathbf{v}_1) para uma determinada distribuição bidimensional de \mathbf{x}_k .

minimalidade do erro quadrático médio entre os vetores \mathbf{x}_k e as suas respectivas projeções em \mathbf{v}_1 (já que sua norma é unitária, apenas sua direção é suficiente para determiná-lo, a menos de uma pouco relevante ambiguidade de sentido). Esta característica da PCA é importante para a compressão de sinais, pois permite uma eliminação de redundância (ao reduzir a quantidade de variáveis) que não compromete em demasia a capacidade de representação dos sinais em questão. Adicionalmente, se os componentes principais forem interpretáveis, a descrição alternativa obtida via PCA costuma ser muito mais simples do que a gerada pelas variáveis principais [81].

O conceito acima exposto de componente principal é generalizado para N componentes principais ($1 \leq N \leq M$) na forma a seguir: se $z_n(k)$ é a n -ésima componente principal, sua expressão é dada por $z_n(k) = \mathbf{v}_n \mathbf{x}_k$, com $\|\mathbf{v}_n\| = 1$ e $E[z_n(k)z_i(k)] = 0$, $1 \leq i \leq n$, sob a restrição de máxima variância de $z_n(k)$.

Logo, os componentes principais podem ser compreendidos quais somas ponderadas dos elementos de \mathbf{x}_k com máxima variância, sob as restrições de que os pesos são normalizados e de decorrelação entre as componentes principais, ou como uma busca por um conjunto de N vetores de uma base ortonormal que geram um subespaço de dimensão N , de sorte que o erro quadrático médio entre \mathbf{x}_k e a sua projeção nesse subespaço é mínima [11]. Em outras palavras, existe uma ordenação das componentes principais de modo que poucas delas (as primeiras) retenham a maior parte

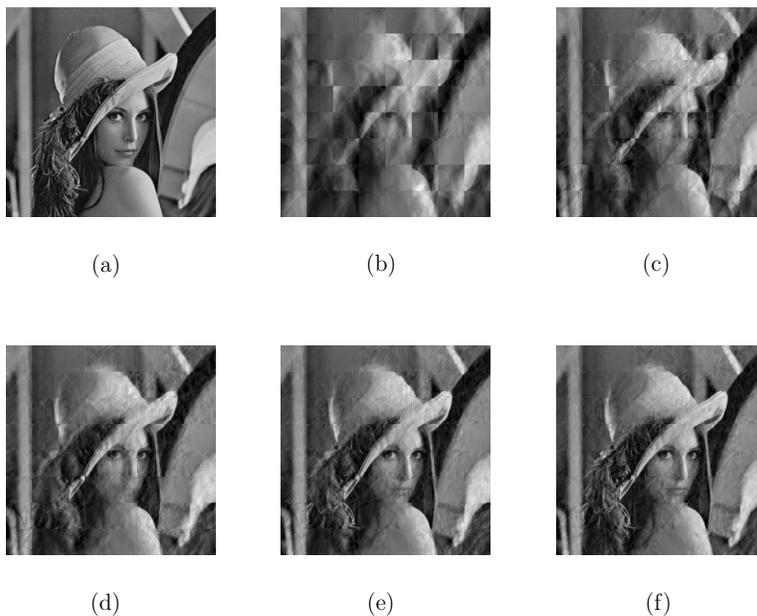


Figura 4.6: Exemplo de compressão de uma imagem via PCA. (a) Imagem original; (b) 5 componentes principais; (c) 10 componentes principais; (d) 15 componentes principais; (e) 20 componentes principais e (f) 25 componentes principais.

da variação presente em todas as variáveis [81]. Sob uma ótica equivalente, a PCA pode ser compreendida qual transformação linear inversível que maximiza a taxa de decrescimento da variância de suas saídas [82].

A Fig. 4.6 explora a capacidade de compressão da PCA. A imagem em tons de cinza 4.6(a) original foi subdividida em 64 blocos de 32×32 pixels, os quais, por meio de uma varredura ao longo das linhas, geraram vetores-linha de 1024 elementos. Estes 64 vetores formaram as linhas de uma matriz, à qual se aplica a PCA. Reter as primeiras componentes principais (anulando as restantes), aplicar a transformada inversa da PCA e gerar novamente imagem nos permite observar a grande capacidade de descrição de que relativamente poucas componentes principais são capazes.

A Fig. 4.7 ilustra a decorrelação da PCA. Cada componente $s_i(k)$ do vetor aleatório bidimensional \mathbf{s}_k foi gerado mediante uma distribuição exponencial (com suporte em \mathbb{R}) i.i.d. (independente e identicamente distribuída). O gráfico $s_2(k) \times s_1(k)$ permite obter uma representação com direções preferenciais paralelas aos eixos das abcissas e das ordenadas, o que é comum em distribuições supergaussianas e independentes. Ao gerarmos dados $\mathbf{x}_k = \mathbf{H}\mathbf{s}_k$, as componentes de \mathbf{x}_k passam a ser correlacionadas, e os ângulos entre as direções mais pronunciadas do gráfico $x_2(k) \times x_1(k)$ não mais respeitam a ortogonalidade. Ao efetuarmos a PCA (dados \mathbf{z}_k), a ortogonalidade volta a ser observada, porém como a decorrelação não implica

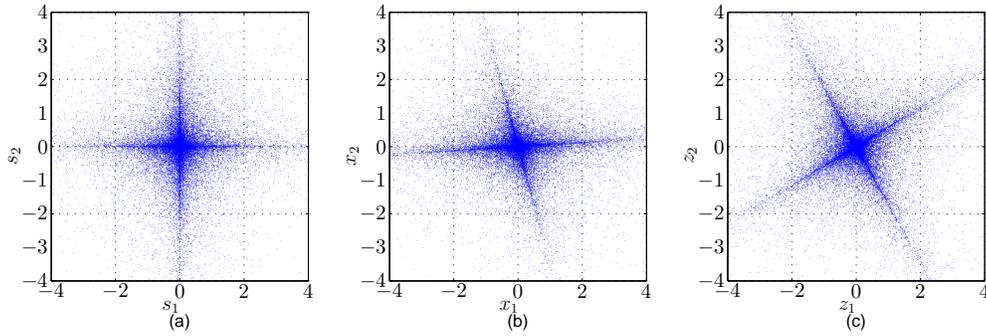


Figura 4.7: Exemplo de transformações num sinal \mathbf{s}_k bidimensional. (a) sinal \mathbf{s}_k ; (b) $\mathbf{x}_k = \mathbf{H}\mathbf{s}_k$, onde \mathbf{H} é uma matriz gerada aleatoriamente e (c) versão esferizada \mathbf{z}_k de \mathbf{x}_k .

independência as direções preferenciais do gráfico $z_2(k) \times z_1(k)$ costumam encontrar-se rotacionadas em relação às direções de $s_2(k) \times s_1(k)$. Uma mera matriz de rotação aplicada aos dados \mathbf{z}_k poderia recuperar as fontes originais, sendo esta uma popular abordagem para a resolução do problema de BSS.

Existem diversos modos para o cálculo das componentes principais, tanto na sua versão *on-line* quanto na *off-line* (a referência [83] apresenta, numa notação unificada, uma ampla revisão da literatura pertinente). A que nos interessa é a que recorre à matriz de autocorrelação \mathbf{R}_x , estimada a partir dos dados \mathbf{x}_k a que temos acesso. Se os vetores-linha \mathbf{v}_i , $1 \leq i \leq N$, são as linhas respectivas da matriz de branqueamento ou esferização \mathbf{V} , pode-se obter uma versão \mathbf{z}_k (descorrelacionada e convenientemente escalada) dos dados \mathbf{x}_k se empregamos a transformação $\mathbf{z}_k = \mathbf{V}\mathbf{x}_k$, sendo [11]:

$$\mathbf{V} = \mathbf{\Lambda}^{-1/2}\mathbf{Q}^T, \quad (4.3)$$

com as matrizes $\mathbf{\Lambda}$ e \mathbf{Q} contendo, respectivamente, os autovalores e autovetores da matriz \mathbf{R}_x .

4.5 Princípios de Separação

O fato de trechos isolados das fontes a recuperar não estarem disponíveis não impede a construção de uma função-custo cuja minimização orientará a busca de parâmetros do sistema de separação. Os princípios de separação motivam e justificam o emprego de uma função-custo específica. A não-gaussianidade da maioria das fontes é uma hipótese comum a quase todos eles, embora tal restrição possa ser contornada caso as fontes apresentem algumas características específicas, tal como a não-estacionariedade.

Um dos princípios mais poderosos reside na hipótese de independência estatísticas das fontes. Segundo [84]:

Esta é uma sólida abordagem quando a modelagem da função de transferência das fontes para os sensores é muito difícil; é inevitável quando nenhuma informação *a priori* acerca desta transferência está disponível. A falta deste conhecimento *a priori* sobre a mistura é compensada por uma hipótese estatisticamente forte porém plausível fisicamente de independência das fontes. O fato de a técnica obtida por este princípio ser chamada de cega não deve ser entendido negativamente; a fraqueza da informação *a priori* é precisamente a força do modelo de separação cega, tornando-o uma ferramenta versátil de exploração da diversidade espacial obtida por um arranjo de sensores.

Sob a perspectiva da maximização da independência das estimativas, observamos a inevitabilidade do problema de permutação, dado que uma permutação qualquer das estimativas não altera a relação de dependência entre elas.

O princípio de independência das fontes é relacionado aos princípios de minimização da informação mútua, maximização da entropia e de máxima verossimilhança, as quais consistem em técnicas associadas à ICA. Estas relações advêm do fato de que princípios de separação distintos (como por exemplo o critério de maximização da entropia e de máxima verossimilhança) podem resultar em algoritmos de separação idênticos [85]. Um mesmo princípio de separação pode, por outro lado, engendrar algoritmos distintos, a depender, dentre outros fatores, do método de otimização da função-custo (por exemplo, gradiente descendente, gradiente natural ou ponto fixo).

Outro princípio de separação explora uma possível esparsidade das fontes, supondo que a maior parte da energia dos sinais localiza-se em poucas amostras dos mesmos. Equivalentemente, podemos dizer que este princípio supõe a supergaussianidade das fontes, a qual pode ser realçada por meio de recurso a alguma transformação esparsificadora.

Esta tese recorre a dois princípios de separação: o da maximização da verossimilhança e o da esparsidade. Este último será empregado numa busca no espaço de mistura (identificação); já o primeiro será empregado (com uma otimização via gradiente natural) numa busca do espaço de separação.

4.6 Estatísticas de Ordem Superior

A não-gaussianidade das fontes é uma hipótese comum à vasta maioria dos algoritmos de separação de fontes. Algoritmos de BSS que se restringem a estatísticas de

segunda ordem em geral precisam recorrer a outras hipóteses adicionais acerca das fontes, como por exemplo a não-estacionariedade (vide, por exemplo, [86]). Logo, a maioria dos algoritmos de ICA emprega estatísticas de ordem superior para angariar seu objetivo. Eis uma importante propriedade que os distingue da maioria dos algoritmos de PCA.

O fato de que momentos de primeira e segunda ordens bastam para especificar as distribuições gaussianas é ilustrado pela assertiva de que, para estas distribuições, a decorrelação implica independência. Sendo a gaussianidade preservada em combinações lineares de variáveis aleatórias gaussianas, o apelo a estatísticas de ordem mais alta é inútil nesta condição - e, neste caso, sem maiores informações a respeito das fontes, o problema de BSS é insolúvel.

O Teorema Central do Limite (amiúde chamado, equivocadamente, de Teorema do Limite Central [87]) justifica o fato de que uma combinação linear de variáveis aleatórias tende a ser mais gaussiana do que cada uma das variáveis aleatórias em separado. Logo, o processo de mistura pode ser compreendido como uma gaussianização das fontes, cuja reversão pode ser empreendida por meio de uma maximização da não-gaussianidade das estimativas. Tal proceder exige uma medida de não-gaussianidade, consistindo a curtose num exemplo de medida passível de ser utilizada. Na prática, devido ao fato de ser pouco robusta a aberrações (*outliers*), a curtose costuma ser preterida a outras medidas, como por exemplo a negentropia [11]. A escolha de uma determinada medida de gaussianidade específica a função *score* $\phi(\cdot)$, futuramente detalhada, que será empregada no processo iterativo.

4.7 Análise de Componentes Independentes via Estimativa de Máxima Verossimilhança

O processo de estimativa de máxima verossimilhança, em resumo, efetua uma busca de parâmetros (desconhecidos) orientada pela maximização da probabilidade das observações (conhecidas); no caso de separação de fontes, é intimamente relacionado à maximização do fluxo de informação [11] (em inglês, *infomax*, de *maximization of information flow*).

No caso de misturas instantâneas, sendo $\mathbf{f}_s(\mathbf{s}_k)$ a distribuição das fontes (supostas independentes), a distribuição conjunta $f_x(\mathbf{x}_k)$ das misturas $\mathbf{x}_k = \mathbf{H}\mathbf{s}_k$ é obtida por:

$$f_x(\mathbf{x}_k) = |\det(\mathbf{H}^{-1})| \prod_{i=1}^N f_{s_i}(s_k(i)), \quad (4.4)$$

onde $\det(\mathbf{A})$ é o determinante da matriz \mathbf{A} .

A maioria das técnicas de ICA não modela as dependências estatísticas entre as

amostras de uma mesma fonte, supondo-as i.i.d.; por esta razão, o índice k , referente às amostras das misturas e das fontes, costuma ser elidido na equação acima. Esta modelagem unidimensional é interessante porque, além de evitar a exigência de um excessivo conhecimento *a priori* das fontes, costuma ser suficiente para um bom desempenho do processo de separação. Importa ressaltar que trata-se apenas de uma hipótese de trabalho que não implica um fracasso do método quando as amostras das fontes não forem i.i.d.; apenas significa que opta-se por não empregar a correlação interna das sequências, embora esta informação possa ser incorporada em algumas técnicas de separação [88].

Sendo-nos desconhecidas as distribuições $f_{s_i}(s_k(i))$ das fontes, as técnicas de ICA frequentemente assumem uma determinada distribuição para as mesmas. Esta modelagem não resulta, em geral, num grande comprometimento do desempenho, desde que a discrepância entre as distribuições real e suposta respeite alguns limites, em geral pouco restritivos. Logo, conquanto o objetivo primordial das técnicas de BSS seja recuperar as fontes, deve-se utilizar um conhecimento *a priori*, mesmo que bastante impreciso, de suas distribuições. Por esta razão, a distribuição suposta pelo algoritmo é considerada um estorvo (em inglês, *nuisance*), dado que o algoritmo não está primordialmente interessado nele. Na prática, basta saber se as fontes são sub ou supergaussianas; no caso em que esta informação não está disponível, é possível estimá-la durante as atualizações do algoritmo [11].

Inspiradas em [84], as Figs. 4.8 e 4.9 apresentam o logaritmo da verossimilhança

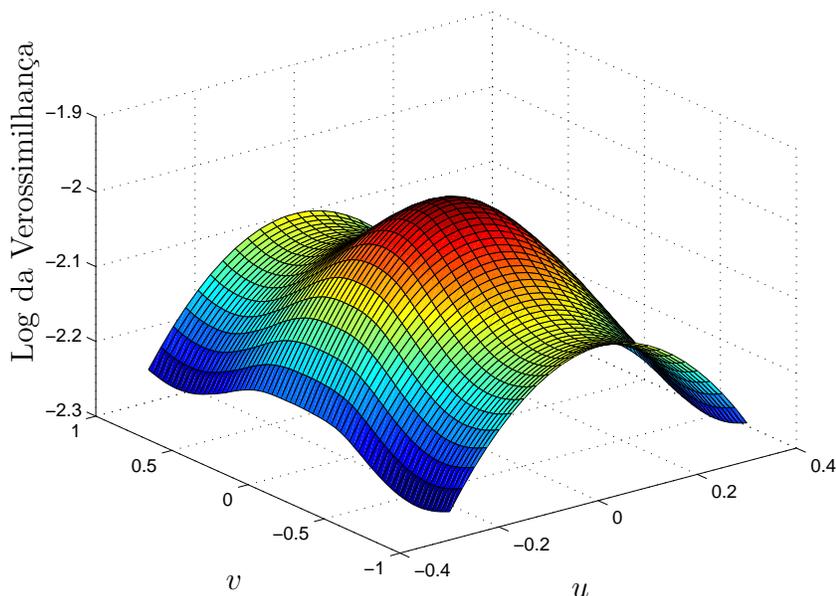


Figura 4.8: Logaritmo da verossimilhança das misturas em função dos parâmetros u e v . Tanto a distribuição real das fontes quanto seu modelo são exponenciais.

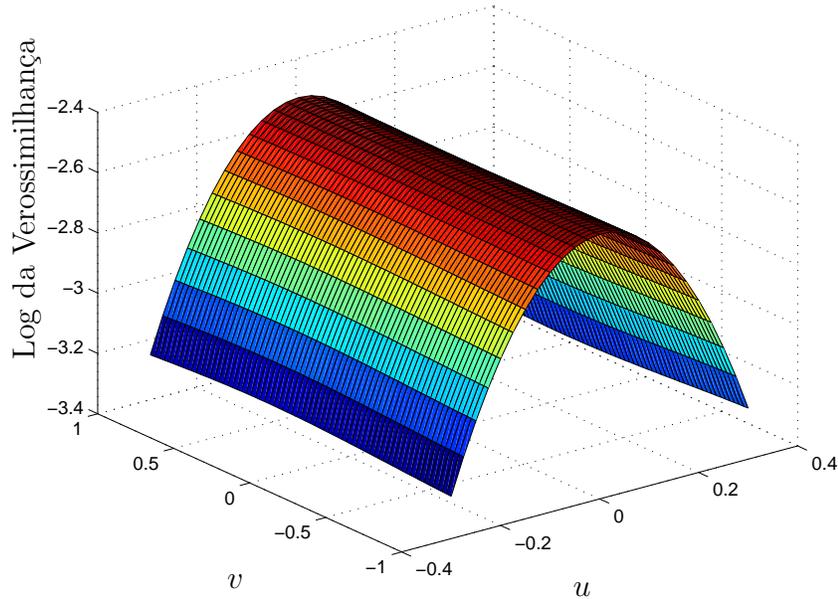


Figura 4.9: Logaritmo da verossimilhança das misturas em função dos parâmetros u e v . Apesar da distribuição real das fontes (exponencial e, portanto, supergaussiana), foi empregado um modelo de distribuição subgaussiana para o cálculo da verossimilhança (uma mistura de gaussianas).

dos dados $\mathbf{x}_k = \mathbf{H}\mathbf{s}_k$, quando a matriz \mathbf{H} é parametrizada como:

$$\mathbf{H}(u, v) = \begin{bmatrix} \cosh(u) & \sinh(u) \\ \sinh(u) & \cosh(u) \end{bmatrix} \begin{bmatrix} \cos(v) & -\sin(v) \\ \sin(v) & \cos(v) \end{bmatrix}, \quad (4.5)$$

Para valores pequenos de u e v , a matriz $\mathbf{H}(u, v)$ pode ser aproximada por:

$$\mathbf{H}(u, v) \approx \mathbf{I} + u \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + v \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad (4.6)$$

de sorte que a parametrização empreendida promove uma exploração da vizinhança da matriz identidade. Quando $u = v = 0$, $\mathbf{H}(u, v)$ torna-se a identidade e as misturas são idênticas às fontes. Portanto, se nosso modelo de distribuição das fontes não é radicalmente distinto da distribuição real das mesmas, espera-se que a verossimilhança (ou o logaritmo dela) apresente um máximo próximo à origem.

A Fig. 4.8 apresenta o comportamento esperado, pois o máximo da verossimilhança encontra-se próximo à origem. O mesmo não se aplica à Fig. 4.9, a qual, devido à falha de modelagem da distribuição real das fontes, apresenta um conjunto não enumerável de máximos. Essas figuras ressaltam a importância de conhecermos o tipo de distribuição das fontes. Tal necessidade não é demasiado

restritiva, já que diversas fontes de sinais encontradas em aplicações importantes apresentam consistentemente dois padrões de distribuição, a saber: supergaussiana ou subgaussiana. Sinais de áudio e de voz, por exemplo, de frequente aderem a uma distribuição supergaussiana.

O fato de a verossimilhança ser máxima quando as estimativas se aproximam das fontes permite-nos eleger o critério de máxima verossimilhança (definida mais à frente) para orientar a busca de uma matriz de separação \mathbf{W} , de sorte que as estimativas $\mathbf{y}_k = \mathbf{W}\mathbf{x}_k = \mathbf{W}\mathbf{H}\mathbf{s}_k$, findo o processo de otimização, maximizem a verossimilhança dos dados a que temos acesso. Tipicamente, uma técnica de otimização iterativa é empregada de modo a continuamente refinar os parâmetros de \mathbf{W} , a partir de uma escolha inicial, a qual pode ser uma matriz de branqueamento obtida via PCA.

4.8 Gradiente Natural como Ferramenta de Otimização

Dentre as diversas técnicas de otimização, a do gradiente descendente com relação a uma parametrização específica da função-custo, como a empregada na derivação do algoritmo LMS, é uma das mais populares em filtragem adaptativa. Porém, quando a estrutura do espaço de parâmetros não respeita a métrica euclidiana, este tipo de gradiente não encerra necessariamente um significado, pois sua direção é dependente da parametrização escolhida.

Fundamentando-se em [89], [90], [11] e, especialmente, em [91], seja a função $\mathcal{F}(\mathbf{w}) = \mathcal{F}(w_0, w_1) = w_0 + (w_1 - 1)^2$. Em um sistema alternativo de parâmetros $\boldsymbol{\theta} = (\theta_0, \theta_1) = (w_0, \frac{w_1}{2})$, a função é escrita como $\mathcal{F}(\boldsymbol{\theta}) = \theta_0 + (2\theta_1 - 1)^2$. Na origem de \mathbb{R}^2 , o gradiente de \mathcal{F} com relação a \mathbf{w} e $\boldsymbol{\theta}$ é dado, respectivamente, por $\frac{\partial \mathcal{F}}{\partial \mathbf{w}} = (1, -2)$ e $\frac{\partial \mathcal{F}}{\partial \boldsymbol{\theta}} = (1, -4)$. Sendo a coordenada w_1 o dobro de θ_1 , $\frac{\partial \mathcal{F}}{\partial \boldsymbol{\theta}}$ corresponde a $(1, -8)$ no espaço tangente das coordenadas \mathbf{w} (diferente de $\frac{\partial \mathcal{F}}{\partial \mathbf{w}}$), o que implica que o gradiente de \mathcal{F} depende da parametrização, conforme ilustrado pela Fig. 4.10.

Quando atentamos para a métrica de Riemann do espaço dos parâmetros $\{\mathbf{w} \in \mathbb{R}^{\bar{q}}\}$, cumpre encontrar um gradiente que independa da parametrização da função-custo, o qual costuma ser denominado de gradiente natural, contravariante ou relativo. A métrica euclidiana recorre a um sistema linear de coordenadas ortonormais para expressar o quadrado do módulo de um pequeno vetor de incremento $d\mathbf{w}$ que conecta \mathbf{w} a $\mathbf{w} + d\mathbf{w}$ como:

$$\|d\mathbf{w}\|^2 = \sum_{i=0}^{\bar{q}-1} (dw_i)^2, \quad (4.7)$$

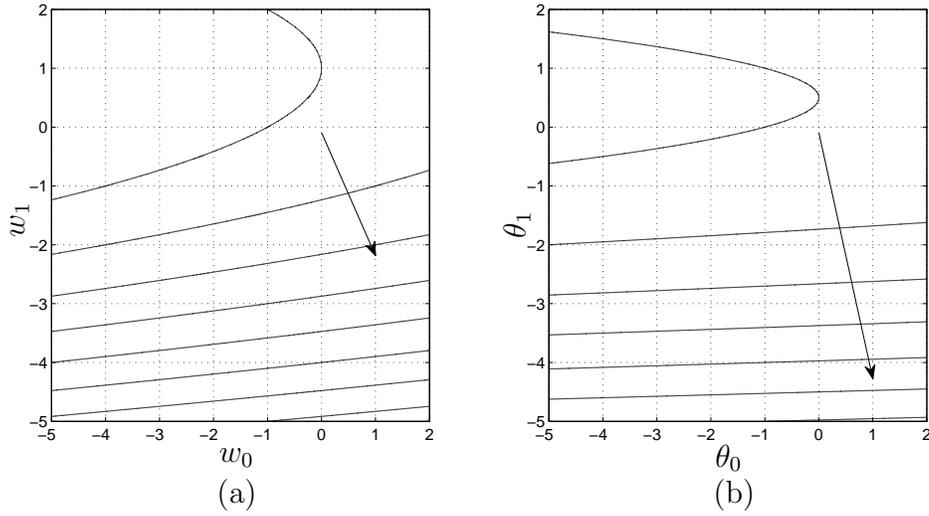


Figura 4.10: O vetor gradiente em duas diferentes parametrizações da mesma função \mathcal{F} . (a) $\mathcal{F}(\mathbf{w})$; (b) $\mathcal{F}(\boldsymbol{\theta})$.

onde dw_i é o i -ésimo componente de $d\mathbf{w}$. Porém, caso o espaço dos parâmetros seja uma variedade curvada (o que ocorre por exemplo com os parâmetros sinápticos em redes neurais), inexistem coordenadas lineares ortonormais [89], e portanto $\|d\mathbf{w}\|^2$ sempre será escrito como:

$$\|d\mathbf{w}\|^2 = \sum_{i,j} \bar{g}_{ij}(\mathbf{w}) dw_i dw_j, \quad (4.8)$$

onde a matriz $\bar{\mathbf{G}}(\mathbf{w}) = (\bar{g}_{ij}(\mathbf{w}))$ de dimensões $\bar{q} \times \bar{q}$ é o tensor da métrica de Riemann, o qual, quando igual à unidade, recai na métrica euclidiana ortonormal.

A direção de descida mais íngreme de uma função $\mathcal{F}(\mathbf{w})$ em \mathbf{w} é mais coerentemente definida como o vetor $d\mathbf{w}$ que minimiza $\mathcal{F}(\mathbf{w} + d\mathbf{w})$ sob a restrição de que $\|d\mathbf{w}\|$ apresenta módulo fixo. Pode-se provar que esta direção num espaço de Riemman é dada por [89]:

$$-\tilde{\nabla}\mathcal{F}(\mathbf{w}) = -\bar{\mathbf{G}}^{-1}(\mathbf{w})\nabla\mathcal{F}(\mathbf{w}), \quad (4.9)$$

onde o sinal negativo deriva da minimização da função e $\nabla\mathcal{F}(\mathbf{w})$ é o gradiente convencional:

$$\nabla\mathcal{F}(\mathbf{w}) = \left[\frac{\partial\mathcal{F}(\mathbf{w})}{\partial w_0}, \frac{\partial\mathcal{F}(\mathbf{w})}{\partial w_1}, \dots, \frac{\partial\mathcal{F}(\mathbf{w})}{\partial w_{\bar{q}-1}} \right]^T. \quad (4.10)$$

Sendo a maximização da verossimilhança (Eq. (4.4)) uma possível orientação para a busca de parâmetros da matriz de separação, a minimização do negativo do logaritmo da verossimilhança (em inglês, *log-likelihood*) é um problema de otimização

equivalente. Neste caso, a função-custo pode ser expressa como:

$$\mathcal{F}(\mathbf{x}, \mathbf{W}) = -\log \det|\mathbf{W}| - \sum_{i=1}^N \mathbb{E}[\log f_i(y_i)], \quad (4.11)$$

onde $\mathbf{y} = \mathbf{W}\mathbf{x}$. Na prática, as distribuições $f_i(y_i)$ nos são desconhecidas, cabendo portanto empregar uma distribuição $\tilde{f}_i(y_i)$ apropriada.

A estrutura de Riemann no espaço de todas as matrizes não-singulares de dimensões $N \times N$ forma um grupo de Lie, cuja invariância pode ser empregada para deduzir o gradiente natural do negativo do logaritmo da verossimilhança (Eq. (4.11)), dando origem à equação de atualização¹ [89]:

$$\mathbf{W}^{\text{nova}} = \mathbf{W}^{\text{antiga}} + \beta (\mathbf{I}_{N \times N} - \mathbb{E}[\boldsymbol{\phi}(\mathbf{y})\mathbf{y}^T]) \mathbf{W}^{\text{antiga}}, \quad (4.12)$$

onde $\boldsymbol{\phi}(\mathbf{y}_k) = [\phi_1(y_1(k)) \quad \phi_2(y_2(k)) \quad \cdots \quad \phi_N(y_N(k))]^T$, com $\phi_i(y_i(k))$ denotando a função *score*, relacionada à distribuição das fontes por meio da fórmula:

$$\phi_i(y_i) = -\frac{d \log f_i(y_i)}{dy_i}. \quad (4.13)$$

Cabe ressaltar que a atualização apresentada na Eq. (4.12) tem algumas variantes (das quais a maioria pode ser consultada em [92]), como por exemplo a seguinte versão [93]:

$$\mathbf{W}^{\text{nova}} = \mathbf{W}^{\text{antiga}} + \beta (\text{diag}(\mathbb{E}[\boldsymbol{\phi}(\mathbf{y})\mathbf{y}^T]) - \mathbb{E}[\boldsymbol{\phi}(\mathbf{y})\mathbf{y}^T]) \mathbf{W}^{\text{antiga}}, \quad (4.14)$$

onde $\text{diag}(\mathbf{v})$ gera uma matriz diagonal, dispondo os elementos de \mathbf{v} ao longo de sua diagonal principal.

Para distribuições supergaussianas não raro emprega-se $\phi_i(y_i) = \tanh(y_i)$ e, para subgaussianas, $\phi_i(y_i) = y_i^3$ [11]. Mesmo quando a Eq. (4.13) não se aplica (por exemplo, em distribuições uniformes), estas escolhas não prejudicam em demasia a qualidade da separação [11].

4.9 Esparsidade

A hipótese central da Análise de Componentes Esparsos (SCA, do inglês *Sparse Component Analysis*) consiste em supor fontes esparsas. Na literatura, diversos conceitos de esparsidade tem sido usados. No contexto de transformações esparsificadoras (como a STFT ou a *wavelet packet*), a esparsidade significa que a maior parte das amostras de cada fonte são próximas de zero ou que, alternativamente,

¹A derivação é razoavelmente longa e, por simplicidade, será omitida.

relativamente poucas amostras concentrem a maior parte da energia de cada fonte. No contexto de representações esparsas (como a oriunda do emprego de dicionários redundantes), a esparsidade significa que a norma l_0 dos coeficientes de uma dada representação é pequena.

Desafortunadamente, muitos sinais de interesse, como sinais de áudio, voz e dados financeiros não apresentam esta propriedade. Este fato não impede o recurso à SCA desde que, através de uma transformação, encontremos uma representação esparsa das fontes [94], representação essa que consiste em importante tema de pesquisa [80].

Assim, é comum ter como primeiro passo da SCA o emprego de uma transformação esparsificadora, pois a projeção das misturas para um espaço apropriado de representação esparsa aumenta a qualidade da separação. Os elementos do espaço de representação não precisam ser linearmente independentes podendo, por exemplo, formar um dicionário redundante [95], estratégia que será contemplada no Cap. 7. Entre as transformações mais utilizadas, temos a STFT [96], a *wavelet packet* [95],[97] e as correspondentes dos algoritmos derivados do *Matching Pursuit* [98]. Após a transformação, a SCA atua nos coeficientes da mesma, tratando-os quais amostras das fontes. Após as estimativas de cada fonte serem obtidas, é necessário, quando utilizamos uma transformação esparsificadora, efetuar a transformada inversa da mesma, pois na realidade nestes casos a SCA estima os coeficientes das fontes no domínio da transformada. Logo, o último passo da SCA em geral efetua a transformação inversa da empregada no primeiro passo, passando as estimativas no domínio da transformação para o domínio original de cada fonte (em geral, no domínio do tempo).

O segundo passo da SCA consiste na estimativa da matriz \mathbf{H} de mistura. Eis uma importante diferença entre a SCA e a ICA, já que, enquanto esta efetua sua busca no espaço das matrizes de separação, aquela almeja estimar diretamente a matriz de mistura. Uma das razões para este proceder reside no fato de \mathbf{H} não possuir inversa, por não ter posto completo nos casos em que $M < N$ (configurações que são o principal foco da SCA).

Após estimada a matriz \mathbf{H} (a menos das ambiguidades de escalamento e permutação), a SCA empreende a reconstrução das fontes. Esta etapa não é trivial como nas configurações determinadas, onde basta-nos multiplicar a matriz das misturas pela estimativa de sua inversa. Em outras palavras, em contextos subdeterminados, um boa estimativa de \mathbf{H} não implica termos razoáveis estimativas das fontes.

Assim, podemos condensar a SCA em quatro passos: i) aplicação de uma transformação esparsificadora; ii) estimativa da matriz \mathbf{H} ; iii) estimativa dos coeficientes das fontes no domínio da transformada ou reconstrução das fontes e iv) aplicação da inversa da transformação empregada em (i). Em geral, a ICA não utiliza as etapas (i)

e (iv) (embora recorrer a transformadas possa ser um procedimento útil [99],[100]), condensando as etapas (ii) e (iii) numa única (já que encontra diretamente uma estimativa da inversa da matriz de misturas, o que basta para reconstruir as fontes). Embora existam alguns métodos de SCA que estimam simultaneamente as fontes e a matriz de mistura [88], a maioria aborda estas tarefas de forma independente, resultando num custo computacional muito menor, sem aparentemente prejudicar a qualidade da separação. Por esta razão, apenas estas técnicas serão abordadas nesta tese.

4.10 Estimativa da Matriz de Mistura

Seja \mathbf{h}^j a j -ésima coluna da matriz de mistura \mathbf{H} . Podemos então expressar \mathbf{x}_k como:

$$\mathbf{x}_k = \sum_{j=1}^N \mathbf{h}^j s_j(k), \quad (4.15)$$

donde se vê que cada vetor \mathbf{x}_k é uma combinação linear das colunas da matriz de misturas. Por exemplo, se a matriz \mathbf{H} é:

$$\mathbf{H} = \begin{bmatrix} 0,3273 & -0,1867 \\ 0,1746 & 0,7258 \end{bmatrix},$$

e temos $\mathbf{s}_k = [1, 5 \ 4]^T$, então $\mathbf{x}_k = [-0,2559 \ 3,1651]^T$, conforme ilustra a Fig. 4.11.

Tendo pressuposta a esparsidade das fontes, espera-se que a maior parte dos

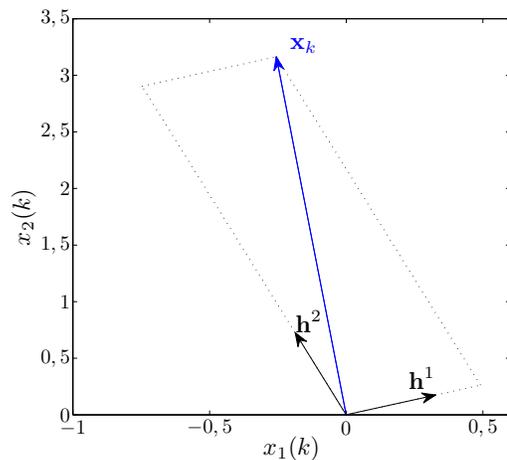


Figura 4.11: Exemplo dos vetores de mistura e da matriz de mistura numa configuração determinada; os vetores em preto são as colunas da matriz de mistura; o vetor \mathbf{x}_k aparece em azul.

coeficientes das fontes sejam próximos de zero. Em outras palavras, temos que em

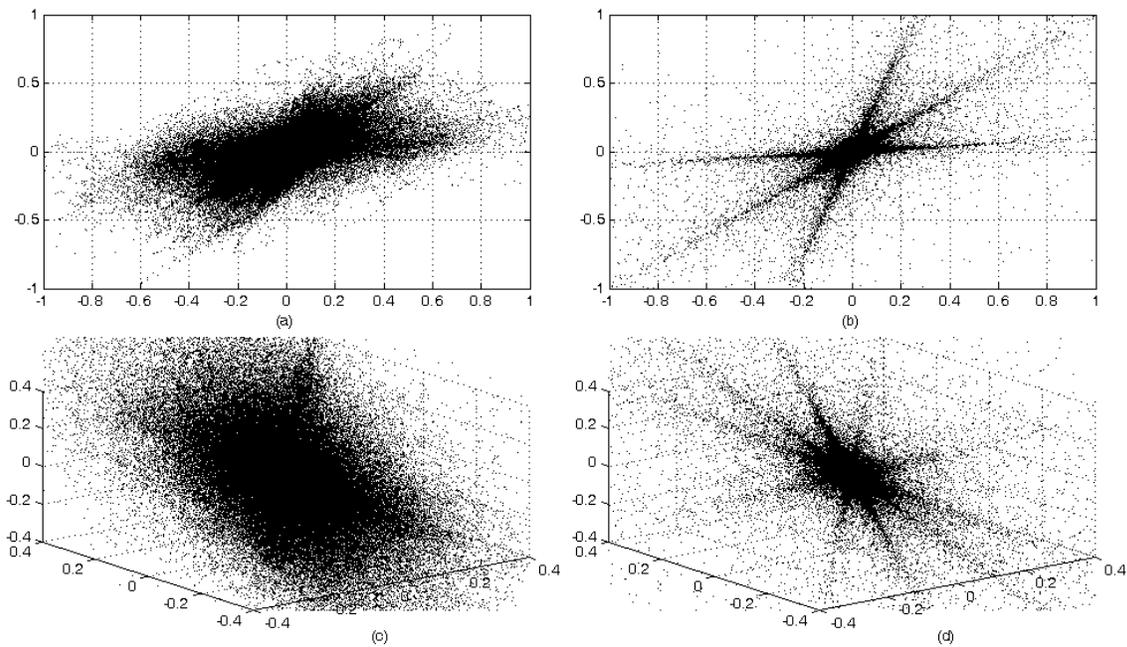


Figura 4.12: Pontos dos vetores \mathbf{x}_k . (a) configuração 2×3 (domínio do tempo); (b) configuração 2×3 (domínio da transformada); (c) configuração 3×5 (domínio do tempo) e (d) configuração 3×5 (domínio da transformada).

cada instante k , em geral, apenas uma fonte (ou, equivalentemente, a coluna da matriz de mistura correspondente) contribui fortemente na combinação linear que gera o vetor \mathbf{x}_k . Esta fonte é chamada de fonte dominante. Isto significa que, caso plotemos todos os vetores \mathbf{x}_k , verificaremos uma tendência de aglomeração destes vetores em direções específicas. Estas direções preferenciais são exatamente as direções das colunas da matriz \mathbf{H} .

A Fig. 4.12 ilustra a formação de retas direcionais nos casos 2×3 e 3×5 (onde $M \times N$ indica M misturas e N fontes), tanto no domínio do tempo quanto no domínio da transformada, para misturas de sinais de voz e música. Nos gráficos Fig. 4.12 (b) e Fig. 4.12 (d), a transformada utilizada foi a *wavelet packet* de 7 níveis com *wavelet* Daubechies 32. Verifica-se que as retas são melhor visualizadas no domínio da transformada, já que os coeficientes desta são mais esparsos do que as amostras das fontes no domínio do tempo.

Se as direções preferenciais dos vetores \mathbf{x}_k coincidem com as direções dos vetores coluna da matriz de mistura, podemos estimar a matriz de mistura a partir do conhecimento destas direções preferenciais, como descrevemos a seguir.

Assumindo 2 misturas e admitindo que somente a i -ésima fonte não é nula no instante k , temos:

$$x_1(k) = \sum_{j=1}^N h_{1j}s_j(k) = h_{1i}s_i(k),$$

$$x_2(k) = \sum_{j=1}^N h_{2j}s_j(k) = h_{2i}s_i(k), \quad (4.16)$$

$$\frac{x_2(k)}{x_1(k)} = \frac{h_{2i}}{h_{1i}} \Rightarrow x_2(k) = \frac{h_{2i}}{h_{1i}}x_1(k).$$

Em outras palavras, quando somente a i -ésima fonte está ativa, $x_2(k)$ torna-se proporcional a $x_1(k)$, com a constante de proporcionalidade dependendo da i -ésima coluna da matriz de mistura. Tendo acesso a \mathbf{x}_k , podemos estimar esta razão. Porém sempre há um grau de liberdade na solução, pois a razão depende de 2 números. Um procedimento comum em SCA para contornar este problema consiste em supor que a primeira linha da matriz de mistura apresenta apenas elementos iguais a 1 (este não é um procedimento problemático, dada a ambiguidade de escalamento - vide Seção 4.2). Assim, chegamos a:

$$x_2(k) = h_{2i}x_1(k). \quad (4.17)$$

Tendo o vetor \mathbf{x}_k por componentes $x_1(k)$ e $x_2(k)$, a tangente do seu ângulo pode ser definida como:

$$\text{tg}(\theta_k) = \frac{x_2(k)}{x_1(k)} = \frac{h_{2i}x_1(k)}{x_1(k)} = h_{2i}, \quad (4.18)$$

de onde vemos que, após estimados os ângulos preferenciais, a tangente destes consiste na estimativa da segunda linha da matriz de mistura². A seguir, veremos um dos métodos mais populares para a estimativa destas direções preferenciais.

4.10.1 Varredura de Zibulevsky

Proposta em [96], a abordagem da Varredura de Zibulevsky consiste num dos primeiros métodos da SCA. É um método simples e altamente competitivo em relação aos mais recentes na literatura. O método consiste na formação de um histograma angular, numa busca dos ângulos do gráfico $x_2 \times x_1$ que possuem maior densidade de pontos. Para a formação do histograma, são escolhidos ângulos teste (tipicamente centenas deles) uniformemente distribuídos no intervalo $[0, \pi]$. Não é necessário contemplar o intervalo $(\pi, 2\pi)$, já que uma reta definida por um ângulo θ equivale a uma definida pelo ângulo $\theta + \pi$ (os ângulos são os únicos parâmetros necessários para determinar as retas, dado que todas elas passam pela origem), bastando que os vetores \mathbf{x}_k cujos ângulos encontrem-se no intervalo $(\pi, 2\pi)$ sejam refletidos simetricamente, de forma que seus ângulos passem a se encontrar no intervalo $[0, \pi]$. Para

²Note que a tangente do ângulo $\frac{\pi}{2}$ é infinita. Esta situação implica que um dos sensores não captura nenhuma informação de uma das fontes, o que significa que não temos uma mistura de todas as fontes em cada sensor. Na prática, esta possibilidade é incomum e não será abordada nesta tese.

cada ângulo de teste θ_t , a varredura calcula a contribuição de todos os ângulos θ_k (associados aos vetores \mathbf{x}_k), somando estas contribuições ao final para gerar a “densidade” de pontos na vizinhança de θ_t . A contribuição de \mathbf{x}_k e de seu respectivo θ_k a θ_t é avaliada segundo a “função potencial” abaixo:

$$\Phi(\theta_t, \theta_k, \mathbf{x}_k) = \begin{cases} |\mathbf{x}_k| \left(1 - \frac{\lambda|\theta_t - \theta_k|}{\pi/4}\right), & \lambda|\theta_t - \theta_k| < \frac{\pi}{4} \\ 0, & \lambda|\theta_t - \theta_k| \geq \frac{\pi}{4} \end{cases}, \quad (4.19)$$

onde λ é um parâmetro arbitrário, geralmente da ordem de dezenas. A função potencial dá maior peso a ângulos θ_k próximos de θ_t (anulando a contribuição quando se encontram muito distantes) e atribui uma confiança maior aos \mathbf{x}_k de maior módulo, dado que o erro angular destes é menos suscetível a ruído.

Na realidade, assim como na maior parte da literatura, a varredura de Zibulevsky, no cálculo da função potencial, despreza os vetores \mathbf{x}_k de módulos baixos, pois estes são mais sensíveis a interferências. Este procedimento possui a vantagem de agilizar muito o processamento (dado que a maioria dos vetores \mathbf{x}_k possuem módulo baixo). Tipicamente, desprezam-se os pontos cujo módulo seja inferior a um décimo do maior módulo de \mathbf{x}_k . Um histograma típico da varredura de Zibulevsky para uma configuração 2×3 é mostrado na Fig. 4.13, sendo que cada um dos três picos associa-se a uma das fontes.

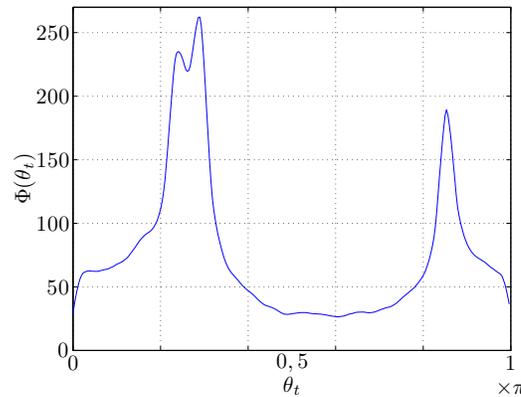


Figura 4.13: Histograma típico obtido pela varredura de Zibulevsky.

4.11 Métricas de Avaliação das Técnicas Não-Supervisionadas de Separação e Desconvolução Cegas de Fontes

Critérios de desempenho possibilitam comparações entre diferentes técnicas e estruturas voltadas para os problemas de separação e desconvolução cega de fontes. Estes critérios podem ser subjetivos (quando fundamentados no julgamento de uma quantidade significativa de pessoas) ou objetivos (caso sua aferição apele apenas a um processamento computacional). Enquanto estes são menos custosos, aqueles costumam agregar maior representatividade às suas estimativas. Os critérios objetivos se desdobram em medidas dependentes do conhecimento das estrutura de mistura e de separação ou em medidas que independem deste conhecimento. Outra classificação possível atenta para as métricas com referência (as quais exigem o conhecimento das fontes) e métricas sem referência (as quais prescindem deste conhecimento). As métricas empregadas nesta tese serão a seguir apresentadas.

4.11.1 Razão Fonte-Interferências (SIR) e Razão Fonte-Distorção (SDR)

Num contexto convolutivo de mistura, esta pode ser modelada como:

$$\mathbf{x} = \mathbf{H} * \mathbf{s} + \boldsymbol{\nu}, \quad (4.20)$$

onde \mathbf{H} é a matriz cujos elementos são os filtros de mistura e o vetor $\boldsymbol{\nu}$ coleta os erros de modelagem e ruídos. Dentre as mais importantes métricas objetivas de avaliação da qualidade de separação encontram-se a razão fonte-interferências (SIR, sigla em inglês de *Source-to-Interferences Ratio*) e a razão fonte-distorções (SDR, sigla em inglês de *Source-to-Distortion Ratio*). [101] propõe uma forma de cálculo da SIR que se tornou muito popular, particularmente após a disponibilização dos respectivos códigos-fonte. Esta formulação é independente das estruturas de mistura e de separação e necessita do conhecimento das fontes, classificando-se portanto como um método com referência.

Por simplicidade, suponhamos a inexistência do problema de permutação; logo, a j -ésima estimativa y_j corresponde à j -ésima fonte s_j . O primeiro passo para o cálculo da SIR e da SDR consiste na decomposição de y_j em quatro componentes:

$$y_j = s_{\text{salvo}} + e_{\text{interf}} + e_{\text{ruído}} + e_{\text{artef}}, \quad (4.21)$$

com s_{alvo} correspondendo a uma versão de s_j modificada por uma distorção permitida e e_{interf} , $e_{\text{ruído}}$ e e_{artef} correspondem, respectivamente, aos termos de erro devidos a interferências, ruídos e artefatos. Se s_j^τ e ν_i^τ correspondem à fonte s_j e ao ruído ν_i atrasados de τ unidades e $\mathcal{P}_{\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q\}}$ $\{\mathbf{y}_j\}$ é a projeção ortogonal de \mathbf{y}_j no subespaço gerado pelos vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q$, a proposta de [101] consiste em calcular os componentes da Eq. (4.21) da seguinte forma:

$$s_{\text{alvo}} = \mathcal{P}_{\{(s_j^\tau)_{0 \leq \tau \leq L-1}\}} \{y_j\}, \quad (4.22)$$

$$e_{\text{interf}} = \mathcal{P}_{\{(s_{j'}^\tau)_{1 \leq j' \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\} - s_{\text{alvo}}, \quad (4.23)$$

$$e_{\text{ruído}} = \mathcal{P}_{\{(s_{j'}^\tau, \nu_i^\tau)_{1 \leq j' \leq N, 1 \leq i \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\} - \mathcal{P}_{\{(s_{j'}^\tau)_{1 \leq j' \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\}, \quad (4.24)$$

$$e_{\text{artef}} = y_j - \mathcal{P}_{\{(s_{j'}^\tau, \nu_i^\tau)_{1 \leq j' \leq N, 1 \leq i \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\}, \quad (4.25)$$

$$(4.26)$$

onde L é o atraso máximo permitido. A Fig. 4.14 apresenta a decomposição de uma estimativa $\underline{\mathbf{y}}_j = [y_j(k) \quad y_j(k+1) \quad y_j(k+2)]$, em termos de projeções em subespaços gerados por $\underline{\mathbf{s}}_1$, $\underline{\mathbf{s}}_2$ e $\underline{\mathbf{v}}_j$; as definições destes vetores são análogas à definição de $\underline{\mathbf{y}}_j$.

Após calculados s_{alvo} e e_{interf} , a SIR e a SDR, robustas à ambiguidade de filtragem, são determinadas da forma a seguir:

$$\text{SIR} = 10 \log_{10} \frac{\|s_{\text{alvo}}\|^2}{\|e_{\text{interf}}\|^2}, \quad (4.27)$$

$$\text{SDR} = 10 \log_{10} \frac{\|s_{\text{alvo}}\|^2}{\|e_{\text{interf}} + e_{\text{ruído}} + e_{\text{artef}}\|^2}. \quad (4.28)$$

Na Fig. 4.14.(a), as medidas de SIR e SDR são de, respectivamente, 5,38 dB e 0,75 dB; já na Fig. 4.14.(b), estas medidas são -4,26 dB e -5,35 dB. Importa notar que o fato de os vetores $\underline{\mathbf{s}}_1$, $\underline{\mathbf{s}}_2$ e $\underline{\mathbf{v}}_j$ gerarem \mathbb{R}^3 implica a inexistência do componente devido a artefatos ($e_{\text{artef}} = 0$) na Fig. 4.14. Estes resultados sugerem que a estimativa $\underline{\mathbf{y}}_j$ provavelmente está relacionada à fonte 1. Para uma avaliação global do processo de separação, os cálculos das métricas de avaliação são efetuados para todas as permutações possíveis das estimativas, escolhendo-se a permutação que gera em média os valores mais altos de SIR ou SDR.

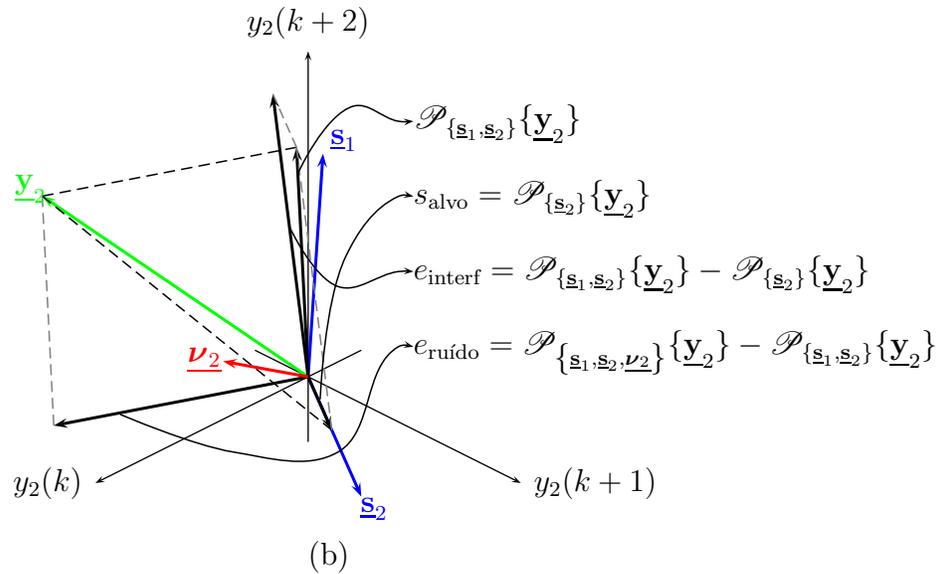
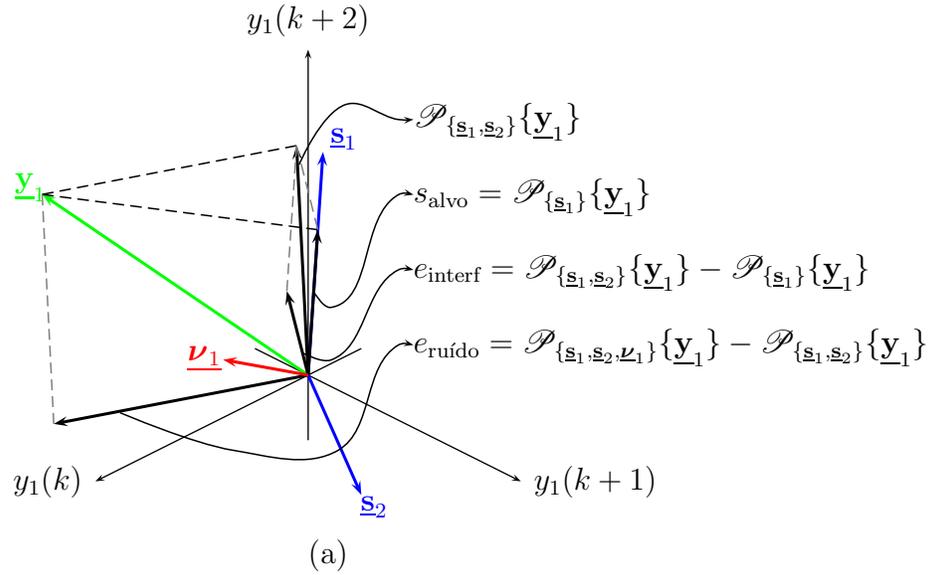


Figura 4.14: A decomposição, num contexto de misturas instantâneas, de um mesmo vetor. (a) supondo que seja uma estimativa da fonte 1; (b) supondo que seja uma estimativa da fonte 2. Os vetores das fontes, do ruído e da estimativa foram mantidos constantes.

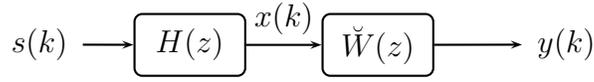


Figura 4.15: Estrutura de desconvolução que será empregada nesta tese. O filtro $H(z)$ (desconhecido) modela a transformação empreendida pelo canal e $\check{W}(z)$ é o filtro equalizador.

4.11.2 Interferência Interssimbólica

Além da necessidade de preservar a estrutura temporal das fontes, o fato de as técnicas de desconvolução cega de fontes poderem atuar num contexto monocal canal encerra outra importante distinção com relação às técnicas de separação cega, pois o principal objetivo das técnicas de desconvolução pode se resumir a reverter (a menos de um atraso e de, possivelmente, uma ambiguidade de escala ou de sinal) o efeito de um canal, cuja distorção é não raro modelada por um filtro FIR.

Sejam $H(z)$ e $\check{W}(z)$ os filtros correspondentes, respectivamente, ao canal e ao equalizador, conforme ilustrado pela Fig. 4.15. A função de transferência $C(z) = H(z)\check{W}(z)$ é a resposta global do processo de filtragem do canal e de desconvolução. A resposta global pode ser utilizada para o cálculo da interferência interssimbólica (ISI, do inglês *intersymbol interference*), a qual é uma métrica de desempenho do processo de equalização, que será objeto de atenção no Cap. 8. Em contraposição à SIR, uma ISI reduzida reflete uma equalização bem-sucedida.

Nesta tese, o cálculo da ISI é efetuado por meio da fórmula [102]:

$$\text{ISI} = 10 \log_{10} \left(\frac{\sum_l c_l^2}{\max_j c_j^2} - 1 \right), \quad (4.29)$$

na qual c_i é o i -ésimo elemento da resposta global (canal + equalizador). Intuitivamente, a ISI reflete o grau de proeminência do pico da resposta global.

4.11.3 Avaliação Subjetiva

É impraticável, dada a complexidade das avaliações subjetivas, aplicá-las a todas as propostas de técnicas para a resolução do problema de separação cega de fontes apresentadas nesta tese. Porém, numa técnica específica proposta no Cap. 5, lançamos mão de uma métrica subjetiva, a qual atuará como complemento da avaliação objetiva, num contexto de separação cega de sinais de voz.

O teste subjetivo implementado tem como primeiro passo a apresentação ao avaliador de uma das misturas captadas pelo microfone, a primeira estimativa e a segunda estimativa (nesta ordem). Cabe então ao avaliador conceder uma nota,

numa escala de 1 (significando “muito perturbador”) a 5 (“imperceptível”).

Quatro diferentes aspectos serão avaliados na escala supracitada:

- O impacto da interferência na compreensão (inteligibilidade) da fonte de interesse;
- A distorção apresentada pela fonte de interesse caso comparada à versão constante na mistura, com o objetivo de avaliar variações drásticas de timbre;
- O nível de ruído de fundo, caso comparado à mistura;
- A qualidade total do sistema (opinião global da qualidade de separação).

No Cap. 5, a avaliação subjetiva será empregada para consolidar a análise da proposta de inicialização das matrizes de separação, empregadas em técnicas de separação de fontes no domínio da frequência, as quais costumam promover uma inserção mais intensa de artefatos do que as técnicas no domínio do tempo.

4.12 Considerações Finais

Este capítulo serve de prelúdio para os Caps. 5 e 6, detalhando o problema da separação cega de fontes e alguns princípios utilizados para resolvê-lo. O Cap. 5 contempla as propostas desta tese concernentes a técnicas de separação no domínio da frequência, enquanto que o Cap. 6 apresenta inovações concernentes a extensões em subbandas de técnicas de separação cega de fontes no domínio do tempo.

Capítulo 5

Separação Cega de Fontes no Domínio da Frequência

5.1 Introdução

Neste capítulo investigamos técnicas de separação cega de fontes que operam no domínio da frequência. Tais técnicas, especialmente devido ao seu reduzido custo computacional, englobam a maioria das abordagens propostas na literatura. Esta família de técnicas de resolução do problema de BSS fundamenta-se em transformações, notadamente a STFT. Importa realçar a arbitrariedade que reside na distinção entre técnicas que apelam a transformações e as que empregam bancos de filtros; talvez seja mais apropriado diferenciá-las no tocante à sua resolução frequencial (baixa ou elevada) [103].

Este capítulo estrutura-se do seguinte modo: a Seção 5.2 introduz o modelo padrão desta família de técnicas, ou seja, a estrutura mais comum de implementação destas técnicas. A Seção 5.3 disserta sobre uma nova forma de inicialização das matrizes de separação. Já a Seção 5.4 detalha uma proposta de inserção de coeficientes interfrequenciais, a qual contorna algumas limitações do modelo padrão. Por sua vez, a Seção 5.5 investiga se o emprego de estimativas cegas de T_{60} , originariamente propostas para registros de um único sinal de voz, podem providenciar informação lateral suficiente, em contextos de misturas de fontes, para permitir uma escolha orientada do tamanho do *frame* da STFT empregada nas técnicas de separação cega. Por fim, a Seção 5.6 elenca as conclusões advindas das seções progressivas.

5.2 Separação Cega de Fontes no Domínio da Frequência

Nas técnicas de BSS no domínio da frequência, as misturas capturadas pelos microfones são primeiramente transformadas para um domínio que permita um acompanhamento da evolução espectral das misturas. Em geral, o mapeamento tempo-frequência empregado recorre à transformada STFT. Devido à redundância desta transformada¹, o conjunto de frequências nas quais o processo de separação será integralmente aplicado restringe-se a $\{0, \frac{f_s}{F}, \dots, \frac{f_s}{2}\}$. Na literatura de separação de fontes, janelas de Hanning são as mais populares, com deslocamentos de $F/2$ ou $F/4$. O fato de estes deslocamentos serem inferiores a F insere as STFTs neste capítulo empregadas na classe das representações redundantes, as quais também são objeto de atenção no Cap. 7. O fato de cada raia estar associada a uma sequência temporal é essencial para o processo de separação, como veremos.

A partir da suposição de que o comprimento F dos quadros da STFT seja suficientemente longo para contemplar a parte principal das respostas ao impulso do sistema de mistura², a hipótese de que o modelo convolutivo pode ser aproximado por um modelo instantâneo em cada frequência é amplamente utilizada, ou seja,

$$\mathcal{X}(\mathbf{f}, k) = \begin{bmatrix} \mathcal{X}_1(\mathbf{f}, k) \\ \mathcal{X}_2(\mathbf{f}, k) \\ \vdots \\ \mathcal{X}_M(\mathbf{f}, k) \end{bmatrix} \approx \mathbf{H}(\mathbf{f})\mathcal{S}(\mathbf{f}, k) = \mathbf{H}(\mathbf{f}) \begin{bmatrix} \mathcal{S}_1(\mathbf{f}, k) \\ \mathcal{S}_2(\mathbf{f}, k) \\ \vdots \\ \mathcal{S}_N(\mathbf{f}, k) \end{bmatrix}. \quad (5.1)$$

Se a aproximação da Eq. (5.1) é modelada como uma igualdade, o problema da separação cega de fontes em ambientes reverberantes exige a resolução de múltiplos problemas de separação de fontes misturadas conforme um modelo instantâneo, o que motiva a busca por matrizes de separação $\mathbf{W}(\mathbf{f})$ ao longo das raias. Embora teoricamente a modelagem seja menos aproximada quando F é razoavelmente superior a M , na prática esta condição não costuma ser satisfeita, porque, concomitantemente ao aumento de coeficientes adaptativos a estimar, quadros longos implicam redução no número de amostras nas raias, de sorte que a escassez de informação disponível colapsa as hipóteses de independência e de média nula em cada componente frequencial; por sua vez, quadros pequenos não permitem ao sistema de separação compensar devidamente os efeitos da reverberação [104].

O emprego do modelo da Eq. (5.1) torna as técnicas de separação cega de fontes

¹Para sinais reais, a STFT apresenta a propriedade de simetria conjugada, a qual implica que $\mathcal{X}_i(\mathbf{f}, k) = \mathcal{X}_i^*(F - \mathbf{f}, k)$, para $\mathbf{f} = 1, \dots, \frac{F}{2} - 1$.

²Seguindo a notação de capítulos anteriores, M denota o comprimento dos filtros do sistema de mistura.

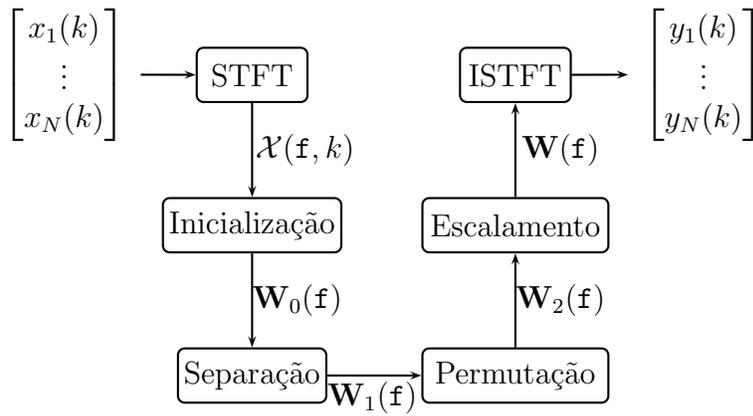


Figura 5.1: Diagrama de blocos das técnicas de separação de fontes no domínio da frequência.

no domínio da frequência atrativas em aplicações que requerem filtros de separação de ordem elevada (como em misturas de sinais de voz em ambientes reverberantes), pois permite que custosas operações de convolução no domínio do tempo se transmutem em produtos (ou misturas instantâneas) ao longo das raias.

O processo de determinação das matrizes de separação $\mathbf{W}(\mathbf{f})$, incumbidas de reverter o efeito das matrizes de mistura $\mathbf{H}(\mathbf{f})$, costuma em geral seguir quatro passos: inicialização, separação, correção de permutação e resolução de escalamento, conforme ilustra a Fig. 5.1. Na Seção 5.4 deste capítulo será proposta uma técnica de determinação dos parâmetros de uma estrutura de separação mais complexa, a qual utiliza informações laterais oriundas de raias adjacentes.

5.2.1 Inicialização

O procedimento padrão de inicialização do processo de separação ao longo das raias se fundamenta no branqueamento (via PCA) das misturas.

Em misturas sobredeterminadas, a estratégia adicional de reter as N primeiras componentes principais (mapeando o número de misturas para o número de estimativas) é utilizada para reduzir a dimensão das misturas. É interessante efetuar esta redução antes da aplicação da ICA (quando temos mais informação relevante para a escolha de um subespaço apropriado), já que a ambiguidade de escalamento distorce os autovalores da matriz de covariância das estimativas [105]. Um método alternativo é a seleção de subespaço fundamentada na geometria dos sensores [106], a qual emprega os dados oriundos dos sensores mais distantes para a separação nas baixas frequências, enquanto que para as altas os dados dos sensores mais próximos são os selecionados. Em [105], verificou-se a maior robustez do emprego da PCA que, ao garantir a maior diferença de fase possível entre as componentes principais, naturalmente enfatiza os sensores mais distantes em baixas frequências (angariando

portanto um desempenho similar ao do método advogado por [106]). Em frequências altas e na presença de ruído, o fato de a PCA recorrer à informação que advém de todos os sensores justifica seu desempenho superior [105].

O primeiro passo para a inicialização via PCA consiste em efetuar uma transformação linear ao longo das raias:

$$\mathbf{Z}(\mathbf{f}, k) = \mathbf{V}(\mathbf{f})\mathbf{X}(\mathbf{f}, k), \quad (5.2)$$

onde $\mathbf{V}(\mathbf{f})$ é uma matriz de branqueamento de dimensões $N \times M$, capaz de compatibilizar o número de misturas com o número de estimativas (ou fontes). As sequências $\mathbf{Z}(\mathbf{f}, k)$ geradas no processo consistem de versões esferizadas de $\mathbf{X}(\mathbf{f}, k)$.

A matriz $\mathbf{V}(\mathbf{f})$ costuma ser determinada via PCA complexo, por meio da seguinte fórmula:

$$\mathbf{V}(\mathbf{f}) = \mathbf{\Lambda}^{-1/2}\mathbf{Q}^H, \quad (5.3)$$

onde as matrizes $\mathbf{\Lambda}$ e \mathbf{Q} , conforme já mencionado no Capítulo 3, contém, respectivamente, os autovalores e autovetores da matriz $\mathbf{R}_x(\mathbf{f}) = \text{E}[\mathbf{X}(\mathbf{f}, k)\mathbf{X}^H(\mathbf{f}, k)]$, com a média amostral operando ao longo dos quadros.

Sob a restrição de estimativas descorrelacionadas, pode-se resumir o processo de separação à busca de uma matriz unitária³ $\mathbf{W}_0(\mathbf{f})$, de modo que as estimativas sejam dadas por $\mathbf{Y}(\mathbf{f}, k) = \mathbf{W}_0(\mathbf{f})\mathbf{Z}(\mathbf{f}, k) = \mathbf{W}_0(\mathbf{f})\mathbf{V}(\mathbf{f})\mathbf{X}(\mathbf{f}, k)$. Existem alguns métodos de ICA (como o FastICA [107]) que exigem um pré-branqueamento das misturas, e mesmo algoritmos de ICA que não requerem este processamento se beneficiam de uma convergência eficiente caso o implementemos [11].

Branquear as misturas e admitir que as matrizes de separação (que atuam sobre os sinais esferizados) são unitárias implica empregar estatísticas de segunda ordem para “reduzir o trabalho à metade” [84], ao inserir restrições nos elementos das matrizes de separação e diminuir o grau de liberdade dos mesmos. Porém, não é evidente que estatísticas de segunda ordem sejam mais confiáveis que as estatísticas de ordem superior [84], o que é ressaltado pelo fato de as sequências $\mathcal{S}_i(\mathbf{f}, k)$, $i = 1, \dots, N$ não serem completamente descorrelacionadas, dado que dispomos de um número limitado de amostras [108].

Uma estratégia muito comum se resume a aplicar o pré-branqueamento, encontrar uma matriz unitária $\mathbf{W}_0(\mathbf{f})$ por meio de algum algoritmo que exija sinais esferizados (que podem apresentar a vantagem de convergência rápida, como o FastICA) e então refinar a solução $\mathbf{W}_0(\mathbf{f})\mathbf{V}(\mathbf{f})$ obtida, empregando-a como inicialização de um algoritmo que não privilegie as estatísticas de segunda ordem (como o Infomax). Uma possível estratégia alternativa inicializa, sem pré-branqueamento, a matriz de

³Uma matriz unitária \mathbf{A} está sujeita à restrição $\mathbf{A}\mathbf{A}^H = \mathbf{I}$.

separação segundo [109]:

$$\mathbf{W}_0(\mathbf{f}) = (\mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}})^{-1}. \quad (5.4)$$

Neste capítulo, na Seção 5.3 detalhamos uma proposta de um novo princípio de inicialização, fundamentado em detecções da direção de chegada das fontes em cada raia.

5.2.2 Separação

Na etapa de separação, a matriz de separação é iterativamente atualizada, conforme algum algoritmo de ICA. Nesta tese empregamos majoritariamente algoritmos de atualização derivados do princípio Infomax, obtidos via gradiente natural. Seja o vetor $\mathcal{Y}(\mathbf{f}, k)$ definido por:

$$\mathcal{Y}(\mathbf{f}, k) = \mathbf{W}(\mathbf{f})\mathcal{X}(\mathbf{f}, k) = \begin{bmatrix} \mathcal{Y}_1(\mathbf{f}, k) \\ \mathcal{Y}_2(\mathbf{f}, k) \\ \vdots \\ \mathcal{Y}_N(\mathbf{f}, k) \end{bmatrix}. \quad (5.5)$$

A função-custo expressa na Eq. (4.11) assume que as estimativas são unidimensionais (ou seja, variam apenas ao longo do tempo). Porém, a STFT mapeia cada mistura (unidimensional) para uma estrutura bidimensional, na qual os coeficientes estão dispostos tanto temporalmente quanto ao longo de componentes frequenciais. Num determinado quadro, a concatenação de todas as raias de uma mesma estimativa gera um vetor com F elementos. Se $\tilde{\mathcal{Y}}_i(k)$ é definido segundo:

$$\tilde{\mathcal{Y}}_i(k) = \begin{bmatrix} \mathcal{Y}_i(1, k) \\ \mathcal{Y}_i(2, k) \\ \vdots \\ \mathcal{Y}_i(F, k) \end{bmatrix}, \quad (5.6)$$

a função-custo da Eq. (4.11) é facilmente generalizada:

$$\mathcal{F}(\mathcal{X}, \mathbf{W}(\mathbf{f})) = - \sum_{\mathbf{f}=1}^F \log \det |\mathbf{W}(\mathbf{f})| - \sum_{i=1}^N \mathbb{E} \left[\log \mathbf{f}_i \left(\tilde{\mathcal{Y}}_i(k) \right) \right], \quad (5.7)$$

onde \mathbf{f}_i passa a ser multidimensional, dado que depende de F raias. Não raro o processo físico responsável pela geração dos sinais das fontes engendra uma interdependência entre as raias de uma mesma fonte num dado quadro. Esta interdependência se manifesta, por exemplo, nos instantes de silêncio (nos quais uma determinada fonte se encontra inativa) ou nas características de timbre de um sinal de voz. A

maioria das estratégias de ICA empregadas na literatura opta por não contemplar as dependências estatísticas existentes entre as raias durante parte do processo de determinação das matrizes $\mathbf{W}(\mathbf{f})$. A otimização empregada para o ajuste destas matrizes costuma ser dividida em estágios, com o primeiro deles se restringindo a adaptar as matrizes de separação $\mathbf{W}_1(\mathbf{f})$ (vide Fig. 5.1) em cada raia de modo independente das demais. Este procedimento torna-se natural à medida em que as componentes frequenciais são supostas independentes e identicamente distribuídas, de modo que $\mathbf{f}_i(\tilde{\mathcal{Y}}_i(k)) = \prod_{\mathbf{f}=1}^F f_i(\mathcal{Y}_i(\mathbf{f}, k))$. Neste caso, a Eq. (5.7) pode ser simplificada para:

$$\begin{aligned} \mathcal{F}(\mathcal{X}, \mathbf{W}(\mathbf{f})) &= - \sum_{\mathbf{f}=1}^F \log \det |\mathbf{W}(\mathbf{f})| - \sum_{i=1}^N \sum_{\mathbf{f}=1}^F \mathbb{E} [\log f_i(\mathcal{Y}_i(\mathbf{f}, k))] \quad (5.8) \\ &= - \sum_{\mathbf{f}=1}^F \left[\log \det |\mathbf{W}(\mathbf{f})| - \sum_{i=1}^N \mathbb{E} [\log f_i(\mathcal{Y}_i(\mathbf{f}, k))] \right], \end{aligned}$$

onde verifica-se que o problema de separação se desmembra em diversos processos mais simples e independentes de otimização nas raias, supostamente não interrelacionadas, tornando célere e estável o processo de convergência [110].

Enquanto que em misturas instantâneas o fato de as estimativas serem versões permutadas das fontes não costuma ser problemático; em contextos reverberantes, contudo, se as matrizes de separação são determinadas independentemente, a coerência entre as raias de uma mesma estimativa não é preservada, pois a ausência de um controle das permutações durante o processo de otimização as tornam aleatórias ao longo das raias. Este fenômeno demanda técnicas exógenas de correção da permutação, geralmente aplicadas após a determinação das matrizes $\mathbf{W}_1(\mathbf{f})$. A Fig. 5.2 ilustra o problema da permutação, para uma configuração de duas fontes. A correção deste problema costuma ser implementada por meio de uma permutação das linhas das matrizes $\mathbf{W}_1(\mathbf{f})$, orientada por alguma heurística que permita aferir a coerência das estimativas. Posteriormente, detalharemos algumas estratégias de correção de permutação.

A equação de atualização da ICA via gradiente natural, para uma modelagem das raias como independentes, se torna:

$$\mathbf{W}^{\text{nova}}(\mathbf{f}) = \mathbf{W}^{\text{antiga}}(\mathbf{f}) + \beta (\mathbf{I} - \mathbb{E} [\phi(\mathcal{Y}(\mathbf{f}, k)) \mathcal{Y}^H(\mathbf{f}, k)]) \mathbf{W}^{\text{antiga}}(\mathbf{f}), \quad (5.9)$$

onde verifica-se ser a otimização de cada matriz $\mathbf{W}(\mathbf{f})$ independente das demais. Em geral, as funções $\phi_i(\cdot)$ que constam em $\phi(\mathcal{Y}(\mathbf{f}, k)) = [\phi_1(\mathcal{Y}_1(\mathbf{f}, k)) \quad \phi_2(\mathcal{Y}_2(\mathbf{f}, k)) \quad \cdots \quad \phi_N(\mathcal{Y}_N(\mathbf{f}, k))]^T$ são idênticas, independentemente do índice i da estimativa, o que justifica a elisão do mesmo.

O fato de $\mathcal{Y}_i(\mathbf{f}, k)$ serem obtidas via STFT engendra a necessidade de generali-

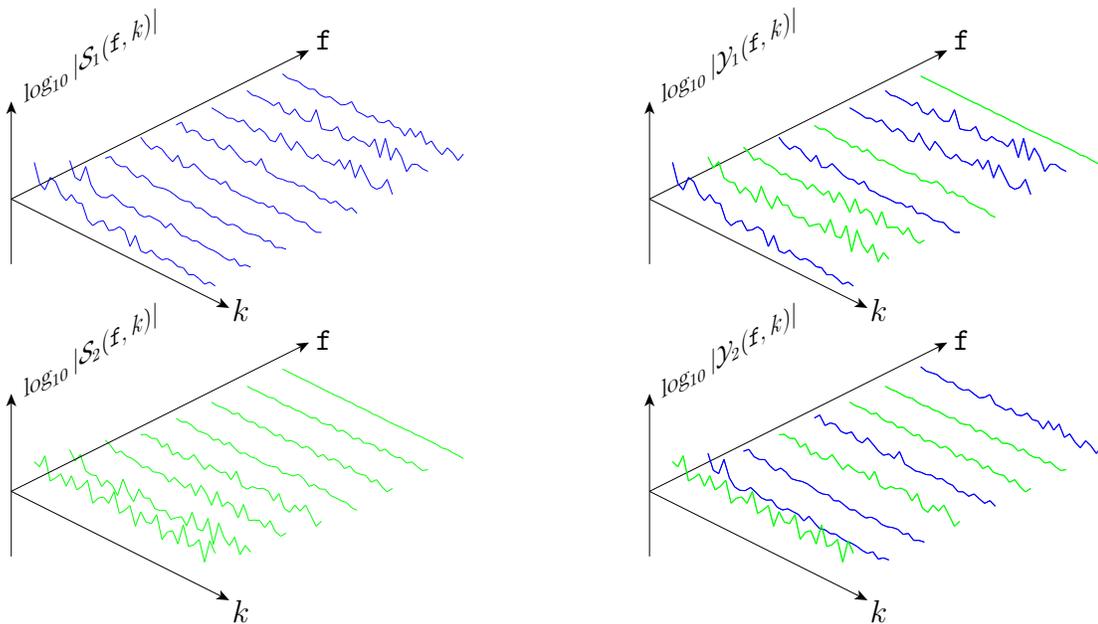


Figura 5.2: Ilustração da importância da resolução de permutação. Em azul, componentes da fonte 1; em verde, componentes da fonte 2. Embora em todas as raias a separação tenha sido bem-sucedida, a falta de consistência na associação das estimativas às fontes corretas degrada de forma intolerável o desempenho do processo de separação. Por simplicidade, supõe-se que o problema de escalamento foi perfeitamente resolvido.

zação das funções *score* para o caso complexo. A função *score* mais popular para sinais de voz é a que emprega a tangente hiperbólica que, para números complexos, é definida por $\tanh(z) = \frac{e^z + e^{-z}}{e^z - e^{-z}}$. Sendo não limitada e indefinida para $z = i(k + \frac{1}{2})\pi$, esta função é inadequada em contextos complexos. Empregando, entre outras, as propriedades de que a função *score* e sua derivada devem ser limitadas, [111] advoga o emprego de:

$$\phi(\mathcal{Y}(\mathbf{f}, k)) = \tanh(\Re(\mathcal{Y}(\mathbf{f}, k))) + i \cdot \tanh(\Im(\mathcal{Y}(\mathbf{f}, k))), \quad (5.10)$$

onde $\Re(z)$ e $\Im(z)$ consistem, respectivamente, nos componentes real e imaginário de z . Verificando que a aplicação da Eq. (5.10), também conhecida como variante “cartesiana”, supõe que a pdf de $\mathcal{Y}_i(\mathbf{f}, k)$ depende da fase (o que não é natural, já que a fase depende da posição da janela, a qual pode ser alterada arbitrariamente), além de exigir uma significativa independência entre as partes real e imaginária e ser passível a problemas de convergência, [112] propõe o recurso a uma função *score* “polar”:

$$\phi(\mathcal{Y}(\mathbf{f}, k)) = \tanh(|\mathcal{Y}(\mathbf{f}, k)|) e^{i\angle\mathcal{Y}(\mathbf{f}, k)}, \quad (5.11)$$

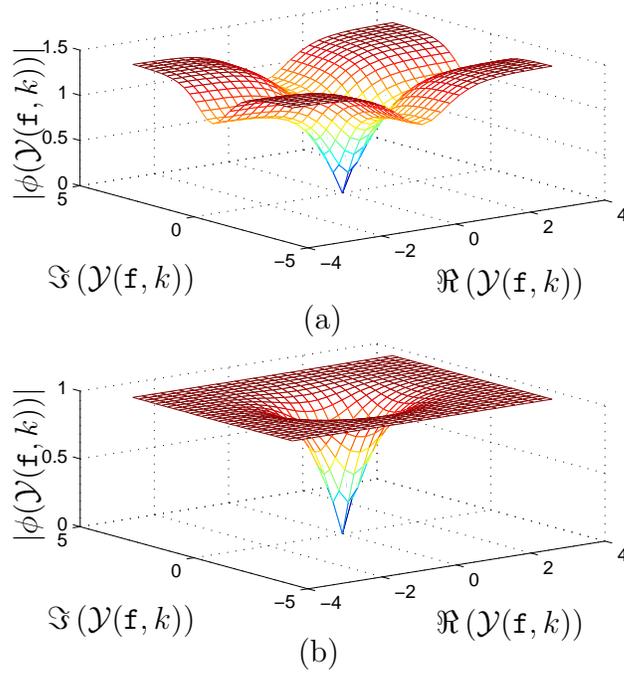


Figura 5.3: Ilustração dos módulos das funções *score* complexas derivadas da tangente hiperbólica real; (a) versão “cartesiana”; (b) versão “polar”.

a qual se mostrou mais robusta do que sua versão “cartesiana”. As duas alternativas são apresentadas na Fig. 5.3, na qual verifica-se uma maior suavidade da versão “polar”, o que contribui para uma convergência mais rápida.

Outra popular opção de função *score* emprega $\phi(\mathcal{Y}(\mathbf{f}, k)) = \frac{\mathcal{Y}(\mathbf{f}, k)}{|\mathcal{Y}(\mathbf{f}, k)|}$. Versões mais sofisticadas incluem funções de densidade de probabilidade adaptativas, como as de Pearson [113].

As supracitadas opções de escolha das funções *score* são produtos de uma modelagem unidimensional, a qual contempla de forma independente todas as raias, ignorando as não raro presentes dependências estatísticas entre as mesmas. Estas dependências são explicitamente levadas em consideração em [114], o qual modela as fontes como multidimensionais ao longo das raias, de modo que, a cada quadro, entende-se que cada fonte engendra um vetor aleatório, formado por coeficientes da STFT. Desse modo, a função *score* se torna multidimensional e dependente da raia. [114] defende a adoção de:

$$\phi_i^{(\mathbf{f})}(\mathcal{Y}_i(1, k), \mathcal{Y}_i(2, k), \dots, \mathcal{Y}_i(F, k)) = \frac{\mathcal{Y}_i(\mathbf{f}, k)}{\sqrt{\sum_{\mathbf{f}=1}^F |\mathcal{Y}_i(\mathbf{f}, k)|^2}}, \quad (5.12)$$

o que modifica a equação de atualização das matrizes $\mathbf{W}(\mathbf{f})$ para:

$$\mathbf{W}^{\text{nova}}(\mathbf{f}) = \mathbf{W}^{\text{antiga}}(\mathbf{f}) + \beta \left(\mathbf{I} - \mathbb{E} \left[\phi^{(\mathbf{f})}(\mathcal{Y}(\mathbf{f}, k)) \mathcal{Y}^H(\mathbf{f}, k) \right] \right) \mathbf{W}^{\text{antiga}}(\mathbf{f}), \quad (5.13)$$

onde $\phi^{(\mathbf{f})}(\mathcal{Y}(\mathbf{f}, k)) = \left[\phi_1^{(\mathbf{f})}(\mathcal{Y}_1(\mathbf{f}, k)) \quad \phi_2^{(\mathbf{f})}(\mathcal{Y}_2(\mathbf{f}, k)) \quad \cdots \quad \phi_N^{(\mathbf{f})}(\mathcal{Y}_N(\mathbf{f}, k)) \right]^T$. O emprego destas funções *score* multidimensionais naturalmente garante um acoplamento entre as matrizes $\mathbf{W}(\mathbf{f})$ (ou equivalentemente, entre os vetores associados a cada uma das estimativas), de modo a prescindir de uma posterior etapa de correção de permutação. Nesta tese, a técnica resultante do recurso a esta função multidimensional será denominada de EDAOR (Exploração de Dependências de Alta Ordem entre Raias).

5.2.3 Correção da Permutação

A otimização das matrizes de separação $\mathbf{W}(\mathbf{f})$ de forma independente não preserva o alinhamento das estimativas, conforme ilustrado pela Fig. 5.2, sendo este fato o principal limitante das técnicas de separação no domínio da frequência, em especial quando o horizonte de reverberação é amplo [115]. Para agrupar os sinais separados (nas raias) que são oriundos da mesma fonte, de modo a tornar efetivo o processo de separação, diversas técnicas foram propostas. Nesta seção, sem pretender esgotar o amplo espectro de contribuições efetuadas na literatura concernentes a este tema, apresentamos concisamente algumas destas técnicas que se destacaram, seja por sua abordagem pioneira, seja por apresentarem conceitos que serão utilizados nas propostas deste capítulo.

Assumindo a exatidão da Eq. (5.1), uma matriz de separação $\mathbf{W}(\mathbf{f})$ seria ideal quando se igualasse à inversa de $\mathbf{H}(\mathbf{f})$. Sendo inerentes ao problema de separação cega de fontes a ambiguidade de escalamento e a permutação, consideramos o processo de separação bem-sucedido quando a condição $\mathbf{W}(\mathbf{f}) = \mathbf{P}(\mathbf{f})\mathbf{D}(\mathbf{f})\mathbf{H}^{-1}(\mathbf{f})$ é satisfeita, onde $\mathbf{D}(\mathbf{f})$ é uma matriz diagonal cujos elementos não-nulos respondem por um escalamento arbitrário e $\mathbf{P}(\mathbf{f})$ é uma matriz de permutação, capaz de ser obtida por meio de permutações arbitrárias das linhas de uma matriz identidade. O problema da permutação no domínio da frequência se manifesta nas variações das matrizes de permutação $\mathbf{P}(\mathbf{f})$ ao longo da frequência [116].

Dentre as diversas soluções propostas na literatura para contornar este problema destacam-se a suavização das matrizes $\mathbf{W}(\mathbf{f})$ ao longo da frequência [117], o emprego de informação geométrica das fontes (distância ou direção de chegada) [118]-[119], a exploração da dependência entre raias de uma mesma fonte (a qual pode ser empregada após os múltiplos processos independentes de separação ao longo das raias [120] ou inserida diretamente no algoritmo de ICA, como o método EDAOR) ou uma combinação das estratégias anteriores [121].

(A) Suavização das Matrizes no Domínio da Frequência

Uma das primeiras propostas para a resolução do problema da permuta-

ção no domínio da frequência advoga o recurso à suavização das matrizes de separação no domínio da frequência [117]. Esta suavização é implementada por meio da restrição do suporte dos filtros de separação no domínio do tempo, de modo que a cada iteração os filtros no domínio da frequência são convertidos para o domínio do tempo, sofrem restrição do seu suporte (tipicamente zerando todos os coeficientes situados após os $F/8$ primeiros coeficientes, com F denotando o comprimento dos filtros envolvidos) e então são novamente transformados para o domínio da frequência. Esta restrição pode ser compreendida, no domínio da frequência, como resultante de uma convolução com uma função sinc de fase linear [122], cujo efeito de suavização garante um acoplamento entre as diferentes raias que, ao evitar que matrizes de separação adjacentes apresentem entre si bruscas mudanças, atenua o surgimento de incoerências nas permutações. Esta técnica foi alvo de investigação em [123], que relatou a permanência de significativo desalinhamento das matrizes de separação em ambientes reais. Isto se dá porque a restrição de suavidade na resposta em frequência, consistindo numa modificação da atualização via gradiente, perturba as soluções intermediárias, disseminando erros de forma descontrolada e não uniforme ao longo das raias [122], o que justifica a pouca popularidade desta estratégia atualmente. Quando se almeja encontrar filtros de separação no domínio do tempo, [122] relata que a restrição do suporte destes durante o processo de separação pode evitar alguns problemas de circularidade e não-causalidade nos mesmos.

(B) Informações Geométrica das Fontes

Inspirando-se no fato de que as fontes estão localizadas em posições distintas, uma família de técnicas de resolução de permutação recorre a informações geométricas para, em cada frequência, determinar a permutação da matriz de separação capaz de preservar o alinhamento das estimativas. Em geral, assume-se um modelo de propagação, seja em campo próximo, seja em campo distante. Por ser mais geral, começaremos por abordar o modelo de campo próximo. Sejam os vetores de posição apresentados na Fig. 5.4. Para fins de estimativa de direção de chegada ou de diferença entre atrasos, usualmente emprega-se um modelo anecoico, que se mostra efetivo quando o horizonte de reverberação é moderadamente baixo [110]. Neste modelo, a componente da i -ésima fonte no j -ésimo sensor apresenta uma atenuação e um atraso, sendo este devido à finitude da velocidade de propagação do som e aquela provocada pela perda de propagação no espaço livre. Embora na literatura a modelagem deste fenômeno seja apresentada para sinais contínuos,

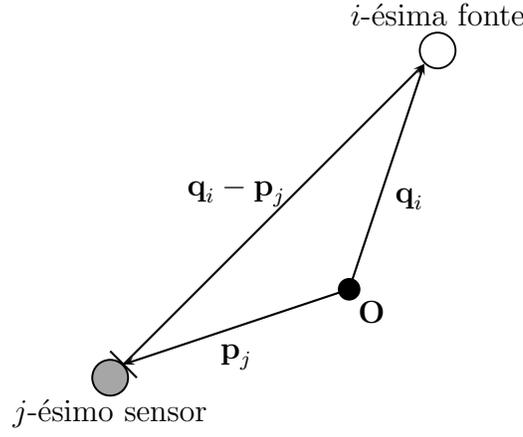


Figura 5.4: Vetores empregados na modelagem de campo próximo.

desenvolveremos versões discretas das derivações presentes em outros artigos, já que, na prática, em sistemas digitais, os sinais são amostrados. A adaptação para sinais discretos permite-nos atentar para algumas sutilezas, inexistentes em contextos contínuos. Esta modelagem nos permite escrever:

$$x_{ji}(k) = h_{ji}s_i(k - \Delta_{ji}), \quad (5.14)$$

onde $x_{ji}(k)$ é a contribuição da i -ésima fonte para o registro do j -ésimo sensor, sendo Δ_{ji} o atraso correspondente. Supondo uma perda por propagação no espaço livre proporcional ao inverso da distância entre a fonte e o sensor, temos:

$$x_{ji}(k) = \frac{\alpha_0}{\|\mathbf{q}_i - \mathbf{p}_j\|} s_i(k - \Delta_{ji}), \quad (5.15)$$

com a constante de proporcionalidade (independente da fonte e do sensor) sendo denotada por α_0 . No domínio da STFT, a razão entre a fonte e sua componente no sensor é dada por:

$$\begin{aligned} \frac{\mathcal{X}_{ji}(\mathbf{f}, k)}{\mathcal{S}_i(\mathbf{f}, k)} &= \frac{\sum_{k'=k-\frac{\mathbb{F}}{2}}^{k+\frac{\mathbb{F}}{2}-1} x_{ji}(k') W_a(k' - k\mathbf{S}) e^{-i2\pi\frac{k'}{\mathbb{F}}} }{\sum_{k'=k-\frac{\mathbb{F}}{2}}^{k+\frac{\mathbb{F}}{2}-1} s_i(k') W_a(k' - k\mathbf{S}) e^{-i2\pi\frac{k'}{\mathbb{F}}}} \\ &= \frac{\frac{\alpha_0}{\|\mathbf{q}_i - \mathbf{p}_j\|} \sum_{k'=k-\frac{\mathbb{F}}{2}-\Delta_{ji}}^{k+\frac{\mathbb{F}}{2}-1-\Delta_{ji}} s_i(k') W_a(k' + \Delta_{ji} - k\mathbf{S}) e^{-i2\pi\frac{k'}{\mathbb{F}}} e^{-i2\pi\frac{k'}{\mathbb{F}}\Delta_{ji}}}{\sum_{k'=k-\frac{\mathbb{F}}{2}}^{k+\frac{\mathbb{F}}{2}-1} s_i(k') W_a(k' - k\mathbf{S}) e^{-i2\pi\frac{k'}{\mathbb{F}}}}. \end{aligned} \quad (5.16)$$

Seja a decomposição dos somatórios do numerador e do denominador da equação acima em duas partes, com uma delas coletando os valores de k' no intervalo $[k - \frac{\mathbb{F}}{2}, k + \frac{\mathbb{F}}{2} - 1 - \Delta_{ji}]$ e a outra com os demais valores. A contribuição

dos termos referentes a estes valores⁴ é, para Δ_{ji} reduzido, pouco significativa, pois, estando estes termos situados nas extremidades do suporte da janela, são ponderados por pesos próximos de zero. Desprezando estes termos e supondo⁵ que $\mathbb{W}(k' + \Delta_{ji} - k\mathbf{S}) \approx \mathbb{W}(k' - k\mathbf{S})$, a Eq. (5.16) pode ser aproximada por:

$$\frac{\mathcal{X}_{ji}(\mathbf{f}, k)}{\mathcal{S}_i(\mathbf{f}, k)} \approx h_{ji}(\mathbf{f}) \approx \frac{\alpha_0 e^{-i2\pi\frac{f}{F}\Delta_{ji}}}{\|\mathbf{q}_i - \mathbf{p}_j\|}, \quad (5.17)$$

onde se recorreu à modelagem presente na Eq. (5.1). A equação acima implica ser a fase de $h_{ji}(\mathbf{f})$ (coeficiente da matriz $\mathbf{H}(\mathbf{f})$) dependente de \mathbf{f} e de Δ_{ji} . Como a razão acima necessita de informação acerca da fonte a estimar, para torná-la útil cabe efetuar desenvolvimentos adicionais, tendo sido propostas para este fim diversas estratégias na literatura. Já que apelam à diversidade espacial das misturas, todas estas estratégias exigem informações de pelos menos dois sensores. Efetuar a razão entre a contribuição da i -ésima fonte para dois sensores distintos resulta em:

$$\frac{\mathcal{X}_{ji}(\mathbf{f}, k)}{\mathcal{X}_{j'i}(\mathbf{f}, k)} = \frac{\mathcal{X}_{ji}(\mathbf{f}, k)}{\mathcal{S}_i(\mathbf{f}, k)} \frac{\mathcal{S}_i(\mathbf{f}, k)}{\mathcal{X}_{j'i}(\mathbf{f}, k)} \approx \frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})} \approx \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{-i2\pi\frac{f}{F}(\Delta_{ji} - \Delta_{j'i})}, \quad (5.18)$$

sob a restrição $j' \neq j$. Este resultado ainda apresenta a dificuldade de exigir conhecimento dos componentes das misturas $\mathcal{X}_{ji}(\mathbf{f}, k)$ e $\mathcal{X}_{j'i}(\mathbf{f}, k)$, conquanto tenhamos acesso tão somente às misturas $\mathcal{X}_j(\mathbf{f}, k)$ e $\mathcal{X}_{j'}(\mathbf{f}, k)$. Esta dificuldade pode ser contornada por meio de estimativas (possivelmente escaladas) de $h_{ji}(\mathbf{f})$ e $h_{j'i}(\mathbf{f})$, como veremos. Dois lugares geométricos podem ser extraídos da razão acima: uma esfera (associada ao termo $\frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|}$) e uma hiperbolóide (obtida da fase da exponencial); a interseção entre elas permite estimar a localização da i -ésima fonte [124]. Os detalhes serão omitidos porque apenas o modelo de campo distante será contemplado nesta tese.

Em [125] há a demonstração de que a popular regra que prescreve um espaçamento máximo entre sensores de metade do comprimento de onda [110] é, em termos rigorosos, inválida no contexto do modelo de campo próximo, pois a transformada de Fourier (no domínio espacial) dos sinais registrados nos sensores não é limitada em banda. Embora em ângulos próximos de 90° este *aliasing* espacial possa ser considerado pouco significativo, importa não negligenciar esta

⁴Os termos são $\sum_{k'=k-\frac{F}{2}-\Delta_{ji}}^{k-\frac{F}{2}-1} s_i(k')\mathbb{W}_a(k' + \Delta_{ji} - k\mathbf{S})e^{-i2\pi\frac{f}{F}k'}e^{-i2\pi\frac{f}{F}\Delta_{ji}}$ (no numerador) e $\sum_{k'=k+\frac{F}{2}-\Delta_{ji}}^{k+\frac{F}{2}-1} s_i(k')\mathbb{W}_a(k' - k\mathbf{S})e^{-i2\pi\frac{f}{F}k'}$ (no denominador).

⁵Esta hipótese é razoável se a janela for suave e se Δ_{ji} for pequeno. Para um distanciamento entre sensores de 10 cm, velocidade do som igual a 340 m/s e numa taxa de amostragem de 8 kHz, o termo Δ_{ji} é próximo de 2, muito menor do que os valores de F tipicamente empregados para separação de fontes.

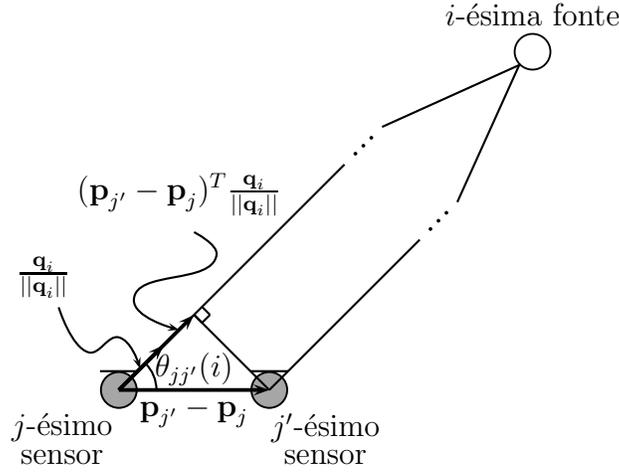


Figura 5.5: Modelo de campo distante.

limitação intrínseca do modelo de campo próximo, ainda mais porque nestes ângulos as direções estimadas tendem a apresentar erros de elevada magnitude [126]. Cabe ressaltar que efetuar uma aproximação dos sensores para reduzir o *aliasing* espacial resulta em perda de desempenho em baixas frequências [110].

O concurso da hipótese adicional que considera desprezível a distância entre os sensores diante da distância das fontes com relação a eles promove algumas simplificações, das quais advém o modelo de campo distante, ilustrado na Fig. 5.5. No modelo de campo distante, os vetores $\mathbf{q}_i - \mathbf{p}_j$ e $\mathbf{q}_i - \mathbf{p}_{j'}$ são considerados paralelos, de modo que a diferença entre os instantes de chegada (TDOA, acrônimo de *time difference of arrival*) $\zeta_{jj'}(i)$ da i -ésima fonte aos sensores j e j' possa ser simplificada para:

$$\zeta_{jj'}(i) = \frac{\Delta_{ji} - \Delta_{j'i}}{f_s} = -\frac{\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|}{cf_s} \approx -\frac{(\mathbf{p}_{j'} - \mathbf{p}_j)^T \mathbf{q}_i}{cf_s \|\mathbf{q}_i\|}, \quad (5.19)$$

onde $\frac{\mathbf{q}_i}{\|\mathbf{q}_i\|}$ denota a direção da i -ésima fonte com relação ao j -ésimo sensor (convencionalmente identificado com a origem do sistema cartesiano) e c é a velocidade de propagação do som. Recorrendo ao fato de que, no modelo de campo distante, $\|\mathbf{q}_i - \mathbf{p}_{j'}\| \approx \|\mathbf{q}_i - \mathbf{p}_j\|$, a aplicação da Eq. (5.19) na Eq. (5.18) resulta em:

$$\frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})} \approx e^{i \frac{2\pi \mathbf{f} (\mathbf{p}_{j'} - \mathbf{p}_j)^T \mathbf{q}_i}{c^F \|\mathbf{q}_i\|}} = e^{i \frac{2\pi f d_{jj'} \cos(\theta_{jj'}(i))}{c^F}}, \quad (5.20)$$

onde $d_{jj'}$ é a distância entre os sensores de índices j e j' e $\theta_{jj'}(i)$ é o ângulo mostrado na Fig. 5.5, associado à direção de chegada (DOA, do inglês *difference of arrival*) da i -ésima fonte. Supondo que as fontes estejam posicionadas em dire-

ções distintas e, por simplicidade, que existem apenas dois sensores, podemos elidir os índices jj' do ângulo e estimar o DOA da i -ésima fonte por meio de:

$$\theta(i) = \arccos \left(\frac{\arg \left(\frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})} \right)}{2\pi \frac{f}{F} c^{-1} d_{jj'}} \right). \quad (5.21)$$

Para que a equação acima se torne útil, importa estimar $\frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})}$. Aplicando técnicas de ICA, a inversão das matrizes de separação $\mathbf{W}(\mathbf{f})$ pode ser empregada para este fim, com a ambiguidade de escalamento sendo automaticamente contornada pela estrutura algébrica da razão $\frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})}$ [121]. Assim, o concurso da informação geométrica da direção de chegada permite que se permutem as linhas das matrizes de separação $\mathbf{W}(\mathbf{f})$ estimadas de modo a se alinharem as direções de chegada, fornecendo portanto um critério de preservação da coerência das estimativas. Em geral, o critério de direção de chegada não é confiável nas baixas frequências (pois a diferença de fase nos sensores é muito pequena) e nas altas frequências (devido à possibilidade de ocorrência de *aliasing* espacial), sendo na prática frequentemente conjugado com outras técnicas, de modo a garantir maior robustez [121]. Outros critérios derivados da direção de chegada podem ser empregados, como padrões de diretividade [127], sendo somente citados doravante os estritamente relacionados ao escopo deste trabalho.

A informação de atraso relativo nos sensores, relacionada com a diferença de fase entre as misturas, pode ser explorada mediante a construção de um mascaramento binário (vide Cap. 1). Embora técnicas fundamentadas em mascaramento binário consistam numa poderosa ferramenta [8] que apresenta como vantagens o baixo custo computacional e a aplicabilidade em contextos subdeterminados, suas estimativas apresentam muitos coeficientes nulos no domínio da transformada, cujas descontinuidades ensejam severas distorções nas estimativas, entre as quais destaca-se um significativo ruído musical [9]. Para reduzir estas distorções, [9] propõe, num contexto de 2 misturas e 3 fontes, o emprego do mascaramento binário para extração de uma fonte das misturas, tornando então a mistura determinada, a qual pode ser separada por algum método derivado da ICA. Para cada ponto na grade tempo \times frequência, a DOA é estimada por meio de [9]:

$$\bar{\theta}(\mathbf{f}, k) = \arccos \left(\frac{\arg \left(\frac{\mathcal{X}_1(\mathbf{f}, k)}{\mathcal{X}_2(\mathbf{f}, k)} \right)}{2\pi \frac{f}{F} c^{-1} d} \right), \quad (5.22)$$

onde d é a distância entre os dois sensores. Observe a semelhança entre esta equação e a Eq. (5.21); o termo $\arg \left(\frac{h_{ji}(\mathbf{f})}{h_{j'i}(\mathbf{f})} \right)$ não pode ser empregado aqui porque, sendo o número de sensores inferior ao de fontes, a matriz de mistura

é degenerada.

Efetuada um histogramas das DOAs em cada frequência, verifica-se a formação de 3 picos, cada qual associado a uma das fontes. Se $\bar{\theta}_i(\mathbf{f})$ for o i -ésimo pico na raia \mathbf{f} , qualquer ponto em (\mathbf{f}, k) será associado à i -ésima fonte, desde que a distância entre $\bar{\theta}(\mathbf{f}, k)$ e $\bar{\theta}_i(\mathbf{f})$ seja menor do que um parâmetro máximo permitido⁶ [9].

Nesta tese há a proposta de uma estratégia alternativa, para contextos determinados, de integração entre as técnicas baseadas em mascaramento binário e as derivadas da ICA, aliando a qualidade da separação destas com uma maior rapidez.

(C) Dependências entre Raias

O mecanismo de produção da fala humana naturalmente gera uma estrutura temporal correlata na magnitude do espectro (ou envoltória), mesmo em raias bem distantes, conforme ilustra a Fig. 5.6. Por esta razão, [120] propõe o emprego da covariância entre os envelopes de raias distintas, definida por:

$$\text{cov}(\vartheta_i, \vartheta_j) = \text{E}[\vartheta_i \vartheta_j] - \text{E}[\vartheta_i] \text{E}[\vartheta_j], \quad (5.23)$$

onde $\vartheta_i(\mathbf{f}, k) = |\mathcal{Y}_i(\mathbf{f}, k)|$, com a média se efetuando ao longo dos quadros. Sensível à potência dos sinais envolvidos, a covariância pode implicar escolhas erradas de permutações [20]; por essa razão, o coeficiente de correlação passou a ser empregado em artigos posteriores, como em [128]:

$$\text{corr}(\vartheta_i, \vartheta_j) = \frac{\text{E}[\vartheta_i \vartheta_j] - \text{E}[\vartheta_i] \text{E}[\vartheta_j]}{\sqrt{(\text{E}[\vartheta_i^2] - (\text{E}[\vartheta_i])^2) (\text{E}[\vartheta_j^2] - (\text{E}[\vartheta_j])^2)}}. \quad (5.24)$$

Ainda que justificado, o emprego dos envelopes $|\mathcal{Y}_i(\mathbf{f}, k)|$ para o cálculo do coeficiente de correlação apresenta problemas de robustez. Apelando à esparsidade (no domínio da frequência) entre sinais de voz, [129] sugere o uso da razão de potência (RP) entre o i -ésimo sinal separado e a potência total dos sinais separados, definida por:

$$\text{RP}_i(\mathbf{f}, k) = \frac{\|\mathbf{a}_i(\mathbf{f}) \mathcal{Y}_i(\mathbf{f}, k)\|^2}{\sum_{k=1}^N \|\mathbf{a}_k(\mathbf{f}) \mathcal{Y}_k(\mathbf{f}, k)\|^2}, \quad (5.25)$$

onde $\mathbf{a}_i(\mathbf{f})$ é a i -ésima coluna de $\mathbf{W}^{-1}(\mathbf{f})$. A razão de potência $\text{RP}_i(\mathbf{f}, k)$ é uma medida situada no intervalo $[0, 1]$ associada ao grau de dominância da i -ésima

⁶Reduzir este parâmetro implica melhorar o desempenho da separação em termos de SIR, porém degradando-a em termos de ruído musical.

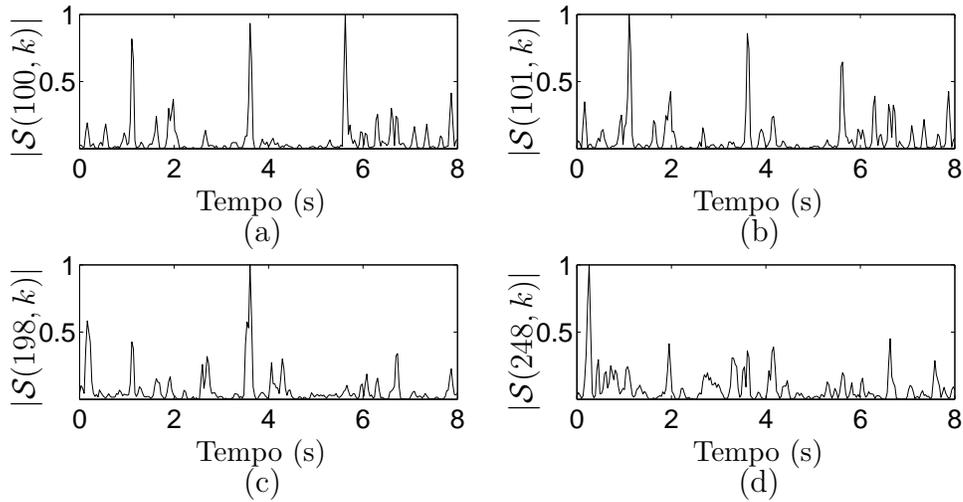


Figura 5.6: Evolução do envelope de um sinal de voz de 8 segundos de duração, com $F = 1024$, $S = 256$ e janela de Hanning. (a) frequência de 773,43 Hz; (b) frequência (adjacente) de 781,25 Hz; (c) frequência harmônica de 1554,68 Hz e (d) frequência não harmônica de 1929,68 Hz.

fonte no ponto (\mathbf{f}, k) , pois seu numerador coleta a soma das energias desta fonte ao longo dos sensores. [129] argumenta que o coeficiente de correlação pode se tornar uma medida mais precisa caso, em detrimento do envelope, empreguemos a razão de potência:

$$\vartheta_i^{\mathbf{f}}(k) = \text{RP}_i(\mathbf{f}, k), \quad (5.26)$$

Escolhida a métrica de avaliação de coerência (por exemplo, o coeficiente de correlação da RP), cabe por fim definir uma técnica para o alinhamento de permutação. Por ter se mostrado bastante robusta em nossas simulações, apresentamos a proposta de [129], a qual será empregada nos experimentos deste trabalho (exceto quando uma função *score* multidimensional tenha sido utilizada, tornando desnecessária a etapa de resolução de permutação).

A escolha das permutações se dá em duas etapas, a global e a local, sendo esta um refinamento do resultado gerado pela primeira.

Seja $\Pi_{\mathbf{f}}$ a permutação a efetuar na \mathbf{f} -ésima raia; sua inicialização é $\Pi_{\mathbf{f}} = [1 \ 2 \ \dots \ N]$. Para a otimização global, deseja-se minimizar a função custo:

$$\mathcal{F}(\mathbf{c}_1, \dots, \mathbf{c}_N, \Pi_1, \dots, \Pi_F) = - \sum_{\mathbf{f}=1}^F \sum_{n=1}^N \text{corr}(\vartheta_i^{\mathbf{f}}(k), \mathbf{c}_k) \Big|_{i=\Pi_{\mathbf{f}}(k)}, \quad (5.27)$$

onde \mathbf{c}_j é um centroide associado à j -ésima estimativa. A etapa de geração dos centroides é efetuada com o valor médio da medida da Eq. (5.26) com o atual

conjunto de permutações $\Pi_{\mathbf{f}}$:

$$\mathbf{c}_j(k) = \frac{1}{F} \sum_{\mathbf{f}=1}^F \vartheta_i^{\mathbf{f}}(k) \Big|_{i=\Pi_{\mathbf{f}}(j)}. \quad (5.28)$$

Já o critério de maximização dos coeficientes de correlação engendra a etapa de otimização das permutações $\Pi_{\mathbf{f}}$ da seguinte forma:

$$\Pi_{\mathbf{f}} \leftarrow \arg \max_{\Pi} \sum_{n=1}^N \text{corr}(\vartheta_i^{\mathbf{f}}, \mathbf{c}_j) \Big|_{i=\Pi_{\mathbf{f}}(j)}. \quad (5.29)$$

As operações de geração dos centroides e de otimização das permutações constituem um processo iterativo cuja parada se dá quando as permutações se mantêm inalteradas ao longo de todas as raias.

O refinamento local das permutações maximiza a soma dos coeficientes de correlação ao longo de um conjunto de frequências selecionadas $\mathbf{G}(\mathbf{f})$ para uma determinada frequência \mathbf{f} :

$$\Pi_{\mathbf{f}} \leftarrow \arg \max_{\Pi} \sum_{\mathbf{g} \in \mathbf{G}(\mathbf{f})} \sum_{n=1}^N \text{corr}(\vartheta_i^{\mathbf{f}}, \vartheta_{i'}^{\mathbf{g}}) \Big|_{i=\Pi_{\mathbf{f}}(n), i'=\Pi_{\mathbf{g}}(n)}. \quad (5.30)$$

O fato de que é particularmente alta a correlação entre raias adjacentes e harmônicas orienta a escolha das raias que constam no conjunto $\mathbf{G}(\mathbf{f})$. Por isso, [129] propõe o emprego de:

$$\mathbf{G}(\mathbf{f}) = \mathbf{G}_0(\mathbf{f}) \cup \mathbf{G}_1(\mathbf{f}), \quad (5.31)$$

com o conjunto $\mathbf{G}_0(\mathbf{f})$ correspondendo às raias adjacentes e $\mathbf{G}_1(\mathbf{f})$ às harmônicas, da forma a seguir:

$$\mathbf{G}_0(\mathbf{f}) = \{\mathbf{f} - 3, \mathbf{f} - 2, \mathbf{f} - 1, \mathbf{f} + 1, \mathbf{f} + 2, \mathbf{f} + 3\}, \quad (5.32)$$

$$\mathbf{G}_1(\mathbf{f}) = \{\text{round}(\mathbf{f}/2 - 1), \text{round}(\mathbf{f}/2), \text{round}(\mathbf{f}/2 + 1), 2\mathbf{f} - 1, 2\mathbf{f}, 2\mathbf{f} + 1\}, \quad (5.33)$$

onde $\text{round}(x)$ aproxima x pelo número inteiro que lhe é mais próximo.

Para cada frequência selecionada, efetua-se a otimização local, a qual é repetida até que nenhuma alteração se verifique nas permutações. Outras estratégias para a resolução do problema de permutação são abordadas em [20], [130] e [131].

5.2.4 Resolução de Escalamento

A recuperação exata da escala dinâmica das raias equivale a resolver o problema de derreverbração cega [110], uma tarefa desafiadora para sinais coloridos como os de voz. Na prática há um relaxamento deste objetivo. A técnica mais popular para a resolução do problema do escalamento foi proposta em [132] e generalizada em [133]. Denominada de Princípio da Distorção Mínima (MDP, do inglês *Minimal Distortion Principle*), ela reduz ao máximo a diferença entre as estimativas e as misturas, de modo que, em contextos de separação de fontes no domínio da frequência, a função de transferência global $\mathbf{W}(\mathbf{f})\mathbf{H}(\mathbf{f})$ seja tão próxima de $\mathbf{H}(\mathbf{f})$ quanto possível. Este princípio obtém um procedimento simples de contornar o problema do escalamento:

$$\mathbf{W}(\mathbf{f}) \leftarrow \text{diag}^0(\mathbf{W}^{-1}(\mathbf{f})) \mathbf{W}(\mathbf{f}), \quad (5.34)$$

onde $\text{diag}^0(\mathbf{X})$ zera todos os elementos que não estão situados na diagonal principal da matriz \mathbf{X} . O MDP deve ser efetuado ao fim do processo de separação via ICA e a consequente resolução de permutação. O princípio da distorção mínima garante que as distorções presentes nas estimativas sejam proporcionais às provocadas pelo sistema de mistura, o que em geral engendra um resultado melhor do que não impor restrições ao escalamento ao longo das raias, tornando o mesmo completamente arbitrário. Alheias ao escopo deste trabalho, diversas estratégias alternativas para a resolução do problema de permutação foram propostas, dentre as quais destacamos: redução do comprimento dos filtros de separação no domínio do tempo [134], normalização dos filtros de separação no domínio da frequência [135], restrição dos elementos diagonais das matrizes de separação para a unidade [117] e emprego dos filtros de separação para a filtragem inversa em cada raia [136].

5.3 Proposta de Procedimento de Inicialização Empregando Estimativas de Direção de Chegada

Embora a maior parte das estruturas de separação cega de fontes no domínio da frequência sejam derivadas do algoritmo clássico de análise de componentes independentes, alguns trabalhos têm adotado soluções diferentes, como o emprego de estimativas de direções de chegada ou atrasos de propagação [137], [9], [138], [139], obtendo dessa forma soluções rápidas para as estimativas das fontes. A desvantagem destas técnicas reside na sua qualidade perceptual inferior.

A abordagem de mascaramento binário apresentada em [139] introduz ruído musical e distorção não-linear, sendo portanto inaceitável como solução final. Entre-

tanto, a partir da separação cega de fonte via mascaramento binário, uma solução aproximada de Wiener pode ser obtida para inicializar os coeficientes do filtro de um método linear de BSS no domínio da frequência.

Nesta seção, combinamos e estendemos os resultados de [139] e o método EDAOR para alcançar a agilidade angariada pelo primeiro e a alta qualidade perceptual obtida pelo último⁷. Para garantir estas qualidades num mesmo algoritmo, adaptamos o método de mascaramento no espaço tempo \times frequência proposto em [139] para obter uma boa solução inicial para os coeficientes do EDAOR, e então substituir o procedimento usual de inicialização via pré-branqueamento pela supracitada solução de Wiener.

A ideia consiste em empregar uma versão modificada do método de mascaramento no espaço tempo \times frequência proposto em [139] como um passo de inicialização para um processo de separação mais robusto em termos globais. Assumindo que os sinais de voz que desejamos separar a partir das misturas são esparsos no domínio da frequência [8], o que significa que na maior parte do tempo cada componente de frequência das misturas contém informação de apenas uma das fontes, a DOA de cada fonte no arranjo de microfones pode ser calculado por meio da STFT. Esta informação de DOA será então usada para grosseiramente separar as fontes.

A informação desejada acerca das DOAs das fontes para cada frequência num cenário⁸ 2×2 pode ser obtido pela Eq. (5.22).

Os ângulos correspondentes aos máximos do histograma de $\bar{\theta}(\mathbf{f}, k)$ consistem em valiosa informação para a obtenção de uma separação de fontes preliminar, dado que podemos comparar a DOA de cada raia em um quadro específico com os ângulos dos picos do histograma e determinar se ela está mais próxima da posição de um pico ou de outro. Este conhecimento é então empregado para selecionar a fonte à qual o ponto no espaço tempo \times frequência deve ser assinalado.

Nos casos em que os picos não são proeminentes (especialmente em baixas e altas frequências), a análise via histograma não é simples. Um modo simples de contornar este problema consiste em efetuar agrupamentos de $\bar{\theta}(\mathbf{f}, k)$, de modo a identificar, para cada quadro, um dos centroides da DOA do sinal, em cada raia.

Definidos os centroides, podemos recorrer ao processo de separação de fontes por meio da seleção da fonte associada a cada ponto do espaço tempo \times frequência. Para isto, obtemos máscaras binárias que são aplicadas à STFT de uma das misturas

⁷Esta abordagem é produto de um trabalho em conjunto com Felipe Sander Pereira Clark. A apresentação desta proposta é fortemente baseada em [140].

⁸O método proposto pode empregar o mesmo número de microfones como fontes (caso determinado) ou mais microfones do que fontes (caso superdeterminado). Limitaremos, sem perda de generalidade, nossa análise para o caso 2×2 .

consoante a seguinte regra:

$$\mathcal{M}_1(\mathbf{f}, k) = \begin{cases} 1, & \text{se } |\bar{\theta}(\mathbf{f}, k) - \bar{\theta}_1(\mathbf{f})| < |\bar{\theta}(\mathbf{f}, k) - \bar{\theta}_2(\mathbf{f})|, \\ 0, & \text{caso contrário} \end{cases}, \quad (5.35)$$

$$\mathcal{M}_2(\mathbf{f}, k) = \overline{\mathcal{M}_1(\mathbf{f}, k)},$$

onde $\bar{\theta}_i(\mathbf{f})$ é o centroide associado ao i -ésimo agrupamento na \mathbf{f} -ésima raia e $\overline{\mathcal{M}_1(\mathbf{f}, k)}$ é inversão binária de $\mathcal{M}_1(\mathbf{f}, k)$. Uma primeira estimativa para o \mathbf{f} -ésima componente em frequência da i -ésima fonte em cada quadro pode então ser obtida por meio da Eq. (1.2).

A técnica descrita acima pode ser interpretada como um método de melhoramento de qualidade de um sinal de voz [141], e sua natureza não-linear gera ruídos musicais e distorções. Para minimizar estes problemas, propomos a linearização da abordagem de mascaramento no domínio da frequência por meio de uma solução ótima de Wiener [54], a qual poderá ser facilmente empregada no contexto atual, pois as misturas e as estimativas das fontes nos estão disponíveis.

Para um cenário 2×2 , uma solução inicial para a matriz de separação na \mathbf{f} -ésima raia do método linear de BSS pode ser obtida como se segue:

$$\begin{aligned} \underline{\mathbf{P}}_i(\mathbf{f}) &= \text{E} \left[\begin{bmatrix} \mathcal{X}_1(\mathbf{f}, k) \\ \mathcal{X}_2(\mathbf{f}, k) \end{bmatrix} \mathcal{Y}_i^*(\mathbf{f}, k) \right], \\ \mathbf{R}(\mathbf{f}) &= \text{E} \left[\begin{bmatrix} \mathcal{X}_1(\mathbf{f}, k) \\ \mathcal{X}_2(\mathbf{f}, k) \end{bmatrix} \begin{bmatrix} \mathcal{X}_1^*(\mathbf{f}, k) & \mathcal{X}_2^*(\mathbf{f}, k) \end{bmatrix} \right], \\ \underline{\mathbf{w}}_i(\mathbf{f}) &= \mathbf{R}^{-1}(\mathbf{f}) \underline{\mathbf{P}}_i(\mathbf{f}), \\ \underline{\mathbf{W}}(\mathbf{f}) &= \begin{bmatrix} \underline{\mathbf{w}}_1(\mathbf{f}) & \underline{\mathbf{w}}_2(\mathbf{f}) \end{bmatrix}, \end{aligned} \quad (5.36)$$

onde $\mathbf{W}(\mathbf{f}) = \underline{\mathbf{W}}^H(\mathbf{f})$ é a matriz de separação 2×2 que combina $\underline{\mathbf{w}}_1(\mathbf{f})$ e $\underline{\mathbf{w}}_2(\mathbf{f})$ numa única matriz, de modo que ambas as fontes podem ser estimadas por meio de:

$$\begin{bmatrix} \mathcal{Y}_1(\mathbf{f}, k) \\ \mathcal{Y}_2(\mathbf{f}, k) \end{bmatrix} = \mathbf{W}(\mathbf{f}) \begin{bmatrix} \mathcal{X}_1(\mathbf{f}, k) \\ \mathcal{X}_2(\mathbf{f}, k) \end{bmatrix}. \quad (5.37)$$

Este procedimento pode ser entendido como uma nova técnica para a estimativa das matrizes $\underline{\mathbf{P}}_i$ de correlação cruzada por meio de mascaramento binário de modo a computar os coeficientes iniciais de um conjunto de matrizes $\mathbf{W}(\mathbf{f})$ que podem ser empregadas como uma inicialização para os coeficientes dos filtros de separação do algoritmo do EDAOR. Espera-se uma convergência mais rápida deste algoritmo com o emprego desta nova inicialização, pois a maior parte das matrizes de separação no domínio da frequência já estão alinhadas, não sendo necessário ao EDAOR corrigir

as permutações ao longo de seu processo iterativo.

5.3.1 Resultados

Nesta seção apresentamos um conjunto de testes realizados num ambiente de dimensões $4,7\text{ m} \times 6,6\text{ m} \times 4,2\text{ m}$, sob a configuração circular apresentada na Fig. 5.7. Todos os experimentos foram conduzidos com duas fontes (de uma locutora feminina e de um locutor masculino, ambos lendo textos em português), cada qual localizado em uma das posições indicadas na Fig. 5.7.

Os sinais foram gravados por dois microfones condensadores Behringer ECM8000, conectados a uma placa Behringer Eurorack MX3242X, cujos canais foram conectados a uma placa de áudio soundcard M-AUDIO Firewire 1814. As gravações foram efetuadas com uma resolução de 16 bits em 44,1 kHz. Os dados foram reamostrados para 8 kHz antes de serem processados pelo sistema de BSS, com o número de raias $F = 2048$ (determinado empiricamente, de modo a angariar os melhores resultados) e por meio de uma janela de Hanning com deslocamento $S = F/4$. O clássico algoritmo Fuzzy C-means [142] foi usado para a formação dos agrupamentos necessários para o processo de mascaramento binário descrito na Seção 5.3.

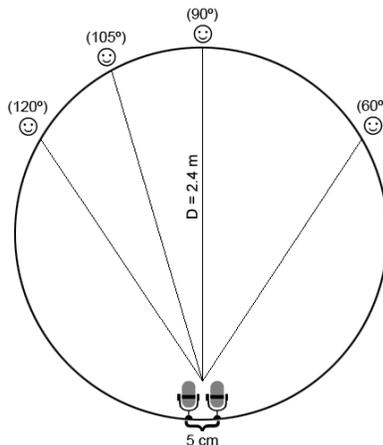


Figura 5.7: Configuração circular empregada para a gravação de misturas de sinais de voz com um arranjo de microfones.

As medidas de desempenho empregadas foram a SIR, o número de iterações necessárias para a convergência do algoritmo e as notas da análise subjetiva para a qualidade global do sistema de separação (obtidas a partir de uma média de 20 avaliações), mostrados nas Tabelas 5.1, 5.2 e 5.3, respectivamente.

A Tabela 5.1 apresenta a SIR obtida por cada estimativa com o algoritmo EDAOR inicializado pelo processo de branqueamento (B-EDAOR), pelo método proposto (M-EDAOR) e com o mascaramento binário sem qualquer otimização posterior (Mascaramento). Três cenários foram considerados por meio de diferentes posicionamentos das fontes, indicados pelos ângulos \angle_1 e \angle_2 medidos a partir do

Tabela 5.1: SIR no ambiente ruidoso e em silêncio.

\angle_1, \angle_2	90°, 120°				60°, 105°				60°, 120°			
	Silencioso		Ruidoso		Silencioso		Ruidoso		Silencioso		Ruidoso	
Mascaramento	9,7	5,8	9,9	7,5	15,3	11,2	10,5	10,8	15,5	11,6	11,4	11,6
B-EDAOR	14,9	13,8	13,7	6,9	18,9	14,8	17,0	5,7	19,1	14,1	14,4	4,2
M-EDAOR	14,8	11,8	8,2	13,1	18,3	14,5	11,5	17,0	18,1	13,9	9,7	16,2

centro do arranjo de microfones, como mostrado na Fig. 5.7. Todos os testes foram conduzidos duas vezes: com o ruído ambiente originado por pessoas trabalhando no ambiente (Ruidoso) e à noite, quando o ambiente estava vazio (Silencioso). O número de iterações N_{it} necessárias para a convergência do EDAOR é apresentada na Tabela 5.2. O tempo proporcional de processamento $T(\%)$, relativo ao B-EDAOR, para a execução completa - no mesmo computador - das tarefas (o que inclui o processo de branqueamento para o B-EDAOR e o mascaramento seguido pela solução de Wiener para o M-EDAOR) também é apresentado.

Tabela 5.2: Número de iterações (N_{it}) para a convergência dos algoritmos e tempo total de execução (T).

\angle_1, \angle_2		90°, 120°		60°, 105°		60°, 120°	
		N_{it}	$T(\%)$	N_{it}	$T(\%)$	N_{it}	$T(\%)$
Silencioso	B-EDAOR	2770	100	630	100	1050	100
	M-EDAOR	150	14,5	150	55,7	230	48,8
Ruidoso	B-EDAOR	1000	100	2770	100	1000	100
	M-EDAOR	120	33,6	130	13,0	90	23,6

Tabela 5.3: Resultados subjetivos no ambiente ruidoso.

\angle_1, \angle_2		90°, 120°		60°, 105°		60°, 120°	
Mascaramento	Interferência	2,2	3,5	2,9	3,7	3,0	4,0
	Distorção	3,6	1,9	3,2	2,1	3,3	2,2
	Ruído	2,9	2,0	2,7	2,5	2,7	2,2
	Global	2,4	1,6	2,7	2,2	2,6	2,5
B-EDAOR	Interferência	4,1	3,9	4,4	4,0	4,4	4,5
	Distorção	3,7	3,7	3,7	3,3	3,9	3,8
	Ruído	2,7	2,6	2,6	2,0	2,6	2,5
	Global	3,2	3,2	3,4	2,9	3,7	3,4
M-EDAOR	Interferência	4,1	3,9	4,2	4,5	4,6	4,4
	Distorção	4,1	3,7	3,9	3,9	4,0	4,1
	Ruído	2,9	2,5	3,0	2,5	3,3	2,6
	Global	3,7	3,2	3,6	3,6	4,0	3,7

A Tabela 5.3 apresenta as notas referentes à avaliação subjetiva (numa escala de 1 a 5) no tocante a quatro importantes aspectos de BSS: interferência⁹, ruído,

⁹Nota 1 significa que a interferência a partir da outra fonte é extremamente perturbadora,

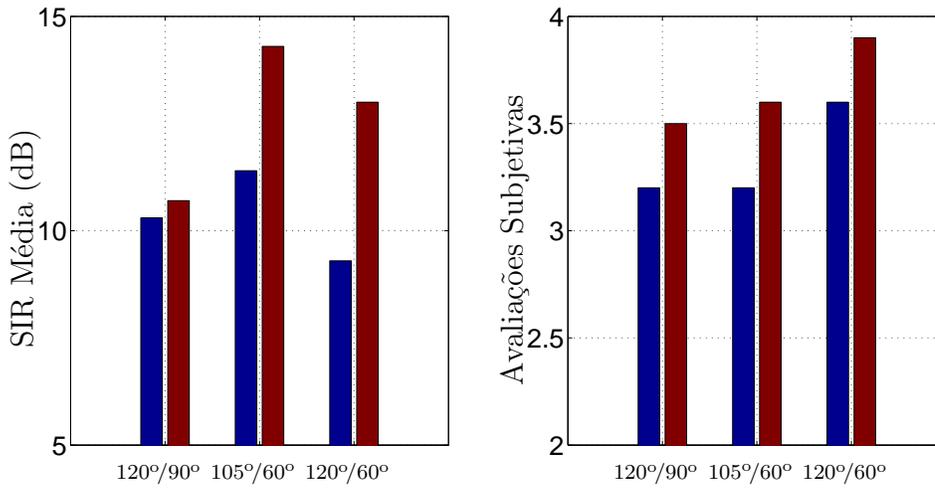


Figura 5.8: Média da SIR (em dB) e das avaliações subjetivas para o ambiente ruidoso. Em azul, B-EDAOR; em vermelho, M-EDAOR.

distorção¹⁰ e qualidade global do sistema¹¹.

Apenas o caso ruidoso, o qual é usual em situações práticas nas quais técnicas de BSS são empregadas, foi submetido a testes subjetivos. A Fig. 5.8 apresenta a SIR média das duas estimativas e as médias das notas subjetivas da qualidade global para os métodos B-EDAOR e M-EDAOR.

A partir das Tabelas 5.1 e 5.2, pode-se observar que no ambiente silencioso o algoritmo M-EDAOR proposto obtém razões fonte-interferência próximas àquelas angariadas pelo algoritmo B-EDAOR, porém com um número de iterações muito menor e com um tempo de processamento total menor. Adicionalmente, as Tabelas 5.1 e 5.3 permitem-nos concluir que em cenários ruidosos as médias da SIR e das avaliações subjetivas são maiores para o M-EDAOR do que para o B-EDAOR, o que indica uma maior robustez daquele.

Estes resultados eram esperados, pois o ruído que compromete as gravações não apresenta um padrão direcional claro e provavelmente não introduziu grandes perturbações na inicialização via direção de chegada. A inicialização via branqueamento, por outro lado, é afetada pelo ruído aditivo. Uma vantagem adicional da abordagem de mascaramento proposta reside na solução prévia do problema de permutação pela associação dos pontos no espaço tempo \times frequência às fontes, por meio dos ângulos de chegada estimados.

Em todos os testes realizados com o algoritmo M-EDAOR, empregamos o fator de aprendizagem $\beta = 0,5$ na Eq. (5.13), sem jamais observar divergência. Por outro

enquanto que 5 implica ser esta inaudível.

¹⁰Nota 1 significa que o algoritmo de BSS adicionou ruídos e distorções intensos, os quais não se encontram presentes nas misturas, enquanto que a nota 5 reflete a ausência de qualquer ruído ou distorção.

¹¹A nota 1 considera o desempenho global do algoritmo ruim e a nota 5 o julga excelente.

lado, o algoritmo B-EDAOR exigiu uma escolha criteriosa de β , geralmente entre 0,01 e 0,02, de modo a evitar divergência. Logo, além do supracitado desempenho superior, o método proposto permite a escolha de valores de β mais altos, com o propósito de incrementar a taxa de convergência significativamente.

5.4 Proposta de Inserção de Coeficientes Adaptativos Interfrequenciais

Em misturas determinadas e sobredeterminadas, se o sistema de mistura é LIT (linear e invariante no tempo), conclui-se que um sistema ótimo de separação também deve apresentar estas propriedades. Embora uma representação fiel de sistemas LIT no domínio da STFT exija filtros cruzados entre as subbandas [143], na maioria das vezes a função de transferência é aproximada como meramente multiplicativa (desprovida de filtros cruzados) no domínio da STFT (modelo MTF, do inglês *Multiplicative Transfer Function*). As técnicas de BSS não constituem exceção à regra. Essa aproximação decerto reduz a potencialidade do sistema de separação.

Em contextos supervisionados, é razoavelmente simples adaptar os coeficientes adaptativos interfrequencias de um sistema LIT [144]. Cabe ressaltar que nem sempre esta modelagem mais complexa é benéfica; o conflito entre o MSE em regime permanente e a taxa de convergência subjaz à escolha do número de coeficientes interfrequenciais [144] e mesmo o comprimento ótimo da janela depende do comprimento do sinal de entrada, bem como da razão sinal-ruído [145].

No entanto, desconhecemos na literatura o emprego de modelos que empregam coeficientes adaptativos cruzados (CACs) em técnicas de separação de fontes. O processo de aprendizado dos coeficientes adaptativos interfrequenciais ótimos na ausência de sinal desejado não é uma tarefa trivial. Uma das novidades deste capítulo reside em propor uma forma de adaptação destes coeficientes.

5.4.1 Estrutura Proposta com Coeficientes Adaptativos Cruzados

A Figura 5.9 ilustra a estrutura proposta, para o caso de 2 fontes e 2 misturas e coeficientes interfrequenciais atuando apenas em subbandas adjacentes. Importa ressaltar o substancial aumento de complexidade do novo sistema de separação, já que o número de coeficientes a estimar é multiplicado por três. Espera-se que este aumento de complexidade (o qual poderá ser atenuado, como veremos posteriormente) seja compensado por uma maior razão fonte-interferência (SIR).

Doravante suporemos, por simplicidade, o caso $N = 2$. Sem os CACs, cada

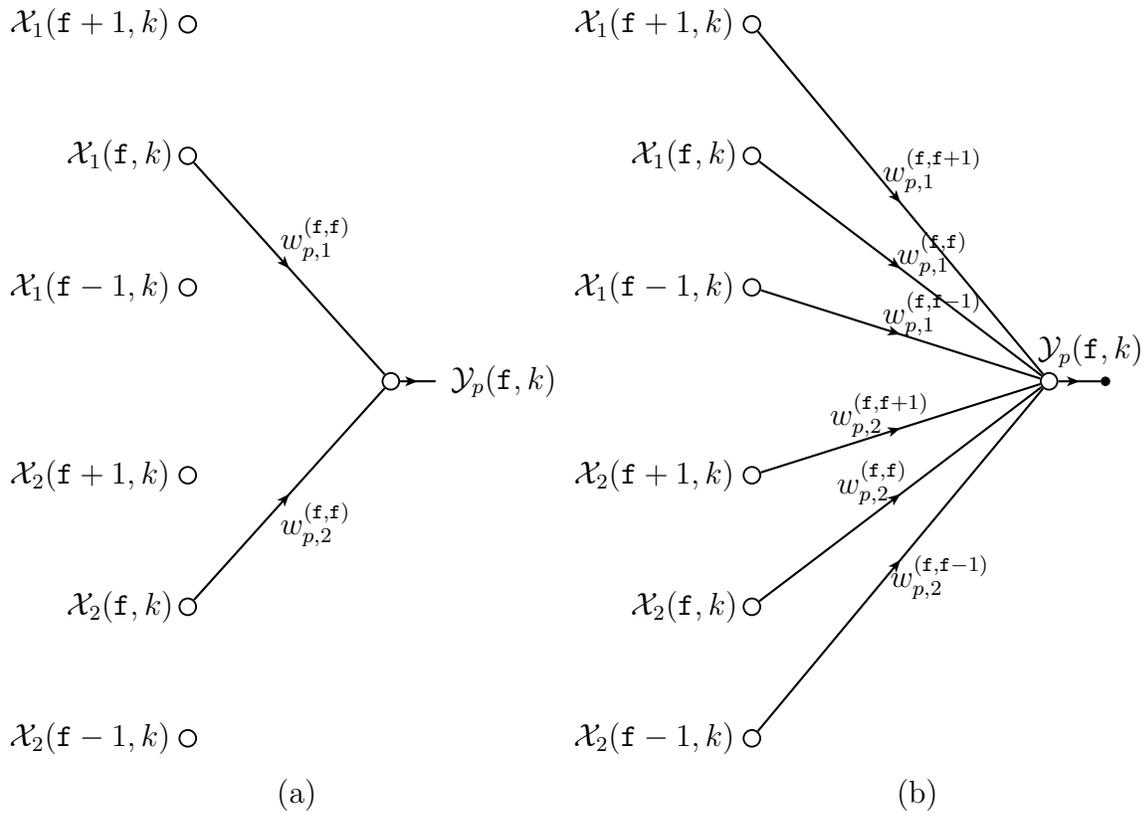


Figura 5.9: Estruturas de sistemas de separação no domínio da STFT. (a) Estrutura padrão; (b) estrutura proposta com coeficientes adaptativos interfrequenciais entre raias adjacentes.

raia exige a determinação de uma matriz $\mathbf{W}(\mathbf{f})$ que apresenta 4 elementos. A inclusão dos CACs requer a determinação de uma matriz $\mathbf{W}_{\text{exp}}(\mathbf{f})$ de dimensões 2×6 , tornando necessária a estimativa de 12 coeficientes.

5.4.2 Restrição do Número de Coeficientes Adaptativos Cruzados

Para evitar um elevado aumento do custo computacional, podemos optar por empregar os coeficientes adaptativos cruzados apenas em algumas raias. Quando os sinais a separar são de voz, é de se esperar um ganho relativo maior da nova estrutura quando os CACs atuam nas baixas frequências, por duas razões: *i*) os sinais apresentam maior energia em baixas frequências, sendo por vezes difícil angariar a informação estatística necessária para a separação nas altas frequências e *ii*) as funções de transferência acústica apresentam predominantemente características passa-baixas, o que requer um processamento mais poderoso nas baixas frequências. Logo, devido às razões acima (além da necessidade de reduzir a carga computacional), a inserção de CACs se restringirá às baixas frequências. O desenvolvimento de um método cego de seleção das raias que exigirão um tratamento diferenciado seria de grande utilidade, e a proposta do artigo [146] pode representar um critério para esta seleção.

5.4.3 Determinação dos Coeficientes Adaptativos da Estrutura Proposta

Definida a estrutura de separação, importa implementar um método efetivo de treinamento dos parâmetros da mesma. O núcleo de nossa proposta consiste em compreender as dimensões da matriz $\mathbf{W}_{\text{exp}}(\mathbf{f})$ como associadas a um processo de mistura sobredeterminado; ou seja, postulamos que o processo de escolha dos coeficientes de $\mathbf{W}_{\text{exp}}(\mathbf{f})$ pode ser eficientemente implementado por meio da resolução de um problema de separação de fontes no qual o número de misturas (6) supera o de fontes a estimar (2). Sendo, por suposição, independentes as fontes a estimar, espera-se que a maximização das independências das duas estimativas torne-as sujeitas a menor grau de interferência mútua. Neste contexto, a inserção dos CACs em tese aumentaria a potencialidade do método de separação.

Se dois sinais são independentes, transformações quaisquer de cada um destes sinais resultarão em sinais (transformados) independentes. A pouca informação *a priori* disponível acerca das fontes aliada ao princípio da maximização da independência (ou algum que lhe seja equivalente) torna o método de BSS incapaz de reduzir as distorções provocadas por transformações dos sinais. Estas distorções são menos

severas em contextos de misturas instantâneas, porém, à medida que o sistema de separação amplia seus graus de liberdade, a uma maior potencialidade de separação costumam corresponder estimativas mais drasticamente distorcidas das fontes. Isso significa que o emprego da nova estrutura implicará uma inserção mais intensa de artefatos e distorções.

Existem diversos métodos de resolução de misturas sobredeterminadas, sendo uma opção postular (a despeito do conhecimento *a priori*) que o número de fontes iguala o de misturas, efetuar a separação e depois condensar, de algum modo, as estimativas “redundantes”, previamente agrupadas por similaridade (vide [78], por exemplo, num contexto de técnicas de separação de misturas convolutivas no domínio do tempo). Optamos por empregar esta técnica, de modo que a primeira etapa da nossa proposta consiste em aplicar a ICA determinada nos 6 sinais associados à f -ésima raia (incluindo as subbandas adjacentes, bem como todas as misturas).

Ao final desta etapa restam-nos seis estimativas, as quais cumpre condensar em duas. A ausência de sinais de referência poderia ser contornada mediante o recurso a técnicas de clusterização. Porém, ainda restaria contornar o problema da permutação ao longo das raias. A realização de agrupamentos independentes em cada raia é incapaz de garantir a coerência dos mesmos ao longo das raias. Logo, nos é necessário adotar um critério confiável que permita-nos associar cada agrupamento a uma das fontes.

Para obtermos os sinais de referência das fontes e a resolução de permutação, optamos por efetuar a separação no domínio da frequência pelo método clássico (modelo MTF + método de resolução de permutação), como uma etapa prévia. A aplicação do método clássico nos dá acesso a uma versão (menos acurada) das estimativas das fontes, a qual nos servirá como “gabarito” para a condensação das seis estimativas (de cada raia onde os CACs serão empregados) que nos cabe agrupar. Em outras palavras, o método clássico orientará a formação de agrupamentos coerentes em todas as raias¹², tornando supervisionada esta operação.

A partir do momento que o método clássico nos fornece estimativas imprecisas das fontes em cada raia, podemos associar diretamente cada uma das seis estimativas (obtidas com a estrutura proposta; logo, cada estimativa está associada a 6 coeficientes adaptativos) a uma das fontes, por meio de algum critério de similaridade. Em nossas simulações, verificamos que este critério prescinde de sofisticação, bastando-nos o módulo do fator de correlação. Assim, por meio da maximização desta medida, associamos cada uma das seis estimativas (mais acuradas) do novo método a uma das duas estimativas prévias do método clássico, formando dois agrupamentos, cada qual relacionado a uma fonte distinta.

¹²Como veremos, outra utilidade do emprego do método clássico de separação será na aplicação do princípio da mínima distorção para atenuar as ambiguidades de escalamento ao longo das raias.

O agrupamento via maximização do módulo do fator de correlação das seis estimativas as rotula como pertencentes à fonte (ou classe) 1 ou à fonte (ou classe) 2, porém cabe condensar cada classe em apenas uma estimativa que, espera-se, contenha uma versão das fontes com menos interferências. Para que esta estimativa condensada se beneficie das informações presentes em todas as estimativas pertencentes à mesma classe e seja produto de uma perda mínima de representação global das mesmas, optamos por empregar a PCA (vide Seção 4.4) para reduzi-las a apenas uma estimativa.

Em nossas simulações, verificamos que não raro algumas das 6 estimativas obtidas pelo método proposto não são significativas, podendo impactar negativamente na qualidade da separação. Para evitar que estimativas pouco significativas sejam associadas a uma das fontes, procede-se a uma eliminação das estimativas de um agrupamento por meio de um limiar de correlação, definido em relação ao máximo módulo do fator de correlação do agrupamento correspondente (tipicamente, o limiar ótimo se localiza entre 5% e 40%). Caso o módulo do fator de correlação esteja abaixo do limiar, a estimativa correspondente será descartada.

A eliminação via limiar de correlação não resolve a necessidade de condensação, já que mais de uma estimativa pode “sobreviver” em cada agrupamento (caso apenas uma reste após a eliminação, a PCA não precisa ser aplicada). A Fig. 5.10 apresenta o diagrama de blocos da parte principal do método proposto. O bloco “LA” corresponde à eliminação por limiar adaptativo.

Por fim, para evitar um escalamento arbitrário das estimativas ao longo das raias, o princípio da distorção mínima foi aplicado nas estimativas do método padrão, de modo a tornar as estimativas destes menos propensas a este problema. Foi efetuado então um escalamento das estimativas do novo método, de modo a igualar sua energia à energia das estimativas correspondentes do método clássico.

5.4.4 Simulações

Como experimento de avaliação da estrutura de separação proposta, testamos duas configurações de mistura, nas quais as fontes são sinais de voz. Em ambas, sinais de voz amostrados em 8 kHz e duração de 10 s foram utilizados. No simulador de funções de transferência acústicas apresentado na Seção 1.3.5, as dimensões do ambiente foram arbitradas em 4 m \times 3 m \times 2,5 m, com dois microfones, distanciados horizontalmente de 5 cm, dispostos simetricamente em torno do centro de massa do ambiente.

Não houve repetição dos sinais (extraídos de <http://www.irisa.fr/metiss/SASSECO7/?show=test>) em diferentes configurações e em cada um delas uma voz era feminina e a outra, masculina. A fonte 2

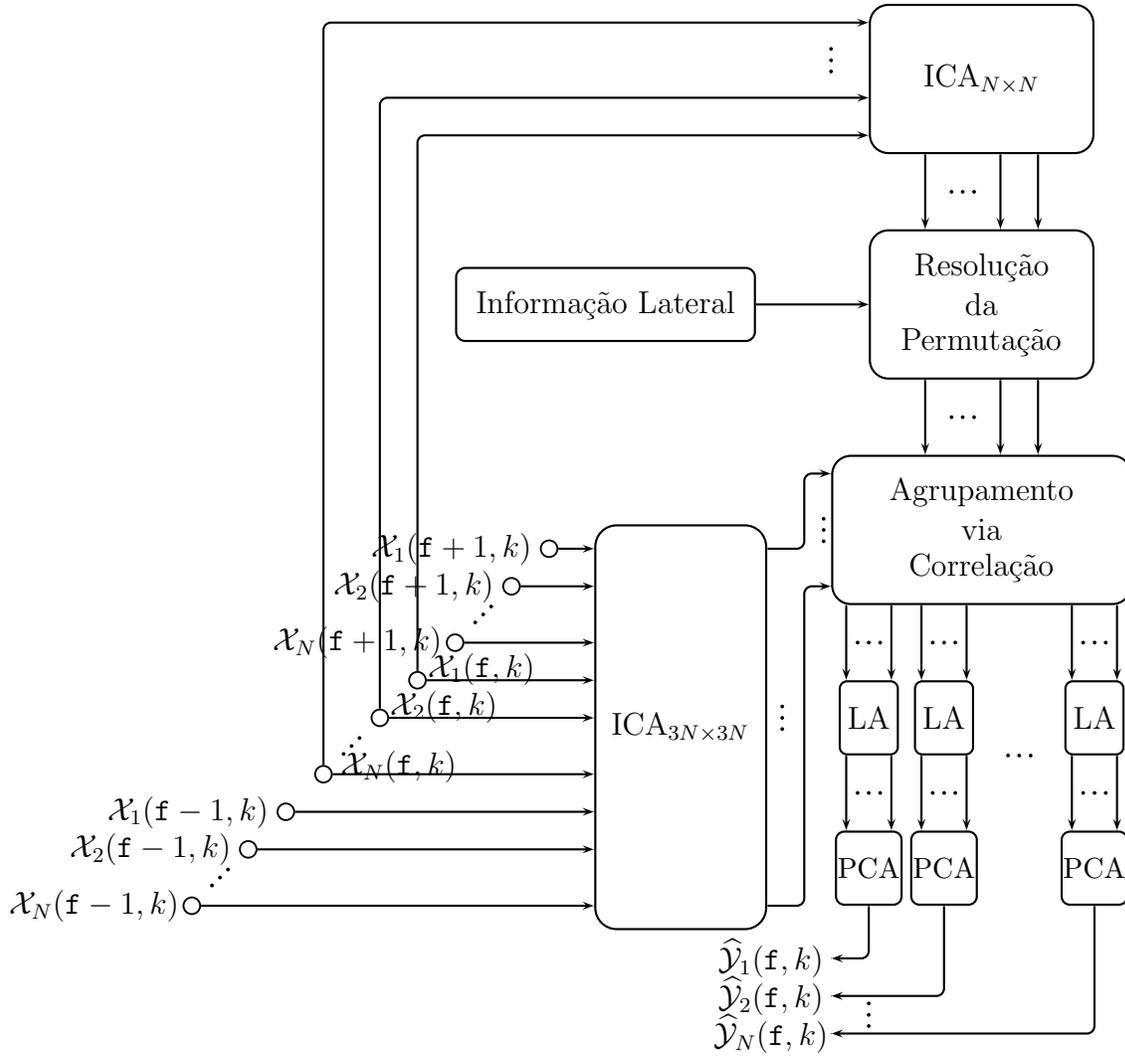


Figura 5.10: Diagrama de blocos do método proposto para a determinação dos coeficientes adaptativos, incluindo os cruzados, os quais são aplicados apenas às subbandas adjacentes.

Tabela 5.4: Especificações adicionais das configurações testadas. \mathcal{N} é o número de raias empregadas para a separação e $\mathcal{N}_{\text{prop}}$ é o número de raias (a partir da segunda) nas quais os CACs foram inseridos. O limiar de correlação é relativo ao maior módulo do coeficiente de correlação de cada agrupamento.

Configuração	\angle_1	T60	\mathcal{N}	$\mathcal{N}_{\text{prop}}$	Limiar de correlação
1	50°	100 ms	512	31	0,1
2	40°	75 ms	512	15	0,35

Tabela 5.5: Arquivos da base de dados referentes às fontes utilizadas em cada caso.

Caso	Fonte 1	Fonte 2
1	female_src_1.wav	female_src_3.wav
2	female_src_2.wav	male_src_3.wav

teve seu ângulo variado de 0° a 90°, enquanto que a fonte 1 não alterou sua posição (ver Tab. 5.4). Ambas as fontes foram posicionadas a uma distância de 1 metro do centro de massa dos microfones. Algumas configurações adicionais dos experimentos são apresentadas na Tab. 5.4. Como método padrão de separação, responsável por gerar uma primeira versão das estimativas das fontes, empregamos como função *score* a \tanh , $\beta = 0,1$, deslocamento de 25% do comprimento do quadro, 200 iterações, inicialização por PCA e resolução de permutação via correlação da razão de potência (RP). A Tab. 5.5 apresenta os arquivos referentes às fontes utilizadas em cada caso.

Observe que, do total de 512 raias em ambos os experimentos, apenas 62 e 30 raias sofreram intervenção dos CACs¹³, o que significa que o aumento do custo computacional do novo método não é excessivo.

A razão fonte-interferência obtida ao longo dos diferentes posicionamentos da fonte 2 (em cada uma das configurações) é ilustrada na Fig. 5.11.

Na maioria dos experimentos, o método proposto obtém ganhos significativos de SIR. Quando o método proposto é vantajoso, obtém-se ganhos médios de SIR de 5,41 dB (configuração 1 / fonte 1), 5,97 dB (configuração 1 / fonte 2), 5,89 dB (configuração 2 / fonte 1) e 4,99 dB (configuração 2 / fonte 2). Quando o método proposto não é vantajoso, a perda da SIR (em média) não é grande: 0,12 dB (configuração 1 / fonte 1), 0,17 dB (configuração 1 / fonte 2) e 0,68 dB (configuração 2 / fonte 2)¹⁴. Algumas estatísticas adicionais encontram-se listadas

¹³O espelhamento das raias, devido à restrição de simetria da STFT, acaba por duplicar o número de raias em que ocorre a intervenção.

¹⁴Para a primeira fonte da segunda configuração, não houve nenhum caso em que o método proposto foi superado pelo método padrão.

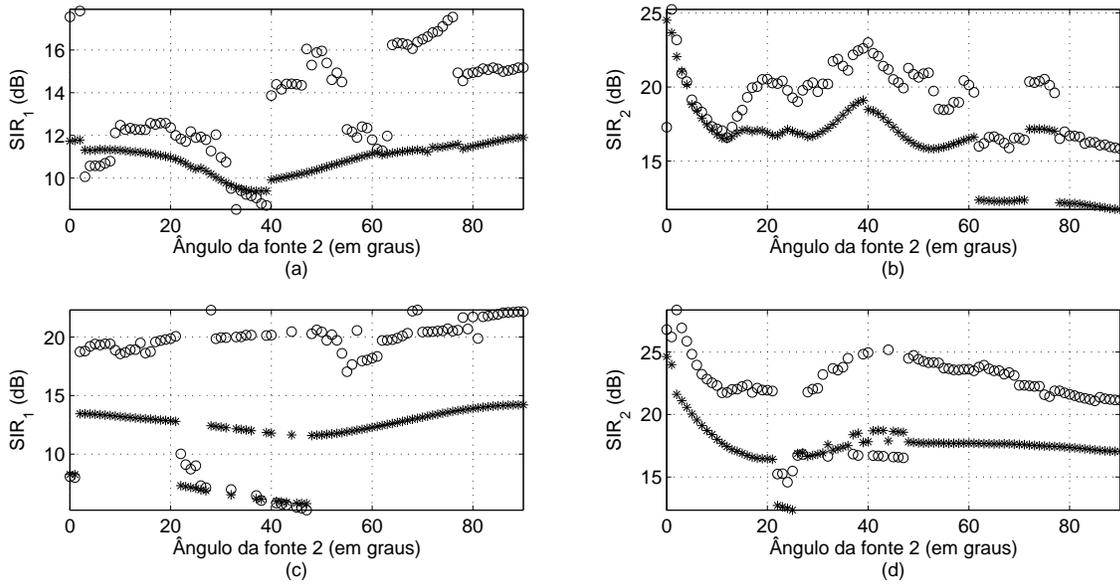


Figura 5.11: Razão fonte-interferência em função do ângulo da fonte 2. Círculos: método proposto; asteriscos: método padrão. (a) e (b): configuração 1; (c) e (d): configuração 2.

Tabela 5.6: Resultados adicionais das configurações testadas. #sup indica o número de vezes (de um total de 91, pois o ângulo da segunda fonte varia de 0° a 90°) nas quais o novo método supera o método padrão.

Configuração	Fonte	SIR médio (padrão)	SIR médio (proposto)	#sup
1	1	10,89	16,12	88
1	2	13,43	19,13	87
2	1	11,74	17,63	91
2	2	17,61	21,94	80

na Tab. 5.6, a qual ilustra que a estrutura que emprega CACs pode contribuir para uma substancial melhora do desempenho das técnicas atualmente empregadas em separação de fontes no domínio da frequência.

Infelizmente as simulações revelam que apenas em baixas frequências há efetiva contribuição do método proposto. Isso dificultou a confecção de testes subjetivos, porque a psicoacústica prevê baixa sensibilidade do ouvido humano nestas frequências.

Outro experimento foi realizado com dois sinais de voz com locutores do sexo feminino, reamostrados para 8 kHz. Empregando o mesmo simulador de funções de transferência da experiência anterior, aumentou-se o tempo de reverberação para 750 ms, muito alto para os padrões atuais de testes de algoritmos de separação cega de fontes e para o qual o desempenho destes ainda não é satisfatório. Inicialmente,

Tabela 5.7: SIR no método padrão e no método com coeficientes interfrequenciais aplicados a todas as raia ($\mathcal{N} = 1024$).

SIR (em dB)	Fonte 1	Fonte 2
Método padrão	8,49	10,79
Com inserção de coeficientes interfrequenciais	9,41	11,81

os coeficientes interfrequência foram aplicados em cada raia separadamente e verificado o valor ótimo do limiar de correlação para cada raia (esta determinação exigiu o conhecimento das fontes, que também foi empregado para resolução do problema de permutação). Posteriormente, o processo de separação foi executado para todas as raia, com o fator de correlação otimizado para a determinação dos coeficientes interfrequências. Os resultados da SIR encontram-se na Tab. 5.7. O fato de o conhecimento das fontes ter sido necessário estabelece um limitante superior de desempenho do método, que resultou em melhoras de aproximadamente 1 dB. Portanto, as simulações realizadas até o momento sugerem que sua contribuição em configurações de alto tempo de reverberação é modesta.

5.5 Proposta de Emprego de Estimativas Cegas do Tempo de Reverberação

O comprimento ótimo da STFT empregada nas técnicas de separação cega de fontes no domínio da frequência é dependente do comprimento M dos filtros responsáveis pela mistura. Na prática, M nos é desconhecido e difícil de estimar. Porém, é estreitamente relacionado ao parâmetro T_{60} , cuja estimativa cega (ou seja, usando apenas o sinal reverberante), em registros de uma única fonte sonora, tem sido alvo de interesse em publicações recentes [147],[148],[149]. Embora a estimativa do T_{60} comumente se destine a avaliar as características acústicas de um ambiente ou para o projeto de um sistema de desreverberação [147], esta seção investiga se o emprego de estimativas de T_{60} , em contextos de misturas de múltiplas fontes, é capaz de prover informação suficiente para que, em algoritmos de BSS, uma escolha cega de F obtenha estimativas de melhor qualidade, caso comparada ao emprego de um valor fixo deste parâmetro. A técnica de estimativa de T_{60} empregada nesta tese é derivada de [148], sendo resumida a seguir.

Primeiramente, uma das misturas é arbitrariamente escolhida e então submetida a uma STFT de comprimento F_T , o qual pode ser diferente do comprimento F da STFT utilizada para a separação. O tempo de reverberação está relacionado com o decaimento da energia ao longo do tempo. Para a raia f_0 e o quadro k_0 , associamos

Tabela 5.8: Intervalo de variação dos parâmetros de cada simulação.

Parâmetro	Limite inferior	Limite superior
T_{60}	25ms	400ms
largura	3m	3,5m
altura	3m	3,5m
comprimento	3m	3,5m
distância entre microfones	4cm	6cm
distância das fontes a cada microfone	0,8m	1,2m
coeficientes de absorção das paredes	0,35	0,65
ângulos das fontes com relação ao centro do arranjo de microfones	-90°	90°

um decaimento, relacionado a $\lambda_h(\mathbf{f}_0, k_0)$, de sorte a utilizar a seguinte aproximação:

$$\ln |\mathcal{X}_i(\mathbf{f}_0, k)|^2 \approx \ln |\mathcal{X}_i(\mathbf{f}_0, k_0)|^2 + \lambda_h(\mathbf{f}_0, k_0)k, \quad (5.38)$$

onde $k \in [k_0, k_0 + N_f - 1]$, N_f é um número arbitrado de quadros consecutivos e $\lambda_h(\mathbf{f}_0, k_0)$ é encontrado por meio do método dos mínimos quadrados. Seja $f_x(\lambda_h)$ a distribuição de λ_h obtida através de todas as raias e quadros. Definindo a variância NS (do inglês *negative-side*) $\sigma_{x^-}^2$ como a variância de uma distribuição simétrica $f_x^-(\lambda_h)$ com a parte negativa idêntica à da distribuição original $f_x(\lambda_h)$ [148]:

$$f_x^-(\lambda_h) = \begin{cases} f_x(\lambda_h), & \text{para } \lambda_h \leq 0 \\ f_x(-\lambda_h), & \text{para } \lambda_h > 0 \end{cases}, \quad (5.39)$$

pode-se associar $\sigma_{x^-}^2$ a T_{60} .

A técnica proposta de escolha de F_T deve ser avaliada em termos estatísticos. Dada a dificuldade de se realizar extensas gravações em diversos ambientes, nesta seção o simulador de funções de transferência apresentado na Seção 1.3.5 é utilizado. Foram efetuadas 1600 misturas para cada frequência de amostragem (8 ou 16 kHz). Cada mistura foi formada por duas fontes, escolhidas aleatoriamente entre 8 sinais de voz (4 locutores femininos e 4 masculinos, falantes de inglês ou japonês), extraídos de <http://www.irisa.fr/metiss/SASSECO7/?show=test>, com duração de 10s, mantidos em 16 kHz ou reamostrados em 8 kHz, nos casos em que a frequência menor de amostragem foi utilizada. Das 1600 misturas, 1250 foram destinadas para teste (avaliação de resultados) e 350 para treinamento dos limiares de decisão. Diversas características de cada simulação foram escolhidas aleatoriamente dentro de intervalos, os quais são especificados pela Tab. 5.8.

A Fig. 5.12 apresenta, para as 1250 misturas de teste, os valores de T_{60} em função do parâmetro $\sigma_{x^-}^2$, onde verifica-se que a relação inversa entre ambas as variáveis é razoavelmente suave num contexto de misturas de sinais de voz, o que permite-nos

aplicar técnicas desenvolvidas originalmente para registros de um único sinal em misturas de sinais de voz.

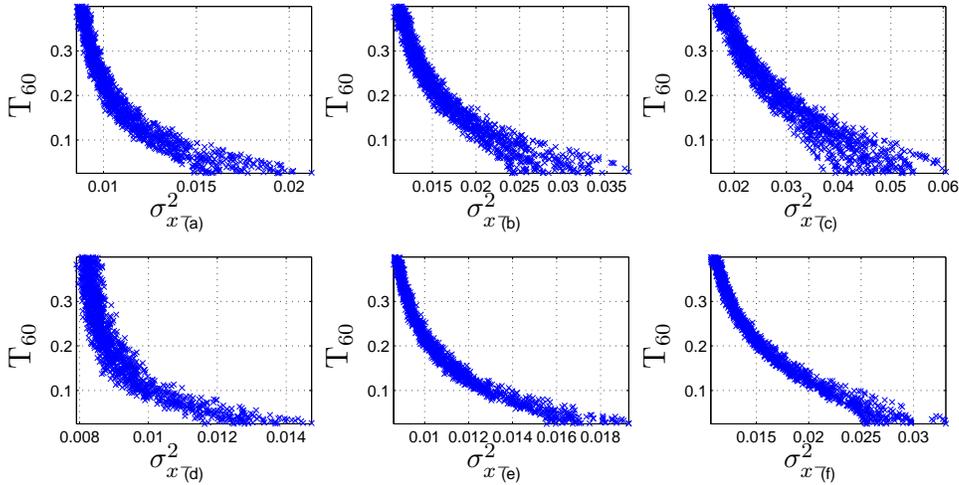


Figura 5.12: T_{60} em função de $\sigma_{x^-}^2$ em misturas de sinais de voz, com janelamento de Hanning com sobreposição de 50% e $N_f = 20$. (a) $f_s = 8$ kHz e $F_T = 32$; (b) $f_s = 8$ kHz e $F_T = 64$; (c) $f_s = 8$ kHz e $F_T = 128$; (d) $f_s = 16$ kHz e $F_T = 32$; (e) $f_s = 16$ kHz e $F_T = 64$ e (f) $f_s = 16$ kHz e $F_T = 128$.

5.5.1 Seleção de F via Limiares de Decisão

Estatisticamente, podemos associar valores baixos de $\sigma_{x^-}^2$ a valores altos de T_{60} e vice-versa. Embora a relação entre T_{60} e $\sigma_{x^-}^2$ não seja linear, a característica monotonicamente decrescente da relação $T_{60} \times \sigma_{x^-}^2$ permite-nos eleger limiares de decisão de $\sigma_{x^-}^2$, de sorte a associar intervalos desta variável à escolha de um valor específico de F.

A um aumento do grau de liberdade do sistema de mistura (correspondente a um valor maior de M), se requer (em média) um aumento do parâmetro F, de sorte a evitar uma degradação de desempenho. Cabe ressaltar não ser tal tendência tão simples, porque na prática a duração finita dos trechos das misturas limita o montante de informação estatística disponível, penalizando o desempenho da separação quando F é elevado [104].

Para a duração dos sinais e os tempos de reverberação testados nesta tese, verifica-se que o aumento de T_{60} (ou a redução de $\sigma_{x^-}^2$) requer, em média, valores maiores de F para que se otimize a SDR. Para a separação, os quatro valores de F testados (512, 1024, 2048 e 4096) para as 350 misturas elencadas no conjunto de treinamento acabam por exigir a definição de 3 limiares, sob a restrição de que intervalos associados a valores maiores de $\sigma_{x^-}^2$ requerem valores mais baixos de F. Para uma dada escolha de F_T , escolhemos os limiares de decisão que maximizam a

SDR média no conjunto de treinamento e então aplicamos estes limiares ao conjunto de teste.

5.5.2 Resultados

Para $f_s = 8$ kHz e $f_s = 16$ kHz, os valores de F_T que apresentaram o melhor resultado foram, respectivamente, 32 e 128, sendo estes os valores utilizados nos resultados que seguem. Para determinação das matrizes de separação em cada raia utilizou-se a PCA como pré-processamento, a Eq. (5.9) como atualização iterativa com a tangente hiperbólica como função *score*, a técnica de resolução de permutação de [129] e o princípio da mínima distorção [133] para atenuar a ambiguidade de escalamento. A Fig. 5.13 apresenta, para cada mistura do conjunto de teste, a escolha ótima de F , confirmando que o aumento de F deve acompanhar a redução de σ_x^2 . A SDR média, obtida no conjunto de teste (formado por 1250 misturas), para ambas as taxas de amostragem, é apresentada na Tab. 5.9, a qual mostra que a estratégia proposta de escolha de F implica ganhos, em média, de 0,5 dB. A escolha ótima refere-se, para cada cada mistura no conjunto de teste, ao valor de F que maximiza a SDR. Na prática tal escolha não é uma opção, já que requer conhecimento de cada fonte em separado para o cálculo da SDR; utilizamos esta informação para determinar o valor ótimo de F (oráculo).

A taxa de acerto da escolha proposta (em relação à escolha ótima) é de

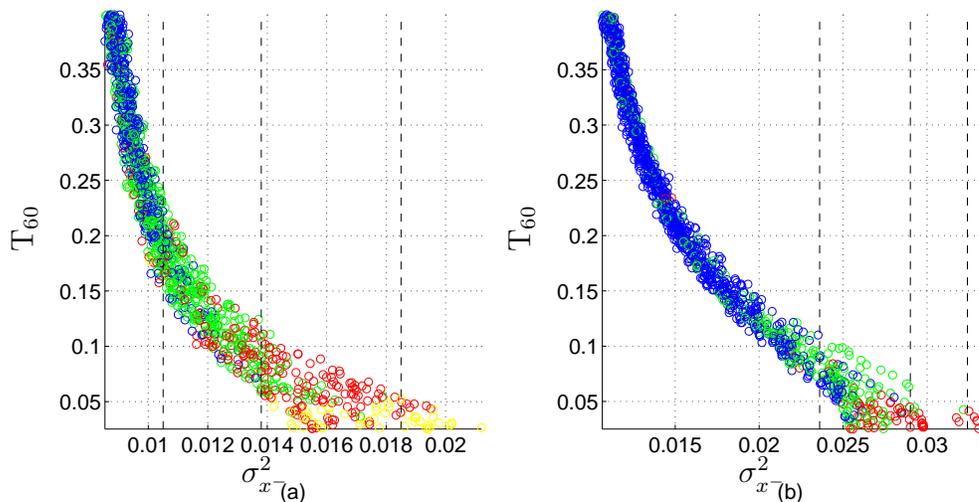


Figura 5.13: Escolha ótima de F para os casos que constam no conjunto de teste. Em amarelo: $F = 512$, em vermelho: $F = 1024$, em verde: $F = 2048$ e em azul, $F = 4096$. As linhas tracejadas assinalam os limiares obtidos pela maximização da SDR média no conjunto de treinamento. Os limiares assinalados determinam quatro intervalos, para os quais a técnica proposta escolheria, da esquerda para a direita, os seguintes valores de F , respectivamente: 4096, 2048, 1024 e 512. (a) $f_s = 8$ kHz e $F_T = 32$; (b) $f_s = 16$ kHz e $F_T = 128$.

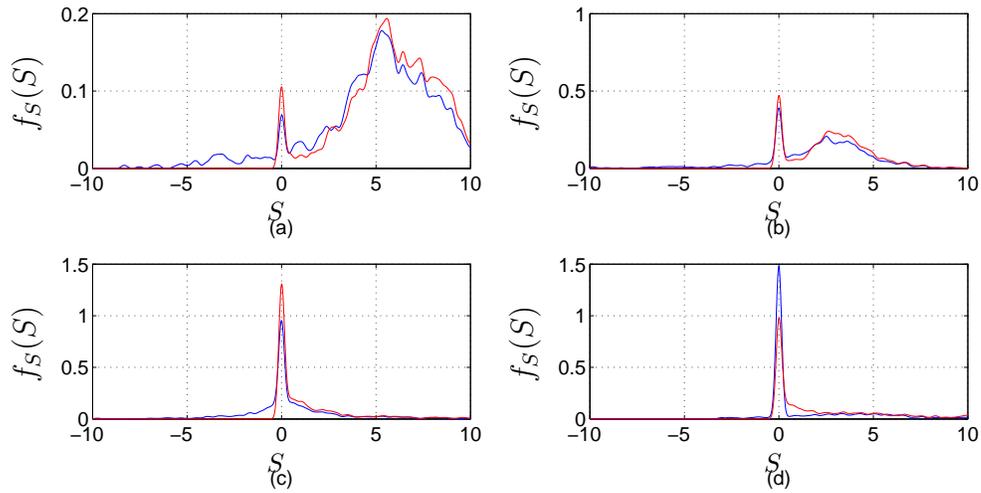


Figura 5.14: Estimativa da função densidade de probabilidade de $S(n)$ para 8 kHz. Em azul: estratégia proposta; em vermelho: escolha ótima. (a) escolha fixa $F = 512$; (b) escolha fixa $F = 1024$; (c) escolha fixa $F = 2048$ e (d) escolha fixa $F = 4096$.

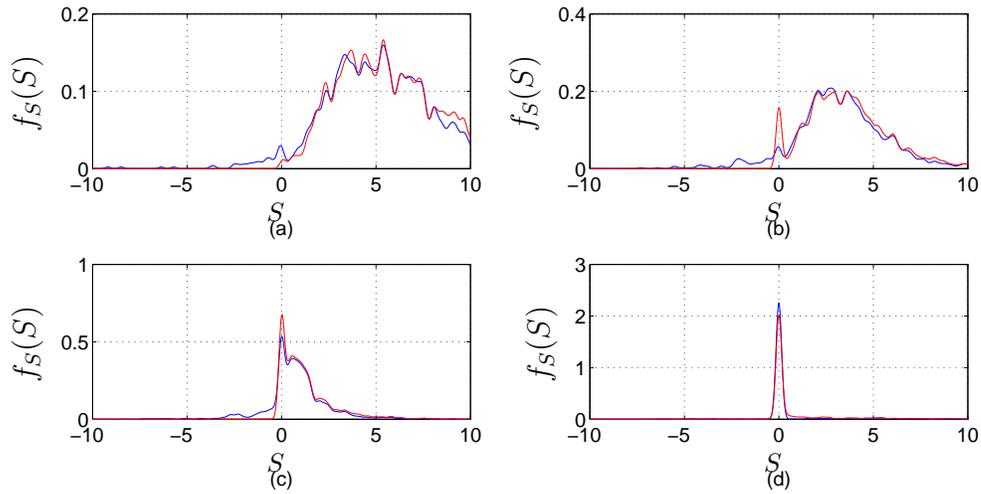


Figura 5.15: Estimativa da função densidade de probabilidade de $S(n)$ para 16 kHz. Em azul: estratégia proposta; em vermelho: escolha ótima. (a) escolha fixa $F = 512$; (b) escolha fixa $F = 1024$; (c) escolha fixa $F = 2048$ e (d) escolha fixa $F = 4096$.

Tabela 5.9: SDR média (em dB) em função de diferentes escolhas de K , no conjunto de teste.

Escolha	$f_s = 8$ kHz	$f_s = 16$ kHz
$F = 512$	9,80	6,18
$F = 1024$	13,10	8,26
$F = 2048$	14,44	10,74
$F = 4096$	12,54	11,00
Proposta	15,07	11,54
Ótima	15,96	12,01

Tabela 5.10: Probabilidade P_b dos casos em que a escolha automática (proposta) de F obtém um desempenho igual ou superior à escolha de um F fixo, para $f_s = 8$ kHz e $f_s = 16$ kHz, avaliada no conjunto de teste.

Escolha	$f_s = 8$ kHz	$f_s = 16$ kHz
F = 512	93,20%	96,72%
F = 1024	87,36%	92,56%
F = 2048	71,36%	85,12%
F = 4096	95,12%	96,96%

Tabela 5.11: Probabilidade P_p da estratégia proposta escolher um valor específico de F, para $f_s = 8$ kHz e $f_s = 16$ kHz, avaliada no conjunto de teste.

Escolha	$f_s = 8$ kHz	$f_s = 16$ kHz
F = 512	1,76%	0,64%
F = 1024	13,68%	1,04%
F = 2048	28,80%	13,52%
F = 4096	55,76%	84,80%

58,32% para $f_s = 8$ kHz e de 78,08% para $f_s = 16$ kHz, ante uma taxa de 25% que seria obtida mediante valores uniformemente aleatórios. A robustez da estratégia proposta pode ser medida pela probabilidade P_b dos casos em que a escolha proposta de F é igual ou superior à escolha de um F fixo, diante da probabilidade P_p de se escolher um valor específico de F. As probabilidades P_b e P_p , expressas em termos percentuais, são apresentadas nas Tabs. 5.10 e 5.11. Para $f_s = 8$ kHz, o valor fixo de F que obtém o melhor resultado é F = 2048, escolha que não supera a estratégia proposta em 71,36% dos casos, embora tal estratégia somente escolha este valor com probabilidade de 28,8%. Para $f_s = 16$ kHz, o mesmo se verifica: a melhor escolha de um valor fixo de F ocorre quando F = 4096, escolha que não supera a estratégia proposta em 96,96% dos casos, com uma probabilidade de 84,8% relacionada às vezes em que a estratégia proposta escolhe tal valor de F (vide Tab. 5.10).

Seja $S(n)$ a diferença entre a SDR obtida pela técnica proposta na n -ésima mistura constante no conjunto de teste e a SDR obtida, nesta mistura, pela escolha de um valor fixo de F para todas as misturas. Idealmente, $S(n)$ deveria ser uma variável aleatória não-negativa. As Figs. 5.14 e 5.15 apresentam a pdf estimada desta variável aleatória para o conjunto de teste, para diferentes frequências de amostragem e para diferentes escolhas fixas de F, tanto para o método de escolha proposto como para a escolha ideal. A pdf foi estimada por meio *kernels* gaussianos com desvio padrão de 0,15. A semelhança entre as pdfs obtidas pela estratégia proposta e as obtidas pela escolha ideal mostram que a escolha proposta de F consegue aproximar razoavelmente a escolha ótima.

5.6 Considerações Finais

Neste capítulo, apresentamos um algoritmo eficiente de inicialização de métodos de BSS no domínio da frequência. Combinando a abordagem de mascaramento binário proposto em [139] com o algoritmo EDAOR [114], foi possível obter uma convergência mais rápida e uma melhor qualidade perceptual. O procedimento de inicialização proposto também tornou o desempenho do algoritmo EDAOR menos vulnerável a ruído e à escolha do fator de aprendizagem, permitindo o emprego de valores relativamente altos, próximos a 0,5. O método proposto pode também prover uma boa solução inicial para outros algoritmos de BSS no domínio da frequência, permitindo atenuar significativamente o problema da permutação.

Uma proposta de inserção de coeficientes adaptativos interfrequenciais foi detalhada, com sua implementação resultando em ganhos razoáveis de SIR para ambientes com tempo de reverberação moderadamente baixo.

A Seção 5.5 mostra que o emprego da informação lateral oriunda da estimativa do T_{60} se revela útil para a escolha do comprimento do quadro da STFT das técnicas de separação cega, em relação a uma escolha fixa deste parâmetro.

Capítulo 6

Estruturas em Subbandas para Técnicas de Separação Cegas de Fontes no Domínio do Tempo

6.1 Introdução

Durante quase toda a década de 1990, a literatura dedicada ao problema de separação de fontes enfocou as configurações instantâneas de misturas. Dado que na prática frequentemente as misturas não aderem a este modelo, o problema de separação cega de fontes misturadas em ambientes reverberantes foi, a pouco e pouco, atraindo maior atenção da comunidade científica.

Quando a saturação nos sensores (e outros efeitos não-lineares indesejados) é desprezível, o processo do registro de sinais de áudio pode ser descrito com precisão aceitável caso modelemos o caminho acústico como um sistema MIMO linear, de modo que cada par fonte-sensor corresponda a um filtro do tipo FIR, conforme ilustrado pela Fig. 4.4.

Diversas estratégias para a separação cega de fontes nestes contextos são possíveis. Neste capítulo, estamos interessados na introdução de melhoras no algoritmo GBSS-SOS [86] (do inglês *Generalization of Blind Source Separation methods based on Second-Order Statistics*), o qual efetua adaptações em filtros FIR de separação diretamente no domínio do tempo. Técnicas de separação cega de fontes no domínio da frequência (alvo de atenção do Cap. 5) são especialmente vantajosas no tocante à redução de custo computacional. Porém, tais técnicas promovem uma significativa inserção de artefatos nas estimativas. Caso seja prioritária a minimização destes artefatos (tendo como contrapartida um aumento do custo computacional), técnicas de separação no domínio do tempo são mais interessantes. Ao longo deste capítulo, supomos ser o número N de fontes idêntico ao de misturas (configuração determi-

nada).

Algoritmos de separação cega de fontes no domínio do tempo almejam encontrar iterativamente filtros $\mathbf{W}_{ij}(z)$ (de comprimento L) de forma cega, de modo que, ao final do processo de otimização, a i -ésima estimativa dada por:

$$y_i(k) = \sum_{n=1}^N \sum_{l=0}^{L-1} w_{in}(l)x_n(k-l), \quad (6.1)$$

seja aproximadamente uma versão (provavelmente filtrada) de uma das fontes.

Uma das técnicas mais robustas propostas na literatura para este problema é o algoritmo GBSS-SOS [86] que, apelando apenas a estatísticas de segunda ordem, supõe duas características das fontes: não estacionariedade e espectro colorido¹. Afortunadamente, sinais de voz (foco de nosso interesse) apresentam estas características, tornando o método aplicável para a separação destes sinais.

A modelagem empregada para o processo do mistura (Fig. 4.4) implica a seguinte relação da i -ésima mistura com relação às fontes:

$$x_i(k) = \sum_{n=1}^N \sum_{m=0}^{M-1} h_{in}(m)s_n(k-m), \quad (6.2)$$

onde, conforme a modelagem do Cap. 1, M é o comprimento das funções de transferência que respondem pelo processo de mistura. A estrutura do processo de mistura e de separação na qual o algoritmo GBSS-SOS [86] está inserido é apresentada na Fig. 6.1.

Tendo por objetivo empregar a flexibilidade de estruturas em subbandas, assim como possível aumento na taxa de convergência advinda do emprego destas estruturas, este capítulo apresenta, na Seção 6.2, uma extensão de um algoritmo de separação cega de fontes no domínio do tempo e explícita (na Seção 6.3) uma estratégia de variação do fator de aprendizagem para evitar a instabilidade deste algoritmo. Na Seção 6.4, uma estratégia de pós-processamento é adotada para incrementar a taxa de convergência do algoritmo. A Seção 6.5 apresenta os resultados do capítulo e a Seção 6.6 encerra-o com as conclusões a ele referentes.

6.2 Estrutura Multitaxas Aplicada ao Algoritmo GBSS-SOS

Algoritmos adaptativos em subbandas permanecem objeto de atenção da comunidade científica, seja em versões não criticamente decimadas [150], seja nas máxima-

¹Neste paradigma a não-gaussianidade poderia ser explorada por meio da inserção de estatísticas de mais alta ordem.

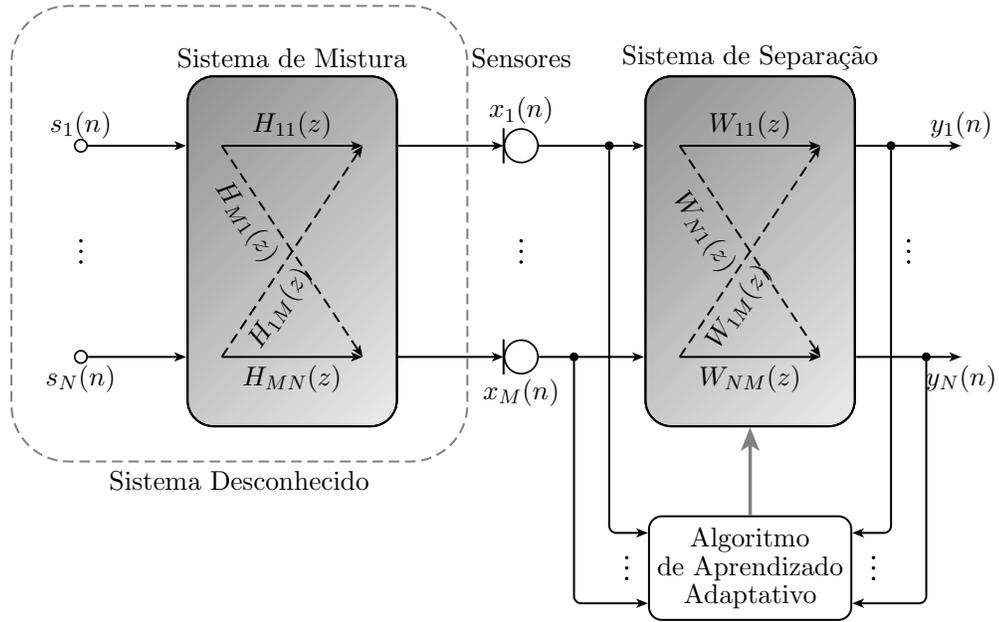


Figura 6.1: Configuração MIMO linear para a solução do problema de BSS determinado e superdeterminado.

mente decimadas [151]. Na abordagem criticamente decimada de [152], os efeitos de *aliasing* foram mitigados por meio da inserção de filtros extras entre subbandas. A estrutura resultante, contudo, incrementa significativamente o custo computacional e reduz a taxa de convergência do algoritmo. Em [153], uma nova estrutura em subbandas criticamente decimada foi derivada da estrutura WT-SF (apresentada na Fig. 3.11), por meio da inserção de um sistema multitaxas (como o da Fig. 1.5) criticamente decimado e de reconstrução perfeita; uma das vantagens desta estrutura reside em atenuar os efeitos do *aliasing* sem recorrer a filtros adaptativos adicionais (numa configuração de \underline{M} subbandas, apenas \underline{M} subfiltros são empregados).

Mais explicitamente, esta estrutura em subbandas foi obtida por meio da inserção de um sistema multitaxas numa versão uniforme da estrutura WT-SF (e utilizando os mesmos filtros de análise) após cada subfiltro esparsado adaptativo e considerando que os filtros de análise são suficientemente seletivos de modo que a sobreposição de suas respostas no domínio da frequência ocorra apenas entre subbandas adjacentes. Nesta estrutura, os filtros adaptativos atuam numa taxa que é \underline{M} vezes menor do que a taxa dos dados de entrada.

Apresentaremos diretamente a estrutura uniforme em subbandas que será empregada para o algoritmo GBSS-SOS, a qual pode ser considerada uma generalização do algoritmo original em banda cheia, já que o mesmo é obtido arbitrando-se $\underline{M} = 1$ e $\underline{H}_0(z) = \underline{F}_0(z) = 1$.

Cada sinal observado $x_i(k)$ é decomposto por meio de filtros $\underline{H}_n(z)$. Os sinais

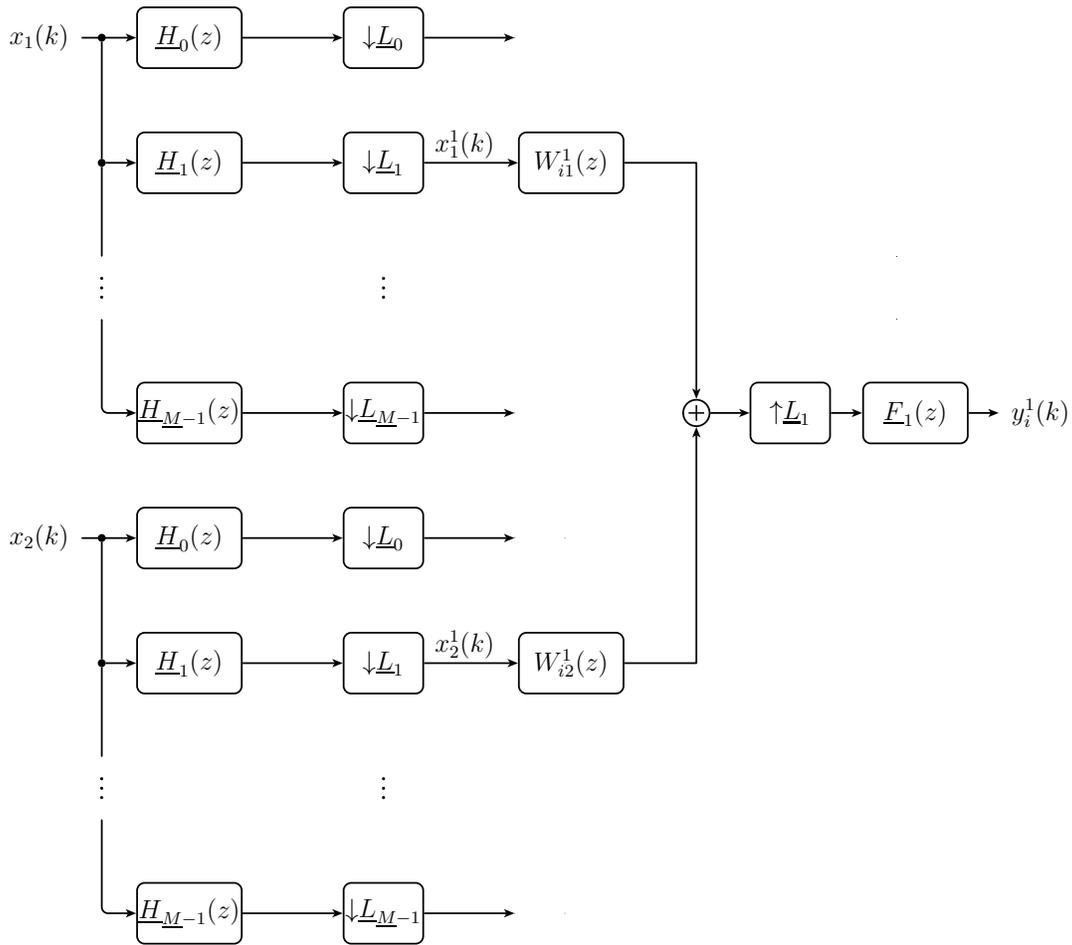


Figura 6.2: Estrutura empregada para a separação em subbandas, para um número de fontes $N = 2$. A geração da i -ésima estimativa $y_i^1(k)$ na segunda subbanda, por meio dos filtros adaptativos $W_{ij}^1(z)$, encontra-se explicitada. A i -ésima estimativa final é encontrada por meio de $y_i(k) = \sum_{n=0}^{M-1} y_i^n(k)$.

resultantes são decimados, consistindo então na entrada dos subfiltros de separação (na n -ésima subbanda) $W_{ji}^n(z)$. Os sinais de saída correspondentes são interpolados e combinados pelos filtros de síntese $E_n(z)$ de modo a promover a restauração das sinais de saída (estimativas das fontes), conforme ilustrado pela Fig. 6.2 a qual supõe, por simplicidade, a configuração TITO (do inglês *Two-Input Two-Output*), adotada no restante deste capítulo.

Os sinais $x_q^n(m)$ observados nas subbandas podem ser expressos por:

$$x_q^n(m) = \underline{\mathbf{h}}_n \mathbf{x}_q^n(m), \quad (6.3)$$

onde

$$\underline{\mathbf{h}}_n = \left[h_n(0) \quad h_n(1) \quad \dots \quad h_n(M_H - 1) \right], \quad (6.4)$$

$$\underline{\mathbf{x}}_q^n(m) = \left[x_q^n(m\underline{M}) \quad x_q^n(m\underline{M} - 1) \quad \dots \quad x_q^n(m\underline{M} - M_{\underline{H}_n} + 1) \right]^T, \quad (6.5)$$

com $M_{\underline{H}_n}$ consistindo no comprimento dos filtros de análise.

A técnica GBSS-SOS propõe a diagonalização da matriz de correlação das estimativas $y_i(k)$ considerando múltiplos blocos com diferentes atrasos relativos, de sorte a promover a decorrelação das estimativas num horizonte temporal limitado (em blocos). Há duas vantagens de se adotar o critério de decorrelação em blocos: a primeira reside num maior número de restrições e a segunda reside no apelo à não-estacionariedade das fontes, característica típica dos sinais de voz (vide Seção 1.3.1). O i -ésimo sinal de saída na n -ésima subbanda é dado por:

$$\underline{y}_j^n(m) = \sum_{p=1}^N \mathbf{w}_{jp}^n \underline{\mathbf{x}}_p^n(m), \quad (6.6)$$

onde $\underline{\mathbf{x}}_p^n(m)$ é o vetor-coluna que contém as últimas L'_n amostras do sinal em subbanda $x_p^n(m)$, onde L'_n é o comprimento dos subfiltros de separação na n -ésima subbanda (sendo L o comprimento dos filtros do algoritmo “equivalente” em banda cheia, uma escolha possível para L'_n seria $L'_n = \frac{L}{M}$, já que estamos detalhando uma estrutura uniforme) e

$$\mathbf{w}_{jp}^n = \left[w_{jp}^n(0) \quad w_{jp}^n(1) \quad \dots \quad w_{jp}^n(L'_n - 1) \right] \quad (6.7)$$

é o vetor que contém os L'_n coeficientes dos subfiltros de separação na n -ésima subbanda. O vetor $\underline{\mathbf{x}}_p^n(m)$ pode ser escrito como:

$$\underline{\mathbf{x}}_p^n(m) = \underline{\mathbf{X}}_p^n(m) \underline{\mathbf{h}}_n^T, \quad (6.8)$$

onde a matriz $\underline{\mathbf{X}}_p^n(m)$ (de dimensões $L'_n \times M_{\underline{H}_n} - 1$) é dada por:

$$\underline{\mathbf{X}}_p^n(m) = \begin{bmatrix} \mathbf{x}_p^n(m\underline{M})^T \\ \mathbf{x}_p^n((m-1)\underline{M})^T \\ \vdots \\ \mathbf{x}_p^n((m-L'+1)\underline{M})^T \end{bmatrix}. \quad (6.9)$$

No algoritmo em subbandas em questão cumpre, assim como na sua versão em banda cheia, gerar blocos dos sinais, que terão comprimento \underline{N}_n na n -ésima subbanda. Se D_n ($1 \leq D_n \leq L'_n$) é o número de correlações com diferentes atrasos relativos (na n -ésima subbanda) considerados no problema de otimização [86], podemos descrever o j -ésimo vetor de saída do bloco de índice l da n -ésima subbanda

como:

$$\underline{\mathbf{y}}_j^n(l) = \sum_{p=1}^N \mathbf{w}_{jp}^n \widehat{\underline{\mathbf{x}}}_p^n(l), \quad (6.10)$$

com a matriz $\widehat{\underline{\mathbf{x}}}_p^n(l)$ de dimensões $L'_n \times \underline{N}_n$ expressa a partir de:

$$\widehat{\underline{\mathbf{x}}}_p^n(l) = \left[\underline{\mathbf{x}}_p^n(lL'_n) \quad \underline{\mathbf{x}}_p^n(lL'_n + 1) \quad \dots \quad \underline{\mathbf{x}}_p^n(lL'_n + \underline{N}_n - 1) \right] \underline{\mathbf{H}}_n^T, \quad (6.11)$$

onde a matriz $\underline{\mathbf{H}}_n$, de dimensões $\underline{N}_n \times \underline{N}_n(M_{\underline{H}_n} - 1)$, apresenta a primeira linha formada pelos coeficientes de $\underline{h}_n(k)$ seguidos de $(\underline{N}_n - 1)M_{\underline{H}_n}$ zeros, com as linhas seguintes obtidas via deslocamentos circulares de $M_{\underline{H}_n}$ posições de uma linha para outra. As matrizes $\underline{\mathbf{Y}}_j^n(m)$ (de dimensões $D_n \times \underline{N}_n$) contém D_n vetores subsequentes de saída, podendo ser escritas como:

$$\underline{\mathbf{Y}}_j^n(l) = \sum_{p=1}^N \underline{\mathbf{W}}_{jp}^n(l) \underline{\mathbf{x}}_p^n(l), \quad (6.12)$$

com

$$\underline{\mathbf{x}}_p^n(l) = \left[\widehat{\underline{\mathbf{x}}}_p^n(l) \quad \widehat{\underline{\mathbf{x}}}_p^n(l-1) \right]^T \quad (6.13)$$

e $\underline{\mathbf{W}}_{jp}^n(l)$ sendo uma matriz do tipo Sylvester de dimensões $D_n \times 2L'_n$ dada por:

$$\underline{\mathbf{W}}_{jp}^n(l) = \begin{bmatrix} w_{jp}^n(0) & w_{jp}^n(1) & \dots & w_{jp}^n(L' - 1) & 0 & \dots & 0 & 0 \\ 0 & w_{jp}^n(0) & w_{jp}^n(1) & \dots & w_{jp}^n(L' - 1) & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & w_{jp}^n(0) & w_{jp}^n(1) & \dots & w_{jp}^n(L' - 1) & 0 \end{bmatrix}. \quad (6.14)$$

Combinando todas as N saídas de cada subbanda, podemos reescrever a Eq. (6.12) de forma concisa como:

$$\underline{\mathbf{Y}}^n(l) = \left[\underline{\mathbf{Y}}_1^n(l) \quad \dots \quad \underline{\mathbf{Y}}_N^n(l) \right]^T = \underline{\mathbf{W}}^n(l) \underline{\mathbf{X}}^n(l), \quad (6.15)$$

onde

$$\underline{\mathbf{X}}^n(l) = \left[\underline{\mathbf{x}}_1^n(l) \quad \dots \quad \underline{\mathbf{x}}_N^n(l) \right]^T, \quad (6.16)$$

$$\underline{\mathbf{W}}^n(l) = \begin{bmatrix} \underline{\mathbf{W}}_{11}^n(l) & \dots & \underline{\mathbf{W}}_{1N}^n(l) \\ \vdots & \ddots & \vdots \\ \underline{\mathbf{W}}_{N1}^n(l) & \dots & \underline{\mathbf{W}}_{NN}^n(l) \end{bmatrix}. \quad (6.17)$$

6.2.1 Algoritmo GBSS-SOS em Subbandas

Empregando soluções independentes para o problema de separação nas diferentes subbandas, a função custo *off-line* do algoritmo GBSS-SOS para a n -ésima subbanda é dada por:

$$\mathcal{F}^n(i) = \sum_{l=1}^{N_{b,n}} \frac{1}{N_{b,n}} \left\{ \log \left(\det \left(\text{bdiag} \left(\underline{\mathbf{Y}}^n(l) [\underline{\mathbf{Y}}^n(l)]^T \right) \right) \right) - \log \left(\det \left(\underline{\mathbf{Y}}^n(l) [\underline{\mathbf{Y}}^n(l)]^T \right) \right) \right\}, \quad (6.18)$$

onde i é o índice da iteração *off-line*, $N_{b,n}$ é o total de blocos (com ou sem sobreposição) considerados na otimização da n -ésima subbanda e $\text{bdiag}(\mathbf{A})$ é o operador que, atuando numa matriz de submatrizes, anula todas as submatrizes que não se encontram localizadas na diagonal principal de \mathbf{A} .

Considerando um sistema TITO e aplicando o método de gradiente natural² [86], obtemos:

$$\underline{\mathbf{W}}^n(i+1) \leftarrow \underline{\mathbf{W}}^n(i) - \beta \nabla_{\underline{\mathbf{W}}^n}^{GN} \mathcal{F}^n(i), \quad (6.19)$$

onde β é o fator de aprendizagem e $\nabla_{\underline{\mathbf{W}}^n}^{GN} \mathcal{F}^n(i)$ é o gradiente natural da função custo apresentada na Eq. (6.18), podendo ser expresso, para $N = 2$, por [86]:

$$\nabla_{\underline{\mathbf{W}}^n}^{GN} \mathcal{F}^n(i) = \sum_{l=1}^{N_{b,n}} \frac{1}{N_{b,n}} \begin{bmatrix} [\mathbf{R}_{11}^n(l)]^{-1} [\mathbf{R}_{21}^n(l)]^T \underline{\mathbf{W}}_{21}^n(i) & [\mathbf{R}_{11}^n(l)]^{-1} [\mathbf{R}_{21}^n(l)]^T \underline{\mathbf{W}}_{22}^n(i) \\ [\mathbf{R}_{22}^n(l)]^{-1} [\mathbf{R}_{12}^n(l)]^T \underline{\mathbf{W}}_{11}^n(i) & [\mathbf{R}_{22}^n(l)]^{-1} [\mathbf{R}_{12}^n(l)]^T \underline{\mathbf{W}}_{12}^n(i) \end{bmatrix}, \quad (6.20)$$

onde $\mathbf{R}_{ij}^n(l) = \underline{\mathbf{Y}}_i^n(l) [\underline{\mathbf{Y}}_j^n(l)]^T$ é uma matriz de dimensões $D_n \times D_n$ que contém as estimativas da correlação dos sinais $\underline{y}_i(m)$ e $\underline{y}_j(m)$ para D_n blocos. As redundâncias da matriz $\underline{\mathbf{W}}_{jp}^n(l)$ (Eq. (6.14)) permitem a simplificação de atualizar apenas a primeira linha desta matriz³. Para redução do custo computacional, as matrizes $[\mathbf{R}_{ij}^n(l)]^{-1}$ na Eq. (6.20) podem ser substituídas pela inversa das matrizes diagonais $\hat{\mathbf{R}}_{ii}^n(l) = \underline{\mathbf{y}}_i^n(l) [\underline{\mathbf{y}}_i^n(l)]^T \mathbf{I}$ (com $\underline{\mathbf{y}}_i^n(l)$ dado na Eq. (6.10)), as quais contém, em sua diagonal principal, as potências do bloco mais recente da i -ésima estimativa $\underline{y}_i^n(m)$ na n -ésima subbanda. Esta simplificação será doravante adotada.

6.3 Proposta de Adoção de Fatores de Aprendizagem Variantes

O algoritmo GBSS-SOS, mesmo na sua versão *off-line*, é suscetível à instabilidade, particularmente em regime permanente. Tal instabilidade pode ser significativa

²Estritamente falando, o gradiente empregado não é o natural, porém foi motivado pelas modificações ensejadas pelo gradiente natural em misturas instantâneas.

³Outras estratégias, não contempladas nesta tese, são possíveis, como empregar a primeira coluna ou a soma de diversas linhas defasadas [154].

mesmo quando o fator de aprendizagem β é suficientemente pequeno para promover uma aprendizagem bastante lenta e suave, de sorte que, neste caso, a convergência excessivamente lenta não é garantia de convergência para uma boa solução. Esta característica indesejável é um empecilho para um emprego efetivo deste algoritmo.

Os objetivos frequentemente conflitantes de obter convergência rápida e alto desempenho em regime permanente podem ser contemplados por meio da variação temporal do fator de aprendizagem, como efetuado, para os algoritmos da família PNLMS, na Seção 3.6. Porém, estratégias bem-sucedidas para outros algoritmos não se revelam proveitosas para o algoritmo GBSS-SOS, porque em geral implementam alterações excessivamente suaves para o fator de aprendizagem. Para que o procedimento de alteração do fator de aprendizagem seja efetivo, cabe impor uma drástica redução da magnitude do fator de aprendizagem assim que uma dinâmica instável for detectada.

Por esta razão, desenvolvemos um mecanismo robusto de adaptação do fator de aprendizagem, de sorte a evitar que um aumento na taxa de convergência se traduza em instabilidade ou perda de desempenho em regime permanente. O emprego de um fator de aprendizagem independente para cada coeficiente adaptativo não se revelou produtivo, porque o acoplamento da dinâmica temporal dos diversos coeficientes de um mesmo subfiltro é muito complexa. Por sua vez, nossas simulações revelam que a adoção de fatores de aprendizagem independentes para cada subbanda e para cada estimativa é adequada para a obtenção dos resultados desejáveis. Note que em cada subbanda existem diversos subfiltros, cada qual com vários coeficientes. Portanto, o fator de aprendizagem adotado para um determinado subfiltro implica que os fatores de aprendizagem de cada um de seus coeficientes é idêntico, na mesma iteração. Assim, no caso em que há duas fontes e duas estimativas, cada subbanda apresenta quatro subfiltros e dois fatores de aprendizagem independentes, já que cada estimativa está associada a dois subfiltros (vide Eq. (6.6)). A seguir, descrevemos as características de nossa proposta.

Seja $\beta_{n,p}$ o fator de aprendizagem referente à p -ésima estimativa na n -ésima subbanda. Enquanto o procedimento de aprendizagem engendrar atualizações suaves nos coeficientes adaptativos, tal fator é ligeiramente aumentado, por uma constante multiplicativa (arbitrada pelo usuário) γ_β , ligeiramente maior do que 1, de modo a acelerar a convergência, num contexto de estabilidade da adaptação. O acompanhamento da instabilidade associada aos subfiltros de separação relativos a $\beta_{n,p}$ é feito via monitoramento da direção de atualização de cada coeficiente adaptativo destes subfiltros. Se esta direção se mantiver em uma elevada fração f_d dos coeficientes, conclui-se pela inexistência de instabilidade e, por conseguinte, $\beta_{n,p}$ é incrementado por um fator γ_β , de modo a acelerar a convergência. Caso haja um número suficiente de coeficientes com direção de atualização variante entre duas

atualizações consecutivas, conclui-se pela ocorrência da instabilidade. Em geral, quando a ocorrência de instabilidade é detectada pela primeira vez, o fator $\beta_{n,p}$ encontra-se excessivamente elevado; reduzi-lo pelo fator γ_β não é suficiente para evitar fortes oscilações de desempenho. Por essa razão, a detecção de instabilidade engendra duas ações distintas, a depender do último valor de $\beta_{n,p}$. Se este valor for superior ao inicial (referente ao começo do processo de adaptação), supõe-se que este é intoleravelmente alto para manutenção do desempenho do algoritmo, e então este parâmetro volta a adquirir seu valor original. Caso, no entanto, a detecção de instabilidade seja associada a um valor de $\beta_{n,p}$ inferior ao original, a redução do mesmo não é crítica, e portanto uma estratégia simples e robusta consiste em reduzi-lo pelo fator γ_β . Para que a estratégia proposta seja adequada, a fração f_d dos coeficientes que mantém sua direção de atualização deve ser superior a 50%, de modo a garantir uma detecção suficientemente precoce da instabilidade.

O emprego de fatores de aprendizagem independentes para cada subbanda permite um ajuste individualizado dos mesmos para a dinâmica de cada subfiltro, tornando a adaptação mais flexível e possivelmente mais rápida do que a do algoritmo GBSS-SOS original no domínio do tempo. A seguir, apresentamos um pseudo-código da técnica de controle de instabilidade proposta.

Controle de Instabilidade

Inicialização (valores típicos)

$$f_d \leftarrow 0,7$$

$$\gamma_p \leftarrow 1,02$$

$$\beta_{n,p}^{(0)} \leftarrow 0,001$$

Processo de Adaptação

Para $n = 1, \dots, \underline{M}$

Para $p = 1, \dots, N$

Para $it = 1, \dots, NumIt$

Para $k = 1, \dots, N$

Atualize \mathbf{w}_{pk}^n segundo (6.19)

Fim

$f_{n,p} \leftarrow$ proporção dos coeficientes adaptativos dos filtros

$\mathbf{w}_{p1}^n, \mathbf{w}_{p2}^n, \dots, \mathbf{w}_{pN}^n$ que não alteraram sua direção de atualização.

$$\text{Se } f_{n,p} > f_d$$

$$\beta_{n,p}^{(it)} \leftarrow \gamma_p \beta_{n,p}^{(it-1)}$$

Senão

$$\text{Se } \beta_{n,p}^{(it-1)} > \beta_{n,p}^{(0)}$$

$$\beta_{n,p}^{(it)} \leftarrow \beta_{n,p}^{(0)}$$

Senão

$$\beta_{n,p}^{(it)} \leftarrow \frac{\beta_{n,p}^{(it-1)}}{\gamma_p}$$

Fim

Fim

Fim

6.4 Estratégia de Pós-Processamento

Técnicas de separação cega de fontes frequentemente apresentam, na estimativa de uma determinada fonte, interferências oriundas de fontes indesejadas. Ainda que tal interferência tenha, em geral, baixa energia, não raro é audível, o que prejudica a qualidade da separação. Para atenuar este problema, técnicas de pós-processamento têm sido propostas [155],[156],[157],[158],[159].

Se $y_i^n(k)$ é a i -ésima estimativa referente à n -ésima subbanda, durante o processo de separação podemos associá-la a dois componentes:

$$y_i^n(k) = y_{i,s}^n(k) + y_{i,c}^n(k), \quad (6.21)$$

onde $y_{i,s}^n(k)$ contém uma versão filtrada da fonte associada à i -ésima estimativa e $y_{i,c}^n(k)$ é um componente indesejado, produto de interferências (*cross-talk*).

Se a magnitude de uma das fontes é muito elevada relativamente às demais, é de se esperar que a estimativa $y_i^n(k)$ referente a esta fonte também apresente magnitude significativamente superior às demais estimativas. Os instantes associados a esta possibilidade são críticos, porque as estimativas não dominantes podem apresentar, devido à insuficiente supressão de interferência da técnica de separação, componentes indesejáveis com energia significativa.

Em especial no transiente, em que a técnica de separação ainda não angariou um desempenho razoável, é interessante atenuar tais componentes indesejáveis. Embora no domínio do tempo esta estratégia também seja possível, ela é mais natural quando se adota uma estrutura em subbandas, pois no mesmo instante fontes distintas podem ser dominantes, em diferentes subbandas.

Inspirada em [156] (que enfoca o caso de técnicas de separação no domínio da frequência), esta seção propõe um mecanismo de redução de componentes relativos a interferências que, atuando de forma contínua desde o início do aprendizado e fun-

damentado em filtros cruzados, permite uma convergência mais rápida da estrutura adaptativa. Toda a explicação será efetuada num contexto de aprendizado *off-line*, capaz de obter desempenhos mais elevados.

Após cada atualização dos subfiltros adaptativos, se encontram instantes em que uma determinada estimativa $y_i^n(k)$ é dominante, formando-se os conjuntos:

$$\mathcal{I}_{n,i} = \{k \mid |y_j^n(k)| < \gamma_p |y_i^n(k)|, \forall j \neq i\}, \quad (6.22)$$

onde γ_p é um parâmetro positivo menor do que a unidade, cuja escolha encerra um compromisso para o projetista: caso seja muito próximo de zero, praticamente não haverá detecção das estimativas dominantes. Se for muito próximo de um, poderá detectar erroneamente uma estimativa como dominante, tendo por resultado distorções e artefatos que impactariam negativamente na qualidade do processo de separação.

Após a formação dos conjuntos $\mathcal{I}_{n,i}$, tem início a etapa de aprendizado de filtros cruzados \mathbf{w}_{ji}^c , de comprimento N_p (arbitrariamente escolhido pelo projetista), que estimam os componentes interferentes. Para cada $k \in \mathcal{I}_{n,i}$, o filtro \mathbf{w}_{ji}^c estima o componente de *cross-talk* oriundo da i -ésima estimativa na j -ésima estimativa, por meio da seguinte equação (para a n -ésima subbanda):

$$\begin{bmatrix} w_{ji}^c(0) & w_{ji}^c(1) & \dots & w_{ji}^c(N_p - 1) \end{bmatrix} \begin{bmatrix} y_i^n(k) \\ y_i^n(k-1) \\ \vdots \\ y_i^n(k-N_p+1) \end{bmatrix} = y_j^n(k) + \epsilon_{ji}(k), \quad (6.23)$$

onde ϵ_{ji} é um erro cuja norma quadrática (ao longo de todos os $k \in \mathcal{I}_{n,i}$) deve ser minimizada. Portanto, a determinação dos coeficientes dos filtros cruzados é efetuada mediante uma técnica de mínimos quadrados, semelhante à descrita na Seção 5.3. Após a determinação dos coeficientes dos filtros \mathbf{w}_{ji}^c , a saída dos mesmos é subtraída das fontes não dominantes, de sorte a nelas atenuar os componentes oriundos de interferências.

6.5 Resultados

Três diferentes casos de mistura são o foco desta seção. Em todos, as funções de transferência acústica envolvidas foram obtidas mediante o simulador apresentado na Seção 1.3.5, para um ambiente de dimensões 4 m \times 3 m \times 2,5 m, cujo centro de massa coincide com o centro de massa do arranjo de microfones. Os sinais de voz foram extraídos de <http://www.irisa.fr/metiss/SASSECO7/?show=test> e reamostrados para 8 kHz. A Tab. 6.1 apresenta as demais especificações dos casos testados.

Tabela 6.1: Características dos casos testados.

Parâmetro	Caso 1	Caso 2	Caso 3
fonte 1	female_src_1	female_src_2	male_src_2
fonte 2	male_src_1	female_src_3	female_src_4
T_{60}	150 ms	100 ms	110 ms
coeficientes de absorção das paredes	0,8	0,85	0,95
distância entre os microfones	5 cm	5 cm	7 cm
distância das fontes até o centro do arranjo de microf.	1 m	1,2 m	1,3 m
ângulo da fonte 1	-45°	-30°	30°
ângulo da fonte 2	45°	-55°	65°

Nos casos testados, o banco de filtros utilizado apresenta $\underline{M} = 4$ subbandas, com bancos de filtros de reconstrução perfeita modulados por cosseno [67],[160]. Cada filtro de análise e de síntese foi obtido a partir de um filtro protótipo de comprimento $M_{\underline{H}} = 64$, determinado a partir da proposta de [161]. O algoritmo GBSS-SOS não demonstrou desempenho razoável com bancos de filtros criticamente decimados. Em nossas simulações, as subbandas mais altas são as mais sensíveis à decimação. Uma escolha dos fatores de decimação que, nos casos testados, obteve um bom desempenho foi:

$$\underline{L}_n = \begin{cases} 2, & \text{para } n = 0, \\ 1, & \text{para } n = 1, 2, 3 \end{cases}, \quad (6.24)$$

de modo que serão estes os fatores de decimação empregados nas simulações desta seção. Em cada subbanda, os filtros de separação têm comprimento 256, de sorte que o total de coeficientes adaptativos da estrutura, correspondente a cada caminho mistura-estimativa, é de 1024. Os fatores de aprendizagem são inicializados com o valor 10^{-3} .

Nas implementações testadas, a SIR média foi avaliada a cada 50 iterações. O algoritmo original GBSS-SOS (sem estrutura em subbandas) foi simulado com 3 diferentes comprimentos de filtros de separação: $L = 512, 1024$ e 2048 . Quando empregado, o controle de instabilidade utilizou os parâmetros $\gamma_{\beta} = 1,02$ e $f_d = 70\%$.

A Fig. 6.3 avalia o impacto da inserção do controle de instabilidade no caso 2, para todas as configurações avaliadas. Esta figura revela a vantagem do emprego de fatores de aprendizagem variantes no tempo, os quais promovem uma convergência mais suave e um melhor desempenho em regime permanente. Em todas as demais simulações estas características se replicaram.

A Fig. 6.4 compara a evolução da SIR média no algoritmo padrão com a do algoritmo em subbandas, nos dois primeiros casos. Em regime permanente, o de-

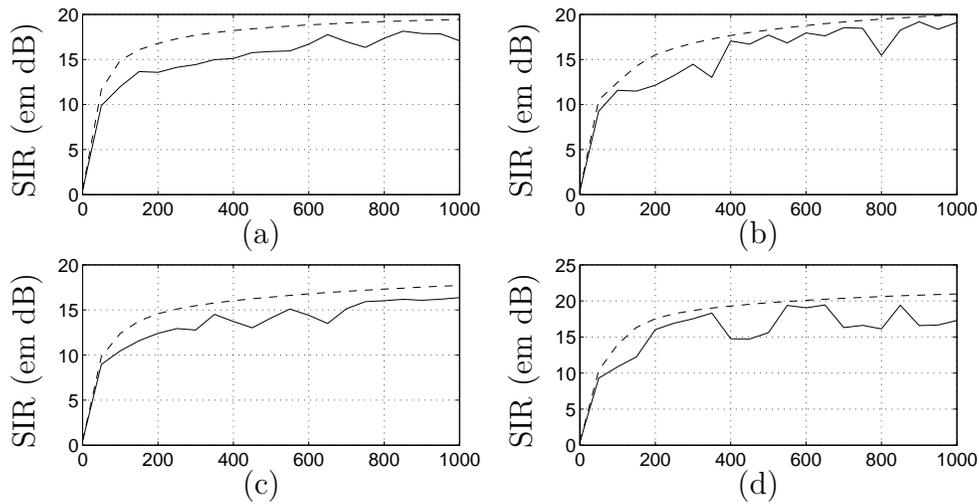


Figura 6.3: SIR média (em dB) obtida pelos algoritmos na versão original (linha sólida) e na versão com variação do fator de aprendizagem (linha tracejada) para controle de instabilidade, no caso 2. (a) algoritmo padrão, com $L = 512$; (b) algoritmo padrão, com $L = 1024$; (c) algoritmo padrão, com $L = 2048$; (d) algoritmo em subbandas.

sempenho do algoritmo em subbandas se revelou superior em ambos os casos. A taxa de convergência no caso 2 foi favorável ao algoritmo padrão, o que não ocorrerá quando avaliarmos o impacto do pós-processamento.

A inserção do pós-processamento resulta em aumento da taxa de convergência e, no caso 2, também implica melhora do desempenho em regime permanente, conforme os resultados expostos na Fig. 6.5. Cabe ressaltar que o algoritmo padrão (sem apelo a bancos de filtros) não experimenta diferença notável de desempenho com o pós-processamento, porque não são muitos os instantes em que apenas uma das fontes encontra-se ativa. A estrutura em subbandas, ao separar o sinal em diferentes faixas de frequências, permite que estes instantes se tornem mais frequentes, pois um sinal de voz pode apresentar energia alta em uma subbanda e energia reduzida em outra, durante determinado período de tempo. A Fig. 6.6 revela que o caso 3, no qual o desempenho do algoritmo de separação não foi bom, o algoritmo de pós-processamento, que pressupõe a coexistência de estimativas razoavelmente separadas, não angaria ganhos de desempenho.

Embora o custo computacional da estrutura em subbandas seja superior ao do algoritmo em banda cheia (devido ao fato de usar decimação não crítica), ela é capaz de obter desempenho superior, aliado a uma taxa de convergência maior. Cumpre notar que o algoritmo em banda cheia costuma experimentar perda de desempenho quando aumentamos o comprimento L dos filtros de separação (e por conseguinte o custo computacional), de modo que a estrutura em subbandas proposta é uma opção interessante quando o objetivo é melhorar o desempenho, mesmo que à custa

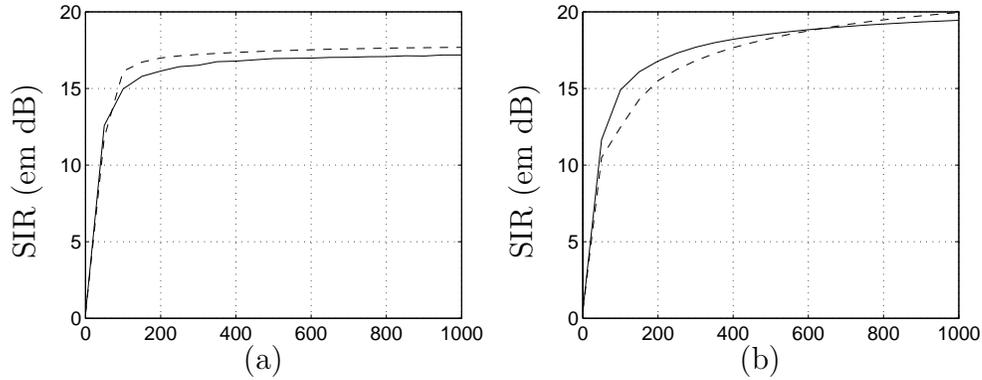


Figura 6.4: SIR média (em dB) obtida pelos algoritmos sem processamento multitaxas (linha sólida) e na versão em subbandas. (a) caso 1 ($L = 512$); (b) caso 2 ($L = 1024$). Foram escolhidos, para o algoritmo padrão, os valores de L que apresentaram melhor desempenho.

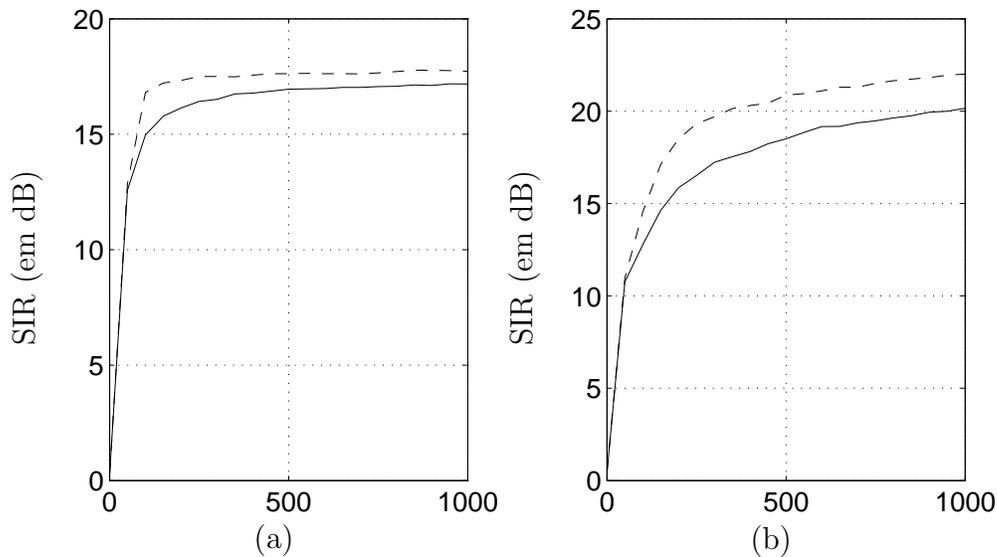


Figura 6.5: SIR média (em dB) obtida pelos algoritmos sem processamento multitaxas (linha sólida) e na versão em subbandas. Em ambas as versões, o pós-processamento proposto foi inserido, com $\gamma_p = 0,8$ e $N_p = 10$. (a) caso 1 ($L = 512$); (b) caso 2 ($L = 1024$). Foram escolhidos, para o algoritmo padrão, os valores de L que apresentaram melhor desempenho.

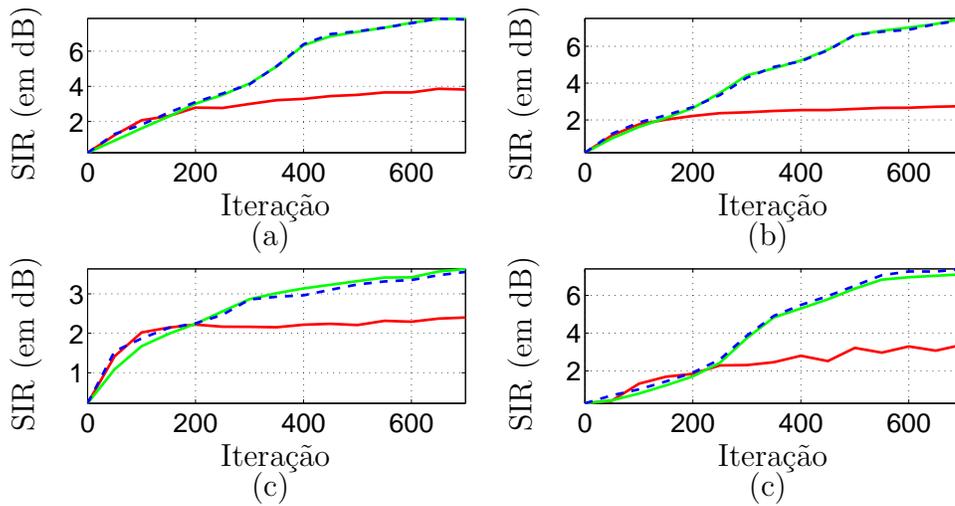


Figura 6.6: SIR média (em dB) obtida pelos algoritmos, para o caso 3. Em vermelho: algoritmo padrão; em verde: algoritmo com controle de instabilidade; em azul tracejado: algoritmo com controle de instabilidade e pós-processamento. Quando utilizado, o pós-processamento utilizou como parâmetros $\gamma_p = 0,8$ e $N_p = 10$. Nesta configuração, o baixo desempenho não permitiu que o pós-processamento ensejasse ganhos. (a) algoritmo em banda cheia ($L = 512$); (b) algoritmo em banda cheia ($L = 1024$); (c) algoritmo em banda cheia ($L = 2048$) e (d) algoritmo em subbandas.

de uma maior complexidade.

O fato de a estrutura em subbandas proposta, aliada ao controle de instabilidade e à estratégia de pós-processamento, utilizar decimação não-crítica não implica necessariamente aumento do número de coeficientes adaptativos. No caso 2, por exemplo, a estrutura em subbandas tem um desempenho que supera o do algoritmo em banda cheia, com o mesmo quantitativo de coeficientes adaptativos. Tal fato se deve à característica de o algoritmo GBSS-SOS, por ser completamente cego, não conseguir adaptar adequadamente um filtro com excessivo número de coeficientes, o que pode ser explorado por meio de processamento em subbandas.

6.6 Considerações Finais

Neste capítulo, o algoritmo GBSS-SOS, proposto para a separação cega de fontes no domínio do tempo, teve uma extensão em subbandas apresentada. Tal extensão, caso aliada às demais propostas aqui elencadas (controle de instabilidade e pós-processamento), é capaz de prover aumento da taxa de convergência e melhora do desempenho em regime permanente. A estrutura em subbandas não é criticamente decimada, o que não necessariamente implica, dadas as especificidades do algoritmo GBSS-SOS, que seu emprego aumente o número de coeficientes adaptativos. O problema de instabilidade a que o algoritmo GBSS está sujeito foi contornado por

meio de fatores de aprendizagem variantes no tempo e a taxa de convergência de sua versão em subbandas foi aumentada com a adoção de uma estratégia de pós-processamento.

Capítulo 7

Identificação do Sistema de Mistura por meio de Técnicas de Matching Pursuit

7.1 Introdução

Para configurações de misturas instantâneas, lineares, sem ruído e para as quais $M \geq N$, o problema de BSS pode ser considerado resolvido [11]. Porém, devido às propriedades físicas (como a finitude da velocidade de propagação dos sinais ou a existência de múltiplos percursos) da maioria das configurações de misturas, tais condições são frequentemente violadas. Neste capítulo, apresentamos uma técnica de BSS que identifica o sistema de mistura em configurações anecoicas, sem a restrição de o número de fontes ser inferior ou igual ao número de misturas. Também contemplaremos as vantagens que um dicionário adaptado aos sinais pode trazer em termos de melhora na acurácia das estimativas.

No Cap. 5, representações redundantes foram empregadas por conveniência: mesmo na ausência de sobreposição entre quadros adjacentes, seria possível efetuar a separação no domínio da frequência (ainda que em geral com degradação de desempenho) caso, por exemplo, empregássemos uma janela retangular. Por outro lado, neste capítulo a habilidade de prover adaptativamente representações esparsas, oriunda do emprego de dicionários redundantes, será fundamental para a técnica utilizada no problema de separação cega de fontes.

Este capítulo se insere no paradigma das técnicas de análise de componentes esparsos (SCA, abordada no Cap. 4), as quais podem, em geral, atuar em configurações subdeterminadas com diversidade espacial ($M > 1$), sendo por isso consideradas mais poderosas do que as técnicas de ICA [95].

Este capítulo está estruturado no seguinte modo: a Seção 7.2 motiva o emprego

de representações redundantes. A Seção 7.3 apresenta o algoritmo *Matching Pursuit* (MP) original e a seguir detalha a nossa proposta que emprega dicionários redundantes multicanais. Já na Seção 7.4, modelamos o processo de mistura anecoica, bem como as simplificações decorrentes das ambiguidades inerentes ao processo de separação cega. A Seção 7.5 detalha como podemos, a partir de nossa abordagem de decomposição multicanal, estimar os coeficientes de atenuação do processo de mistura, em conjunto com os atrasos. No contexto de misturas instantâneas, a Seção 7.6 investiga se o emprego de dicionários adaptados, caso sejam disponíveis trechos isolados das fontes a separar, implica um desempenho superior na identificação. Por fim, a Seção 7.7 elenca as conclusões finais deste capítulo.

7.2 Motivação para o Emprego de Representações Redundantes

Em expansões (ou representações) não redundantes, o sinal é escrito em termos de uma combinação linear de elementos que, sendo linearmente independentes, formam uma base não raro ortonormal. Para propósitos de codificação, bases eficientes são aquelas que permitem uma representação compacta com uma baixa complexidade computacional [162]. Porém, é conhecido o fato de que alguns sinais não podem ser representados eficientemente numa base ortonormal, como demonstra o popular exemplo de que nem impulsos nem senóides expressam adequadamente o comportamento de uma mistura de impulsos e de senóides [163]. Exemplos como esse estimularam pesquisas por técnicas mais elaboradas de análise de sinais.

Em contextos não-redundantes, para reduzir a diluição da informação nos elementos da base e obter melhor compactação de energia, diversas estratégias têm sido propostas na literatura para adaptar a base às estatísticas de um sinal. No entanto, esta adaptação é efetuada para um sinal particular. Como resultado, mesmo que esta adaptação seja eficiente para representar sinais com propriedades estacionárias, não terá eficácia relevante quando aplicada a sinais mais complexos e não estacionários, como sinais de voz [162].

Não sendo obrigatório o emprego de uma base, podemos, por exemplo, utilizar duas bases ortonormais para representação de um sinal de dimensão L ; neste caso, escolher os coeficientes da representação implica a solução de um conjunto indeterminado de L equações com $2L$ variáveis desconhecidas. Esse maior grau de liberdade, oriundo do descarte da unicidade da decomposição, permite que restrições adicionais orientem a escolha da representação, como por exemplo a minimização do suporte (ou seja, número de coeficientes não-nulos) [164], tornando as técnicas redundantes de compressão por síntese capazes de atingir taxas de compressão mais elevadas do

que as estratégias de codificação via transformadas [165].

Esse raciocínio motiva uma técnica interessante e crescentemente popular de obtenção de representação esparsas que, apelando a conjuntos redundantes (ou dicionários) de formas de onda pré-definidas (também conhecidas como átomos), não raro permitem uma representação compacta dos sinais de interesse. É sabido que a probabilidade de se encontrar uma aproximação esparsa em geral aumenta com a redundância do dicionário [166]. No contexto de separação de fontes, a informação angariada pelos sensores é altamente redundante, em parte por conter múltiplas versões correlacionadas do mesmo fenômeno físico, de modo que a informação relevante sobre os processos subjacentes que geram as observações (misturas) apresenta uma dimensionalidade muito menor do que a dos dados adquiridos [167].

Por essa razão, nos últimos anos houve uma grande atividade no campo de aproximações esparsas. Isto é parcialmente motivado pelo potencial das técnicas relacionadas para tarefas típicas em processamento de sinais como análise, redução de dimensionalidade, retirada de ruído ou compressão, nas quais é necessária uma representação própria dos conjuntos de dados por meio de subespaços de dimensionalidade reduzida, os quais são adaptados tanto para as características dos sinais quanto para a tarefa de processamento em particular [167]. A principal vantagem desta classe de técnicas é a completa liberdade de projeto do dicionário, o qual pode ser eficientemente adaptado para emular as estruturas do sinal [166].

Em expansões redundantes, o sinal é escrito em termos de uma combinação linear de um subconjunto de vetores extraídos de um dicionário redundante. Estes vetores não são mais linearmente independentes, o que acarreta perda da unicidade da expansão e aumento da complexidade computacional. No entanto, entre todas as expansões válidas, pode-se ter interesse em alguma que otimize algum critério, como por exemplo a minimização da norma l_0 (equivalente ao número de coeficientes não nulos) da expansão, sob a restrição de conservação de energia. Pretendendo extrair estruturas coerentes e significativas do sinal, tais expansões podem ser empregadas para potencializar a análise e a filtragem de sinais [162],[168]. Representações esparsas também apresentam vantagens em termos de compressão de sinais e/ou em termos de compreensão de seus processos subjacentes [164], que consistem, no contexto desta tese, nas contribuições das fontes (não observadas) nas misturas (observadas). A redução de dimensionalidade é mais efetiva quando se atenta para o que gera a diferença de dimensionalidade entre os processos físicos e os registros dos mesmos; ainda que o número de fenômenos possíveis seja muito grande, o número de causas observadas por sensores num dado instante é muito menor. Em outras palavras, os processos observados são esparsos no conjunto de todas as causas possíveis; embora o número de subespaços de representação seja alto, apenas alguns poucos conterão amostras dos dados medidos pelos sensores. Ao identificar estes poucos

subespaços, encontramos uma representação num espaço de menor dimensão [167].

Encontrar a representação mais esparsa de um sinal em um dado dicionário redundante é, em geral, um difícil problema NP (do inglês *Nondeterministic Polynomial Time*). Mesmo algoritmos subótimos (ou que empregam heurísticas subótimas) implicam alto custo computacional. O emprego dessas heurísticas permite encontrar representações que, sendo suficientemente bem-sucedidas, não se distanciam em demasia da otimalidade. Essas estratégias reduzem a complexidade computacional por meio do relaxamento do problema de otimização original [166], o que permite esquivar-se da elevada complexidade oriunda de uma busca combinatorial.

Uma destas técnicas mais populares é o *Matching Pursuit* (MP) [169] o qual, partindo de um dicionário redundante, define uma decomposição adaptativa de um sinal. Esta decomposição progressivamente isola as estruturas do sinal coerentes com o dicionário adotado por meio de projeções ortogonais nos átomos do dicionário, o que provê uma representação adaptativa na qual os coeficientes mais significativos são extraídos primeiramente [162]. Esta propriedade adaptativa do MP é adequada para a SCA, pois motiva uma técnica de identificação do sistema de mistura que, sem apelo à inversão deste sistema, pode reconstruir as fontes mesmo quando tal sistema é degenerado.

Outras técnicas de aproximação esparsas poderiam ser empregadas no lugar do MP, como por exemplo o *Orthogonal Matching Pursuit* (OMP) [170], o *Optimized Orthogonal Matching Pursuit* (OOMP) [171] e o *Basis Pursuit* (BP) [172], porém nossas simulações preliminares revelaram que o aumento de complexidade computacional exigido por algumas destas técnicas não se traduz em uma melhora correspondente na robustez das estimativas; por essa razão, concentramo-nos neste capítulo na técnica MP.

A maior parte deste capítulo detalhará técnicas que atuam no paradigma da SCA e que, por não atuarem no espaço de separação como as técnicas de ICA, podem se aplicar a um contexto subdeterminado (que se caracteriza por apresentar um número de fontes superior ao de sensores) o qual, dada a menor disponibilidade de informação, é mais desafiador. Por outro lado, a modelagem do processo de mistura, embora mais complexa que a empregada nas técnicas padrão de ICA [11], não abrange as misturas convolutivas, foco dos Caps. 5 e 6.

7.3 Decomposição via Dicionários Redundantes: Algoritmo MP

Relacionado às técnicas estatísticas de regressão via *Projection Pursuit* [173], o algoritmo MP é uma estratégia sub-ótima para a obtenção de uma representação

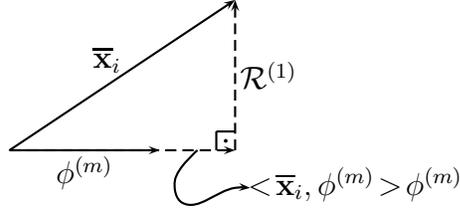


Figura 7.1: Exemplo de projeção ortogonal de um sinal $\bar{\mathbf{x}}_i$ no átomo $\phi^{(m)}$.

adaptativa esparsa de um sinal. Apesar da progressiva decomposição em termos de uma combinação linear de alguns dos elementos constantes no dicionário, a aproximação empregada pelo MP é não-linear [169], pois a escolha (e a ordem) dos componentes não satisfaz a propriedade da superposição.

Seja um dicionário $\mathcal{D} = \{\phi^{(m)}\}$, com $m \in \{1, \dots, \#D\}$ composto de átomos de comprimento L e norma unitária, com o m -ésimo átomo denotado por $\phi^{(m)}$. Cumpra ao MP decompor um determinado vetor-coluna $\bar{\mathbf{x}}_i$ (de comprimento L), de forma iterativa. Efetuando-se uma projeção ortogonal no m -ésimo átomo, a decomposição se torna:

$$\bar{\mathbf{x}}_i = \langle \bar{\mathbf{x}}_i, \phi^{(m)} \rangle \phi^{(m)} + \mathcal{R}^{(1)}, \quad (7.1)$$

onde $\langle \mathbf{x}, \mathbf{y} \rangle$ é o produto interno entre dois vetores-coluna \mathbf{x} e \mathbf{y} (de modo que $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$) e $\mathcal{R}^{(1)}$ é o resíduo resultante da primeira iteração. Multiplicando-se à esquerda por $\phi^{(m)T}$ e empregando a propriedade dos átomos terem norma unitária, temos:

$$\phi^{(m)T} \bar{\mathbf{x}}_i = \langle \bar{\mathbf{x}}_i, \phi^{(m)} \rangle + \phi^{(m)T} \mathcal{R}^{(1)} \Rightarrow \langle \mathcal{R}^{(1)}, \phi^{(m)} \rangle = 0, \quad (7.2)$$

donde decorre que o resíduo $\mathcal{R}^{(1)}$ é ortogonal ao átomo $\phi^{(m)}$, conforme ilustrado pela Fig. 7.1. Daí conclui-se:

$$\|\bar{\mathbf{x}}_i\|^2 = |\langle \bar{\mathbf{x}}_i, \phi^{(m)} \rangle|^2 + \|\mathcal{R}^{(1)}\|^2, \quad (7.3)$$

de sorte que, para reduzir a energia do resíduo cabe-nos escolher m (ou, equivalentemente, o átomo do dicionário) de modo a maximizar o termo $|\langle \bar{\mathbf{x}}_i, \phi^{(m)} \rangle|^2$. Almejando (ainda que de forma sub-ótima) maximizar a taxa de redução dos resíduos, o procedimento iterativo proposto no algoritmo MP [169] para a determinação

de uma soma ponderada de J átomos de um dicionário redundante monocanal \mathcal{D} é:

Algoritmo MP Padrão

Inicialização

$$\mathcal{R}^{(0)} \leftarrow \bar{\mathbf{x}}_i$$

Laço Principal

Para $j = 1, 2, \dots, J$

Escolha $\phi^{(m_j)} := \arg \max_{\phi \in \mathcal{D}} |\langle \mathcal{R}^{(j-1)}, \phi \rangle|$

$$\mathcal{R}^{(j)} \leftarrow \mathcal{R}^{(j-1)} - \langle \mathcal{R}^{(j-1)}, \phi^{(m_j)} \rangle \phi^{(m_j)}$$

Fim

Findo o processo iterativo, o vetor $\bar{\mathbf{x}}_i$ está decomposto segundo:

$$\bar{\mathbf{x}}_i = \sum_{j=1}^J \langle \mathcal{R}^{(j-1)}, \phi^{(m_j)} \rangle \phi^{(m_j)} + \mathcal{R}^{(J)}. \quad (7.4)$$

As Figs. 7.2 e 7.3 apresentam geometricamente o processo de decomposição via projeções ortogonais, empreendido pelo MP, para $L = 3$ e $L = 512$ (respectivamente); interpretar cada produto interno ou cálculo de correlação como um procedimento de filtragem [162] contribui para elucidar o tipo de filtragem adaptativa implementada pelo algoritmo MP.

Não sendo garantido *a priori* o alinhamento temporal entre as formas de onda dos átomos constantes no dicionário e as encontradas nos sinais a decompor, normalmente diversos deslocamentos temporais de uma mesma forma de onda são inseridos no dicionário \mathcal{D} , o que contribui para aumentar sobremaneira sua cardinalidade (número de elementos) $\#\mathcal{D}$. Como consequência deste procedimento, a complexidade da iteração MP, que já é relevante, costuma aumentar significativamente [166]. Uma estratégia para atenuar este problema consiste em efetuar convoluções circulares entre cada átomo e o sinal (ou resíduo) a decompor, operação que pode ser agilizada mediante o emprego de FFTs (versões rápidas da DFT, do inglês *Fast Fourier Trans-*

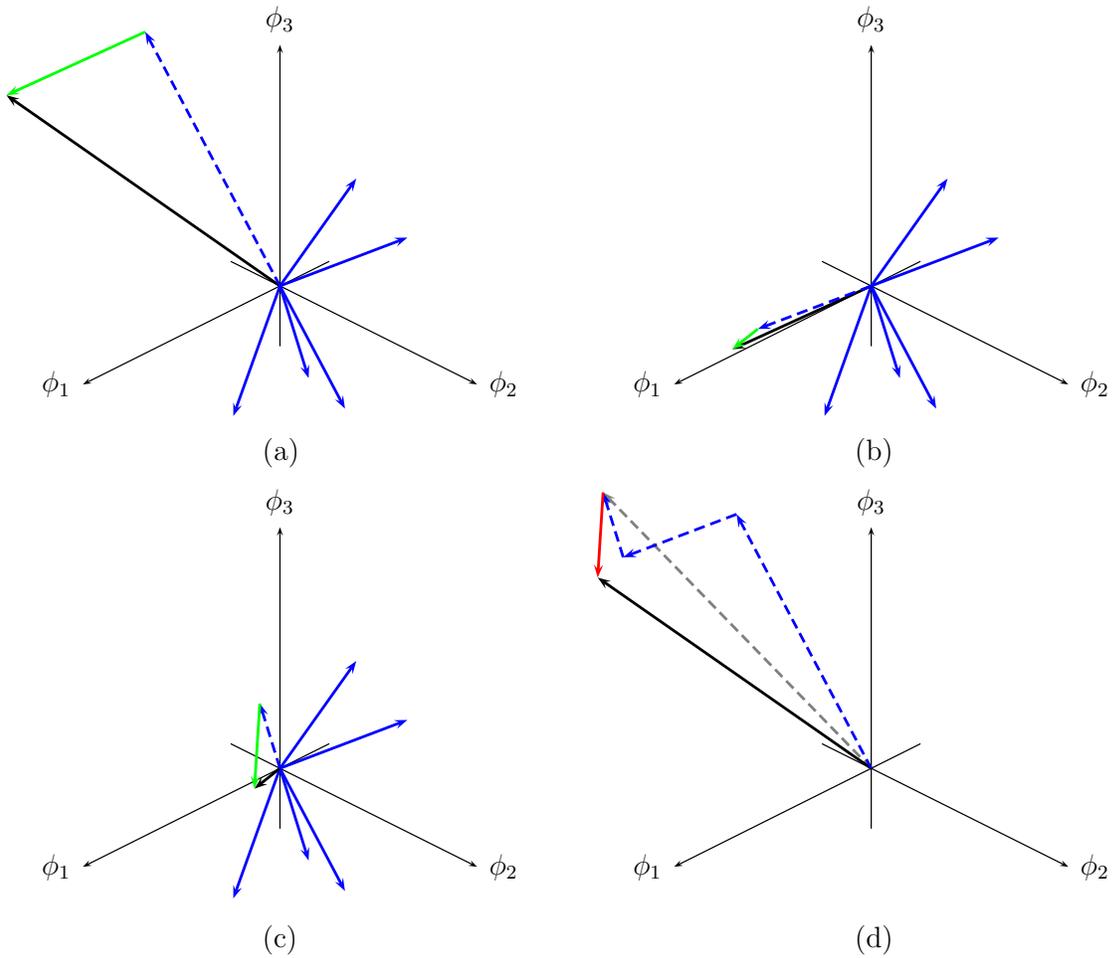


Figura 7.2: Exemplo de decomposição via MP de um sinal tridimensional ($L = 3$), com um dicionário composto por 5 átomos (vetores em azul), cada qual com componentes ϕ_i , para $i \in \{1, 2, 3\}$, os quais não devem ser confundidos com o i -ésimo átomo $\phi^{(i)}$. Em verde, os resíduos de cada iteração. Os vetores tracejados em azul indicam as projeções do sinal (na primeira iteração) ou dos resíduos (nas iterações posteriores) no átomo que apresenta o maior produto interno. (a) Primeira iteração com sinal original em preto; (b) segunda iteração, com o resíduo $\mathcal{R}^{(1)}$ em preto; (c) terceira iteração, com o resíduo $\mathcal{R}^{(2)}$ em preto e (d) aproximação (em cinza tracejado) do sinal original (em preto), com o resíduo final $\mathcal{R}^{(3)}$ da representação assinalado em vermelho. Ao contrário deste exemplo ilustrativo, em geral o número de iterações é bem inferior a L .

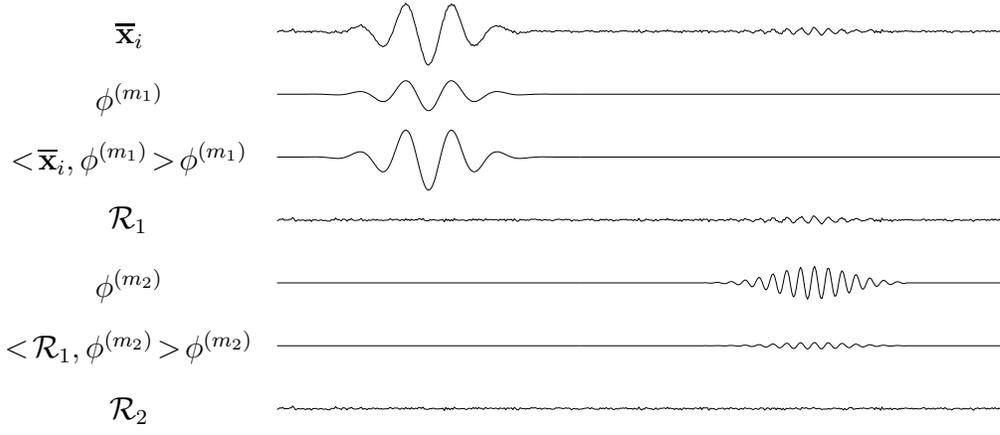


Figura 7.3: Exemplo de decomposição via MP de um sinal de dimensão 512, no qual a maior parte da energia do sinal é extraída em apenas duas iterações.

form). Deste modo, o algoritmo MP pode ser implementado da seguinte forma:

Algoritmo MP com FFTs

Inicialização

$$\mathcal{R}^{(0)} \leftarrow \bar{\mathbf{x}}_i$$

Laço Principal

Para $j = 1, 2, \dots, J$

Para $i = 1, 2, \dots, \#\mathcal{D}$

$$\phi^{(i)} := \text{IFFT}\{\text{FFT}\{\mathcal{R}^{(j-1)}\} \odot \text{FFT}\{\phi^{(i)}\}\}$$

$$l_i := \arg \max_l \{|\phi^{(i)}(l)|\}$$

$$\kappa_i := \phi^{(i)}(l_i)$$

Fim

Escolha $m_j := \arg \max_u |\kappa_u|$

$$\mathcal{R}^{(j)} \leftarrow \mathcal{R}^{(j-1)} - \kappa_{m_j} \text{CIRCSHIFT}\{\check{\phi}^{(m_j)}, l_{m_j}\}$$

Fim

onde $\text{FFT}\{\mathbf{u}\}$ e $\text{IFFT}\{\mathbf{v}\}$ consistem na DFT e na sua inversa, aplicadas aos vetores \mathbf{u} e \mathbf{v} , \odot é o produto elemento a elemento entre vetores, $\check{\phi}^{(m_j)}$ é a versão revertida no tempo do átomo $\phi^{(m_j)}$ e $\text{CIRCSHIFT}(\mathbf{u}, n)$ implementa um deslocamento circular no vetor \mathbf{u} de n unidades. Importa ressaltar que o recurso a FFTs para redução de complexidade computacional do MP não é a única estratégia possível; outras estratégias, alheias ao escopo deste trabalho, são, entre outras, o emprego de dicionários em subbandas [162], a poda do dicionário via quantização vetorial [174] e a otimização analítica da fase ótima [175] que, embora tenha sido proposta para átomos contínuos, pode ser generalizada para átomos discretos [176].

Embora, na nossa formulação, o número J de iterações seja arbitrado *a priori*, é

possível implementar um critério adaptativo de parada para o MP, como por exemplo o cálculo de uma média móvel da razão entre as energias do sinal e do ruído, cuja aproximação de um valor médio estacionário (dependente do dicionário e de L) implica a incoerência do dicionário com o resíduo, de modo que este, se resumindo (aproximadamente) a um ruído branco e gaussiano, não exija decomposições adicionais [176],[177],[178].

Uma escolha popular de dicionários consiste no emprego de dicionários parametrizados [169] obtidos via versões moduladas, transladadas e escalonadas de uma função $\bar{\phi}(\cdot)$, de modo que:

$$\phi^{(m)}(k) = c_m \bar{\phi} \left(\frac{k - \vartheta_1}{\vartheta_2} \right) \cos(2\pi\vartheta_3(k - \vartheta_1) + \vartheta_4), \quad (7.5)$$

onde ϑ_1 é responsável pelo deslocamento temporal, ϑ_2 pelo escalonamento, ϑ_3 pela frequência de modulação, ϑ_4 pela fase, c_m é uma constante de normalização escolhida de modo a tornar unitária a norma de todos os átomos, e m é o índice do átomo que define, sem ambiguidade, os valores de ϑ_i , para $i = 1, \dots, 4$. A Fig. 7.4 apresenta alguns átomos e a influência da escala e da frequência de modulação na estrutura temporal dos mesmos. Em geral, não utilizaremos versões transladadas dos átomos (relacionadas a diferentes valores de ϑ_1), porque o emprego de FFTs no cômputo dos produtos internos naturalmente fornece estes deslocamentos (ainda que circulares).

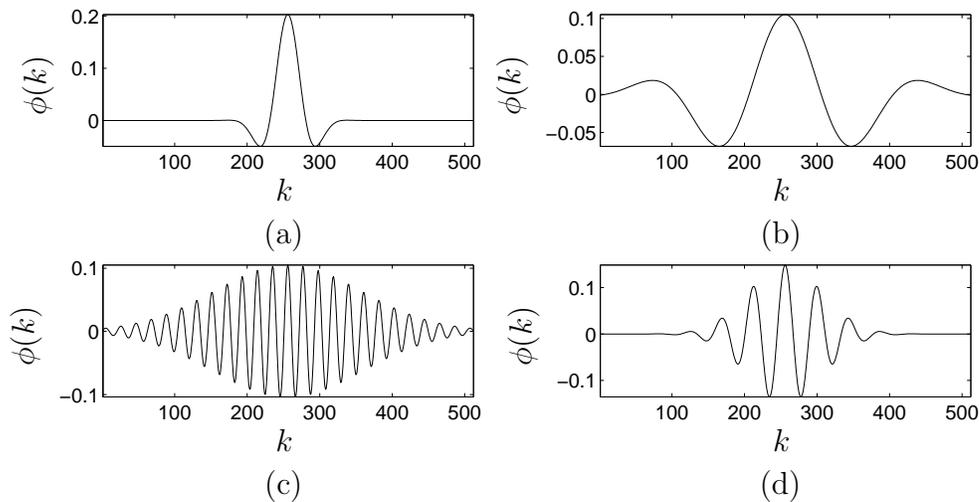


Figura 7.4: Exemplos de átomos de um dicionário com $L = 512$ e $\bar{\phi}(k) = e^{-\pi k^2}$. (a) $\vartheta_3 = 0,0101$ e $\vartheta_2 = 64$; (b) $\vartheta_3 = 0,0101$ e $\vartheta_2 = 256$; (c) $\vartheta_3 = 0,0477$ e $\vartheta_2 = 256$ e (d) $\vartheta_3 = 0,0226$ e $\vartheta_2 = 128$.

7.4 O Processo Anecoico de Mistura

Num processo de mistura anecoico, as misturas $x_j(k)$ ($1 \leq j \leq M$) podem ser expressas em função das fontes como $x_j(k) = \sum_{i=1}^N h_{ji} s_i(k - \Delta_{ji})$, onde h_{ji} e Δ_{ji} são, respectivamente, o coeficiente de atenuação e o atraso existentes no caminho da i -ésima fonte para o j -ésimo sensor. O processo de mistura pode ser condensado por uma matriz \mathbf{H} de dimensões $M \times N$, cujos elementos são os filtros $H_{ij}(z) = h_{ij} z^{-\Delta_{ij}}$, de modo que podemos escrever $\mathbf{X}(z) = \mathbf{H}(z)\mathbf{S}(z)$, com as fontes e as misturas, no domínio da transformada \mathcal{Z} , dispostas nas N linhas da matriz \mathbf{S} e nas M linhas da matriz \mathbf{X} , respectivamente.

Sabe-se que as ambiguidades de filtragem e permutação, oriundas da ausência de informações concernentes às fontes, permitem-nos considerar bem-sucedida uma separação mesmo quando as estimativas são versões filtradas e possivelmente permutadas das fontes, conforme visto no Cap. 4. Por isso, seja $\tilde{s}_i(k)$ a referência da i -ésima fonte a recuperar, a qual expressamos em função da fonte original $s_i(k)$, como $\tilde{s}_i(k) = \frac{s_i(k + \Delta_{1i})}{h_{1i}}$. Podemos modelar então o processo de mistura como $\mathbf{X}(z) = \tilde{\mathbf{H}}(z)\tilde{\mathbf{S}}(z)$, onde a matriz $\tilde{\mathbf{H}}(z)$ pode ser obtida da matriz $\mathbf{H}(z)$ fazendo-se $\tilde{H}_{ij}(z) = \frac{h_{ij}}{h_{1j} z^{-\Delta_{1j}}}$. Deste modo, doravante consideraremos, sem perda de generalidade para o problema de BSS, que a matriz $\tilde{\mathbf{H}}(z)$ de dimensões $M \times N$ que condensa o processo de mistura é:

$$\tilde{\mathbf{H}}(z) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tilde{h}_{21} z^{-\tilde{\Delta}_{21}} & \tilde{h}_{22} z^{-\tilde{\Delta}_{22}} & \ddots & \tilde{h}_{2N} z^{-\tilde{\Delta}_{2N}} \\ \vdots & \ddots & \ddots & \vdots \\ \tilde{h}_{M1} z^{-\tilde{\Delta}_{M1}} & \tilde{h}_{M2} z^{-\tilde{\Delta}_{M2}} & \cdots & \tilde{h}_{MN} z^{-\tilde{\Delta}_{MN}} \end{bmatrix}, \quad (7.6)$$

com $\tilde{\Delta}_{ij} = \Delta_{ij} - \Delta_{1j}$. A ambiguidade de filtragem nos permite entender $\tilde{s}_i(k)$ como a k -ésima amostra da i -ésima fonte. Um procedimento bastante semelhante de normalização da primeira linha da matriz de misturas costuma ser utilizado nas técnicas de análise de componentes esparsos [96]. A principal diferença por nós adotada em relação a esses métodos reside no fato de contemplarmos o caso anecoico, o que implica a necessidade de incluir os atrasos em nossa análise.

As técnicas de BSS podem efetuar a busca de sua solução no espaço de mistura (almeçando identificar o sistema de mistura) ou no espaço de separação (quando tentam otimizar parâmetros de um sistema de separação que resulte em estimativas fidedignas das fontes). Tipicamente, técnicas de análise de componentes independentes atuam no espaço de separação, enquanto que técnicas de análise de componentes esparsos se articulam para identificar (a menos das ambiguidades supracitadas) o sistema de mistura.

7.5 Emprego do MP para BSS

Como vimos, no contexto de decomposições MP a unicidade da representação é descartada, dando lugar a um procedimento iterativo que, sem prejuízo à conservação de energia, pretende encontrar uma representação esparsa de um determinado sinal. A redundância permite, à custa de um aumento da complexidade computacional, uma representação mais estruturada e atenta às idiossincrasias dos sinais. Estando os átomos do dicionário bem localizados no espaço tempo-frequência, podemos empregar a diversidade temporal-espectral normalmente existente entre as diferentes fontes e supor que um átomo que descreva razoavelmente bem a mistura está associado, na maior parte das vezes, a apenas uma das fontes. Supondo a decorrelação das fontes, esta é uma hipótese razoável [179]. Esta hipótese nos permite estimar os parâmetros do sistema de mistura, inserindo a técnica resultante sob o paradigma da análise de componentes esparsos, vista brevemente no Cap. 4.

Na etapa de reconstrução, podemos determinar qual fonte está associada a um determinado átomo empregando a diversidade espacial, já que as fontes alcançam sensores distintos em instantes diferentes e com atenuações distintas. Embora não tenhamos contemplado a etapa de reconstrução, vale notar que [98] propõe, para configurações de misturas instantâneas, o emprego de átomos multicanais para essa etapa.

7.5.1 Nova Proposta de um Algoritmo MP Multicanal para BSS

Para a resolução do problema de BSS, o emprego do algoritmo MP padrão para cada mistura de forma independente é proposto em [179], num contexto instantâneo e estéreo ($M = 2$). A razão entre os produtos internos de um mesmo átomo selecionado em ambas as misturas (ainda que em iterações distintas) provê uma estimativa pontual de um coeficiente de atenuação do sistema de mistura. Uma desvantagem dessa abordagem reside na redução do número de estimativas efetivas, já que não raro átomos selecionados em um canal não são escolhidos para descrever o outro, impossibilitando o processo de estimativa do sistema de mistura nos componentes “órfãos”, o que prejudica o processo de reconstrução das fontes. Estes inconvenientes podem ser contornados por dicionários maiores, os quais potencializam a coerência entre os átomos e as características dos sinais. De todo modo, se espera uma menor taxa de reutilização dos átomos em decomposições independentes ao longo das misturas.

Por esta razão, uma estratégia interessante é procurar os átomos considerando simultaneamente as diversas misturas a que temos acesso. A proposta que detalharemos a seguir emprega átomos multicanais e atua na etapa de estimativa dos

parâmetros do sistema de mistura; suas relações com a abordagem de [180] serão posteriormente elucidadas. O algoritmo multicanal resultante para decomposição das misturas é apresentado a seguir.

Algoritmo MP Multicanal com FFTs Proposto

Inicialização

Para $q = 1, 2, \dots, M$
 $\mathcal{R}_q^{(0)} \leftarrow \bar{\mathbf{x}}_q$

Fim

Laço Principal

Para $j = 1, 2, \dots, J$

Para $i = 1, 2, \dots, \#\mathcal{D}$

$\phi_q'^{(i)} := \text{IFFT}\{\text{FFT}\{\mathcal{R}_q^{(j-1)}\} \odot \text{FFT}\{\phi^{(i)}\}\}$

Fim

$(l_i^{(1)}, l_i^{(2)}, \dots, l_i^{(M)}) := \arg \max_{(l^{(1)}, l^{(2)}, \dots, l^{(M)})} \left\{ \sum_{q=1}^M \left| \phi_q'^{(i)}(l^{(1)}, l^{(2)}, \dots, l^{(M)}) \right|^p \right\}$

Para $q = 1, 2, \dots, M$

$\kappa_i^{(q)} := \phi_q'^{(i)}(l_i^{(q)})$

Fim

Escolha $m_j := \arg \max_u \sum_{q=1}^M \left| \kappa_u^{(q)} \right|^p$

Para $q = 1, 2, \dots, M$

$\mathcal{R}_q^{(j)} \leftarrow \mathcal{R}_q^{(j-1)} - \kappa_{m_j}^{(q)} \text{CIRCSHIFT} \left\{ \phi_{m_j}^{(q)}, l_{m_j}^{(q)} \right\}$

Fim

Fim

Almejando selecionar o mesmo átomo¹ para descrever as misturas, uma forma possível de generalização do algoritmo MP padrão, o qual maximiza $|\phi'^{(i)}(l)|$, seria a maximização de

$$\Phi_M^{(i)}(\mathbf{p}) = \sum_{q=1}^M \left| \phi_q'^{(i)}(l^{(1)}, l^{(2)}, \dots, l^{(M)}) \right|^p, \quad (7.7)$$

onde $\Phi_M^{(i)}(\mathbf{p})$ é relacionada à norma l_p das projeções do i -ésimo átomo nas misturas e o parâmetro \mathbf{p} insere uma flexibilidade no algoritmo resultante, a qual poderá ser empregada para incrementar seu desempenho. No caso de MP monocanal, maximizar $|\phi'^{(i)}(l)|$ equivale a maximizar $|\phi'^{(i)}(l)|^2$, porém em contextos multicanais essa equivalência não se sustenta, permitindo a inserção deste parâmetro adicional. Vale enfatizar que a otimização deste parâmetro será efetuada no tocante à qualidade das estimativas resultantes. Assim, uma redução elevada da taxa de redução de energia dos resíduos a cada iteração não é, nesta etapa, primordial.

¹Cabe ressaltar que o átomo selecionado pode se encontrar temporalmente deslocado entre as misturas.

A maximização de $\Phi_M^{(i)}(\mathbf{p})$ depende da escolha da norma l_p , a qual pode, por exemplo, implicar otimizações no tocante ao número de elementos não-nulos ($\mathbf{p} = 0$), à soma dos valores absolutos ($\mathbf{p} = 1$), à norma euclidiana ($\mathbf{p} = 2$) e ao maior coeficiente ($\mathbf{p} \rightarrow \infty$). Ressalte-se o fato de que as projeções nos átomos são sempre efetuadas mediante a norma l_2 (por meio de correlações temporais), porém a seleção do átomo/atraso a utilizar emprega como critério a norma l_p das projeções ao longo dos canais (ou seja, a norma l_p é calculada em termos espaciais). Enquanto a minimização da norma l_1 implica penalizar soluções pouco esparsas, a maximização desta norma tende a reforçar projeções mais equilibradas espacialmente. Já a escolha da norma l_2 tende a realçar os casos em que a projeção em um canal apresenta uma magnitude razoavelmente superior à outra.

A estratégia proposta é estreitamente relacionada à preconizada por [180], com algumas distinções importantes: o emprego de FFTs, a generalização para M misturas (sem a restrição $M = 2$), o emprego de técnicas aplicáveis às situações não-instantâneas (embora o arcabouço teórico de [180] contemple a decomposição de misturas anecoicas, suas simulações abordam apenas misturas instantâneas, não indicando como estimar os atrasos), a flexibilidade advinda da escolha do parâmetro \mathbf{p} e o estudo de características da decomposição e dos átomos que impactam a qualidade da estimativa, conforme veremos posteriormente.

7.5.2 Estimativa Conjunta dos Coeficientes de Atenuação e dos Atrasos

A técnica MP multicanal proposta pode ser empregada para estimar simultaneamente os coeficientes de atenuação e de atraso, caso os átomos apresentem um comprimento não muito reduzido. O objetivo do uso do MP multicanal justifica-se por esperarmos que, caso o termo $\Phi_M^{(i)}(\mathbf{p})$ seja elevado para um determinado átomo, este átomo esteja majoritariamente associado a uma das fontes. Se $\phi^{(m_j)}$ é o átomo selecionado na iteração, podemos aproximar a primeira mistura e a q -ésima mistura por meio de:

$$\begin{aligned}\bar{\mathbf{x}}_1 &\approx \kappa_{m_j}^{(1)} \text{CIRCSHIFT} \left\{ \ddot{\phi}^{(m_j)}, l_{m_j}^{(1)} \right\}, \\ \bar{\mathbf{x}}_q &\approx \kappa_{m_j}^{(q)} \text{CIRCSHIFT} \left\{ \ddot{\phi}^{(m_j)}, l_{m_j}^{(q)} \right\}.\end{aligned}\tag{7.8}$$

Se o átomo predominantemente descreve a i -ésima fonte, temos que $l_{m_j}^{(q)} - l_{m_j}^{(1)} \approx \tilde{\Delta}_{qi}$ e, portanto:

$$\gamma_h^{(q)} = \frac{\kappa_{m_j}^{(q)}}{\kappa_{m_j}^{(1)}} \approx \tilde{h}_{qi}.\tag{7.9}$$

No momento, pretendemos apresentar uma técnica de estimativa conjunta dos coeficientes de atenuação e dos atrasos de um sistema de mistura, por meio da criação de um histograma bidimensional, cuja construção deve também depender dos deslocamentos dos átomos ao longo das misturas. Seja a diferença de deslocamento (circular) entre a q -ésima e a primeira mistura do átomo selecionado, definidas por:

$$\gamma_{\Delta}^{(q)} = l_{m_j}^{(q)} - l_{m_j}^{(1)}. \quad (7.10)$$

O histograma bidimensional² apresenta o seguinte formato, para um coeficiente de atenuação candidato \hat{h}_{2i}^c e um atraso candidato $\hat{\Delta}_{2i}^c$:

$$\Phi_b \left(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c \right) = \sum_{n=1}^z \Theta_b \left(\gamma_h^{(2)}(n), \gamma_{\Delta}^{(2)}(n), \hat{h}_{2i}^c, \hat{\Delta}_{2i}^c \right). \quad (7.11)$$

Uma escolha que se revelou interessante para a construção de um histograma fundamenta-se em *kernels* gaussianos. Assim:

$$\Theta_b \left(\gamma_h^{(2)}(n), \gamma_{\Delta}^{(2)}(n), \hat{h}_{2i}^c, \hat{\Delta}_{2i}^c \right) = \frac{1}{\sigma_{\Delta} \sigma_h} e^{-\frac{(\gamma_h^{(2)}(n) - \hat{h}_{2i}^c)^2}{2\sigma_{\Delta}^2} - \frac{(\gamma_{\Delta}^{(2)}(n) - \hat{\Delta}_{2i}^c)^2}{2\sigma_h^2}}, \quad (7.12)$$

onde σ_{Δ} e σ_h são os desvios-padrão dos *kernels* dos atrasos e dos coeficientes de atenuação, respectivamente. Os atrasos e os coeficientes de atenuação estimados são os associados aos picos da função $\Phi_b \left(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c \right)$.

Dada a alta complexidade de cada iteração MP, estudamos nesta tese a influência que diversos parâmetros da decomposição MP apresentam sobre a robustez da estimativa resultante. Os impactos dos seguintes fatores serão avaliados: o comprimento do átomo, o número da iteração MP, a escala ϑ_2 do átomo, o valor de p , a frequência ϑ_3 do átomo, e a interferência de cada componente, conceito este que será a seguir explicado.

Sendo uma estratégia subótima, alguns dos átomos selecionados pelo MP podem não contribuir para uma representação dotada de significado. Esta degradação de eficiência, a qual também pode ocorrer por descasamentos entre o sinal a decompor e o dicionário, acaba por aumentar o número de átomos necessários para modelar um sinal para um dado erro. Não raro alguns átomos são selecionados apenas para corrigir os erros introduzidos por uma seleção progressiva equivocada. Por essa razão, os artigos [181],[182] apresentam o cálculo da quantidade de interferência, seja do tipo construtiva (quando o componente basicamente modela o sinal) ou destrutiva (quando o componente tem por principal função corrigir os erros progressivos de modelagem).

²A partir de agora, simplificaremos para o caso $M = 2$.

O conceito de interferência baseia-se na diferença de energia provocada pela inserção de um novo átomo no modelo; se ele remove partes de outros componentes ou se é (quase) ortogonal a estes componentes [182]. Esta consideração motiva a seguinte definição: dois vetores \mathbf{u} e \mathbf{v} interferem construtivamente entre si se $\|\mathbf{u} + \mathbf{v}\|^2 > \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, interferindo-se destrutivamente se $\|\mathbf{u} + \mathbf{v}\|^2 < \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, o que pode ser visualizado na Fig. 7.5. Dessa forma a diferença

$$\|\mathbf{u} + \mathbf{v}\|^2 - (\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) \quad (7.13)$$

quantifica o montante de interferência mútua entre \mathbf{u} e \mathbf{v} , com o sinal desta diferença identificando o tipo de interferência; caso este seja positivo, a interferência é construtiva; caso seja negativo, a interferência é destrutiva³. Assim, [182] define a interferência do n -ésimo átomo selecionado numa representação de J átomos como:

$$\begin{aligned} \mathcal{I}(n) = & \frac{1}{2} \left[\left\| \sum_{j=1}^J \langle \mathcal{R}^{(j-1)}, \phi^{(m_j)} \rangle \phi^{(m_j)} \right\|^2 \right. \\ & \left. - \left(\left\| \sum_{j=1, j \neq n}^J \langle \mathcal{R}^{(j-1)}, \phi^{(m_j)} \rangle \phi^{(m_j)} \right\|^2 + \left\| \langle \mathcal{R}^{(n-1)}, \phi^{(m_n)} \rangle \phi^{(m_n)} \right\|^2 \right) \right]. \end{aligned} \quad (7.14)$$

7.5.3 Avaliação de Consistência das Estimativas

Esta seção pretende analisar as estimativas conjuntas dos coeficientes de atenuação e dos atrasos pelo método MP multicanal proposto, avaliando o impacto de alguns parâmetros da decomposição na qualidade destas estimativas. Os testes efetuados utilizaram sinais de voz oriundos da base de dados *TSP Speech Database*⁴, amostrados a 8 kHz. Como os sinais apresentam duração curta (em torno de 2s), concatenamos os sinais de um mesmo locutor, de modo a tornar a duração dos registros compatível com a necessária para os experimentos; diversos testes foram efetuados, com os locutores (de um total de 20, entre 10 masculinos e 10 femininos) sendo escolhidos aleatoriamente em cada teste.

Foram testadas 10 diferentes configurações de mistura do tipo 2×2 (2 fontes e 2 misturas), num total de 10 configurações. Para cada configuração, 4 combinações de comprimento L dos átomos e \mathbf{p} foram testadas: $(L, \mathbf{p}) \in \{(256, 1), (256, 2), (512, 1), (512, 2)\}$, gerando um total de 40 testes. A duração das misturas é de 19,2s, gerando um total de 300 blocos (sem superposição) para $L = 512$

³Vide Fig. 7.6 para um exemplo ilustrativo de interferência destrutiva. [182] demonstra que o MP, quando o número de iterações tende ao infinito, gera um montante de interferência construtiva idêntico ao de destrutiva.

⁴<http://www-mmsep.ece.mcgill.ca/Documents/Data/index.html>.

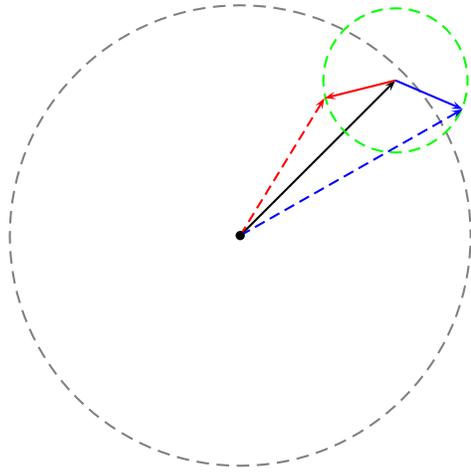


Figura 7.5: Exemplo gráfico do conceito de interferência. Ao vetor preto adicionamos dois vetores (o azul e o vermelho) de normas equivalentes ao raio do círculo verde. Se o vetor resultante for externo ao círculo cinza (caso do vetor azul), a soma é construtiva; sendo interno, a soma é destrutiva (caso do vetor vermelho).

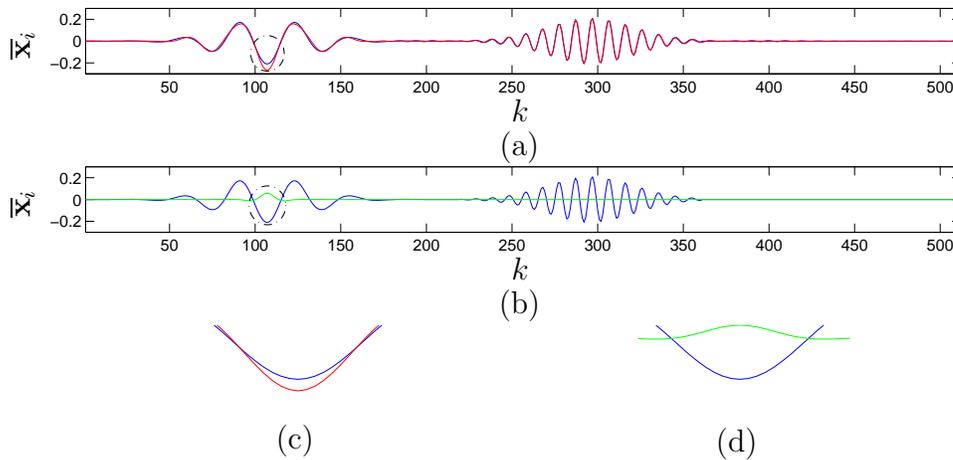


Figura 7.6: Exemplo real de decomposição MP com o surgimento de componentes que não modelam o sinal original, cujos surgimentos se devem apenas à redução de erros progressivos de modelagem. (a) sinal original (em azul) e a aproximação obtida com os 6 primeiros componentes (em vermelho); (b) sinal original (em azul) e o sétimo componente (em verde) extraído via MP, o qual, por não ter relação com o sinal original, apresenta interferência destrutiva; (c) ampliação do trecho assinalado em (a); (d) ampliação do trecho assinalado em (b).

e 600 blocos para $L = 256$. O número de iterações MP por bloco foi arbitrado em 10. Os atrasos da contribuição de cada fonte para a segunda mistura com relação à contribuição para a primeira mistura foram inteiros escolhidos aleatoriamente no intervalo $[-10, 10]$, com os coeficientes de atenuação para a primeira mistura arbitrados em 1 e os da segunda mistura reais e aleatórios no intervalo $[-5, 5]$.

Os átomos foram criados mediante o emprego de $\bar{\phi}(k) = e^{-\pi k}$, com $\vartheta_1 = L/2$, $\vartheta_2 = 2^j$, com $j \in \{1, 2, \dots, \log_2 L\}$, com ϑ_3 assumindo 200 valores uniformemente distribuídos no intervalo $[0, 1/2]$ e $\vartheta_4 \in \{0, \pi/6, 2\pi/6, \dots, 11\pi/6\}$.

A construção do histograma bidimensional (via função Φ_b , apresentada na Eq. (7.11)) se mostrou bastante robusta, permitindo uma ampla faixa de escolha dos parâmetros σ_Δ e σ_h . As Figs. 7.7-7.10 apresentam alguns dos exemplos gerados, nos quais o pico do histograma (sempre bem definido) coincide com os valores dos coeficientes de atenuação do sistema de mistura, num comportamento que se mostrou estável ao longo de todos os testes⁵.

O elevado custo computacional do MP motiva o estudo da influência de alguns

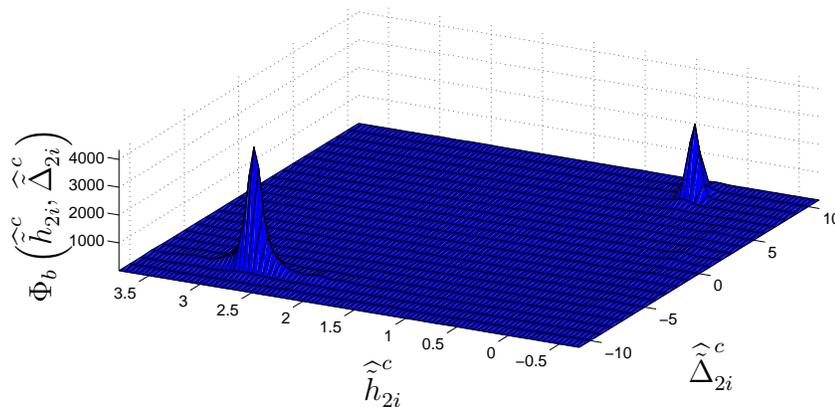


Figura 7.7: Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $\mathbf{p} = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = 0,30798z^9$ e $\tilde{H}_{22}(z) = 2,7917z^{-8}$.

parâmetros da decomposição na qualidade da estimativa pois, se pudermos aferir a qualidade das estimativas \hat{h}_{qi} ou $\hat{\Delta}_{qi}$, menos iterações podem ser necessárias para que o processo de estimação seja bem-sucedido.

Algumas métricas de avaliação foram testadas, como por exemplo a entropia da variável aleatória bidimensional $(\gamma_h^{(2)}(n) - \tilde{h}_{2i_n}, \gamma_\Delta^{(2)}(n) - \tilde{\Delta}_{2i_n})$, onde i_n foi selecionado de modo a reduzir a distância entre a estimativa pontual e os valores reais de atenuação e de atraso; em outras palavras, associa-se a cada estimativa pontual o

⁵Cabe ressaltar que esse comportamento desejável somente se revela, em nossas simulações, quando o tamanho dos átomos é maior ou igual a 256. Por isso, os testes apresentados não empregam átomos com comprimentos inferiores a este número.

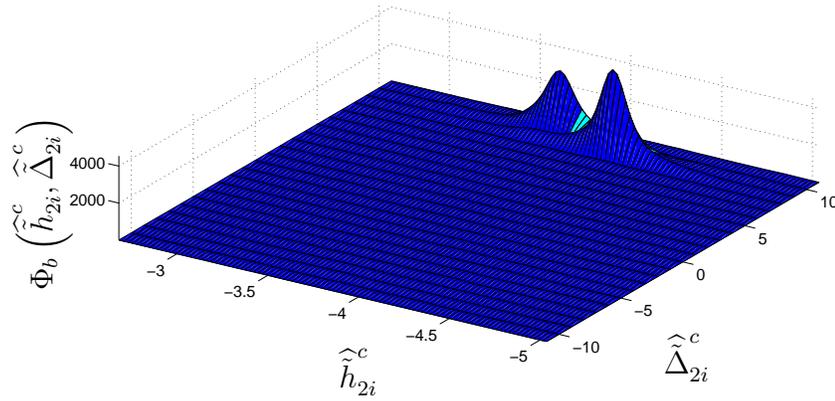


Figura 7.8: Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = -4,0355z^9$ e $\tilde{H}_{22}(z) = -3,6803z^{10}$.

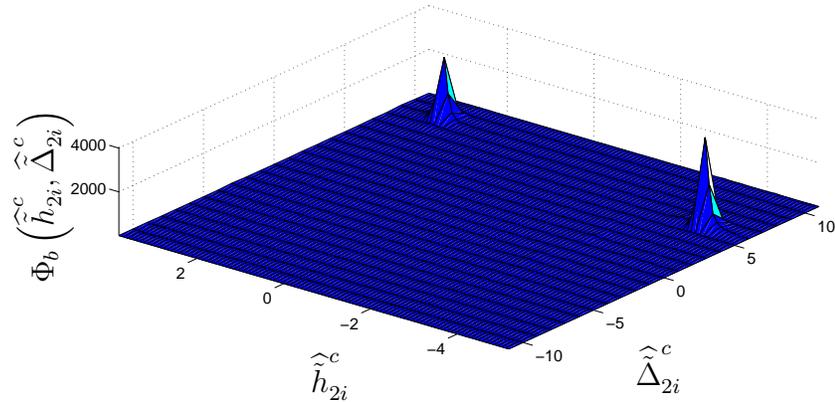


Figura 7.9: Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = 2,8023z^9$ e $\tilde{H}_{22}(z) = -4,1887z^6$.

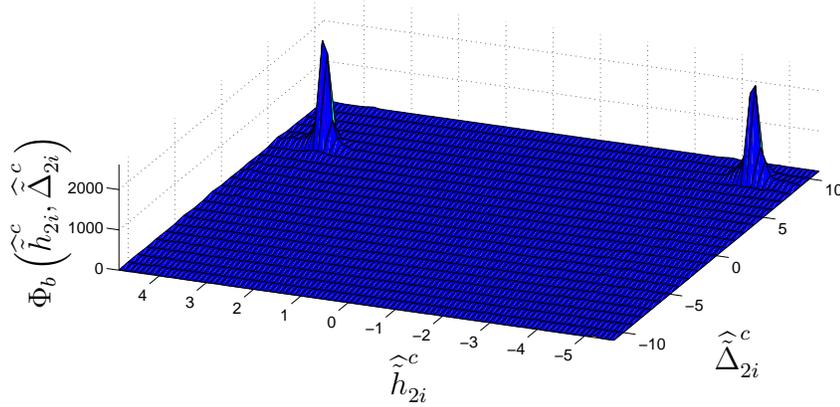


Figura 7.10: Histograma construído a partir do cálculo de $\Phi_b(\hat{h}_{2i}^c, \hat{\Delta}_{2i}^c)$ para $L = 512$ e $p = 2$, com $\sigma_h = 0,05$ e $\sigma_\Delta = 0,3$. As funções de transferência a identificar são $\tilde{H}_{21}(z) = -4,6226z^9$ e $\tilde{H}_{22}(z) = 3,8517z^6$.

valor correto mais próximo, subtraindo-se ambos. Se todas as estimativas pontuais coincidirem com algum valor real do sistema de mistura, a entropia da variável aleatória $(\gamma_h^{(2)}(n) - \tilde{h}_{2i_n}, \gamma_\Delta^{(2)}(n) - \tilde{\Delta}_{2i_n})$ é nula.

Porém a medida de entropia apresentou grande variabilidade entre diferentes testes, dificultando uma análise comparativa. Um critério mais simples é a variância da variável aleatória $\Gamma_h(n)$, definida por:

$$\Gamma_h(n) = \gamma_h^{(2)}(n) - \tilde{h}_{2i_n}, \quad (7.15)$$

onde $i_n = \arg \min_{i_n} |\gamma_h^{(2)}(n) - \tilde{h}_{2i_n}|$. Sendo σ_Γ^2 a variância de $\Gamma_h(n)$, a redução de sua magnitude reflete uma melhor estimativa dos parâmetros do sistema de mistura. Doravante, esta métrica será utilizada para avaliar o impacto de algumas propriedades da decomposição.

A Tabela 7.1 apresenta o valor de σ_Γ^2 para as quatro combinações testadas de p e L , com 10 iterações por bloco, ao longo das 10 configurações simuladas. Os resultados revelam que átomos maiores implicam melhores estimativas, o que já é de certo modo esperado, porque átomos maiores aumentam a liberdade da modelagem do MP, permitindo uma aderência mais acurada com relação à estrutura temporal das fontes. Outra constatação, menos óbvia, é que a escolha $p = 1$ gera estimativas melhores do que $p = 2$. A adoção da norma l_2 , embora possa ser ótima em termos de redução do erro quadrático médio da reconstrução, se revelou menos interessante na etapa (anterior à reconstrução das fontes) de estimativa dos parâmetros do sistema de mistura. Como o emprego de $p = 1$ não implica aumento de custo computacional, este resultado indica que esta escolha é superior ao procedimento de se arbitrar $p = 2$.

Tabela 7.1: Valores de σ_F^2 para diferentes combinações de p e L (média das 10 configurações testadas).

	$L = 256$	$L = 512$
$p = 1$	3,72	3,17
$p = 2$	4,01	3,39

A Tab. 7.2 apresenta o percentual de componentes selecionados ao longo das 10 configurações, em função de L , p , número da iteração e dos fatores de escala. Esta tabela permite-nos concluir que a quantidade de componentes selecionados em iterações MP aumenta com o crescimento da escala, com a diferença sendo gradativamente reduzida à medida que se aumenta o número da iteração. Arbitrar $p = 2$ tem por efeito privilegiar levemente componentes de escalas maiores. Já a Tab. 7.3 apresenta o percentual de componentes selecionados ao longo das 10 configurações, em função de L , p , número da iteração e de intervalos de frequência, mostrando que componentes de frequências mais baixas predominam, com a assimetria se diluindo um pouco com o aumento do número de iterações. Esta tabela também ressalta um leve favorecimento dos componentes no intervalo de frequências mais baixas quando $p = 1$.

Para a escala dos átomos selecionados, os resultados constantes na Tab. 7.4 indicam que os átomos de pequena escala (em especial $\log_2(\vartheta_2) = 1$) apresentam estimativas de pior qualidade. Segundo a Tab. 7.2, componentes da escala mais baixa são raramente selecionados, principalmente nas primeiras iterações que, como veremos, são as mais confiáveis. Isso indica que provavelmente a eliminação dos átomos da escala mais baixa não implicaria uma significativa redução da acurácia das estimativas, tendo como vantagem a redução do custo computacional.

Como o número de frequências ϑ_3 testadas é muito grande (200 valores distintos), não é possível angariar estatísticas robustas para cada frequência. Por isso, efetuamos uma divisão das frequências em cinco intervalos. Os resultados constantes na Tab. 7.5 revelam que os átomos mais confiáveis são os de frequência mais elevada, em especial para $\vartheta_3 \geq 0,3$. Com as frequências se dá o inverso do que ocorre com o fator de escala: em geral, os componentes selecionados apresentam frequência baixa, e o grupo menos numeroso (frequências altas) é justamente o que fornece as melhores estimativas. Esta conclusão pode significar que a exclusão dos átomos de frequência mais baixa pode acarretar grande redução do número global de iterações MP, já que evitaria muitas decomposições que geram estimativas de pior qualidade. Interessantemente, a escolha $p = 1$, ainda que privilegie a escolha de componentes de frequências mais baixas (menos confiáveis) é superior à escolha $p = 2$; uma possível razão para este fato reside na sua preferência por componentes de escalas mais elevadas, em especial as escalas “médias” ($\log_2(\vartheta_2) = 3$ ou 4), que

Tabela 7.2: Percentual de componentes selecionados em função de L , p , número da iteração e de $\log_2(\vartheta_2)$.

$L = 256$ e $p = 1$											
$\log_2(\vartheta_2)$ /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
1	0,53	0,68	0,86	0,98	1,31	1,71	1,73	2,03	2,45	2,51	1,48
2	1,25	1,91	2,98	4,03	4,5	5,06	5,93	6,33	6,68	6,43	4,51
3	3,5	5,23	7,23	7,68	9,43	9,7	10,41	10,75	10,53	12,06	8,65
4	7,35	9,15	9,93	12,21	11,91	12,66	13,45	13,85	13,91	13,63	11,8
5	9,55	10,25	12,01	12,38	12,91	14,03	14,28	14,33	14,03	15,55	12,93
6	7,41	8,38	11,55	12,46	13,85	14,05	13,83	13,21	15,53	14,88	12,51
7	13,25	13,48	13,63	13,55	14,11	13,2	12,48	12,33	11,85	10,61	12,85
8	57,15	50,9	41,78	36,68	31,95	29,56	27,86	27,15	25	24,3	35,23
$L = 256$ e $p = 2$											
$\log_2(\vartheta_2)$ /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
1	0,48	0,73	0,76	1,21	1,03	1,63	1,76	2,1	2,14	2,66	1,45
2	1,23	2	2,9	3,7	4,21	4,61	5,53	5,88	6,08	5,73	4,19
3	3,43	5,09	7,11	7,55	8,46	9,6	9,55	10,11	9,71	10,76	8,14
4	7,04	8,75	9,33	11,01	11,56	11,83	12,43	13,1	13,05	13,15	11,12
5	9,48	10,03	11,65	12,51	12,7	13,55	13,9	13,7	14,18	15,1	12,68
6	7,51	8,31	11,68	12,66	13,9	13,95	14,11	13,85	15,55	14,48	12,6
7	13,46	13,56	14,24	13,85	14,45	14	13,45	13,06	12,4	11,83	13,43
8	57,33	51,5	42,3	37,48	33,66	30,81	29,25	28,18	26,86	26,26	36,36
$L = 512$ e $p = 1$											
$\log_2(\vartheta_2)$ /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
1	0,23	0,43	0,7	0,86	1	0,9	0,7	0,96	1,4	1,3	0,85
2	1,06	1,43	1,7	2,13	3	3,26	3,96	4,23	4,06	5,06	2,99
3	3,03	3,63	5,36	6,33	6,76	7,56	7,2	8,33	9,03	8,86	6,61
4	4,66	7,53	8,03	9,43	9,66	10,23	11,86	11,06	11,13	11,6	9,52
5	7,8	8,43	10,86	10,33	11,1	11,06	12,63	11,8	12,73	13,03	10,98
6	4,56	6,53	8,66	9,86	10,66	11,2	12,2	13,83	13,06	14,13	10,47
7	19,96	21,2	20,6	22,76	24,96	23,1	22,8	22,5	22,63	22,2	22,27
8	29,76	25,96	22,63	19,06	16,63	16,63	14,66	13,9	13,16	11,33	18,37
9	28,9	24,83	21,43	19,2	16,2	16,03	13,96	13,36	12,76	12,46	17,91
$L = 512$ e $p = 2$											
$\log_2(\vartheta_2)$ /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
1	0,26	0,3	0,76	1,06	0,8	0,7	0,86	0,96	0,86	0,93	0,75
2	1,1	1,3	1,86	2,23	2,53	3,2	3,36	3,86	4,4	4,03	2,79
3	2,76	3,86	4,76	5,66	6,83	6,63	7,16	7,03	8,26	8,59	6,16
4	4,46	6,7	7,46	8,16	8,96	10,36	9,5	10,16	10,23	10,23	8,62
5	7,6	8,43	10,6	10,46	10,56	10,93	12,43	11,56	12,96	13,73	10,93
6	4,6	6,26	8,76	9,73	11,26	11,16	12,83	13,86	12,76	14,16	10,54
7	19,7	21,4	20,96	22,93	24,93	22,8	23,46	23,73	23,6	22,93	22,64
8	30,1	26,73	22,86	19,96	17,26	17,5	15,46	14,63	14,43	11,76	19,07
9	29,4	25	21,93	19,76	16,83	16,7	14,9	14,16	12,46	13,6	18,47

Tabela 7.3: Percentual de componentes selecionados em função de L , p , número da iteração e de intervalos de ϑ_3 .

$L = 256$ e $p = 1$											
Intervalo de ϑ_3 /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
[0, 1/10)	82,4	76,1	68,48	62,93	58,1	55,48	52,23	49,55	48,13	46,18	59,96
[1/10, 2/10)	9,55	12,83	17,38	20,34	21,46	22,78	24,31	25,03	26,5	25,83	20,6
[2/10, 3/10)	2,91	4,68	6,55	8,16	9,85	10,85	11,46	12,21	12,23	13,61	9,25
[3/10, 4/10)	1,95	3,03	3,8	4,85	6,35	6,63	7,11	8,21	8,08	8,68	5,87
[4/10, 1/2]	3,18	3,35	3,78	3,7	4,23	4,25	4,86	4,98	5,05	5,68	4,3
$L = 256$ e $p = 2$											
Intervalo de ϑ_3 /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
[0, 1/10)	82,36	75,98	67,95	62,7	57,5	54,9	51,68	49,1	47,2	44,98	59,43
[1/10, 2/10)	9,6	12,71	17,73	20,25	21,86	22,58	24,73	24,95	26,35	26,36	20,71
[2/10, 3/10)	2,93	4,81	6,6	8,36	10,13	11,08	11,45	12,36	12,41	14,18	9,43
[3/10, 4/10)	1,91	3,01	3,91	4,96	6,18	7,08	7,46	8,16	8,43	9,06	6,02
[4/10, 1/2]	3,18	3,46	3,8	3,71	4,31	4,34	4,66	5,41	5,6	5,4	4,39
$L = 512$ e $p = 1$											
Intervalo de ϑ_3 /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
[0, 1/10)	86,53	81,33	76,2	71,86	67,43	64,66	60,83	59,93	57,86	54,36	68,1
[1/10, 2/10)	8	11,16	14,43	16,6	18,4	19,53	22,46	21,5	22,13	23,93	17,81
[2/10, 3/10)	1,9	2,73	4,03	5,6	7,06	7,7	8,13	8,96	9,3	10,93	6,63
[3/10, 4/10)	1,03	2,1	2,46	3	4,43	4,56	5,06	5,83	6,53	6,36	4,14
[4/10, 1/2]	2,53	2,66	2,86	2,93	2,66	3,53	3,5	3,76	4,16	4,4	3,3
$L = 512$ e $p = 2$											
Intervalo de ϑ_3 /Número da Iteração	1	2	3	4	5	6	7	8	9	10	Total
[0, 1/10)	86,43	81,16	75,56	71,5	67,3	65,1	59,63	59,26	55,86	54,63	67,64
[1/10, 2/10)	8	11,1	15,06	17	18,46	19,53	22,6	21,56	23,46	23,83	18,06
[2/10, 3/10)	2	2,7	4,2	5,4	7,2	7,76	8,26	9,26	10	10,63	6,74
[3/10, 4/10)	0,96	2,23	2,36	3,26	4,29	4,4	5,56	5,86	6,33	6,76	4,2
[4/10, 1/2]	2,6	2,8	2,8	2,83	2,73	3,2	3,93	4,03	4,33	4,13	3,34

Tabela 7.4: Valor de σ_F^2 em função da escala $\log_2(\vartheta_2)$ (média das configurações testadas).

$\log_2(\vartheta_2)$	$L = 256$	$L = 512$
1	7,41	5,75
2	4,49	3,35
3	3,19	2,43
4	3,41	2,99
5	3,89	3,26
6	4,27	3,77
7	4,07	3,27
8	3,74	3,33
9	-	3,35

Tabela 7.5: Valor de σ_F^2 em função da frequência ϑ_3 (média das configurações testadas).

Intervalo de ϑ_3	$L = 256$	$L = 512$
$[0, 1/10)$	4,04	3,55
$[1/10, 2/10)$	3,58	2,83
$[2/10, 3/10)$	3,76	2,54
$[3/10, 4/10)$	3,49	2,27
$[4/10, 1/2]$	3,15	2,45

são, segundo a Tab. 7.4, mais confiáveis.

A avaliação do impacto da interferência exige contornar um problema: o sinal da interferência pode se alterar ao longo das misturas. Contornamos este problema assumindo que o átomo apresenta interferência construtiva global se sua interferência é construtiva em todas as misturas. Caso comparemos as estimativas oriundas dos átomos que apresentam interferência construtiva global com os que não a apresentam, o valor de σ_F^2 é superior nos átomos que não apresentam esta propriedade. O valor de⁶ σ_F^2 neste último caso é de 3,93, enquanto que para o primeiro caso é de 3,18. Isso significa que componentes que se destinam a consertar erros de modelagem das iterações anteriores engendram estimativas menos precisas.

Por fim, consideremos o efeito do número da iteração para a qualidade da estimativa. A Tab. 7.6 revela uma tendência de a qualidade da estimativa piorar com o aumento do número da iteração. Este resultado confirma o esperado teoricamente, já que os primeiros componentes são mais representativos que os últimos [183]. Isso significa que aumentar o número de iterações pode não resultar em estimativas finais mais robustas.

⁶Os valores de σ_F^2 fornecidos são médias ao longo de todos os experimentos.

Tabela 7.6: Valor de σ_{Γ}^2 em função do número da iteração que extraiu o componente (média das configurações testadas).

Número da Iteração	σ_{Γ}^2
1	0,68
2	0,72
3	0,74
4	0,76
5	0,77
6	0,86
7	0,83
8	0,85
9	0,82
10	0,91

7.5.4 Comparação com o DUET

Para comparar o desempenho da técnica MP proposta, utilizamos o DUET (do inglês *Degenerate Unmixing Estimation Technique*) [8], provavelmente o algoritmo mais popular de separação de sinais de voz em contextos anecoicos e subdeterminados. O DUET emprega a disjunção, no domínio da STFT, entre diferentes sinais de voz para, por meio de um mascaramento binário, recuperar cada fonte em separado. O DUET também cria um histograma bidimensional, cujos picos estão associados a diferentes fontes e a diferentes pares atenuação \times atraso. Porém, o DUET apresenta duas restrições fortes: os atrasos devem ser pequenos (difícilmente ultrapassando duas amostras) e os coeficientes de atenuação devem ser positivos. Por outro lado, o DUET apresenta a vantagem de estimar atrasos não inteiros, o que seria obtido por nossa técnica apenas com um grande aumento do custo computacional (por meio de interpolação das misturas).

Por questão de justiça, simulamos 40 novas configurações aleatórias de misturas, cada qual com 4 fontes e 2 misturas, nas quais os atrasos (inteiros) se restringem ao intervalo $[-1,1]$ e os coeficientes de atenuação (reais) encontram-se no intervalo $[0,5]$. As demais características da simulação, como a base de dados, são idênticas às da configuração anterior. Como o DUET estima atrasos reais, as estimativas dos mesmos foram aproximadas para os inteiros mais próximos.

Para estimar os parâmetros do sistema de mistura, testamos a técnica proposta com $L \in \{256, 512\}$ e $p \in \{1, 2\}$. A Tab. 7.7, que exhibe a taxa de acerto dos atrasos, de um total de 160 (40 misturas \times 4 fontes por mistura), revela que a técnica de separação cega que recorre ao MP proposta apresenta uma excelente recuperação dos atrasos, com 100% de acerto.

Para as estimativas dos coeficientes de atenuação, elegemos o MSE como métrica de aferição de qualidade. A Tab. 7.8 apresenta o MSE para as 40 misturas do DUET

Tabela 7.7: Taxa de acerto dos atrasos ao longo das 40 misturas efetuadas.

Técnica	Taxa de acerto (%)
DUET	89,38
MP ($L = 256$, $p = 1$)	100
MP ($L = 512$, $p = 1$)	100
MP ($L = 256$, $p = 2$)	100
MP ($L = 512$, $p = 2$)	100

Tabela 7.8: MSE das estimativas dos coeficientes de atenuação, ao longo das 40 misturas.

Técnica	MSE
DUET	0,537
MP ($L = 256$, $p = 1$)	0,1164
MP ($L = 512$, $p = 1$)	0,3326
MP ($L = 256$, $p = 2$)	0,1296
MP ($L = 512$, $p = 2$)	0,5203

e da técnica proposta. Esta tabela nos permite concluir que a técnica proposta apresenta melhor qualidade nas estimativas dos coeficientes de atenuação, cabendo ressaltar o excelente desempenho da escolha $p = 1$, o que é compatível com nossa análise pregressa acerca da norma l_1 . Por outro lado, observamos uma degradação de desempenho com átomos maiores, possivelmente relacionada à limitada faixa dos atrasos envolvidos (de no máximo uma amostra).

7.6 Aplicação da Aprendizagem de Dicionários a Técnicas Semi-cegas de Separação

Embora constituam uma escolha popular, dicionários paramétricos (como o dicionário empregado nas simulações anteriores) são oriundos de formulações analíticas artificiais que frequentemente não garantem uma boa modelagem (ou uma decomposição eficiente) dos sinais sob análise. Ainda que tal ineficiência intrínseca de modelagem possa em parte ser contornada por meio do aumento da cardinalidade do dicionário, o aumento do custo computacional e uma possível perda de qualidade nas estimativas dos parâmetros do sistema de mistura são possíveis consequências do emprego deste tipo de dicionários.

Conquanto tradicionalmente as pesquisas em representações esparsas tenham se concentrado na decomposição de sinais com relação a um dado dicionário, recentemente o projeto de dicionários adaptados aos sinais (área também conhecida como *aprendizagem de dicionários*) tem sido uma preocupação relevante na literatura

[184],[185]. O emprego de dicionários adaptados é motivado pelo fato de que uma compreensão mais adequada das causas subjacentes ao fenômeno físico capturado pelos sensores, ou dos subespaços onde os dados se encontram, é crucial para a extração da informação relevante que está contida em dados multidimensionais [167].

Se no problema de separação de fontes temos previamente acesso a trechos isolados das fontes que não constam na mistura a separar, podemos empregar o conhecimento oriundo destes trechos para modelar cada fonte, de modo a facilitar o processo de separação. Este procedimento, por tornar semi-cego o problema de separação, é menos geral do que as técnicas cegas. Todavia, apresenta maior capacidade de separação, o que justifica o contínuo interesse por ele na comunidade científica [79],[186].

Uma modelagem possível consiste em adaptar um dicionário para cada fonte em particular. Embora a adaptação de dicionários em geral seja feita no domínio do tempo, não há dificuldade técnica em estendê-la para o domínio da frequência. A principal vantagem do uso de dicionários no domínio da frequência consiste na possibilidade de se trabalhar com representações redundantes dos blocos de sinais a decompor (por exemplo, utilizando-se diversos quadros adjacentes e com sobreposição, no domínio da STFT).

Esta seção investiga se o emprego de dicionários adaptados permite obter um desempenho superior na etapa de identificação do sistema de mistura em contextos de misturas instantâneas. Esse desempenho superior poderia se traduzir num menor número de decomposições MP, capaz de reduzir o custo computacional do processo de identificação.

Métodos de aprendizagem de dicionários comumente alternam diferentes estratégias de otimização, guardando profundas relações com técnicas de clusterização [187],[188], particularmente a quantização vetorial (VQ, do inglês *Vector Quantization*) do algoritmo generalizado de Lloyd [189] (AGL, também conhecido como *K-means*). Cada iteração deste algoritmo pode ser dividida em duas etapas de otimização. Primeiramente, a classificação ótima (correspondente ao melhor vetor no livro de códigos que representa cada vetor do conjunto de treinamento) é encontrada, utilizando um determinado livro de códigos. Posteriormente, o livro de códigos é atualizado por meio da classificação já existente e do conjunto de treinamento. Este novo livro de códigos é garantidamente superior ao progresso, de modo que o AGL caminha continuamente rumo a um mínimo (possivelmente local) [187].

No AGL, cada vetor do conjunto de treinamento é associado a um vetor do livro de códigos. No contexto de aprendizado de dicionários, o dicionário é o livro de códigos e cada vetor deste livro é um átomo. A estratégia mais popular de aprendizado de dicionário alterna de modo iterativo a etapa de codificação esparsa (a partir do dicionário atual) dos exemplos constantes na base de dados e a etapa de atualizar o

dicionário de modo a representar mais fielmente os dados.

Assim, a representação (extremamente esparsa) admitida pelo AGL consiste na seleção de apenas um átomo do dicionário (o mais próximo do sinal), sem contudo escalá-lo pelo produto interno entre ambos, como faz o MP⁷. Como as técnicas de aprendizado de dicionário associam a cada sinal uma soma ponderada de átomos, podemos encará-las como generalizações do problema de clusterização [190].

7.6.1 Algoritmo K -SVD

Seja $\mathbf{D}_{L \times \#D}$ a matriz que contém, em suas colunas, os átomos (de comprimento L) do dicionário \mathcal{D} . A partir de um conjunto $\{\bar{\mathbf{x}}_i\}_{i=1}^Z$ de Z exemplos de sinais de comprimento L , cabe ao algoritmo de aprendizado de dicionário buscar uma matriz \mathbf{D} de modo a maximizar a esparsidade da representação. Em termos de otimização, tal tarefa pode ser expressa, idealmente, como [191]:

$$\arg \min_{\mathbf{D}, \mathbf{A}} \mathcal{J}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2, \text{ sujeito a } \|\mathbf{a}_i\|_0 \leq k_0, \forall 1 \leq i \leq Z, \quad (7.16)$$

onde \mathbf{a}_i é a i -ésima coluna de \mathbf{A} , $\|\cdot\|_F$ é a norma de Frobenius associada a uma métrica de distância entre \mathbf{X} e $\mathbf{D}\mathbf{A}$, e $k_0 \ll L < \#D$.

Uma das técnicas associada ao estado da arte em aprendizado de dicionários é o K -SVD [190], cujo nome advém da conjunção da estratégia do algoritmo K -means com a fatoração por decomposição em valores singulares (SVD, do inglês *Singular Value Decomposition*). Esta técnica será utilizada neste capítulo para adaptação de dicionários, sendo portanto resumidamente descrita a seguir.

A primeira etapa da técnica K -SVD é o estágio de codificação. Sendo esta etapa empreendida a partir de um dicionário fixo, o problema de otimização (7.16) se transforma em:

$$\arg \min_{\mathbf{A}} \mathcal{J}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \sum_{i=1}^Z \|\bar{\mathbf{x}}_i - \mathbf{D}\mathbf{a}_i\|_2^2, \text{ sujeito a } \|\mathbf{a}_i\|_0 \leq k_0, \forall 1 \leq i \leq Z. \quad (7.17)$$

Como $k_0 \ll L$, tal problema de otimização admite apenas matrizes \mathbf{A} que apresentem colunas esparsas. Tal restrição pode ser garantida caso recorrarmos a técnicas MP, que automaticamente associam a cada bloco de dados $\bar{\mathbf{x}}_i$ um conjunto de átomos.

A próxima etapa consiste na atualização do dicionário. Se na etapa de codificação esparsa cada vetor de dados $\bar{\mathbf{x}}_i$ for aproximado por uma combinação linear de k_0

⁷Existe uma versão do AGL no qual o coeficiente que multiplica o átomo selecionado pode diferir da unidade[189].

átomos, podemos escrever:

$$\bar{\mathbf{x}}_i = \sum_{j=1}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{(\mathcal{I}^0(i,j))} + \bar{e}_i, \quad (7.18)$$

onde \bar{e}_i é o erro residual, $a_{i,j}$ é o elemento correspondente da matriz \mathbf{A} e $\mathcal{I}^0(i,j)$ é o índice do j -ésimo átomo usado para descrever $\bar{\mathbf{x}}_i$. Para atualizar o k -ésimo átomo $\phi^{(k)}$, poder-se-ia minimizar a seguinte norma quadrática:

$$\min_{\phi^{(k)}, \bar{\mathbf{a}}_k} \sum_{i=1}^Z \left\| \bar{\mathbf{x}}_i - \sum_{j=1, \mathcal{I}^0(i,j) \neq k}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{(\mathcal{I}^0(i,j))} - a_{k,i} \phi^{(k)} \right\|_2^2, \quad (7.19)$$

onde $\bar{\mathbf{a}}_k$ é a k -ésima linha de \mathbf{A} . O problema de otimização (7.19) é facilmente tratável, porém não é adequado para o aprendizado de dicionários, porque preenche $\bar{\mathbf{a}}_k$ com elementos não-nulos. Isto equivale a impor que cada átomo $\phi^{(k)}$ modele todos os blocos de dados. Tal imposição não é coerente com representações esparsas, pois nessas apenas alguns átomos do dicionário descrevem cada vetor $\bar{\mathbf{x}}_i$. Seja $\mathcal{I}^1(i)$ o conjunto dos índices dos átomos utilizados na etapa progressiva de decomposição MP para descrever o i -ésimo bloco $\bar{\mathbf{x}}_i$. Para contornar o problema, a técnica K -SVD restringe a otimização do k -ésimo átomo aos blocos $\bar{\mathbf{x}}_i$ nos quais $k \in \mathcal{I}^1(i)$. Na etapa de adaptação de dicionário, o K -SVD implementa portanto a seguinte otimização para $\phi^{(k)}$:

$$\min_{\phi^{(k)}, \bar{\mathbf{a}}_k} \sum_{i=1, k \in \mathcal{I}^1(i)}^Z \left\| \bar{\mathbf{x}}_i - \sum_{j=1, \mathcal{I}^0(i,j) \neq k}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{(\mathcal{I}^0(i,j))} - a_{k,i} \phi^{(k)} \right\|_2^2, \quad (7.20)$$

a qual pode ser eficientemente resolvida via técnica SVD [190]. Na etapa de atualização do dicionário, o algoritmo K -SVD apresenta a particularidade de também atualizar os coeficientes não-nulos de \mathbf{A} (o que não ocorre com outras técnicas de aprendizagem de dicionários, como o MOD⁸[192]). [191] propõe que a etapa de atualização do dicionário do K -SVD seja executada repetidas vezes, de modo a acelerar a convergência. Esta proposta apresenta o mérito de não incrementar em demasia o custo computacional, pois o mesmo se concentra na etapa de decomposição MP.

7.6.2 Resultados

Nesta seção, utilizaremos sinais de voz extraídos da base de dados *The Open Speech Repository*⁹, amostrados a 8 kHz. O emprego desta base de dados específica se deve

⁸Do inglês *Method of Optimal Directions*.

⁹Acessível em http://www.voiptroubleshooter.com/open_speech.

ao fato de que contempla diferentes idiomas, cada qual com poucos locutores. Cada locutor apresenta diversas locuções as quais, caso reunidas, fornecem informação suficiente para a adaptação de dicionários. Nos ateremos, nesta seção, aos idiomas **fr**, **us** e **uk**, que correspondem a diversas falas em francês, inglês americano e inglês britânico, respectivamente. Cabe ressaltar que os trechos de fala de cada idioma são constituídos majoritariamente por um único locutor (o mesmo utilizado nos testes), embora possa haver, por falta de dados suficientes, sinais oriundos de locutores diferentes. A existência de diferentes locutores torna o problema mais difícil, pois polariza a adaptação rumo ao aprendizado de um determinado idioma (independente de locutor), que é mais difícil do que modelar as particularidades de um locutor específico. Nas simulações desta seção, adapta-se um dicionário distinto para cada idioma.

Para explorar a redundância dos sinais de voz, optamos por adaptar dicionários no domínio da STFT, com janela de Hanning, sobreposição de 50% entre quadros adjacentes, os quais apresentam comprimento $F = 256$. Definindo \mathcal{Q} o número de quadros adjacentes utilizados para formar um átomo, o número de amostras de cada átomo é $L = \left(\frac{\mathcal{Q}+1}{2}\right) F$. A cardinalidade dos dicionários é linearmente dependente do fator de redundância \mathcal{N} , segundo a fórmula:

$$\#D = \mathcal{N}L. \quad (7.21)$$

Quando $\mathcal{N} = 1$, o dicionário não é redundante, apresentando cardinalidade equivalente ao comprimento dos átomos. Em nossos testes, a adaptação de dicionários foi efetuada com as combinações $\mathcal{Q} \times \mathcal{N}$, com $\mathcal{Q} \in \{1, 2, 3\}$ e $\mathcal{N} \in \{1, 2, 4, 8\}$. As Figs. 7.11-7.13 apresentam alguns exemplos de átomos extraídos dos dicionários adaptados para diferentes locutores.

Nesta seção, comparamos a qualidade das estimativas dos coeficientes de atenuação do sistema de mistura em contextos instantâneos (sem atrasos), utilizando a técnica MP padrão (sem apelo a FFTs). Havendo um dicionário adaptado ($\mathcal{D}_{\text{adaptado}}$) para cada fonte que consta na mistura, deve-se eleger um critério para definir qual dicionário fornecerá o átomo utilizado em uma iteração MP de modo a aproximar o resíduo de um dado bloco $\bar{\mathbf{x}}_i$, formado pela concatenação de blocos de todas as misturas. Optamos por adotar uma estratégia competitiva, na qual os dicionários competem para apresentar o átomo mais correlacionado com o bloco de sinal. A partir da seleção do átomo e do dicionário, é possível gerar uma estimativa pontual $\gamma_h(n)$ para algum coeficiente \tilde{h}_{ji} .

Espera-se que o MSE das estimativas advindas com o emprego do dicionário adaptado seja inferior ao de outros dicionários. Para efetuarmos tal comparação, escolhemos empregar como dicionários alternativos o paramétrico ($\mathcal{D}_{\text{paramétrico}}$), já

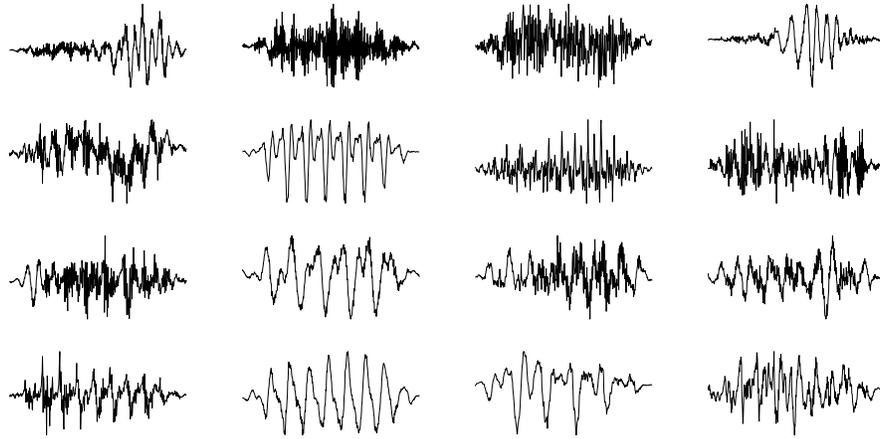


Figura 7.11: Exemplos de átomos extraídos do dicionário adaptado para o idioma `fr`, com $Q = 3$ e $\mathcal{N} = 8$.

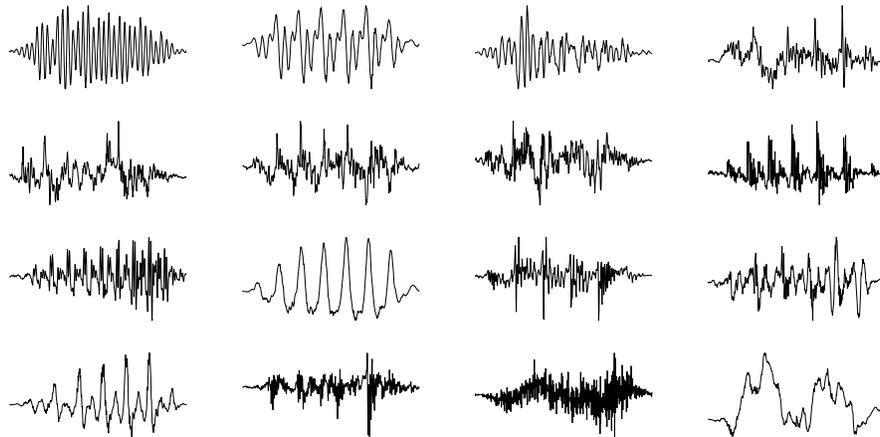


Figura 7.12: Exemplos de átomos extraídos do dicionário adaptado para o idioma `us`, com $Q = 3$ e $\mathcal{N} = 8$.

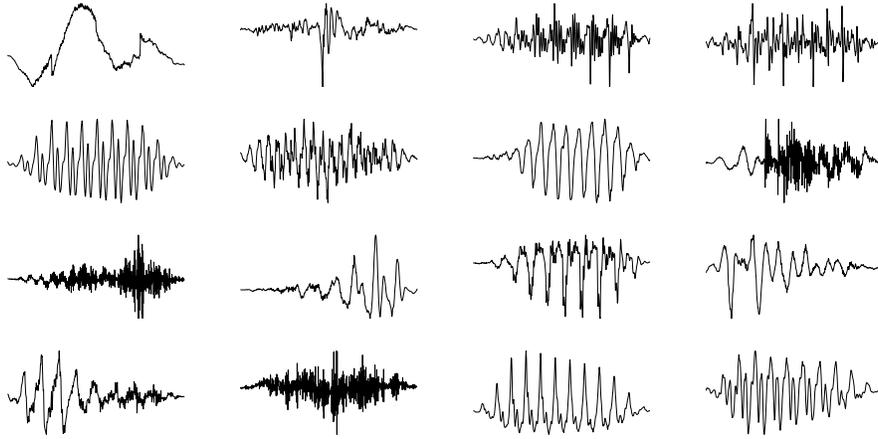


Figura 7.13: Exemplos de átomos extraídos do dicionário adaptado para o idioma uk, com $\mathcal{Q} = 3$ e $\mathcal{N} = 8$.

utilizado neste capítulo e um dicionário cuja formação dos átomos se dá por meio da extração aleatória de alguns trechos da base de dados de treinamento ($\mathcal{D}_{\text{amostras}}$). A variação de \mathcal{Q} (e, por conseguinte, de L) é fácil de se impor a estes dicionários alternativos. Porém, por razões de justiça na comparação, cumpre igualar a cardinalidade dos diferentes tipos de dicionários, o que exige um critério de seleção de átomos do dicionário paramétrico, já que a cardinalidade deste dicionário (para um dado valor de L) pode ser muito maior do que a cardinalidade do dicionário aprendido.

A formação de um novo dicionário paramétrico de menor cardinalidade pode se dar a partir da extração aleatória de átomos do dicionário paramétrico padrão. Porém, este procedimento pode prejudicar o desempenho deste novo dicionário, porque os átomos selecionados podem não modelar razoavelmente os sinais. Este problema pode ser contornado mediante a seleção de átomos que descrevam melhor os sinais envolvidos. A análise do impacto dos parâmetros dos átomos na confiabilidade dos mesmos (empreendida na Seção 7.5.3) concorre para embasar um procedimento seletivo, a partir da constatação de que átomos que apresentam escalas altas e frequências baixas são menos confiáveis. A partir desta constatação, adotamos o procedimento de selecionar apenas átomos que apresentem escala $\log_2(\vartheta_2) < 4$ e frequência $\vartheta_3 \geq 0,3$. Se após este procedimento a cardinalidade do dicionário paramétrico ainda superar a cardinalidade do dicionário adaptado, optamos por efetuar uma retirada aleatória de átomos do dicionário paramétrico, até que se estabeleça a equivalência do número de átomos entre diferentes dicionários.

Nas simulações a seguir, as locuções de um mesmo idioma foram concatenadas em um mesmo arquivo, com os primeiros 20 s do mesmo reservados para os testes de

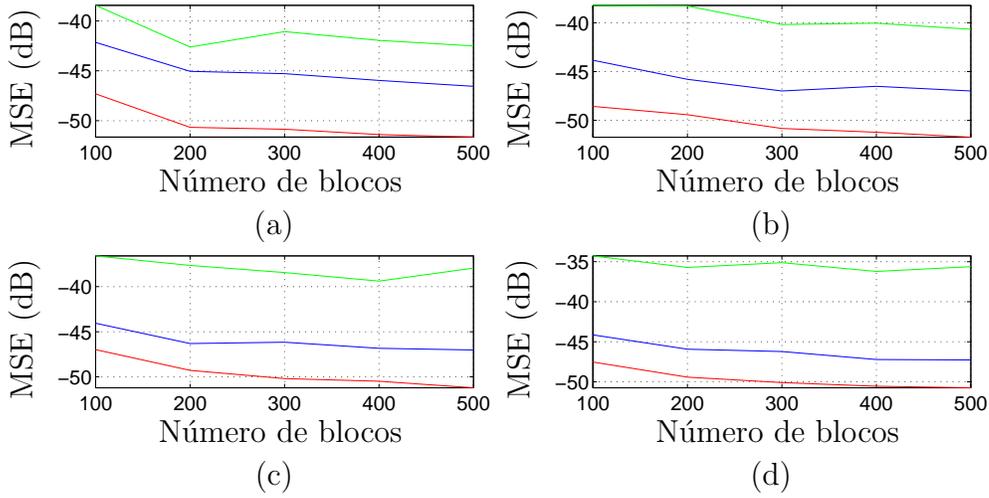


Figura 7.14: MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 1$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

separação e os 500 s (um pouco mais de 8 minutos) restantes destinados ao aprendizado de dicionário. A partir do processo de mistura, diversos blocos de cada mistura, de comprimento idêntico à dimensionalidade dos dicionários, foram aleatoriamente selecionados. Em cada bloco, aplicou-se apenas uma iteração MP, capaz de gerar uma estimativa pontual de um dos coeficientes de atenuação da segunda linha de $\tilde{\mathbf{H}}$.

Para cada combinação $(\mathcal{Q}, \mathcal{N})$ foram efetuados 200 procedimentos de estimação, cada um deles diferindo do outro nos blocos utilizados para estimativa (aleatoriamente selecionados), no dicionário paramétrico (pois os átomos aleatoriamente retirados do mesmo para impor a igualdade no número de átomos entre diferentes dicionários variou em cada procedimento de estimação) e no dicionário cujos átomos consistem em trechos dos sinais de treinamento (aleatoriamente escolhidos). Para cada procedimento de estimação, os parâmetros extraídos da aplicação do MP aos diversos blocos foram utilizados como entrada para a varredura de Zibulevsky (vide Seção 4.10.1), com 10000 ângulos de teste θ_t , uniformemente espaçados no intervalo $[0, \pi]$. Para cada procedimento de estimação, calculou-se o quadrado da norma euclidiana dos erros entre os parâmetros reais e os estimados. Em alguns poucos casos (menos de 0,5% das ocorrências) o procedimento de estimação falhou na seleção de picos, gerando erros muito grandes. Para evitar que estas falhas comprometam a interpretação dos resultados médios, os resultados que apresentaram soma quadrática dos erros superior a 0,1 foram descartados. O MSE foi calculado a partir da média do quadrado da norma quadrática dos erros ao longo de cada configuração avaliada, a menos dos resultados descartados.

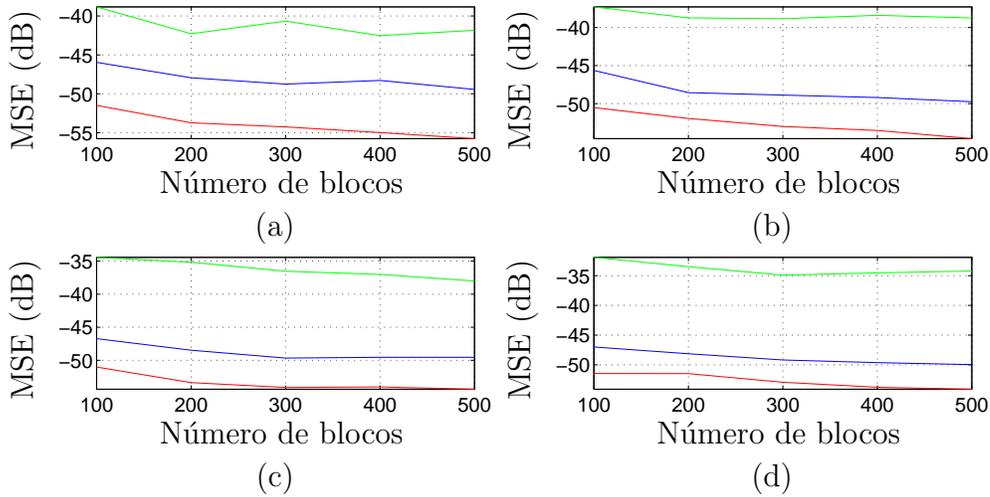


Figura 7.15: MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 2$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

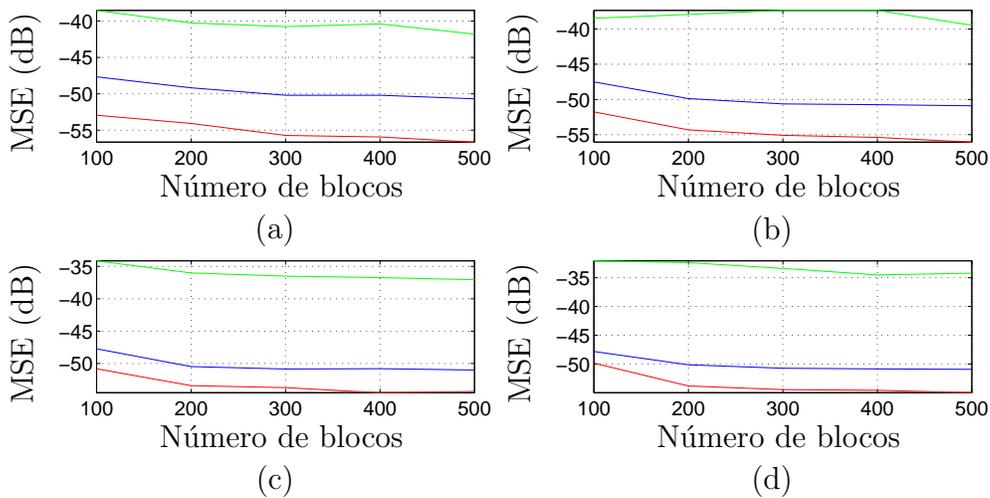


Figura 7.16: MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $\mathcal{Q} = 3$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

A. Caso determinado

A partir da matriz de mistura

$$\tilde{\mathbf{H}} = \begin{bmatrix} 1 & 1 \\ -0,5 & 0,7 \end{bmatrix}, \quad (7.22)$$

foi gerada uma mistura a partir dos sinais de teste oriundos dos idiomas **fr** e **uk**. As possibilidades testadas para o número de blocos pertencem ao conjunto $\{100, 200, 300, 400, 500\}$. A varredura de Zibulevsky foi empregada com parâmetro $\lambda = 1$. As Figs. 7.14-7.16 apresentam o MSE das configurações testadas, revelando que o uso de dicionários adaptados apresenta um desempenho consistentemente superior ao dos demais dicionários avaliados. Estas figuras também revelam que o desempenho do processo de identificação tende a melhorar com o aumento do número de blocos, já que este aumento disponibiliza mais informação para a construção do histograma. Iremos analisar os resultados com mais detalhes após apresentar a simulação do caso subdeterminado, de modo a poder facilmente comparar ambos os casos.

B. Caso subdeterminado

A partir da matriz de mistura

$$\tilde{\mathbf{H}} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0,2 & 0,7 \end{bmatrix}, \quad (7.23)$$

foi gerada uma mistura a partir dos sinais de teste oriundos dos idiomas **fr**, **uk** e **us**. O número de blocos foi fixado em 500, já que com menos blocos a variância do erro quadrático ainda é muito alta (tal não ocorre no caso determinado, mais simples). A varredura de Zibulevsky foi empregada com parâmetro $\lambda = 10$. A Fig. 7.17 apresenta o MSE das configurações testadas, revelando que o uso de dicionários adaptados apresenta um desempenho quase sempre superior (à exceção de uma das configurações) ao dos demais dicionários avaliados.

Análise dos resultados

As Tabs. 7.9 e 7.10 apresentam os resultados dos casos determinado e subdeterminado, respectivamente. As informações constantes nestas tabelas nos permitem inferir, para as simulações efetuadas:

- o caso determinado é mais simples do que o subdeterminado, superando-o em desempenho, segundo o esperado, já que o número de parâmetros a estimar

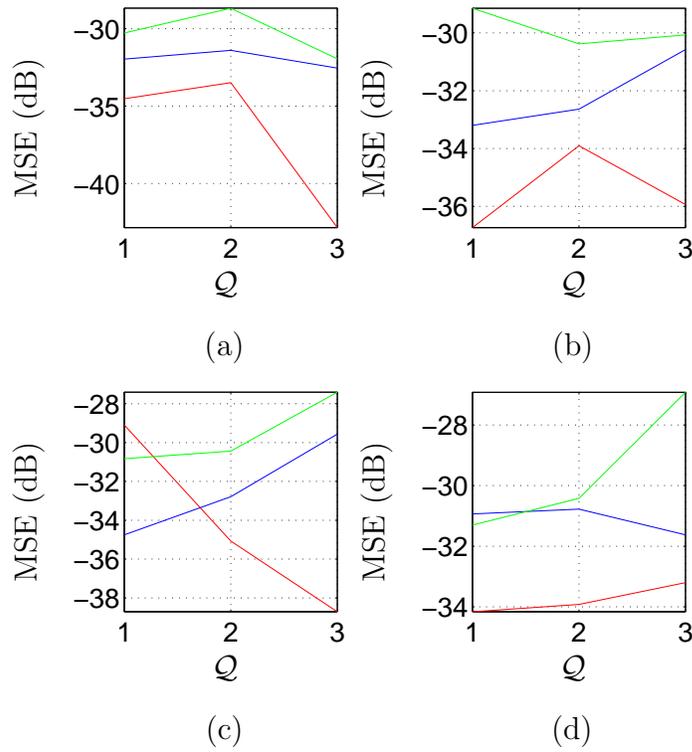


Figura 7.17: MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para o caso subdeterminado e com 500 blocos decompostos via MP. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

\mathcal{N}	\mathcal{Q}	$\mathcal{D}_{\text{adaptado}}$	$\mathcal{D}_{\text{amostras}}$	$\mathcal{D}_{\text{paramétrico}}$
1	1	-51,65	-46,55	-42,5
1	2	-55,76	-49,43	-41,84
1	3	-56,63	-50,68	-41,79
2	1	-51,71	-46,98	-40,67
2	2	-54,6	-49,75	-38,74
2	3	-56,06	-50,9	-39,47
4	1	-51,2	-47,01	-37,94
4	2	-54,37	-49,55	-38
4	3	-54,31	-51	-37,05
8	1	-50,73	-47,26	-35,61
8	2	-54,1	-49,95	-34,2
8	3	-54,99	-50,92	-34,21

Tabela 7.9: MSE (em dB) obtido com emprego de diferentes dicionários, para o caso determinado, para diferentes fatores de redundância \mathcal{N} e número de quadros \mathcal{Q} , com 500 blocos decompostos via MP.

\mathcal{N}	\mathcal{Q}	$\mathcal{D}_{\text{adaptado}}$	$\mathcal{D}_{\text{amostras}}$	$\mathcal{D}_{\text{paramétrico}}$
1	1	-34,52	-31,96	-30,27
1	2	-33,48	-31,4	-28,68
1	3	-42,84	-32,55	-31,94
2	1	-36,73	-33,19	-29,14
2	2	-33,9	-32,63	-30,38
2	3	-35,93	-30,57	-30,07
4	1	-29,09	-34,75	-30,83
4	2	-35,07	-32,78	-30,43
4	3	-38,7	-29,56	-27,4
8	1	-34,16	-30,93	-31,3
8	2	-33,91	-30,77	-30,41
8	3	-33,2	-31,62	-26,91

Tabela 7.10: MSE (em dB) obtido com emprego de diferentes dicionários, para o caso subdeterminado, para diferentes fatores de redundância \mathcal{N} e número de quadros \mathcal{Q} , com 500 blocos decompostos via MP.

é inferior. Comparando o desempenho do mesmo dicionário nos casos determinado e subdeterminado, naquele se verifica um desempenho superior, em média, de 18,7 dB, no dicionário adaptado. Tal melhora é de 17,3 dB para $\mathcal{D}_{\text{amostras}}$ e de 8,6 dB para $\mathcal{D}_{\text{paramétrico}}$.

- o dicionário aprendido quase sempre supera os demais, pois é mais adaptado às peculiaridades dos sinais envolvidos. No caso determinado, ele supera $\mathcal{D}_{\text{amostras}}$ e $\mathcal{D}_{\text{paramétrico}}$, na média, em 4,7 dB e 15,4 dB, respectivamente. No caso subdeterminado, tais melhoras são de, respectivamente, 3,23 dB e 5,3 dB. Isso significa que, em contextos onde há uma base de dados para treinamento das fontes disponíveis e não se dispõe de muitos recursos computacionais para adaptar um dicionário, é mais interessante formar um dicionário a partir de trechos da base de dados do que empregar um dicionário paramétrico. Havendo maior disponibilidade destes recursos, a adaptação de dicionários pode ser uma opção mais adequada.
- o dicionário formado por trechos extraídos dos sinais de treinamento apresenta desempenho superior ao dicionário paramétrico, o que não é surpreendente, porque este deriva de uma formulação *a priori* artificial. No caso determinado, tal melhora é de, em média, 10,6 dB. Já no contexto subdeterminado, o MSE de $\mathcal{D}_{\text{amostras}}$ encontra-se, em média, 2 dB acima do MSE obtido pelo dicionário paramétrico.
- para o caso determinado, com os dicionários $\mathcal{D}_{\text{adaptado}}$ e $\mathcal{D}_{\text{amostras}}$, o desempenho aumenta com o número de quadros modelados por cada átomo. Nos

demais casos, não há uma tendência nítida de aumento ou piora do MSE com a mudança de \mathcal{Q} . Este resultado não é de fácil interpretação, devendo ser futuramente aprofundado.

7.7 Considerações Finais

Neste capítulo, uma estratégia de estimativa conjunta dos coeficientes de atenuação e dos atrasos relativos das contribuições das fontes às misturas, em contextos anecoicos, foi desenvolvida. Esta estratégia, que foi favoravelmente comparada ao método DUET, emprega a construção de histogramas bidimensionais, cujos picos são associados aos parâmetros corretos do sistema de mistura. Diversos parâmetros da decomposição foram avaliados. As conclusões principais oriundas desta avaliação, para os cenários simulados, foram: átomos maiores implicam estimativas mais acuradas; a norma l_1 parece ser mais interessante do que a norma l_2 (em termos espaciais); átomos de escala muito baixa não são confiáveis; átomos de frequência elevada apresentam confiabilidade maior e átomos que contribuem para modelar os sinais (em detrimento dos que consertam erros de modelagem) costumam apresentar resultados superiores. O emprego de dicionários adaptados para determinados locutores, passíveis de serem criados quando se encontra disponível uma base de dados das fontes para treinamento, foi avaliado como capaz de prover um desempenho superior (na identificação de sistemas de mistura) ao de outros dicionários. Em contextos de misturas determinadas, foram obtidos ganhos acima de 15 dB, caso comparados a dicionários paramétricos, muito populares na literatura.

Capítulo 8

Análises de um Algoritmo de Desconvolução Cega

Este capítulo se propõe a analisar alguns aspectos de uma técnica de desconvolução cega de fontes em ambientes reverberantes. Mais desafiadoras do que as instantâneas, as configurações convolutivas, dada a ubiquidade do fenômeno dos múltiplos percursos, apresentam um repertório mais amplo de aplicações, justificando o intenso interesse da comunidade acadêmica neste tópico.

Enquanto técnicas de separação cega de fontes são consideradas bem-sucedidas quando as estimativas são versões filtradas das respectivas fontes, o objetivo das técnicas de desconvolução é mais restrito neste aspecto, pois a estas cabe também recuperar a estrutura temporal do sinal, uma restrição relevante em alguns contextos, como o de comunicações digitais¹. Isto justifica o fato de muitas técnicas de desconvolução cega apresentarem aplicações em comunicações de dados, como por exemplo as técnicas de Bussgang [193], que incluem como casos particulares os algoritmos de Sato [194] e de Godard (ou de módulo constante) [195],[196].

Uma outra distinção importante reside no fato de que técnicas cegas de desconvolução podem ser aplicadas a um contexto monocal, no qual apenas as interferências entre símbolos da mesma fonte devem ser removidas ou atenuadas, inexistindo interferências oriundas de outras fontes. Podemos, grosso modo, dividir as técnicas de equalização em dois grandes grupos: as que atuam no espaço de identificação da função de transferência dispersiva (modelo do canal) ou as que agem no espaço de inversão desta função de transferência. Neste capítulo estamos primordialmente interessados nesta segunda estratégia, a qual procura filtros (em geral, do tipo FIR) que, atuando sobre o sinal recebido, desfaçam ou atenuem o efeito do canal.

Em algumas aplicações, como em comunicações digitais, é possível inserir um trecho no sinal original que seja previamente conhecido. Este conhecimento *a priori*

¹Em contextos instantâneas, técnicas de separação cega também não corrompem a estrutura temporal das fontes.

possibilita a inserção de técnicas de filtragem adaptativa supervisionada. Quando não há (ou é indesejável) um acesso a trechos das fontes, cumpre recorrer a técnicas de processamento autodidata, as quais são poderosas por também não exigirem um conhecimento acerca da função de transferência do canal.

Embora haja na literatura muitos artigos dedicados à análise de desempenho de algoritmos de processamento supervisionado de sinais, relativamente poucas destas análises contemplam técnicas de processamento cego de separação/desconvolução de fontes. Como exemplos, abordagens analíticas de alguns desses algoritmos podem ser encontradas em [197],[198]; comparações de desempenho via simulações e médias do tipo Monte Carlo são efetuadas em [199],[200]. Dentre as abordagens analíticas, a maioria trata do caso de misturas instantâneas. Este número reduzido de análises reflete a maior dificuldade de tratamento analítico das técnicas cegas. Duas são as características basicamente responsáveis por essa dificuldade: não-linearidade e impossibilidade de desacoplamento do problema multidimensional em diversos problemas unidimensionais independentes [78]. Mesmo na técnica de equalização CMA (do inglês *constant modulus algorithm*), mais simples do que o algoritmo contemplado neste capítulo, encontram-se problemas analíticos desafiadores, que requerem o recurso à topologia (no caso de sinais de entrada coloridos [201]), à geometria (volume das regiões de convergência [202]), à caracterização da região de dispersão [203] e à determinação de classes específicas de canais [204], ou o recurso à análise via preservação de energia [205]. Essas razões, entre outras, justificam a dificuldade de obtenção de fórmulas analíticas que estimem a ISI ou o limite superior do fator de aprendizagem, por exemplo.

Este capítulo contempla a técnica de desconvolução cega e de separação de fontes proposta em [102]. Embora esta técnica pretenda recuperar exatamente a estrutura temporal do sinal, mesmo quando não é bem-sucedida no tocante à equalização pode, em contextos multicanais, efetuar uma separação de sinais de voz. Desafortunadamente, não foi possível estender os resultados analíticos aqui apresentados para contextos reais de sinais de voz, dada a grande dificuldade de análise deste algoritmo na versão multicanal e o fato de que uma precisa caracterização estatística destes sinais é muito complexa para ser inserida nos modelos aqui detalhados. De todo modo, a não-gaussianidade destes sinais foi levada em consideração em todas as análises presentes neste capítulo, de sorte a garantir uma certa aderência das características de sinais de voz aos modelos aqui estudados. Esperamos que algumas análises efetuadas neste capítulo possam ser, após algumas adaptações, generalizadas para contextos mais realistas.

Entre os aspectos estudados neste capítulo, encontram-se a equivariância (Seção 8.2.1), as estimativas dos coeficientes em regime permanente (Seções 8.2.2, 8.2.4, 8.2.5 e 8.2.6) e o comportamento no transiente (Seção 8.2.3).

8.1 Algoritmo DCGN

Um dos algoritmos mais eficazes para desconvolução cega de fontes é apresentado em [102], sendo tratado neste trabalho pela sigla DCGN (de “Desconvolução Cega por meio de Gradiente Natural”). O algoritmo DCGN é uma evolução do algoritmo original apresentado em [206], cujo gradiente natural apresentava um viés que reduzia o desempenho final, porque a atualização dos parâmetros incluía termos oriundos de amostras de sinais de entrada as quais não influenciavam a função custo. Estes termos adicionais foram introduzidos ao se calcular o “gradiente natural”; uma forma diferente de truncamento da sequência de entrada elimina estes termos prejudiciais (vide [102] para maiores detalhes).

Como as análises deste capítulo se concentram na configuração monocanal deste algoritmo, por motivos de clareza apenas a apresentação da versão monocanal do algoritmo será efetuada.

8.1.1 Versão monocanal do algoritmo DCGN

Seja \mathbf{h} o filtro de comprimento M correspondente ao canal e $s(k)$ o sinal (ou fonte) a recuperar. Logo, no receptor, o sinal recebido $x(k)$ é uma versão de $s(k)$ corrompida por interferência interssimbólica. Seja $\check{\mathbf{w}}_k$ o filtro adaptativo de comprimento L do equalizador na k -ésima iteração, cuja saída $y(k)$, idealmente, deve ser uma estimativa da fonte desprovida do efeito dispersivo do canal. A estrutura do algoritmo de desconvolução foi apresentada na Fig. 4.15. Se definimos:

$$\check{\mathbf{w}}_k = \left[\check{w}_k(0) \quad \check{w}_k(1) \quad \cdots \quad \check{w}_k(L-1) \right], \quad (8.1)$$

então a saída $y(k)$ pode ser escrita como:

$$y(k) = \check{\mathbf{w}}_k \mathbf{x}_k, \quad (8.2)$$

com \mathbf{x}_k definido na Eq. (2.1). A função custo do algoritmo DCGN é descrita por:

$$\mathcal{F}(\check{W}(z)) = -\frac{1}{2\pi i} \oint \log |\check{W}(z)| z^{-1} dz - \mathbb{E}[\log f(y(k))], \quad (8.3)$$

onde $\check{W}(z) = \sum_{l=0}^{L-1} \check{w}_k(l) z^{-l}$ e $f(\cdot)$ é a pdf da fonte, conforme apresentado no Cap. 4. Esta função-custo pode ser alternativamente colocada como [207]:

$$\mathcal{F}(\check{\mathbf{w}}) = -\frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{l=0}^{L-1} \check{w}(l) e^{-i\theta l} \right| d\theta - \mathbb{E}[\log f(y(k))]. \quad (8.4)$$

Pode-se demonstrar que esta função-custo é, a menos de uma constante independente do equalizador, proporcional à informação mútua da sequência de sinais de saída $\{y(k)\}$, quando $f(\cdot)$ é a pdf da fonte [102]. A equação de atualização do algoritmo DCGN emprega a otimização via gradiente natural, podendo ser explicitamente escrita por [102]:

$$\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_k + \beta [\check{\mathbf{w}}_k - \phi(y(k))\check{\mathbf{z}}_k^T], \quad (8.5)$$

onde $\phi(y(k))$ é a função *score* definida na Eq. (4.13), $\check{\mathbf{z}}_k = \check{\mathbf{R}}_k \mathbf{x}_k$, com $\check{\mathbf{R}}_k$ sendo uma matriz simétrica expressa por:

$$\check{\mathbf{R}}_k = \begin{bmatrix} \check{r}_k(0) & \check{r}_k(1) & \cdots & \check{r}_k(L-1) \\ \check{r}_k(-1) & \check{r}_k(0) & \cdots & \check{r}_k(L-2) \\ \vdots & \ddots & \ddots & \vdots \\ \check{r}_k(-L+1) & \check{r}_k(-L+2) & \cdots & \check{r}_k(0) \end{bmatrix}, \quad (8.6)$$

onde $\check{r}_k(l) = \sum_{p=0}^{L-1-|l|} \check{w}_k(p)\check{w}_k(p+|l|)$. [102] sustenta ser difícil justificar teoricamente o desempenho do algoritmo, dada a enorme dificuldade na obtenção de uma análise estatística completa de seu transiente e comportamento em regime permanente em configurações genéricas. Embora a tarefa seja complexa, neste capítulo são empregadas algumas derivações que elucidam analiticamente algumas características deste algoritmo em determinadas configurações.

8.2 Análises da Versão Monocanal do algoritmo DCGN

Nesta seção, algumas hipóteses a respeito das fontes são formuladas, de modo a permitir uma abordagem analítica. Primeiramente, suas amostras são supostas i.i.d.. Embora, num contexto de separação de fontes, [88] sugira que a hipótese de amostras iid não inutiliza o algoritmo cego quando a fonte apresenta dependência estatística entre amostras (tratando-se apenas de uma opção por não empregar esta dependência), [208] mostra o efeito desta dependência no deslocamento dos mínimos da superfície da função-custo. Como nas simulações da análise as fontes respeitarão esta hipótese, não nos preocuparemos com estas distorções.

8.2.1 Equivariância

Um algoritmo de desconvolução ou separação cega de fontes é dito ser equivariante caso seu desempenho e convergência independam do canal. Para o caso instantâ-

neo e (super)determinado, alguns algoritmos apresentam esta desejável propriedade (para um exemplo, vide [90]), excetuando-se os casos degenerados de uma matriz de mistura singular.

Já em configurações convolutivas, muitos algoritmos são considerados equivariantes, embora muitas destas afirmativas sejam menos fundamentadas em sólida base matemática do que em heurísticas extensões do caso instantâneo (por exemplo, vide [86]). O algoritmo DCGN, embora empregue uma otimização via gradiente natural, não é equivariante em sentido estrito, dada a dependência de seu desempenho final com relação ao filtro do canal.

Nesta seção mostraremos que, num caso simples (mas não trivial) e empregando um conceito mais amplo da propriedade de equivariância, podemos concluir que esta é respeitada pelo DCGN.

Sendo \mathbf{c}_k o vetor que contém as respostas combinadas do canal e do filtro equalizador, sejam as seguintes hipóteses:

- Hipótese I: $M = L = 2$;
- Hipótese II: $h(0)h(1) \neq 0$, apresentando $h(0)$ e $h(1)$ de mesma ordem de magnitude e invariância no tempo;
- Hipótese III: o algoritmo converge para uma solução sem atraso e de baixa ISI, ou seja, $\lim_{k \rightarrow \infty} |c_k(0)| \gg \lim_{k \rightarrow \infty} |c_k(d)|$ para $d \neq 0$;
- Hipótese IV: o fator de aprendizagem β é reduzido o suficiente para que os termos $c_k(i)$ sejam próximos o suficiente de seus valores esperados e permitindo aproximar os coeficientes adaptativos como independentes do sinal de entrada, após a convergência.

Seja $\mathbf{c}_k = [c_k(0) \ c_k(1) \ c_k(2)]$ o vetor que contém os coeficientes do filtro que condensa as operações do canal e do filtro equalizador, ou seja, $\mathbf{c}_k = [h(0)\check{w}_k(0) \ h(0)\check{w}_k(1) + h(1)\check{w}_k(0) \ h(1)\check{w}_k(1)]$. Admitida a primeira hipótese, podemos desmembrar a equação de atualização do vetor $\check{\mathbf{w}}$ em duas equações:

$$\begin{aligned} \check{w}_{k+1}(0) &= \check{w}_k(0) + \beta\check{w}_k(0) - \beta\phi(y(k))\check{w}_k^2(0)x(k) - \beta\phi(y(k))\check{w}_k^2(1)x(k) \\ &\quad - \beta\phi(y(k))\check{w}_k(0)\check{w}_k(1)x(k-1), \\ \check{w}_{k+1}(1) &= \check{w}_k(1) + \beta\check{w}_k(1) - \beta\phi(y(k))\check{w}_k(0)\check{w}_k(1)x(k) - \beta\phi(y(k))\check{w}_k^2(0)x(k-1) \\ &\quad - \beta\phi(y(k))\check{w}_k^2(1)x(k-1), \end{aligned} \tag{8.7}$$

A Hipótese III nos permite aplicar o limite $k \rightarrow \infty$ de modo a, após algumas manipulações matemáticas, obtermos:

$$\begin{aligned}\Delta c(0) &= \beta c(0) - \beta \frac{c^2(0)}{h(0)} \phi(y(k))x(k) - \beta \frac{h(0)c^2(2)}{h^2(1)} \phi(y(k))x(k) \\ &\quad - \beta \frac{c(0)c(1)}{h(1)} \phi(y(k))x(k-1), \\ \Delta c(2) &= \beta c(2) - \beta \frac{c(0)c(2)}{h(0)} \phi(y(k))x(k) - \beta \frac{h(1)c^2(0)}{h(0)} \phi(y(k))x(k) \\ &\quad - \beta \frac{c^2(2)}{h(1)} \phi(y(k))x(k-1),\end{aligned}\tag{8.8}$$

onde elidimos os índices k nos termos $c_k(i)$ e $\Delta c(i) = c_{k+1}(i) - c_k(i)$. A partir da Hipótese IV associamos as equações acima às seguintes equações determinísticas:

$$\begin{aligned}c(0) &= \text{E} \left[\phi(y(k)) \left\{ \left[\frac{c^2(0)}{h(0)} + \frac{h(0)}{h^2(1)} c^2(2) \right] x(k) + \frac{1}{h(1)} c(0)c(2)x(k-1) \right\} \right], \\ c(2) &= \text{E} \left[\phi(y(k)) \left\{ \frac{1}{h(0)} c(0)c(2)x(k) + \left[\frac{h(1)}{h^2(0)} c^2(0) + \frac{1}{h(1)} c^2(2) \right] x(k-1) \right\} \right].\end{aligned}\tag{8.9}$$

Cumpra agora recorrer a algumas simplificações. Em nossas simulações, caso inicializemos \check{w} como $[1 \ 0]$ e $h(0)$ seja maior do que $h(1)$ (o que é uma situação comum, já que as versões atrasadas do sinal tendem a estar mais atenuadas), $c(0)$ costuma ser muito maior do que $c(1)$ e $c(2)$ (admitindo convergência do algoritmo para uma boa solução). Supondo que isso ocorra, $y(k)$ deve ser muito próximo (a menos da constante multiplicativa $c(0)$) de $s(k)$ ou, em outras palavras, o sistema canal + equalizador não introduz atrasos². Estas considerações tornam as hipóteses admitidas como relativamente naturais, embora caiba ressaltar a forte restrição engendrada pela Hipótese III, mesmo para canais de fase mínima.

Sendo iid as amostras de $s(k)$, $y(k)$ deve ser, aproximadamente, independente de $s(k-1)$ e de $s(k-2)$, o que nos permite efetuar a seguinte aproximação:

$$\text{E} [\phi(y(k))s(k-i)] \approx 0, \text{ para } i = 1, 2.\tag{8.10}$$

Como $x(k-i) = h(0)s(k-i) + h(1)s(k-i-1)$, chegamos a:

$$\begin{aligned}c(0) &\approx c^2(0)\text{E} [\phi(y(k))s(k)] + \frac{h^2(0)}{h^2(1)} c^2(2)\text{E} [\phi(y(k))s(k)] \\ &= \left[c^2(0) + \frac{h^2(0)}{h^2(1)} c^2(2) \right] \text{E} [\phi(y(k))s(k)].\end{aligned}\tag{8.11}$$

²Na prática, esta condição em geral não se verifica.

Pelas Hipóteses II e III, podemos desprezar o termo $\frac{h^2(0)}{h^2(1)}c^2(2)$. Assim, chegamos a:

$$c(0) \approx c^2(0)E[\phi(y(k))s(k)], \quad (8.12)$$

de onde obtemos:

$$c(0) \approx \pm \frac{1}{E[\phi(y(k))s(k)]}, \quad (8.13)$$

onde a ambiguidade de escalamento responde pelo sinal \pm . A fórmula acima não é suficiente para a determinação de $c(0)$, já que $y(k)$ também depende de $c(1)$ e $c(2)$. Porém, a Hipótese III permite-nos supor que o equalizador é bem-sucedido, de modo que podemos empregar a aproximação $y(k) \approx c(0)s(k)$ e então obtemos o seguinte resultado:

$$c(0) \approx \frac{1}{E[\phi(c(0)s(k))s(k)]}. \quad (8.14)$$

Supondo conhecidas as estatísticas das fontes (o que é possível quando, por exemplo, é conhecida a modulação de um sinal transmitido) e a função ϕ , o termo $E[\phi(c(0)s(k))s(k)]$ pode ser calculado para qualquer valor de $c(0)$, seja por meio de resultados analíticos, seja, na impossibilidade destes serem efetuados, via médias de Monte Carlo.

O resultado expresso na Eq. (8.14) significa que o valor para o qual o coeficiente adaptativo de maior magnitude $c_k(0)$ converge (condição $k \rightarrow \infty$) praticamente independe da resposta do canal. Esta robustez é uma característica desejável e associada à equivariância. Além deste resultado não ser facilmente generalizável para outras configurações, para o caso em questão verificamos que esta robustez não ocorre com $c_k(1)$ e $c_k(2)$, o que significa que estamos longe de elucidar exatamente em que sentido o algoritmo pode ser associado à propriedade de equivariância. De todo modo, a Eq. (8.14) revela que, de alguma forma, podemos interpretar a técnica DCGN como atendendo à equivariância na configuração sem atraso.

A Fig. 8.1 permite uma comparação da estimativa oriunda da Eq. (8.14) com a obtida pelo algoritmo DCGN. Sendo perfeita a estimativa, a reta vertical vermelha deve coincidir com o ponto de interseção da curva em azul e da reta. Para os dois casos apresentados nesta figura, a Eq. (8.14) revelou boa concordância com os resultados experimentais. A evolução da ISI destes casos é apresentada na Fig. 8.2.

8.2.2 Estimativa de Primeira Ordem dos Parâmetros Finais do Algoritmo DCGN

Estimar o comportamento em regime permanente é uma tarefa complexa para o algoritmo DCGN, dada a não linearidade de sua equação de atualização. Nesta

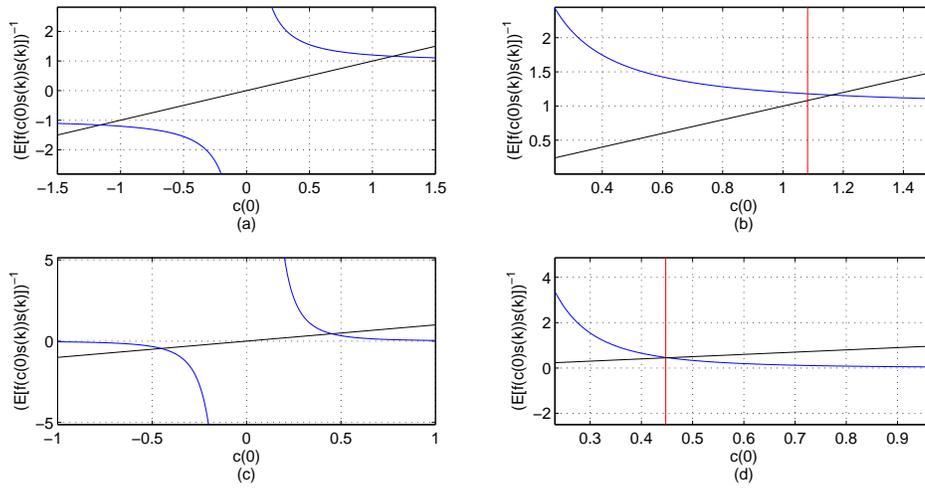


Figura 8.1: Avaliação da estimativa da Eq. (8.14), com inicialização $\check{\mathbf{w}}_0 = [1 \ 0]$, $\beta = 10^{-5}$ e 5×10^5 iterações. O sinal $s(k)$ foi gerado a partir da distribuição $f(s) = \frac{1}{2}e^{-|s|}$. A reta identidade está em preto e o termo $\frac{1}{E[f(c(0)s(k))s(k)]}$ (calculado via médias de Monte Carlo) em azul. Em vermelho, assinala-se o valor real para o qual o algoritmo converge (calculado por meio da média das últimas 3×10^5 iterações). (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0, 3]$; (b) ampliação de parte do gráfico em (a); (c) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$ (d) ampliação de parte do gráfico de (c).

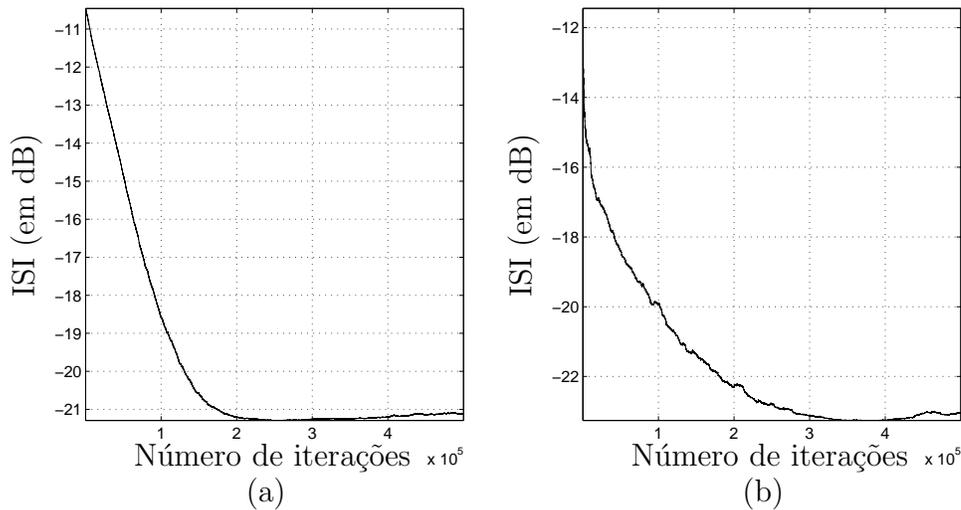


Figura 8.2: Evolução da ISI nos testes efetuados na Fig. 8.1, com inicialização $\check{\mathbf{w}}_0 = [1 \ 0]$ e $\beta = 10^{-5}$. (a) função $\phi(y(k)) = \tanh(y(k))$ e filtro do canal $\mathbf{h} = [1 \ 0, 3]$; (b) função $\phi(y(k)) = y^3(k)$ e filtro do canal $\mathbf{h} = [-1, 5 \ 0, 4]$.

subseção apresentamos uma técnica que emprega a linearização da função $\phi(\cdot)$. A análise a ser efetuada associa uma determinada configuração a um sistema não linear de equações, cuja solução resulta em estimativas dos parâmetros $\check{w}_k(i)$ para os quais o algoritmo converge. Uma desvantagem adicional da análise, em comparação com a do Cap. 3, consiste na variação de desempenho final do algoritmo à medida que se varia o canal, cujo conhecimento então se torna necessário para o prosseguimento da análise. Sejam as hipóteses:

- Hipótese I: O algoritmo converge para uma solução de bom desempenho (ou, equivalentemente, baixa ISI);
- Hipótese II: A fonte apresenta amostras i.i.d., de média zero, variância σ_s^2 e distribuição simétrica;
- Hipótese III: o fator de aprendizagem é suficientemente reduzido de modo a podermos substituir, em regime permanente, os parâmetros $\check{w}_k(i)$ por seus valores esperados, supondo-os também independentes do sinal de entrada.

Por simplicidade, efetuaremos a estimativa de primeira ordem (EPO) para o caso $M = L = 2$ e atraso final unitário (ou seja, $y(k) \approx c(1)s(k-1)$), para não tornar as equações demasiado grandes; não obstante, ao contrário da análise de equivariância, a EPO não se restringe a este caso, como poderá ser observado nas simulações. Uma desvantagem da EPO é a exigência de se arbitrar um atraso final do algoritmo (associado ao pico da função de transferência global - canal + equalizador), em geral desconhecido *a priori*.

Importa enfatizar o relaxamento das hipóteses com relação à análise de equivariância efetuada anteriormente. Impondo a condição $k \rightarrow \infty$ e empregando as Hipóteses I e III, temos:

$$\begin{aligned} 0 &\approx \check{w}(0) - (\check{w}^2(0) + \check{w}^2(1))\text{E}[\phi(y(k))x(k)] - \check{w}(0)\check{w}(1)\text{E}[\phi(y(k))x(k-1)], \\ 0 &\approx \check{w}(1) - (\check{w}^2(0) + \check{w}^2(1))\text{E}[\phi(y(k))x(k-1)] - \check{w}(0)\check{w}(1)\text{E}[\phi(y(k))x(k)], \end{aligned} \quad (8.15)$$

onde a elisão do índice k revela a condição de regime permanente. A principal dificuldade na resolução das equações acima reside nos termos não lineares $\text{E}[\phi(y(k))x(k)]$ e $\text{E}[\phi(y(k))x(k-1)]$, os quais serão, na EPO proposta, linearizados.

Sendo unitário o atraso final e empregando a Hipótese I, temos $y(k) = c(0)s(k) + c(1)s(k-1) + c(2)s(k-2) \approx c(1)s(k-1)$, o que nos permite utilizar uma linearização de Taylor do termo $\phi(y(k))$:

$$\phi(y(k)) \approx \phi(c(1)s(k-1)) + c(0)\phi'(c(1)s(k-1))s(k) + c(2)\phi'(c(1)s(k-1))s(k-2), \quad (8.16)$$

onde $\phi'(\cdot)$ é a derivada de $\phi(\cdot)$.

Empregando as Hipóteses II e III, o termo $E[\phi(y(k))x(k)]$ pode ser expresso por:

$$E[\phi(y(k))x(k)] = h(1) \overbrace{E[\phi(c(1)s(k-1))s(k-1)]}^{\chi_1(c(1))} + h(0)c(0) \overbrace{E[\phi'(c(1))s(k-1)]}^{\chi_2(c(1))} \overbrace{E[s^2(k)]}^{\sigma_s^2}. \quad (8.17)$$

De forma semelhante, chegamos a:

$$E[\phi(y(k))x(k-1)] = h(0)\chi_1(c(1)) + h(1)c(2)\chi_2(c(1))\sigma_s^2. \quad (8.18)$$

Importa ressaltar o fato de que $\chi_1(c(1))$ e $\chi_2(c(1))$ são dependentes das estatísticas da fonte $s(k)$. Estas aproximações nos permitem reescrever as Eqs. (8.15) da forma a seguir:

$$\begin{aligned} 0 &\approx \check{w}(0) - (h(1)\check{w}^2(0) + h(1)\check{w}^2(1) + h(0)\check{w}(0)\check{w}(1))\chi_1(c(1)) \\ &\quad - (h^2(0)\check{w}^3(0) + h^2(0)\check{w}(0)\check{w}^2(1) + h^2(1)\check{w}(0)\check{w}^2(1))\chi_2(c(1))\sigma_s^2, \quad (8.19) \\ 0 &\approx \check{w}(1) - (h(0)\check{w}^2(0) + h(0)\check{w}^2(1) + h(1)\check{w}(0)\check{w}(1))\chi_1(c(1)) \\ &\quad - (h^2(1)\check{w}^2(0)\check{w}(1) + h^2(1)\check{w}^3(1) + h^2(0)\check{w}^2(0)\check{w}(1))\chi_2(c(1))\sigma_s^2, \end{aligned}$$

onde utilizamos o fato de que $c(0) = h(0)\check{w}(0)$ e $c(2) = h(1)\check{w}(1)$ (por simplicidade, optamos por não expressar $c(1)$ como $h(0)\check{w}(1) + h(1)\check{w}(0)$).

Encontrar as soluções do sistema de equações acima implica estimar os parâmetros $\check{w}(0)$ e $\check{w}(1)$, o que nos permitiria, sem recursos a simulações ou a médias do tipo Monte Carlo, obter teoricamente o desempenho do algoritmo numa determinada configuração (ou seja, para determinados valores de $h(0)$ e $h(1)$). O sistema de equações é não linear, o que dificulta uma análise teórica genérica.

Caso particular: Distribuições Exponenciais

Em geral, o sistema de equações (8.19) não apresenta soluções analíticas. Por isso, investigaremos um caso particular de $\phi(\cdot)$ e da pdf da fonte $s(k)$, que permite a determinação destas soluções. Esta análise também servirá para explicitar a metodologia do cálculo de $\chi_1(c(1))$ e de $\chi_2(c(1))$.

Suponhamos que a distribuição de $s(k)$ seja exponencial, de média zero e de parâmetro $\kappa > 0$:

$$f(s) = \frac{1}{2\kappa} e^{-\frac{|s|}{\kappa}}. \quad (8.20)$$

Escolhendo a função *score* ϕ de forma ótima (Eq. (4.13)), encontramos:

$$\phi(s) = -\frac{d \log\left(\frac{1}{2\kappa}\right)}{ds} + \frac{d \frac{|s|}{\kappa}}{ds} = \frac{\text{sign}(s)}{\kappa}. \quad (8.21)$$

Neste caso particular, é possível obter analiticamente o termo $\chi_1(c(1))$:

$$\begin{aligned}\chi_1(c(1)) &= E[s(k-1)\phi(c(1)s(k-1))] = \int_{-\infty}^{\infty} s \cdot \frac{\text{sign}(c(1)s)}{\kappa} \frac{1}{2\kappa} e^{-\frac{|s|}{\kappa}} \\ &= \frac{1}{2\kappa^2} \int_{-\infty}^{\infty} s \cdot \text{sign}(c(1)s) e^{-\frac{|s|}{\kappa}} ds,\end{aligned}\quad (8.22)$$

onde a modelagem das amostras de $s(k)$ como iid permite a elisão do índice k . Para $c(1) > 0$ temos:

$$\chi_1(c(1)) = \frac{1}{2\kappa^2} \left[- \int_{-\infty}^0 s e^{\frac{s}{\kappa}} ds + \int_0^{\infty} s e^{-\frac{s}{\kappa}} ds \right] = 1. \quad (8.23)$$

Para o caso $c(1) < 0$, temos que $\chi_1(c(1)) = -1$; daí:

$$\chi_1(c(1)) = \text{sign}(c(1)). \quad (8.24)$$

Sendo $\phi(s) = \frac{\text{sign}(s)}{\kappa}$, temos que $\phi'(s) = 0$. Isso implica $\chi_2(c(1)) = 0$ (exceto na descontinuidade presente em $s = 0$; porém esta condição possui probabilidade nula de ocorrência, o que nos permite desconsiderá-la). Daí, o sistema de equações (8.19) pode ser simplificado³:

$$\begin{aligned}h(1)\check{w}^2(0) + (h(0)\check{w}(1) - 1)\check{w}(0) + h(1)\check{w}^2(1) &\approx 0, \\ h(0)\check{w}^2(1) + (h(1)\check{w}(0) - 1)\check{w}(1) + h(0)\check{w}^2(0) &\approx 0,\end{aligned}\quad (8.25)$$

A partir das equações abaixo:

$$\check{w}(0) = \frac{1 - h(0)\check{w}(1) \pm \sqrt{1 + h^2(0)\check{w}^2(1) - 2h(0)\check{w}(1) - 4h^2(1)\check{w}^2(1)}}{2h(1)}, \quad (8.26)$$

$$\check{w}(1) = \frac{1 - h(1)\check{w}(0) \pm \sqrt{1 + h^2(1)\check{w}^2(0) - 2h(1)\check{w}(0) - 4h^2(0)\check{w}^2(0)}}{2h(0)}, \quad (8.27)$$

é possível explicitar de forma analítica a solução deste sistema de equações, por meio de substituição. O resultado final foi omitido devido à sua grande extensão.

Obviamente, a solução trivial $\check{w}(0) = \check{w}(1) = 0$ deve ser excluída, bem como eventuais soluções complexas, que podem surgir para valores maiores de M e L .

O sistema de equações (8.26)-(8.27) apresenta uma solução para um caso particular que ilustra uma desejável propriedade do algoritmo. Seja o caso $h(0) = 0$ e $h(1) \neq 0$, na qual o canal apenas introduz um atraso unitário e um fator de escalamento $h(1)$ no sinal. Nesta configuração, manter o atraso unitário na saída do equalizador sem alterar a fonte significa termos $\check{w}(0) = 1/h(1)$ e $\check{w}(1) = 0$. Admi-

³Supomos aqui o caso $c(1) > 0$; para analisar a possibilidade $c(1) < 0$, basta inverter o sinal dos parâmetros $\check{w}(0)$ e $\check{w}(1)$.

tindo $\check{w}(1) = 0$, a Eq. (8.26) (descartando a solução trivial) fornece-nos justamente o valor $\check{w}(0) = 1/h(1)$, conforme esperado. O mesmo comportamento se observa caso façamos $h(0) \neq 0$ e $h(1) = 0$, quando o valor ideal do vetor $\check{\mathbf{w}}$ seria $[0 \ 1/h(0)]$ (supondo atraso unitário). Como esta característica exige uma condição muito particular do canal (ou seja, que o mesmo apresente apenas um componente não nulo), não podemos associá-la à equivariância.

Experimentos

Seja o caso $M = L = 3$. Como ilustração, apresentamos duas das três equações a ele associadas (com as configurações da Fig. 8.3):

$$\begin{aligned} h(2)\check{w}^2(0) + (h(1)\check{w}(1) + h(0)\check{w}(2) - 1)\check{w}(0) + h(2)\check{w}^2(1) + h(2)\check{w}^2(2) + h(1)\check{w}(1)\check{w}(2) &= 0. \\ h(0)\check{w}^2(2) + (h(2)\check{w}(0) + h(1)\check{w}(1) - 1)\check{w}(2) + h(1)\check{w}(0)\check{w}(1) + h(0)\check{w}^2(0) + h(0)\check{w}^2(1) &= 0. \end{aligned} \tag{8.28}$$

A resolução de um sistema de equações não lineares apresenta diversas dificuldades. Embora técnicas padrão de resolução destas equações (empregadas na instrução `solve` do MATLAB) gerem as soluções desejadas, existem também outras soluções que não refletem qualquer comportamento real do algoritmo (soluções espúrias). Esta dificuldade levanta a questão sobre como selecionar as soluções que fazem sentido. Mais à frente tais questões serão objeto de maior atenção, numa modelagem mais sofisticada (que não emprega a aproximação EPO). No momento, foram selecionadas as soluções do sistema que mais se aproximam do comportamento real do algoritmo.

Sejam quatro possíveis escolhas para o filtro do canal com $M = 3$:

$$\begin{aligned} \mathbf{h}^{(1)} &= \begin{bmatrix} 0,19687 & -1,5556 & 0,087869 \end{bmatrix}, \\ \mathbf{h}^{(2)} &= \begin{bmatrix} 0,11839 & -0,78388 & -0,18287 \end{bmatrix}, \\ \mathbf{h}^{(3)} &= \begin{bmatrix} -0,37657 & 1,3414 & 0,073948 \end{bmatrix}, \\ \mathbf{h}^{(4)} &= \begin{bmatrix} -0,072526 & 0,39283 & -0,0047669 \end{bmatrix}, \end{aligned}$$

A Fig. 8.3 apresenta a evolução dos coeficientes e sua comparação com a solução obtida para o sistema de equações (8.28), mostrando que a modelagem se revelou razoavelmente acurada (a análise supôs um atraso de duas amostras na saída do equalizador).

A EPO também permite encontrar soluções de configurações de ordem mais

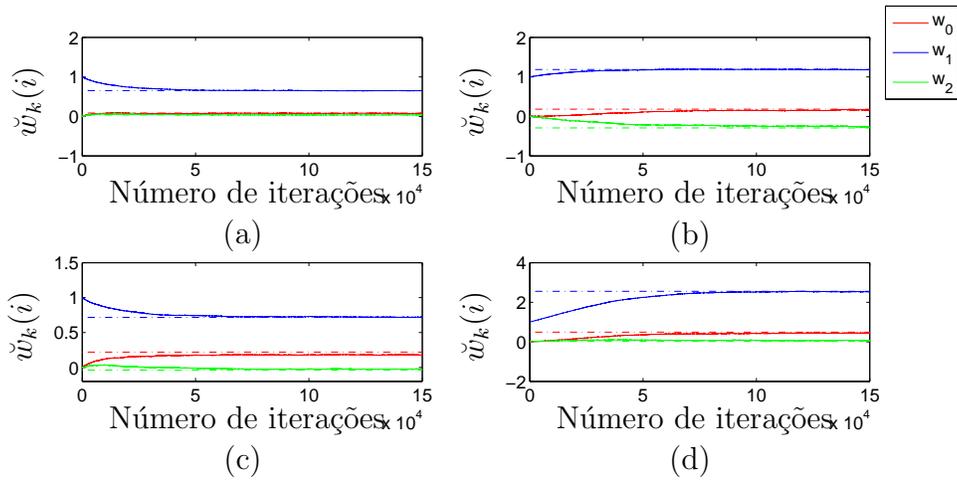


Figura 8.3: Evolução dos parâmetros $\check{w}_k(0)$, $\check{w}_k(1)$ e $\check{w}_k(2)$ para os quatro casos do experimento com $M = L = 3$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 1 \ 0]$. As linhas horizontais refletem o valor teórico. (a) filtro $\mathbf{h}^{(1)}$; (b) filtro $\mathbf{h}^{(2)}$; (c) filtro $\mathbf{h}^{(3)}$ e (d) filtro $\mathbf{h}^{(4)}$.

elevada. Seja o caso $M = 7$ com as seguintes escolhas para o filtro do canal:

$$\begin{aligned} \mathbf{h}^{(1)} &= [0,0235 \ 0,3727 \ 0,3812 \ 2,5901 \ -0,6070 \ -0,0815 \ -0,1358], \\ \mathbf{h}^{(2)} &= [-0,2830 \ -0,1560 \ 0,2387 \ 1,4773 \ -0,0941 \ -0,1025 \ -0,0717], \\ \mathbf{h}^{(3)} &= [-0,1841 \ -0,0863 \ 0,0870 \ 1,4511 \ 0,1498 \ -0,2256 \ 0,1428], \\ \mathbf{h}^{(4)} &= [0,1092 \ 0,2614 \ -0,1463 \ -1,2872 \ -0,4135 \ -0,2158 \ -0,0624]. \end{aligned}$$

Modelando o atraso como de 6 amostras, as soluções da EPO são comparadas com os resultados finais do algoritmo nas Figs. 8.4-8.7, as quais demonstram a acurácia da estimativa gerada pela EPO. A evolução da ISI nestes casos também é apresentada, mostrando que a Hipótese I é satisfeita em todos eles.

Crítica da EPO

Ainda que permita uma estimativa dos valores para os quais o equalizador converge, a EPO proposta apresenta duas características indesejáveis: exige a convergência do algoritmo para uma baixa ISI (o que nem sempre ocorre) para efetuar a linearização e necessita de um conhecimento *a priori* do atraso final do sistema canal + equalizador, o qual depende tanto da inicialização do filtro quanto do canal (em geral, desconhecido).

A propriedade de gerar diversas soluções não é tão problemática, porque em geral, em nossas simulações, a escolha da solução que gera a menor ISI coincide com

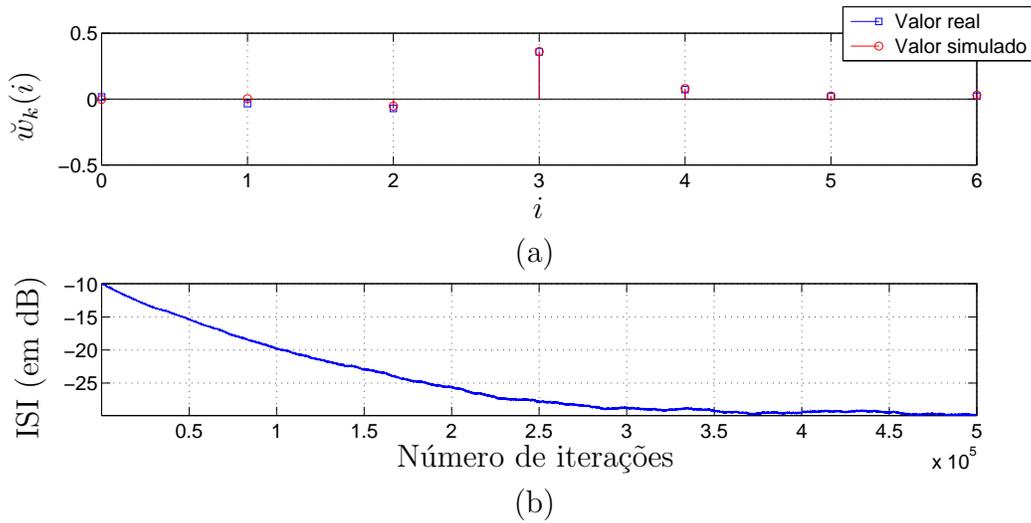


Figura 8.4: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(1)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

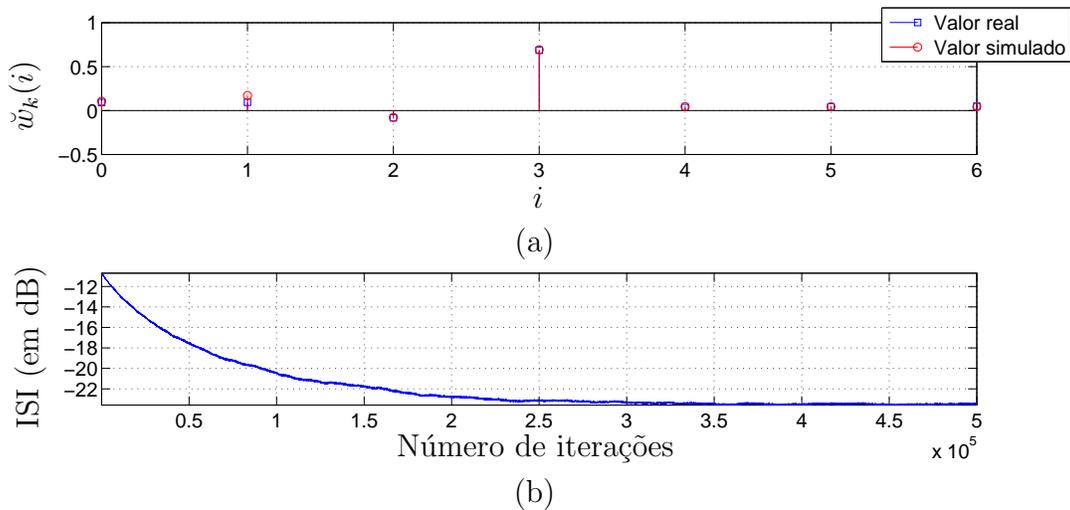


Figura 8.5: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(2)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

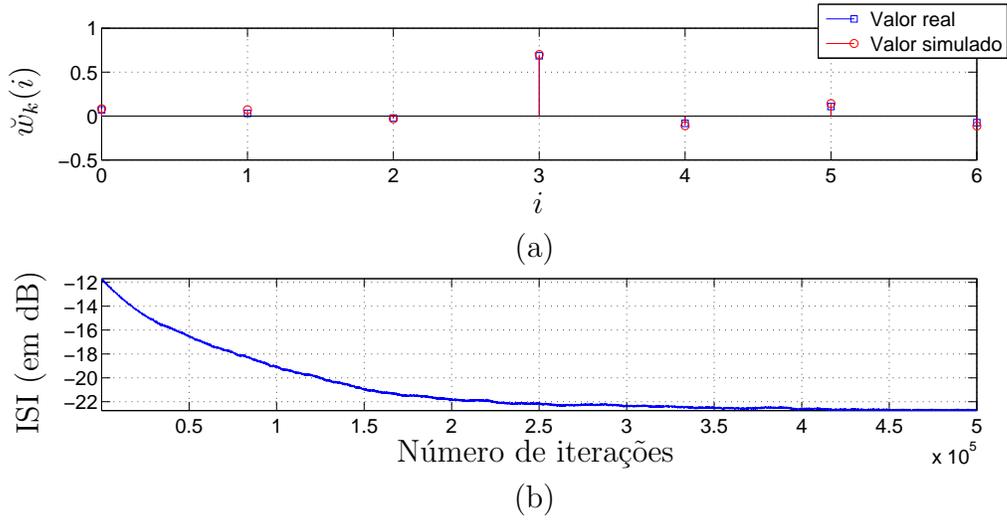


Figura 8.6: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(3)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

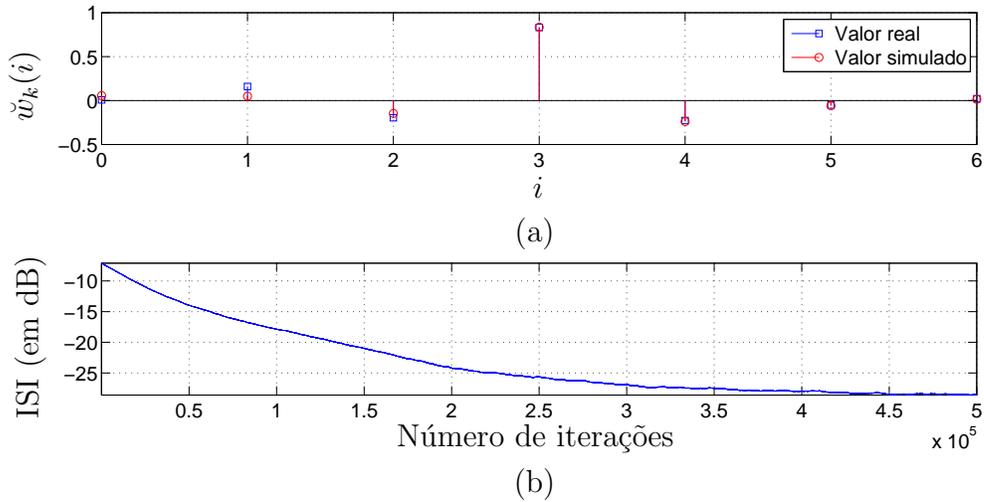


Figura 8.7: Análise da técnica EPO aplicada ao canal $\mathbf{h}^{(4)}$ do experimento com $M = L = 7$. Os parâmetros utilizados foram $\beta = 5 \times 10^{-5}$, $f(s)$ exponencial com $\kappa = 0, 1$, $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$ e vetor $\check{\mathbf{w}}$ inicializado como $[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. (a) comparação dos valores reais com os estimados via EPO; (b) evolução da ISI (em dB).

a solução desejada. E, além disso, mais à frente uma técnica de escolha cega da solução correta será efetuada.

Para contornar estas características indesejáveis, elaboramos uma técnica mais sofisticada, intitulada de EAO (estimativa de alta ordem). Como transição entre a EPO e a EAO, apresentamos uma outra técnica proposta que se destina a rastrear (por meio de equações recursivas) o comportamento do transiente do algoritmo DCGN, como efetuamos para o algoritmo IMPNLMS no Cap. 3. Tanto a análise de transiente quanto a EAO não exigem que a solução final para a qual o algoritmo converge apresente baixa interferência interssimbólica.

8.2.3 Análise de Transiente do Algoritmo DCGN

A eliminação da hipótese de bom desempenho final do algoritmo permite-nos, em alguns casos específicos, não empregar a aproximação da Eq. (8.16). A análise de transiente em questão recorre a apenas duas hipóteses:

- Hipótese IV: O valor de β é suficientemente pequeno para permitir desprezar a variância de $\check{w}_k(i)$, de modo que podemos interpretar a evolução dos coeficientes adaptativos como determinística;
- Hipótese V: os filtros $\check{w}_k(i)$ independem de \mathbf{x}_k .

A Hipótese IV é mais forte do que as comumente empregadas em análise de algoritmos de filtragem adaptativa supervisionada. Tal se deve à intensa não-linearidade das equações de atualização, a qual, para ser contornada, exige o emprego de asunções fortes, cuja justificativa se dará por meio da grande concordância observada com os resultados experimentais.

Por simplicidade⁴, seja o caso $M = L = 2$, cujas equações de atualização encontram-se em (8.7). Interessados na configuração de transiente, não utilizaremos a condição $\lim_{k \rightarrow \infty}$. Aplicando o valor esperado nas equações de atualização supracitadas, temos:

$$\begin{aligned}
 E[\check{w}_{k+1}(0)] &= E[\check{w}_k(0)] + \beta E[\check{w}_k(0)] - \beta E[\check{w}_k^2(0)\phi(y(k))x(k)] - \beta E[\check{w}_k^2(1)\phi(y(k))x(k)] \\
 &\quad - \beta E[\check{w}_k(0)\check{w}_k(1)\phi(y(k))x(k-1)], \tag{8.29} \\
 E[\check{w}_{k+1}(1)] &= E[\check{w}_k(1)] + \beta E[\check{w}_k(1)] - \beta E[\check{w}_k(0)\check{w}_k(1)\phi(y(k))x(k)] \\
 &\quad - \beta E[\check{w}_k^2(0)\phi(y(k))x(k-1)] - \beta E[\check{w}_k^2(1)\phi(y(k))x(k-1)].
 \end{aligned}$$

⁴O fato de ilustrarmos a aplicação da técnica para um caso simples serve apenas para propósitos didáticos, não inutilizando seu emprego em contextos mais complexos.

As Hipóteses IV e V podem ser empregadas para transformar as equações acima em:

$$\begin{aligned}
\check{w}_{k+1}(0) &= \check{w}_k(0) + \beta\check{w}_k(0) - \beta\check{w}_k^2(0)\text{E}[\phi(y(k))x(k)] - \beta\check{w}_k^2(1)\text{E}[\phi(y(k))x(k)] \\
&\quad - \beta\check{w}_k(0)\check{w}_k(1)\text{E}[\phi(y(k))x(k-1)], \\
\check{w}_{k+1}(1) &= \check{w}_k(1) + \beta\check{w}_k(1) - \beta\check{w}_k(0)\check{w}_k(1)\text{E}[\phi(y(k))x(k)] \\
&\quad - \beta\check{w}_k^2(0)\text{E}[\phi(y(k))x(k-1)] - \beta\check{w}_k^2(1)\text{E}[\phi(y(k))x(k-1)].
\end{aligned} \tag{8.30}$$

Na EPO, os termos $\text{E}[\phi(y(k))x(k-i)]$ são calculados por meio da linearização de $\phi(\cdot)$, supondo que $k \rightarrow \infty$ e que o algoritmo tende para uma boa solução. Como tais hipóteses não se adequam à análise de transiente, pode-se delas prescindir por meio do cálculo analítico destes termos. Algumas configurações específicas de pdf das fontes e função $\phi(\cdot)$ permitem este cálculo, como a pdf uniforme com $\phi(y) = y^3$ ou uma pdf exponencial associada a $\phi(y) = \frac{\text{sign}(y)}{\kappa}$ (configuração do próximo experimento). No entanto, em geral os resultados geram fórmulas muito mais complexas (e substancialmente maiores) do que as derivadas da análise EPO.

Experimentos

Como exemplo de aplicação da análise de transiente, seja o caso $M = L = 2$, fontes com pdf exponencial (Eq. (8.20)) e $\phi(y(k)) = \frac{\text{sign}(y(k))}{\kappa}$. O termo $\text{E}[\phi(y(k))x(k)]$ pode então ser calculado por:

$$\begin{aligned}
\text{E}[\phi(y(k))x(k)] &= \frac{h(0)}{4\kappa^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\frac{c(1)}{c(0)}s_1 - \frac{c(2)}{c(0)}s_2}^{\infty} s_1 e^{-\frac{|s_0|+|s_1|+|s_2|}{\kappa}} ds_0 ds_1 ds_2 \\
&\quad + \frac{h(1)}{4\kappa^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\frac{c(1)}{c(0)}s_1 - \frac{c(2)}{c(0)}s_2}^{\infty} s_2 e^{-\frac{|s_0|+|s_1|+|s_2|}{\kappa}} ds_0 ds_1 ds_2,
\end{aligned} \tag{8.31}$$

onde, por motivos de concisão, $s_i = s(k-i)$.

O resultado final dos termos $\text{E}[\phi(y(k))x(k-i)]$ apresenta grande extensão, sendo por isso omitido. As Figs. 8.8 e 8.9 apresentam a comparação entre os resultados simulado e teórico a respeito dos coeficientes adaptativos e da ISI. A Figura 8.10 apresenta a evolução (simulada e teórica) dos coeficientes adaptativos numa outra escolha da pdf das fontes e da função *score* $\phi(\cdot)$. Estas figuras mostram a precisão do cálculo recursivo empregado para a estimativa média da convergência, seja em configurações com β (relativamente) pequeno e elevado, independentemente de o algoritmo tender a uma boa solução ou não.

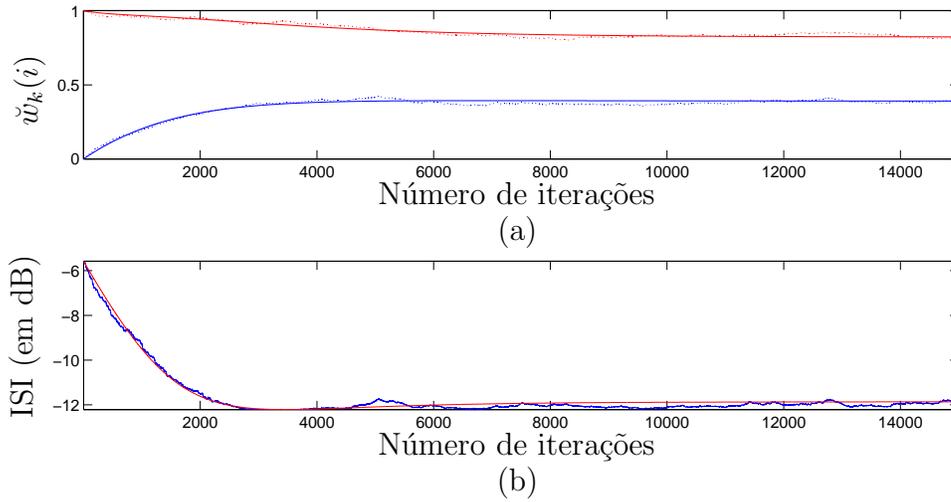


Figura 8.8: Aplicação da análise de transiente para o algoritmo DCGN $\mathbf{h} = [0, 95 \ -0, 5]$, com fontes amostradas segundo uma pdf exponencial ($\kappa = 1$), $\beta = 0,5 \times 10^{-3}$ e $\phi(y(k)) = \text{sign}(y(k))$. (a) comparação da evolução real (em linha tracejada) do algoritmo com a estimada (em linha sólida). A cor vermelha refere-se a $\check{w}_k(0)$ e a azul, a $\check{w}_k(1)$. (b) comparação da evolução da ISI (em dB) real (em azul) e estimada (em vermelho).

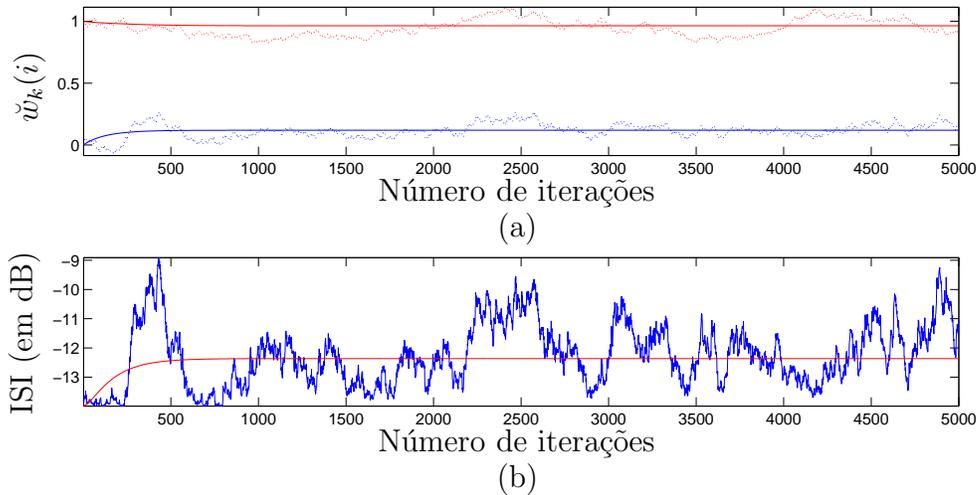


Figura 8.9: Aplicação da análise de transiente para o algoritmo DCGN para um canal $\mathbf{h} = [0, 2 \ -1]$, com fontes amostradas segundo uma pdf exponencial ($\kappa = 1$), $\beta = 0,5 \times 10^{-2}$ e $\phi(y(k)) = \text{sign}(y(k))$. (a) comparação da evolução real (em linha tracejada) do algoritmo com a estimada (em linha sólida). A cor vermelha refere-se a $\check{w}_k(0)$ e a azul, a $\check{w}_k(1)$. (b) comparação da evolução da ISI (em dB) real (em azul) e estimada (em vermelho).

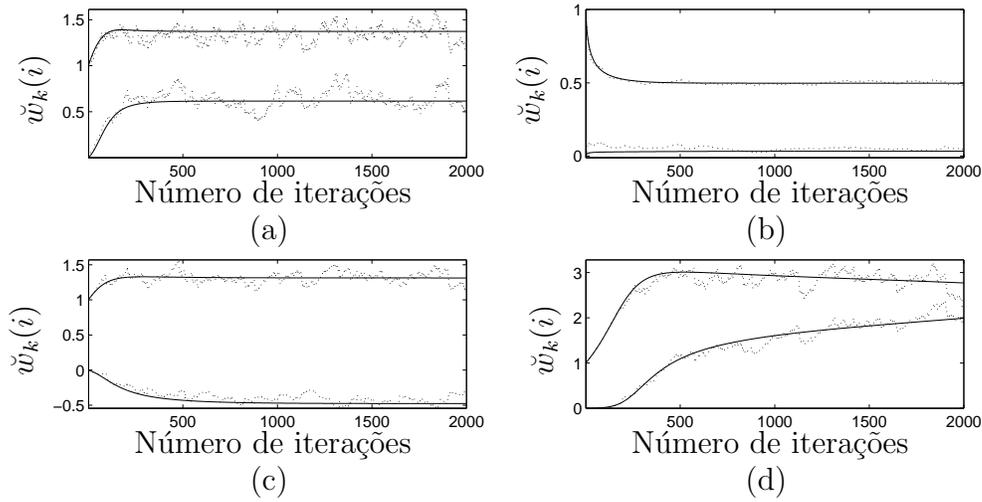


Figura 8.10: Evolução dos coeficientes $\check{w}_k(i)$ ($\mathbf{M} = L = 2$) de uma simulação (linha tracejada) do algoritmo DCGN, comparada com a análise de transiente (linha sólida), com as amostras da fonte distribuídas segundo uma pdf uniforme no intervalo $[-1, 1]$ e $\phi(y(k)) = y^3(k)$ (com 2000 iterações). (a) $\beta = 10^{-2}$ e $\mathbf{h} = [-0, 5 \ 1]$; (b) $\beta = 2 \times 10^{-3}$ e $\mathbf{h} = [-0, 2 \ 3]$; (c) $\beta = 6 \times 10^{-3}$ e $\mathbf{h} = [-1 \ -0, 5]$ e (d) $\beta = 5 \times 10^{-3}$ e $\mathbf{h} = [0, 4 \ -0, 3]$.

8.2.4 Método EDO

A técnica de análise de transiente detalhada apresenta semelhanças com o método EDO⁵ [209],[210] de análise do comportamento da convergência de algoritmos iterativos. O método EDO é bastante geral (por exemplo, não se restringe a algoritmos que empregam atualização derivada de gradientes) e provém do amplo arcabouço fornecido pela intitulada “teoria de aproximação estocástica”. Como diz o nome, o método EDO associa a equação de atualização de um algoritmo adaptativo a uma equação diferencial ordinária, de modo que a convergência desta é acoplada à convergência do algoritmo, seja fracamente (quando o fator de aprendizado é constante, mesmo que reduzido [211]), seja fortemente (quando o fator de aprendizado tende assintoticamente a zero, segundo a proposta original [209],[210]).

Uma das vantagens do método EDO reside no emprego de amplos resultados disponíveis na literatura para equações diferenciais ordinárias, o que explica parte do interesse que desperta na comunidade científica. Embora algumas sutilezas possam prejudicar sua acurácia (por exemplo, a influência da parametrização [212]), verificamos que, para algumas configurações simples do algoritmo DCGN, o método EDO engendra resultados compatíveis com os obtidos via simulação.

Conquanto o método EDO seja obtido com alto rigor matemático, apresentaremos uma motivação informal de sua derivação. Seja a Eq. (8.5), abaixo transcrita

⁵A sigla significa “Equações Diferenciais Ordinárias”.

por conveniência:

$$\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_k + \beta [\check{\mathbf{w}}_k - \phi(y(k))\check{\mathbf{z}}_k^T].$$

Sendo o primeiro termo do lado direito desta equação ($\check{\mathbf{w}}_k$) obtido previamente pelo algoritmo DCGN, a equação acima pode ser reescrita da forma:

$$\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_{k-1} + \beta [\check{\mathbf{w}}_{k-1} - \phi(y(k-1))\check{\mathbf{z}}_{k-1}^T] + \beta [\check{\mathbf{w}}_k - \phi(y(k))\check{\mathbf{z}}_k^T]. \quad (8.32)$$

A aplicação deste procedimento recursivo K vezes gera:

$$\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_{k-K} + \sum_{i=k-K}^k \beta [\check{\mathbf{w}}_i - \phi(y(i))\check{\mathbf{z}}_i^T]. \quad (8.33)$$

Uma breve manipulação matemática nos permite escrever:

$$\frac{\check{\mathbf{w}}_{k+1} - \check{\mathbf{w}}_{k-K}}{K\beta} = \frac{1}{K} \sum_{i=k-K}^k [\check{\mathbf{w}}_i - \phi(y(i))\check{\mathbf{z}}_i^T]. \quad (8.34)$$

O termo no lado direito pode ser compreendido como uma média e, assumindo β muito pequeno, que \mathbf{w}_i pode ser considerado praticamente constante e ergodicidade dos processos envolvidos, podemos associar a equação diferencial ordinária:

$$\frac{d\mathbf{w}(t)}{dt} = \check{\mathbf{w}}(t) - \mathbb{E} [\phi(y(k))\check{\mathbf{z}}_k^T] \quad (8.35)$$

à evolução dos parâmetros $\check{\mathbf{w}}_k(i)$ do algoritmo. Como a abordagem supõe um valor de β muito pequeno, não é capaz de prover limites superiores para o mesmo. Algumas dessas aproximações, para serem precisamente elaboradas ou contornadas, requerem técnicas matemáticas avançadas, como teoria da medida e equações diferenciais estocásticas. A dinâmica evolutiva da solução da EDO resultante não pode ser associada à taxa de convergência do algoritmo, pois a EDO independe do valor de β (apenas assume-se ser este muito pequeno). Porém, o valor que esta solução assume para $t \rightarrow \infty$ é uma forma de estimar o resultado final do algoritmo.

Nossas simulações verificam que, a despeito da intensa não-linearidade das equações do algoritmo DCGN, a solução numérica da equação diferencial ordinária (Eq. (8.35)) gera estimativas precisas (em contextos simples), conforme apresenta a Fig. 8.11. As equações diferenciais não são apresentadas devido à sua grande extensão.

8.2.5 Estimativa Analítica Precisa em Regime Permanente

A análise de transiente efetuada para o algoritmo DCGN contorna duas limitações da abordagem EPO, pois, ao mesmo tempo em que não exige o conhecimento prévio do atraso final do sistema canal + equalizador, prescinde da convergência do algoritmo

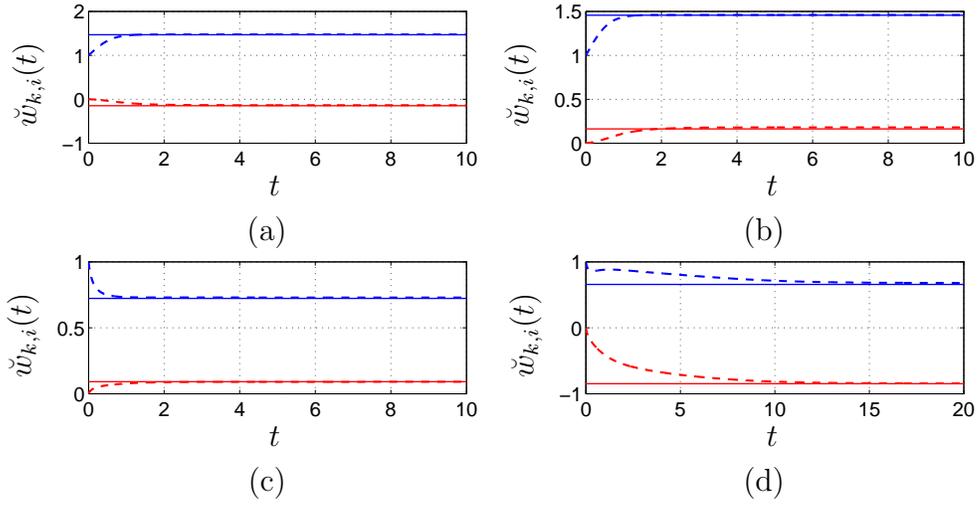


Figura 8.11: Comparação das funções $\check{w}_{k,i}(t)$ (em tracejado) com o valor final dos parâmetros $\check{w}_k(0)$ e $\check{w}_k(1)$ (em linha sólida; todos foram obtidos com uma simulação com 25000 iterações, obtendo-se a média das últimas 5000 iterações e $\beta = 3 \times 10^{-3}$, $f(y(k)) = y^3(k)$ e $s(k)$ com amostras iid e pdf uniforme no intervalo $[-1,1]$). (a) filtro do canal $\mathbf{h} = [1 \ 0, 15]$; (b) filtro do canal $\mathbf{h} = [-1 \ 0, 2]$; (c) filtro do canal $\mathbf{h} = [2 \ -0, 4]$ e (d) filtro do canal $\mathbf{h} = [0, 4 \ 1, 5 \ 1]$. Para todas as simulações arbitrou-se $L = 2$, com a inicialização $\check{\mathbf{w}}_0 = [0 \ 1]$.

para uma solução de baixa ISI.

Para estimar os valores finais do filtro equalizador, efetuamos o limite $k \rightarrow \infty$ ⁶. Supondo convergência, podemos admitir que $\check{w}_{k+1}(i) = \check{w}_k(i)$, simplificando as equações recursivas da análise de transiente e eliminando sua dependência com relação a β . Para a configuração contemplada na Eq. (8.30), por exemplo, resultado final é:

$$\begin{aligned}
 0 &= \check{w}_k(0) - \check{w}_k^2(0)E[\phi(y(k))x(k)] - \check{w}_k^2(1)E[\phi(y(k))x(k)] \\
 &\quad - \check{w}_k(0)\check{w}_k(1)E[\phi(y(k))x(k-1)], \\
 0 &= \check{w}_k(1) - \check{w}_k(0)\check{w}_k(1)E[\phi(y(k))x(k)] \\
 &\quad - \check{w}_k^2(0)E[\phi(y(k))x(k-1)] - \check{w}_k^2(1)E[\phi(y(k))x(k-1)].
 \end{aligned} \tag{8.36}$$

Um sistema de equações não-lineares é gerado a partir das igualdades acima por meio do cálculo analítico dos termos $E[\phi(y(k))x(k-i)]$. Em algumas configurações, o cálculo das soluções deste sistema de equações é possível, exceto quando o conjunto de soluções pertence a uma variedade de dimensão não-nula (possibilidade que foge ao escopo desta tese). Em geral, estas soluções não são enumeráveis quando a função *score* é não-linear e os coeficientes do sistema de mistura são variáveis (sendo estes

⁶Equivalentemente, poderíamos, partindo da análise EDO, impor a condição $t \rightarrow \infty$; porém nesta seção damos preferência à formulação discreta.

fixos, mesmo o emprego de funções *score* não lineares não necessariamente prejudica o encontro das soluções, como veremos na próxima subsecção). Por outro lado, sendo esta linear, a aproximação linear não é necessária, com o resultado sendo preciso mesmo quando a hipótese de bom desempenho final do algoritmo não é satisfeita. Verificaremos a precisão da análise resultante por meio de um experimento.

Experimento

Sejam fontes cujas amostras seguem uma pdf uniforme no intervalo $[-1, 1]$. Se o algoritmo DCGN for aplicado com a função *score*⁷ parametrizada por $f(y(k)) = \alpha y(k)$ para o caso $M = 2$, um possível conjunto de soluções para o sistema de equações⁸ é descrito por:

$$\begin{aligned}\lim_{k \rightarrow \infty} \check{w}_k(0) &= \sqrt{\frac{3}{\alpha (h_0^4 + h_0^2 h_1^2 + h_1^4)}} h_1, \\ \lim_{k \rightarrow \infty} \check{w}_k(1) &= -\sqrt{\frac{3}{\alpha (h_0^4 + h_0^2 h_1^2 + h_1^4)}} h_0.\end{aligned}$$

Para avaliar a precisão das fórmulas acima, testamos diferentes configurações do algoritmo DCGN, com a pdf da fonte seguindo a estipulada pelo cálculo teórico e 40000 iterações do método (com a média dos coeficientes das últimas 10000 iterações para a estimativa empírica dos valores finais do algoritmo). Quatro valores de α foram testados: $1/2$, 1 , $3/2$ e 4 . O filtro do canal teve seu segundo coeficiente variado, com o primeiro fixo em $h_0 = 1$. Os resultados podem ser vistos nas Figs. 8.12 e 8.13, as quais demonstram a ampla precisão das estimativas teóricas.

Não temos conhecimento de nenhuma publicação científica que proveja soluções analíticas para os valores em regime permanente para os quais tendem o filtro equalizador do algoritmo DCGN em nenhuma configuração, como as derivadas neste capítulo. Estas soluções ajudam a elucidar o comportamento do algoritmo, bem como orientar a escolha do parâmetro α de modo a maximizar o desempenho do algoritmo em algumas configurações. Infelizmente, as equações tornam-se demasiado complexas para configurações complicadas (M e L elevados, ou funções *score* não lineares), o que torna estas fórmulas de valor menos prático do que teórico.

⁷Escolhemos a função *score* linear porque ela permite um resultado parametrizável em função de variáveis associadas aos coeficientes do sistema de mistura, o que nos permite a confecção de gráficos mais representativos; sendo estes coeficientes conhecidos, a solução, em alguns casos, pode ser encontrada mesmo que empreguemos funções *score* não lineares, como mostrará o próximo experimento.

⁸Em verdade, existem 9 soluções possíveis para o sistema resultante. Uma delas é a solução nula, facilmente descartada. As outras ocorrem em pares simétricos, dada a ambiguidade de sinal. Na próxima seção detalharemos um procedimento capaz de selecionar as soluções “realistas”.

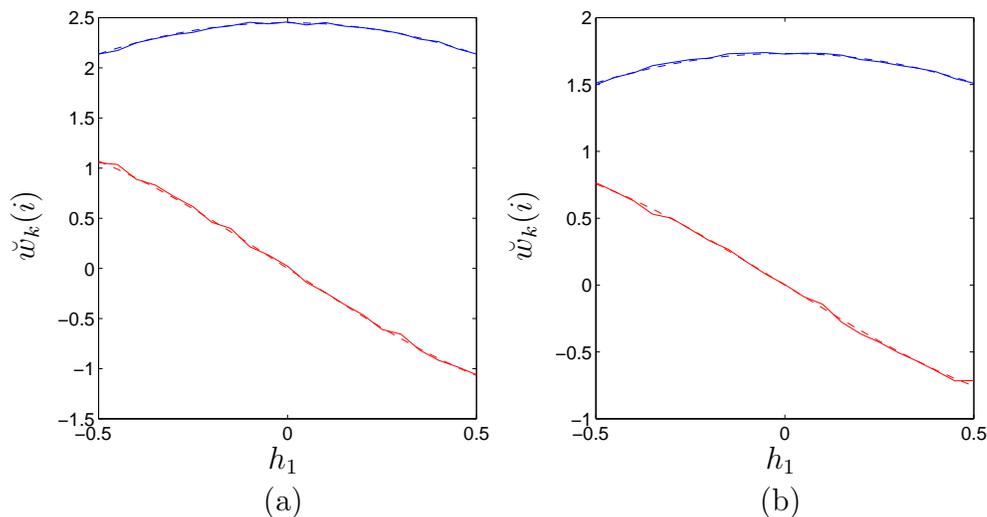


Figura 8.12: Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função *score* parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $\check{w}_k(0)$; em azul: valores de $\check{w}_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 0,5$ e (b) $\alpha = 1$.

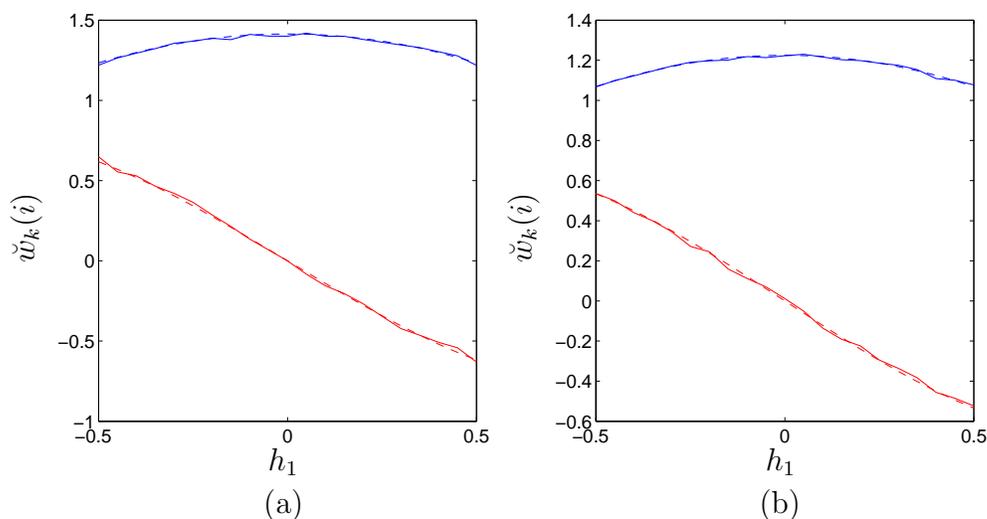


Figura 8.13: Comparação dos valores para os quais os coeficientes adaptativos do algoritmo DCGN tendem com os valores teóricos, com função *score* parametrizada por $f(y(k)) = \alpha y(k)$, amostras das fontes amostradas segundo uma pdf uniforme e coeficiente do canal h_1 variando no intervalo $[-1/2, 1/2]$. Em vermelho: valores de $\check{w}_k(0)$; em azul: valores de $\check{w}_k(1)$. A linha sólida apresenta os valores experimentais e a linha tracejada, os teóricos. (a) $\alpha = 1,5$ e (b) $\alpha = 2$.

8.2.6 Critério de Seleção das Soluções

Em geral, as soluções analíticas de sistema de equações se apresentam em grande número. Sendo espúrias boa parte dessas soluções, dado que não traduzem uma característica real do algoritmo, cumpre encontrar algum critério de seleção.

Várias possibilidades foram testadas, com apenas uma se destacando pela precisão. O critério de maximização da ISI, ainda que interessante, apenas consegue selecionar uma das soluções possíveis, descartando as que poderiam ocorrer, por exemplo, com outras inicializações.

O comportamento qualitativo do sistema dinâmico associado à EDO pode ser estudado, por exemplo, via cálculo do jacobiano do lado direito da Eq. (8.35) [213]. Este procedimento não fornece resultados robustos, possivelmente por ignorar o efeito de β ; por isso apelamos à proposta de [214], intitulada formulação determinística discreta no tempo (DDT, do inglês *Deterministic Discrete-Time*).

A formulação DDT atenta diretamente para a natureza discreta da atualização do algoritmo DCGN (Eq. (8.5)):

$$\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_k + \beta [\check{\mathbf{w}}_k - \phi(y(k))\check{\mathbf{z}}_k^T] = \mathbf{g}(\check{\mathbf{w}}_k), \quad (8.37)$$

onde a função multidimensional $\mathbf{g}(\check{\mathbf{w}}_k)$ incorpora diretamente o efeito do fator de aprendizado. Seguindo o método indireto de Lyapunov [215], o estudo dos autovalores do valor esperado do jacobiano de $\mathbf{g}(\check{\mathbf{w}}_k)$ nos pontos de equilíbrio (ou seja, $\check{\mathbf{w}}_{k+1} = \check{\mathbf{w}}_k$) pode ser empreendido para determinar a estabilidade dos mesmos, pois, na possibilidade de os módulos de todos os autovalores de $\mathbf{g}(\check{\mathbf{w}}_k)$ serem inferiores a um, podemos supor a estabilidade do ponto de equilíbrio (ou, em outras palavras, admitir a solução em questão do sistema de equações como factível).

Experimento

Seja o caso em que a pdf das fontes é uniforme no intervalo $[-1,1]$, a função *score* é $f(y(k)) = y^3(k)$, $M = L = 2$ e $\mathbf{h} = [0 \ 0, 1]$. Seja \mathbf{J}_g o jacobiano de $\mathbf{g}(\check{\mathbf{w}}_k)$. Colocando

parametricamente em função de β , h_0 , h_1 , w_0 e w_1 , os elementos da matriz J_g são:

$$\begin{aligned}
J_g(1, 1) &= -\beta h_0^4 w_0^4 - (13\beta h_0^4 w_0^2 w_1^2)/5 - (8\beta h_0^4 w_1^4)/15 - (16\beta h_0^3 h_1 w_0^3 w_1)/3 \\
&\quad - (18\beta h_0^3 h_1 w_0 w_1^3)/5 - (10\beta h_0^2 h_1^2 w_0^4)/3 - (38\beta h_0^2 h_1^2 w_0^2 w_1^2)/5 \\
&\quad - (8\beta h_0^2 h_1^2 w_1^4)/5 - (16\beta h_0 h_1^3 w_0^3 w_1)/5 - (58\beta h_0 h_1^3 w_0 w_1^3)/15 \\
&\quad - \beta h_1^4 w_0^4 - (13\beta h_1^4 w_0^2 w_1^2)/5 - (8\beta h_1^4 w_1^4)/15 + \beta + 1, \\
J_g(1, 2) &= -(26\beta h_0^4 w_0^3 w_1)/15 - (32\beta h_0^4 w_0 w_1^3)/15 - (4\beta h_0^3 h_1 w_0^4)/3 \\
&\quad - (27\beta h_0^3 h_1 w_0^2 w_1^2)/5 - \beta h_0^3 h_1 w_1^4 - (76\beta h_0^2 h_1^2 w_0^3 w_1)/15 \\
&\quad - (32\beta h_0^2 h_1^2 w_0 w_1^3)/5 - (4\beta h_0 h_1^3 w_0^4)/5 - (29\beta h_0 h_1^3 w_0^2 w_1^2)/5 \\
&\quad - (5\beta h_0 h_1^3 w_1^4)/3 - (26\beta h_1^4 w_0^3 w_1)/15 - (32\beta h_1^4 w_0 w_1^3)/15, \\
J_g(2, 1) &= -(32\beta h_0^4 w_0^3 w_1)/15 - (26\beta h_0^4 w_0 w_1^3)/15 - (5\beta h_0^3 h_1 w_0^4)/3 \\
&\quad - (29\beta h_0^3 h_1 w_0^2 w_1^2)/5 - (4\beta h_0^3 h_1 w_1^4)/5 - (32\beta h_0^2 h_1^2 w_0^3 w_1)/5 \\
&\quad - (76\beta h_0^2 h_1^2 w_0 w_1^3)/15 - \beta h_0 h_1^3 w_0^4 - (27\beta h_0 h_1^3 w_0^2 w_1^2)/5 \\
&\quad - (4\beta h_0 h_1^3 w_1^4)/3 - (32\beta h_1^4 w_0^3 w_1)/15 - (26\beta h_1^4 w_0 w_1^3)/15, \\
J_g(2, 2) &= -(8\beta h_0^4 w_0^4)/15 - (13\beta h_0^4 w_0^2 w_1^2)/5 - \beta h_0^4 w_1^4 \\
&\quad - (58\beta h_0^3 h_1 w_0^3 w_1)/15 - (16\beta h_0^3 h_1 w_0 w_1^3)/5 - (8\beta h_0^2 h_1^2 w_0^4)/5 \\
&\quad - (38\beta h_0^2 h_1^2 w_0^2 w_1^2)/5 - (10\beta h_0^2 h_1^2 w_1^4)/3 - (18\beta h_0 h_1^3 w_0^3 w_1)/5 \\
&\quad - (16\beta h_0 h_1^3 w_0 w_1^3)/3 - (8\beta h_1^4 w_0^4)/15 - (13\beta h_1^4 w_0^2 w_1^2)/5 - \beta h_1^4 w_1^4 + \beta + 1.
\end{aligned}$$

O número de soluções ($\check{w}(0), \check{w}(1)$) do sistema de equações resultante (omitido devido à sua extensão) é igual a 25. Após retirar as soluções simétricas (oriundas da ambiguidade de sinal), trivial (nula) e complexas, temos:

$$S_{\check{w}} = \{(0.0898, 1.4862), (1.4915, 0.1486), (-0.8504, 0.8373), (0.9200, 0.9435)\}.$$

Logo, restam-nos 4 possíveis soluções. Os autovalores do jacobiano de $\mathbf{g}(\check{\mathbf{w}}_k)$ referentes às 4 soluções (para $\beta = 0, 1$)⁹ são, respectivamente:

$$\{(0.8387, 0.6), (0.6, 0.8363), (0.6, 1.0860), (1.0792, 0.6)\}.$$

Logo, detectamos que apenas duas soluções (as duas primeiras) apresentam módulos dos autovalores inferiores a 1. E coerentemente, em nossas simulações, o algoritmo converge para uma das duas soluções (a menos de ambiguidades de sinal). O principal fator que determina a solução para a qual o algoritmo converge reside na inicialização do vetor $\check{\mathbf{w}}$, embora um fator de aprendizagem muito elevado

⁹Em nossas simulações, qualquer valor de β inferior a 0,5 não impacta a escolha das soluções factíveis. Como na prática os fatores de aprendizagem adquirem valores muito menores do que 0,5, basta-nos arbitrar qualquer valor para β no intervalo $(0, 1/2)$.

possa, em alguns casos, alterar o comportamento final do algoritmo. De todo modo, o procedimento acima ilustra como podemos, de forma cega, eliminar as soluções espúrias.

8.3 Considerações Finais

Este capítulo concentrou-se na análise do algoritmo DCGN. A Seção 8.2.1 elucida características deste algoritmo que podem ser associadas à desejável propriedade da equivariância. As estimativas dos coeficientes em regime permanente são alvo de análise nas Seções 8.2.2, 8.2.4 e 8.2.5 as quais, devido à grande complexidade, são menos passíveis de aplicação em configurações complexas. Uma análise de transiente foi empreendida na Seção 8.2.3, a qual revelou grande concordância com os resultados experimentais. Por fim, para contornar o problema das soluções espúrias das estimativas dos coeficientes em regime permanente, a Seção 8.2.6 apresenta um critério de seleção de soluções válidas.

Ao longo de todo este capítulo, foram obtidas fórmulas analíticas inéditas que modelam razoavelmente bem o comportamento real do algoritmo, cujas desejáveis propriedades têm sido justificadas, na literatura, menos por resultados teóricos do que por simulações e médias de Monte Carlo. Este capítulo consiste numa contribuição a uma fundamentação mais sólida, respaldada em resultados analíticos, das características deste algoritmo.

Capítulo 9

Conclusões

Neste capítulo objetivamos efetuar um panorama das contribuições efetuadas nesta tese. Conquanto tais informações se encontrem dispersas ao longo dos capítulos precedentes, a compilação das mesmas neste capítulo pretende promover uma visão global do trabalho.

O Cap. 2 apresenta os algoritmos proporcionais de filtragem adaptativa. Foi apresentada uma revisão bibliográfica, a qual apontou o algoritmo IMPNLMS como um dos mais competitivos algoritmos desta família. Por fim, uma interpretação geométrica da atualização proporcional encerra o capítulo, que é prelúdio do Cap. 3, o qual concentra-se na análise e na extensão do algoritmo IMPNLMS. Na Seção 3.2, uma análise do transiente do MSE do algoritmo IMPNLMS foi proposta. Esta análise guarda semelhanças com as efetuadas em [51],[52], porém concomitantemente apresenta significativas distinções. Em primeiro lugar, [51],[52] supõem que o sinal de entrada seja branco. Esta hipótese é violada por diversos sinais na prática, dentre eles sinais de voz. Quando o sinal de entrada é colorido, as características de convergência tornam-se bastante diferentes e portanto não cabe restringir a análise para este caso. O algoritmo sob análise também difere; [51] contempla um algoritmo PNLMS simplificado, enquanto que [52] se concentra sobre o algoritmo IPNLMS [71]. Já nossa análise é voltada para o algoritmo IMPNLMS [37], mais sofisticado que o IPNLMS no tocante ao uso de uma medida de esparsidade, na estimativa automática do parâmetro $\alpha(k)$ e no emprego da função $F(\cdot)$. Outra característica inovadora da nossa análise é a introdução de técnicas (adaptadas) oriundas de [58], [59], [55] e [60], as quais obtêm precisão significativamente superior para filtros adaptativos curtos (pois não utilizam a aproximação da Eq. (3.27)), permitindo configurações em que L seja pequeno possam ser devidamente contempladas com acuradas análises teóricas. Na Seção 3.3, a análise por balanceamento de energia preconizada por [64] foi empregada para estimar o MSE em regime permanente dos algoritmos da família PNLMS. O resultado coincide com o apresentado em [36], porém com duas vantagens: nossa análise é mais simples e emprega menos hipóteses. Por exemplo, a

suposição acerca da gaussianidade de $\mathbf{x}(k)$ se mostrou desnecessária. Na Seção 3.4 uma estrutura em subbandas não uniforme proposta para os algoritmos de SIRE teve alguns aspectos teóricos elucidados, particularmente a respeito do MSE em regime permanente e a influência do número de subbandas no intervalo admissível (no tocante à convergência) para o fator de aprendizagem β . Ao final do capítulo, técnicas de variação do fator de aprendizagem foram empregadas para encontrar um compromisso adequado entre a taxa de convergência e o desempenho em regime permanente.

O Cap. 4 inaugura a parte não supervisionada da tese, enfocando alguns tópicos importantes dos problemas de separação e equalização cega de fontes, como por exemplo a modelagem do processo de mistura (Seção 4.3), a PCA (Seção 4.4), os princípios de separação (Seção 4.5) e a análise de componentes independentes por meio de estimativa de verossimilhança (Seção 4.7).

As técnicas de separação cega de fontes no domínio da frequência são contempladas no Cap. 5. Técnicas situadas no estado da arte para resolver o problema da permutação são detalhadas, assim como novas propostas. Entre estas, um novo procedimento de inicialização das matrizes de separação (Seção 5.3) que, incorporando a informação de direção de chegada das fontes, obtém resultados semelhantes à inicialização padrão, porém com um número de iterações muito menor, sem experimentar instabilidade, mesmo com fatores de aprendizagem considerados elevados (como $\beta = 0,5$). Uma estrutura que insere coeficientes interfrequenciais é proposta na Seção 5.4.1, bem como uma estratégia para adaptação destes coeficientes. Esta estrutura obteve, em nossas simulações, um ganho de 5 dB na SIR, para sinais gravados num ambiente com T_{60} na faixa de 100 ms. Por fim, técnicas de estimativa cega do tempo de reverberação, originalmente propostas para gravações de uma única fonte, são estendidas para o caso de termos misturas de fontes, na Seção 5.5.1. A informação lateral, correlacionada ao tempo de reverberação foi então empregada para seleção automática do tamanho do quadro da STFT das técnicas de separação cega. Tal seleção automática demonstrou elevada taxa de acerto.

O Cap. 6 trata do problema de separação cega de fontes no domínio do tempo. Um algoritmo situado no estado da arte neste tema é o GBSS-SOS, o qual é estendido para uma formulação em subbandas, a qual é capaz de obter maiores taxas de convergência. A propensão à instabilidade deste algoritmo é contornada via recurso a fatores de aprendizagem variantes no tempo, controlados por um método efetivo de detecção precoce da instabilidade. A taxa de convergência é também incrementada por meio de uma técnica de pós-processamento.

O problema de separação de fontes em misturas anecoicas é contemplado no Cap. 7. Uma estratégia de estimativa conjunta dos coeficientes de atenuação e dos atrasos relativos das contribuições das fontes às misturas, em contextos anecoici-

cos, foi também desenvolvida. Esta estratégia emprega a construção de histogramas bidimensionais, associados a decomposições esparsas das misturas. Os picos do histograma são associados aos parâmetros do sistema de mistura. A identificação dos atrasos, pelo método proposto, apresentou desempenho superior a do popular método DUET. Diversos parâmetros da decomposição foram avaliados, cujas conclusões principais são resumidas a seguir:

1. átomos maiores tendem a produzir estimativas melhores;
2. o emprego da norma l_1 (em termos espaciais) para seleção do átomo se mostrou superior ao emprego da norma l_2 , uma escolha popular em técnicas de separação cega de fontes;
3. átomos de escala muito baixa são menos comuns e implicam degradação da qualidade da estimativa;
4. átomos de frequência elevada são menos frequentes e apresentam confiabilidade superior;
5. átomos que apresentam interferência construtiva global são mais confiáveis do que os demais.

No contexto de separação semi-cega de fontes, a possibilidade de inserção de conhecimento de trechos isolados das fontes no dicionário (via adaptação do mesmo) foi estudada, revelando desempenho (do erro quadrático médio das estimativas dos coeficientes do sistema de mistura) superior em 15 dB com relação ao emprego de dicionários paramétricos, em contextos de mistura determinada.

Por fim, no Cap. 8 o algoritmo DCGN, projetado para efetuar desconvolução cega de sinais, foi objeto de análises teóricas. A Seção 8.2.1 pode esclarecer que, num sentido amplo, podemos associar a este algoritmo as desejáveis propriedades de equivariância. Uma análise de primeira ordem da estimativa final (EPO) do algoritmo foi apresentada na Seção 8.2.2, com resultados razoavelmente precisos, caso suas hipóteses sejam satisfeitas. Uma análise de transiente foi empreendida na Seção 8.2.3, sendo sua relação com a abordagem via EDO desenvolvida na Seção 8.2.4. Como as hipóteses da técnica EPO são não raro muito fortes (por exemplo, supor que o algoritmo convirja para uma solução de bom desempenho), uma proposta mais sofisticada que prescinde destas aproximações é detalhada na Seção 8.2.5. Esta proposta empregou a análise de transiente das seções pregressas, supondo convergência do algoritmo e impondo $k \rightarrow \infty$, para analisar o comportamento do algoritmo em regime permanente. Como a cada configuração é associado um sistema de equações não-lineares com múltiplas soluções, surge a necessidade de um critério que permita aferir as soluções válidas, o que foi empreendido via análise DDT na Seção 8.2.6.

As inovações apresentadas são passíveis de diversos desdobramentos em trabalhos futuros. Entre esses, poderíamos citar:

1. aplicação da estimativa do T_{60} e projeto de estruturas em subbandas criticamente decimadas para algoritmos de separação no domínio do tempo;
2. aprendizado de dicionários com a restrição de penalizar dicionários que, modelando fontes distintas, apresentem similaridade entre si;
3. generalização da análise do algoritmo DCGN para o caso multicanal.

Referências do Autor em Temas Relacionados à Tese

Observação: As publicações anteriores a junho de 2009 foram omitidas da lista abaixo.

Artigos de Conferências

- P1** D. B. Haddad, M. R. Petraglia e P. B. Batalheiro. Análises Exata e Aproximada de um Algoritmo de Deconvolução Cega, IX Congresso de Redes Neurais, 2009, Ouro Preto. Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Artificial (CD-ROM), 2009. v. 1. p. 1-5.
- P2** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro. Equivariância e desempenho de um algoritmo de deconvolução cega. In: IX Congresso de Redes Neurais, 2009, Ouro Preto. Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Artificial (CD-ROM), 2009. v. 1. p. 1-5.
- P3** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro. Frequency-Domain Blind Source Separation employing a Non-Uniform DFT. In: International Telecommunications Symposium, 2010, Manaus. Proceedings of International Telecommunications Symposium, 2010. p. 1-4.
- P4** M. R. Petraglia, D. B. Haddad. New Adaptive Algorithms for Identification of Sparse Impulse Responses - Analysis and Comparisons. In: Seventh International Symposium on Wireless Communication Systems, 2010, York. Proceedings of the Seventh International Symposium on Wireless Communication Systems, 2010. p. 1-5.
- P5** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro. Convergência e Desempenho de um Algoritmo de Desconvolução Cega. In: CBA 2010, 2010, Bonito. XVIII Congresso Brasileiro de Automática, 2010.
- P6** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro. Direção de Chegada e Filtragem Adaptativa Supervisionada Aplicadas à Separação Cega de Fontes. In: CBA 2010, 2010, Bonito. XVIII Congresso Brasileiro de Automática, 2010. p. 1-6.
- P7** J. C. Pires Filho, M. R. Petraglia, D. B. Haddad. Avaliação de Técnicas de Separação Cega de Fontes na Identificação de Instrumentos Musicais. In: CBA 2010, 2010, Bonito. XVIII Congresso Brasileiro de Automática, 2010. p. 1-6.

- P8** P. B. Batalheiro, M. R. Petraglia, D. B. Haddad. Subband Blind Source Separation with Critically Sampled Filter Banks. In: IWSSIP 2010, 2010, Rio de Janeiro. 17th International Conference on Systems, Signals and Image Processing, 2010. p. 1-4.
- P9** J. C. Pires Filho; M. R. Petraglia. Assessment of Blind Source Separation Techniques in the Identification of Musical Instruments. In: IWSSIP 2010, 2010, Rio de Janeiro. Proceedings of the 17th International Conference on Systems, Signals and Image Processing, 2010. p. 1-6.
- P10** E. F. S. Filho, D. B. Haddad, Luiz A. L. de Almeida. Thermal Hysteresis Characterization Through Blind Deconvolution. In: IWSSIP 2010, 2010, Rio de Janeiro. 17th International Conference on Systems, Signals and Image Processing, 2010. p. 1-4.
- P11** D. B. Haddad, L. Lovisolo, J. C. Pires Filho, M. R. Petraglia. Emprego de Dicionários Redundantes e de Minimização da Informação Mútua para a Separação Cega de Fontes em Ambientes Anecoicos com Atrasos Fracionários. In: X Congresso Brasileiro de Inteligência Computacional, 2011, Fortaleza. Anais do X Congresso Brasileiro de Inteligência Computacional, 2011. p. 1-8.
- P12** P. B. Batalheiro, M. R. Petraglia, D. B. Haddad. Subband Blind Source Separation Considering Acoustic Reverberation Characteristics. In: 2011 IEEE International Symposium on Circuits and Systems, 2011, Rio de Janeiro. Proceedings of 2011 IEEE International Symposium on Circuits and Systems, 2011.
- P13** D. B. Haddad, M. R. Petraglia, L. Lovisolo, P. B. Batalheiro, J. C. Pires Filho. Emprego de Técnicas de Matching Pursuit para a Identificação dos Parâmetros de um Sistema de Mistura Anecoico. XI Congresso de Engenharia de Áudio da AES Brasil, 2013. p. 55-62.
- P14** D. B. Haddad, M. R. Petraglia, L. Lovisolo, P. B. Batalheiro. Aprendizado de Dicionários para a Separação Semi-cega de Fontes por Matching Pursuit. XXXI Simpósio Brasileiro de Telecomunicações. *Aceito para publicação*, 2013.

Artigos em Periódicos

- P15** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro. On IIR Filters and Convergence Acceleration for Convolutional Blind Source Separation. Lecture Notes in Computer Science, v. 1, p. 267-273, 2009.

- P16** P. B. Batalheiro, M. R. Petraglia, D. B. Haddad. Online Subband Blind Source Separation for Convolutional Mixtures Using a Uniform Filter Bank with Critical Sampling. *Lecture Notes in Computer Science*, v. 1, p. 211-218, 2009.
- P17** F. S. P. Clark, M. R. Petraglia, D. B. Haddad. A New Initialization Method for Frequency-Domain Blind Source Separation Algorithms. *IEEE Signal Processing Letters*, v. 18, p. 343-346, 2011.
- P18** D. B. Haddad, M. R. Petraglia, P. B. Batalheiro, J. C. Pires Filho. Análise de Componentes Esparsos: Separação Cega de Fontes e Aplicações. *Learning and Nonlinear Models*, v. 1, p. 1-17, 2012.
- P19** E. F. Simas Filho, J. M. de Seixas, N. N. Moura, D. B. Haddad, J. M. Faier, M. C. S. Albuquerque. Independent Component Analysis and Blind Signal Separation: Theory, Algorithms and Applications. *Learning and Nonlinear Models*, v. 1, p. 1-19, 2012.
- P20** M. R. Petraglia e D. B. Haddad. Mean-Square Error and Stability Analysis of a Subband Structure for the Rapid Identification of Sparse Impulse Responses. *Digital Signal Processing (Elsevier)*, p. 1-6, 2012.

Referências Bibliográficas

- [1] TYGEL, A. F. *Métodos de Fatoração de Matrizes Não-negativas para Separação de Sinais Musicais*. Dissertação de mestrado, COPPE - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Dez. 2009.
- [2] CAVALCANTE, C. C., ROMANO, J. M. T. “On the Relationships between MMSE and Information-Theoretic-Based Blind Criterion for Minimum BER Filtering”, *Proceedings of the 8th International Conference ICA 2009*, pp. 17–24, Mar. 2009.
- [3] DENES, P., PINSON, E. *The Speech Chain: The Physics and Biology of Spoken Language*. 1 ed. New York, Worth Publishers, 1993.
- [4] LEE, T., CHING, P. C. “Dealing with Imperfections in Human Speech Communication with Advanced Speech Processing Techniques”, *2011 10th International Symposium on Signals, Circuits and Systems (ISSCS)*, pp. 1–6, Jun. 2011.
- [5] MARIANI, J. “Recent Advances in Speech Processing”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1989)*, v. 1, pp. 429 – 440, May 1989.
- [6] PARANAGUÁ, E. D. S. *Segmentação Automática do Sinal de Voz para Sistemas de Conversão Texto-Fala*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Mar. 2012.
- [7] ESQUEF, P. A. A., BISCAINHO, L. W. P. “Spectral-Based Analysis and Synthesis of Audio Signals”. In: Pérez-Meana, H. (Ed.), *Advances in Audio and Speech Signal Processing: Technologies and Applications*, 1 ed., cap. 3, Hershey, Idea Group, 2007.
- [8] YILMAZ, O., RICKARD, S. “Blind Separation of Speech Mixtures via Time-Frequency Masking”, *IEEE Transactions on Signal Processing*, v. 52, n. 7, pp. 1830–1847, Jul. 2004.

- [9] ARAKI, S., MAKINO, S., BLIN, A., et al. “Underdetermined Blind Separation for Speech in Real Environments with Sparseness and ICA”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2004)*, v. 3, pp. 881–884, May 2004.
- [10] QUATIERI, T. F. *Discrete-Time Speech Signal Processing - Principles and Practice*. 1 ed. Upper Saddle River, Prentice Hall, 2002.
- [11] HYVÄRINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. 1 ed. New York, John Wiley & Sons, 2001.
- [12] MORETTIN, P. A., TOLOI, C. M. C. *Análise de Séries Temporais*. 2 ed. São Paulo, Edgard Blucher, 2006.
- [13] MARTIN, R. “Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors”, *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 5, pp. 845–856, Sept. 2005.
- [14] GAZOR, S., ZHANG, W. “Speech Probability Distribution”, *IEEE Signal Processing Letters*, v. 10, n. 7, pp. 204–207, Jul. 2003.
- [15] ALLEN, J. B. “How Do Humans Process and Recognize Speech”, *IEEE Transactions on Speech and Audio Processing*, v. 2, n. 4, pp. 567–577, Oct. 1994.
- [16] LONG, M. *Architectural Acoustics*. 1 ed. San Diego, Elsevier Academic Press, 2006.
- [17] KINGSBURY, B. E. D. *Perceptually Inspired Signal-Processing Strategies for Robust Speech Recognition in Reverberant Environments*. Tese de D.Sc., University of California, Berkeley, CA, USA, 1998.
- [18] EVEREST, F. A., POHLMANN, K. *Master Handbook of Acoustics*. 4 ed. New York, McGraw-Hill, 2000.
- [19] LIU, J., MALVAR, H. “Blind Deconvolution of Reverberated Speech Signals via Regularization”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2001)*, v. 5, pp. 3037–3040, May 2001.
- [20] DE M. LAPORTE, L. V. *Algoritmos de Separação Cega de Sinais de Áudio no Domínio da Frequência em Ambientes Reverberantes: Estudo e Comparações*. Dissertação de mestrado, COPPE-UFRJ, Rio de Janeiro, Rio de Janeiro, Brasil, Oct. 2010.

- [21] LEHMAN, E. A., JOHANSSON, A. M., NORDHOLM, S. “Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model”, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 159–162, Oct. 2007.
- [22] LEHMAN, E. A., JOHANSSON, A. M. “Prediction of Energy Decay in Room Impulse Responses Simulated with an Image-Source Model”, *The Journal of The Acoustic Society of America*, v. 124, n. 1, pp. 269–277, Jun. 2008.
- [23] ATLAS, L., DUHAMEL, P. “Time-Frequency Masking Method Using Wavelet Transform for BSS Problem”, *IEEE Signal Processing Magazine*, v. 16, n. 1, pp. 16–31, Jan. 1999.
- [24] MILIĆ, L. *Multirate Filtering for Digital Signal Processing*. 1 ed. New York, Hershey, 2009.
- [25] ESTEBAN, D., GALAND, C. “Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1977)*, pp. 191–195, May 1977.
- [26] BUTTERWECK, H. J. “A Steady-State Analysis of the LMS Adaptive Algorithm Without Use of the Independence Assumption”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1995)*, v. 2, pp. 1404 – 1407, May 1995.
- [27] DINIZ, P. S. R. *Adaptive Filtering - Algorithms and Practical Implementation*. 2 ed. Boston/London/Dordrecht, Kluwer, 2002.
- [28] WIDROW, B. “Thinking about Thinking: the Discovery of the LMS Algorithm”, *IEEE Signal Processing Magazine*, v. 22, n. 1, pp. 100–106, Jan. 2005.
- [29] SAYED, A. H., RUPP, M. “Robustness Issues in Adaptive Filtering”. In: *The DSP Handbook*, The Electrical Engineering Handbook Series, CRC Press, pp. 20.1–20.20, 2010.
- [30] HASSINI, B., SAYED, A. H., KAILATH, T. “ H^∞ Optimality of the LMS Algorithm”, *IEEE Transactions on Signal Processing*, v. 44, n. 2, pp. 267–280, Feb. 1996.
- [31] REY, H., VEGA, L. R., TRESSENS, S., et al. “Optimum Variable Explicit Regularized Affine Projection Algorithm”, *Proceedings of the IEEE Inter-*

national Conference on Acoustic, Speech and Signal Processing (ICASSP 2006), pp. 197–200, May 2006.

- [32] BENESTY, J., REY, H., VEGA, L. R., et al. “A Nonparametric VSS NLMS Algorithm”, *IEEE Signal Processing Letters*, v. 13, n. 10, pp. 581–584, Oct. 2006.
- [33] JELFS, B., MANDIC, D. P., BENESTY, J. “A Class of Adaptively Regularised PNLMS Algorithms”, *Proceedings of the 15th International Conference on Digital Signal Processing (DSP 2007)*, pp. 19–22, Aug. 2007.
- [34] BENESTY, J., HUANG, Y., CHEN, J. “An Exponentiated Gradient Adaptive Algorithm for Blind Identification of Sparse SIMO Systems”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2004)*, v. 2, pp. 829–832, May 2004.
- [35] STOER, J., BULIRSCH, R. *Introduction to Numerical Analysis*. 3 ed. New York, Springer, 2002.
- [36] DUTTWEILER, D. L. “Proportionate Normalized Least-Mean Squares Adaptation in Echo Cancelers”, *IEEE Transactions on Speech and Audio Processing*, v. 8, n. 5, pp. 508–518, Aug. 2000.
- [37] LIU, L., FUKUMOTO, M., SAIKI, S. “An Improved Mu-Law Proportionate NLMS Algorithm”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2008)*, pp. 3797 – 3800, Apr. 2008.
- [38] ETTER, D. M., STEARNS, S. D. “Adaptive Estimation of Time Delays in Sampled Data Systems”, *IEEE Transactions on Acoustic, Speech and Signal Processing*, v. 29, n. 3, pp. 582–587, Jun. 1981.
- [39] SUGIYAMA, A., IKEDA, S., HIRANO, A. “A Fast Convergence Algorithm for Sparse-Tap Adaptive FIR Filters Identifying an Unknown Number of Dispersive Regions”, *IEEE Transactions on Signal Processing*, v. 50, n. 12, pp. 3008 – 3017, Dec. 2002.
- [40] CHANDRA, G. B., MITRA, A. “A Fast Adaptive Echo Canceler with Leaky Proportionate NLMS Algorithm”, *IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007)*, pp. 611–615, 2007.

- [41] XU, L., LI, P. “A Set-Membership Approach to Improved PNLMS Algorithm”, *International Conference on Intelligent Control and Information Processing*, pp. 54–57, Aug. 2010.
- [42] GAY, S. “An Efficient, Fast Converging Adaptive Filter for Network Echo Cancellation”, *Proc. of the 32nd Asilomar Conference on Signals, Systems and Computers (ACSSC’98)*, pp. 94–98, Nov. 1998.
- [43] DENG, H., DOROSLOVACKI, M. “Improving Convergence of the PNLMS Algorithm for Sparse Impulse Response Identification”, *IEEE Signal Processing Letters*, v. 12, n. 3, pp. 181–184, Mar. 2005.
- [44] DENG, H., DOROSLOVACKI, M. “Proportionate Adaptive Algorithms for Network Echo Cancellation”, *IEEE Transactions on Signal Processing*, v. 54, n. 3, pp. 1794–1803, Apr. 2006.
- [45] COUCH, L. *Digital and Analog Communication Systems*. 5 ed. New Jersey, Englewood Cliffs, NJ: Prentice Hall, 1997.
- [46] HUANG, Y., BENESTY, J., CHEN, J. *Acoustic MIMO Signal Processing (Signals and Communications Technology)*. 1 ed. New York, Springer, 2006.
- [47] PALEOLOGU, C., BENESTY, J., CIOCHINA, S. “An Improved Proportionate NLMS Algorithm Based on the l_0 Norm”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2010)*, pp. 309–312, Mar. 2010.
- [48] YUKAWA, M., UTSCHICK, W. “Proportionate Adaptive Algorithm for Non-sparse Systems Based on Krylov Subspace and Constrained Optimization”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2009)*, pp. 3121–3124, Abr. 2009.
- [49] AL-NAFFOURI, T. Y., SAYED, A. H. “Transient Analysis of Data-Normalized Adaptive Filters”, *IEEE Transactions on Signal Processing*, v. 51, n. 3, pp. 639–652, Mar. 2003.
- [50] LIMA, M. V. S. *Análise do Algoritmo Set-Membership Affine Projection*. Dissertação de mestrado, COPPE-UFRJ, Rio de Janeiro, Rio de Janeiro, Brasil, Sept. 2009.
- [51] WAGNER, K. T., DOROSLOVACKI, M. I. “Towards Analytical Convergence Analysis of Proportionate-Type NLMS Algorithms”, *IEEE International*

Conference on Acoustics Speech and Signal Processing (ICASSP 2008), pp. 3825–3828, Mar. 2008.

- [52] LOGANATHAN, P., HABETS, E. A. P., NAYLOR, P. A. “Performance Analysis of IPNLMS for Identification of Time-Varying Systems”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2010)*, pp. 317–320, Mar. 2010.
- [53] DENG, H., DOROSLOVACKI, M. “On Convergence of Proportionate-Type NLMS Adaptive Algorithms”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006)*, v. 3, pp. 105–108, May 2006.
- [54] HAYKIN, S. *Adaptive Filter Theory*. 3 ed. New Jersey, Prentice Hall, 1996.
- [55] LOBATO, E. M., TOBIAS, O. J., SEARA, R. “Stochastic model for the NLMS algorithm with correlated Gaussian data”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006)*, v. 3, pp. 760–763, May 2006.
- [56] BARRAULT, G., COSTA, M. H., BERMUDEZ, J. C. M., et al. “A New Analytical Model for the NLMS Algorithm”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2005)*, v. 4, pp. 41–44, Mar. 2005.
- [57] EWEDA, E. “A new approach for analyzing the limiting behavior of the normalized LMS algorithm under weak assumptions”, *Signal Processing*, v. 89, n. 11, pp. 2143–2151, Nov. 2009.
- [58] RUPP, M. “The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes”, *IEEE Transactions on Signal Processing*, v. 41, n. 3, pp. 1149–1160, Mar. 1993.
- [59] LOBATO, E. M., TOBIAS, O. J., SEARA, R. “Stochastic modeling of the transform domain LMS algorithm for a time-varying environment”, *Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2005)*, pp. 1–4, Sept. 2005.
- [60] KOLODZIEJ, J. E., TOBIAS, O. J., SEARA, R., et al. “On the constrained stochastic gradient algorithm: model, performance and improved version”, *IEEE Transactions on Signal Processing*, v. 57, n. 4, pp. 1304–1315, Apr. 2009.

- [61] DAS C. DE SOUZA, F., TOBIAS, O. J., SEARA, R., et al. “Stochastic Model for the Mean Weight Evolution of the IAF-PNLMS Algorithm”, *IEEE Transactions on Signal Processing*, v. 58, n. 11, pp. 5895–5901, Aug. 2010.
- [62] LOBATO, E. M., TOBIAS, O. J., SEARA, R. “Stochastic Modeling of the Transform-Domain ϵ LMS Algorithm for Correlated Gaussian Data”, *Proceedings of the 6th International Conference on Telecommunications (ITS 2006)*, pp. 912–917, Sept. 2006.
- [63] DE ESTUDO 15 DA ITU-T, G. *Digital Network Echo Cancellers (Recommendation)*. Relatório Técnico G.168, ITU-T, 2004.
- [64] YOUSEF, N. R., SAYED, A. H. “A Unified Approach to the Steady-State and Tracking Analyses of Adaptive Filters”, *IEEE Transactions on Signal Processing*, v. 49, n. 2, pp. 314–324, Feb. 2001.
- [65] PETRAGLIA, M. R., BARBOZA, G. “Improved PNLMS Algorithm Employing Wavelet Transform and Sparse Filters”, *Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008)*, pp. 1–8, Aug. 2008.
- [66] PETRAGLIA, M. R., MITRA, S. K. “Adaptive FIR Filter Structure Based on the Generalized Subband Decomposition of FIR Filters”, *IEEE Transactions on Circuits and Systems - II: Analog and Digital Processing*, v. 40, n. 6, pp. 354–362, Jun. 1993.
- [67] VAIDYANATHAN, P. P. *Multirate systems and filter banks*. 1 ed. New Jersey, Prentice Hall, 1993.
- [68] PETRAGLIA, M. R. “New Algorithms for the Rapid Identification of Sparse Impulse Responses”. In: *Signal Processing*, 1 ed., cap. 1, Rijeka: In-Tech, 2010.
- [69] GOLLAMUDI, S., NAGARAJ, S., KAPOOR, S., et al. “Set-Membership Filtering and a Set-Membership Normalized LMS Algorithm with An Adaptive Step Size”, *IEEE Signal Processing Letters*, v. 5, n. 5, pp. 111–114, May 1998.
- [70] NI, J., LI, F. “A Variable Step-Size Matrix Normalized Subband Adaptive Filter”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 6, pp. 1290–1299, Aug. 2010.
- [71] BENESTY, J., GAY, S. L. “An improved PNLMS algorithm”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2002)*, v. 2, pp. 1881–1884, May 2002.

- [72] MØLLER, A. R. *Hearing - Anatomy, Physiology, and Disorders of the Auditory System*. 2 ed. Boston, Elsevier, 2006.
- [73] CAO, X., LIU, R. “General Approach for Blind Source Separation”, *IEEE Transactions on Signal Processing*, v. 44, n. 3, pp. 562–571, Mar. 1996.
- [74] NACEUR, M., LOGHMARI, M. A., BOUSSEMA, M. R. “The Contribution of the Sources Separation Method in the Decomposition of Mixed Pixels”, *IEEE Transactions on Geoscience and Remote Sensing*, v. 42, n. 11, pp. 2642–2653, Nov. 2004.
- [75] ANEMULLER, J., SEJNOWSKI, T., MAKEIG, S. “Complex Independent Component Analysis of Frequency-domain Electroencephalographic Data”, *Neural Networks*, v. 16, n. 9, pp. 1311–1323, Nov. 2003.
- [76] ACERNESE, F., CIARAMELLA, A., MARTINO, S. D., et al. “Neural Networks for Blind-Source Separation of Stromboli Explosion Quakes”, *IEEE Transactions on Neural Networks*, v. 14, n. 1, pp. 167–175, Jan. 2003.
- [77] ERDOGAN, A. T. “Globally Convergent Deflationary Instantaneous Blind Source Separation Algorithm for Digital Communication Signals”, *IEEE Transactions on Signal Processing*, v. 55, n. 5, pp. 2182–2192, May 2007.
- [78] BOURGEOIS, J., MINKER, W. *Time-Domain Beamforming and Blind Source Separation - Speech Input in the Car Environment*. 1 ed. New York, Springer, 2007.
- [79] BENAROYA, L., BIMBOT, F., GRIBONVAL, R. “Audio Source Separation With a Single Sensor”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 1, pp. 191–199, Jan. 2006.
- [80] CICHOCKI, A., ZDUNEK, R., PHAN, A. H., et al. *Nonnegative Matrix and Tensor Factorizations - Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. 1 ed. Singapore, John Wiley & Sons, 2009.
- [81] JOLLIFFE, I. T. *Principal Components Analysis*. 2 ed. New York Berlin Heidelberg, Springer-Verlag, 2002.
- [82] HAYKIN, S. *Neural Networks*. 2 ed. New Jersey, Prentice Hall, 1999.
- [83] DE O. E SOUZA FILHO, J. B. *Classificação Neural de Sinais de Sonar Passivo*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Jul. 2007.

- [84] CARDOSO, J.-F. “Blind Signal Separation: Statistical Principles”, *Proceedings of the IEEE*, v. 86, n. 10, pp. 2009 – 2025, Oct. 1998.
- [85] CARDOSO, J.-F. “Infomax and Maximum Likelihood for Blind Source Separation”, *IEEE Signal Processing Letters*, v. 4, n. 4, pp. 112–114, Apr. 1997.
- [86] BUCHNER, H., AICHNER, R., KELLERMAN, W. “A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures based on Second-Order Statistics”, *IEEE Transactions on Speech and Audio Processing*, v. 1, n. 11, pp. 120–134, Jan. 2005.
- [87] JAMES, B. R. *Probabilidade: Um Curso em Nível Intermediário*. 3 ed. Rio de Janeiro, IMPA, 2004.
- [88] FÉVOTTE, C., GODSILL, S. J. “A Bayesian Approach for Blind Separation of Sparse Sources”, *IEEE Transactions on Audio and Speech Processing*, v. 14, n. 6, pp. 2174–2188, Nov. 2006.
- [89] AMARI, S.-I. “Natural Gradient Works Efficiently in Learning”, *Neural Computation*, v. 10, n. 2, pp. 251–276, Feb. 1998.
- [90] CARDOSO, J.-F., LAHELD, B. H. “Equivariant Adaptive Source Separation”, *IEEE Transactions on Signal Processing*, v. 44, n. 12, pp. 3017 – 3030, Dec. 1996.
- [91] FUKUMIZU, K. “Geometry of Neural networks: Natural Gradient for Learning”. In: *Handbook of Biological Physics, vol. 4: Neuro-informatics and Neural Modelling*, 1 ed., cap. 17, Elsevier, 2001.
- [92] CICHOCKI, A., AMARI, S.-I. *Adaptive Blind Signal and Image Processing - Learning Algorithms and Applications*. 1 ed. West Sussex, John Wiley & Sons, 2002.
- [93] AMARI, S.-I., CHEN, T. P., CICHOCKI, A. “Nonholonomic Orthogonal Learning Algorithm for Blind Source Separation”, *Neural Computation*, v. 12, n. 6, pp. 1463–1484, Jun. 2000.
- [94] GRIBONVAL, R., LESAGE, S. “A Survey of Sparse Component Analysis for Blind Source Separation: Principles, Perspectives, and New Challenges”, *ESANN Proceedings*, pp. 323–330, Apr. 2006.
- [95] KISILEV, P., ZIBULEVSKY, M., ZEEVI, Y. Y. “A Multiscale Framework for Blind Separation of Linearly Mixed Signals”, *Journal of Machine Learning Research*, v. 4, pp. 1–26, Dec. 2003.

- [96] BOFILL, P., ZIBULEVSKY, M. “Underdetermined Blind Source Separation Using Sparse Representations”, *Signal Processing*, v. 81, pp. 2353–2362, Nov. 2001.
- [97] LI, Y., AMARI, S., CICHOCKI, A., et al. “Underdetermined Blind Source Separation Based on Sparse Representation”, *IEEE Transactions on Signal Processing*, v. 54, n. 2, pp. 423–437, Feb. 2006.
- [98] LESAGE, S., KRSTULOVIC, S., GRIBONVAL, R. “Underdetermined Source Separation: Comparison of Two Approaches Based on Sparse Decompositions”, *ICA*, pp. 633–640, Mar. 2006.
- [99] WANG, F., LI, H., LI, R. “Applications of Wavelet to Independent Component Analysis”, *Proceedings of the 6th World Congress on Intelligent Control and Automation*, pp. 2930–2934, Oct. 2006.
- [100] AZZERBONI, B., FINOCCHIO, G., IPSALE, M., et al. “A New Approach to Detection of Muscle Activation by Independent Component Analysis and Wavelet Transform”, *Lecture Notes in Computer Science*, v. 2486, pp. 109–116, Sept. 2002.
- [101] VINCENT, E., GRIBONVAL, R., FÉVOTTE, C. “Performance Measurement in Blind Audio Source Separation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 4, pp. 1462–1469, Jul. 2006.
- [102] DOUGLAS, S. C., SAWADA, H., MAKINO, S. “Natural Gradient Multi-channel Blind Deconvolution and Speech Separation Using Causal FIR Filters”, *IEEE Transactions on Audio and Speech Processing*, v. 13, n. 1, pp. 92–104, Jan. 2005.
- [103] PAINTER, T., SPANIAS, A. “Perceptual Coding of Digital Audio”, *Proceedings of the IEEE*, v. 88, n. 4, pp. 451–512, Apr. 2000.
- [104] ARAKI, S., MUKAI, R., MAKINO, S., et al. “The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutional Mixtures of Speech”, *IEEE Transactions on Speech and Audio Processing*, v. 11, n. 2, pp. 109–116, Mar. 2003.
- [105] WINTER, S., SAWADA, H., MAKINO, S. “Geometrical Understanding of the PCA Subspace Method for Overdetermined Blind Source Separation”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2003)*, v. 2, n. 2, pp. 769 – 772, Jun. 2003.

- [106] SAWADA, H., ARAKI, S., MUKAI, R., et al. “Blind Source Separation with Different Sensor Spacing and Filter Length for each Frequency Range”, *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 465–474, Nov. 2002.
- [107] HYVÄRINEN, A. “Fast and Robust Fixed-point Algorithm for Independent Component Analysis”, *IEEE Transactions on Neural Networks*, v. 10, n. 3, pp. 626–634, May 1999.
- [108] MAKINO, S., SAWADA, H., MUKAI, R., et al. “Blind Source Separation of Convulsive Mixtures of Audio Signals in Frequency Domain”. In: *Topics in Acoustic Echo and Noise Control*, 1 ed., cap. 3, Springer, 2006.
- [109] MEI, T., YIN, F., WANG, J. “Blind Source Separation Based on Cumulants With Time and Frequency Non-Properties”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, n. 6, pp. 1099–1108, Aug. 2009.
- [110] SAWADA, H., ARAKI, S., MUKAI, R., et al. “Grouping Separated Frequency Components by Estimating Propagation Model Parameters in Frequency-Domain Blind Source Separation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 5, pp. 1592–1603, Jul. 2007.
- [111] SMARAGDIS, P. J. *Information Theoretic Approaches to Source Separation*. Dissertação de mestrado, Massachusetts Institute of Technology, Boston, Jun. 1997.
- [112] SAWADA, H., MUKAI, R., ARAKI, S., et al. “Polar Coordinate Based Non-linear Function for Frequency-Domain Blind Source Separation”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2002)*, v. 1, pp. 1001 – 1004, May 2002.
- [113] SOLVANG, H. K., NAGAHARA, Y., ARAKI, S., et al. “Frequency-Domain Pearson Distribution Approach for Independent Component Analysis (FD-Pearson-ICA) in Blind Source Separation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, n. 4, pp. 639–649, May 2009.
- [114] KIM, T., ATTIAS, H. T., LEE, S.-Y., et al. “Blind Source Separation Exploiting Higher-Order Frequency Dependencies”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 1, pp. 70–79, Jan. 2007.

- [115] IKRAM, M. Z., MORGAN, D. R. “Permutation Inconsistency in Blind Speech Separation: Investigation and Solutions”, *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 1, pp. 1–13, Jan. 2005.
- [116] MAZUR, R., MERTINS, A. “A Method for Filter Equalization in Convolutional Blind Source Separation”, *Proceedings of the 9th ICA International Conference ICA 2010*, pp. 328–336, Sept. 2010.
- [117] PARRA, L., SPENCE, C. “Convolutional Blind Separation of Non-Stationary Sources”, *IEEE Transactions on Acoustic, Speech and Signal Processing*, v. 8, n. 3, pp. 320–327, May 2000.
- [118] SARUWATARI, H., KURITA, S., TAKEDA, K., et al. “Blind Source Separation Combining Independent Component Analysis and Beamforming”, *EURASIP J. Appl. Signal Processing*, v. 2003, n. 11, pp. 1135–1146, Nov. 2003.
- [119] MUKAI, R., SAWADA, H., ARAKI, S., et al. “Near-Field Frequency-Domain Blind Source Separation for Convolutional Mixtures”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2004)*, v. IV, pp. 49–52, May 2004.
- [120] ANEMÜLLER, J., KOLLMEIER, B. “Amplitude Modulation Decorrelation for Convolutional Blind Source Separation”, *Proc. ICA*, pp. 215–220, Jun. 2000.
- [121] SAWADA, H., MUKAI, R., ARAKI, S., et al. “A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation”, *IEEE Transactions on Speech and Audio Processing*, v. 12, n. 5, pp. 530–538, Sept. 2004.
- [122] ROBLEDO-ARNUNCIO, E., JUANG, B. “Issues in Frequency Domain Blind Source Separation - A Critical Revisit”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2005)*, v. 5, pp. 281 – 284, Mar. 2005.
- [123] IKRAM, M. Z., MORGAN, D. R. “Exploring Permutation Inconsistency in Blind Separation of Speech Signals in a Reverberant Environment”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2000)*, v. 2, pp. 1041–1044, Jun. 2000.
- [124] SAWADA, H., MUKAI, R., ARAKI, S., et al. “Multiple Source Localization Using Independent Component Analysis”, *Antennas and Propagation Society International Symposium 2005*, v. 4B, pp. 81–84, Jul. 2005.

- [125] ABHAYAPALA, T. D., KENNEDY, R. A., WILLIAMSON, R. C. “Spatial Aliasing for Near-Field Sensor Arrays”, *Electronic Letters*, v. 35, n. 10, pp. 764–765, mai. 1999.
- [126] MUKAI, R., SAWADA, H., ARAKI, S., et al. “Frequency-Domain Blind Source Separation of Many Speech Signals Using Near-Field and Far-Field Models”, *EURASIP Journal on Applied Signal Processing*, v. 2006, pp. 1–13, Jan. 2006.
- [127] KURITA, S., SARUWATARI, H., KAJITA, S., et al. “Evaluation of Blind Signal Separation Method Using Directivity Pattern Under Reverberant Conditions”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2000)*, v. 5, pp. 3140–3143, Jun. 2000.
- [128] SCHOBEN, D. W. E., SOMMEN, P. C. W. “A Frequency Domain Blind Signal Separation Method Based on Decorrelation”, *IEEE Transactions on Signal Processing*, v. 50, n. 8, pp. 1855–1865, Ago. 2002.
- [129] SAWADA, H., ARAKI, S., MAKINO, S. “Measuring Dependence of Bin-wise Separated Signals for Permutation Alignment in Frequency-domain BSS”, *IEEE International Symposium on Circuits and Systems (ISCAS 2007)*, pp. 3247–3250, May 2007.
- [130] MAZUR, R., MERTINS, A. “A Sparsity Based Criterion for Solving the Permutation Ambiguity in Convolutional Blind Source Separation”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2011)*, pp. 1996 – 1999, Mai. 2011.
- [131] WANG, L., DING, H., YIN, F. “A Region-Growing Permutation Alignment Approach in Frequency-Domain Blind Source Separation of Speech Mixtures”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 3, pp. 549–557, Mar. 2011.
- [132] MATSUOKA, K., NAKASHIMA, S. “Minimal Distortion Principle for Blind Source Separation”, *Proc. of ICA 2001*, pp. 722–727, Jun. 2001.
- [133] MATSUOKA, K. “Minimal Distortion Principle for Blind Source Separation”, *Proc. 41st SICE Annu. Conf.*, v. 4, pp. 2138–2143, Aug. 2002.
- [134] MAZUR, R., MERTINS, A. “Using the Scaling Ambiguity for Filter Shortening in Convolutional Blind Source Separation”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2009)*, pp. 1709–1712, Apr. 2009.

- [135] SMARAGDIS, P. “Blind Separation of Convolved Mixtures in the Frequency Domain”, *Neurocomputing*, v. 22, n. 1-3, pp. 21–34, Nov. 1998.
- [136] IKEDA, S., MURATA, N., ZIEHE, A. “An Approach to Blind Source Separation based on Temporal Structure of Speech Signals”, *Neurocomputing*, v. 42, n. 1-4, pp. 1–24, Oct. 2001.
- [137] YASHITA, M., HAMADA, N. “Time-Frequency Masking Method Using Wavelet Transform for BSS Problem”, *2006 IEEE Region 10 Conference (TENCON 2006)*, pp. 1–4, Nov. 2006.
- [138] SAWADA, H., ARAKI, S., MUKAI, R., et al. “Blind Extraction of a Dominant Source from Many Mixtures of Many Sources Using ICA and Time-Frequency Masking”, *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS 2005)*, v. 6, pp. 5882–5885, May 2005.
- [139] ARAKI, S., SAWADA, H., MUKAI, R., et al. “Underdetermined Sparse Source Separation of Convulsive Mixtures with Observation Vector Clustering”, *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS 2006)*, pp. 3594–3597, May 2006.
- [140] CLARK, F. S. P., PETRAGLIA, M. R., HADDAD, D. B. “A New Initialization Method for Frequency-Domain Blind Source Separation Algorithms”, *IEEE Signal Processing Letters*, v. 18, n. 6, pp. 343–346, Jun. 2011.
- [141] ARAKI, S., MAKINO, S., SAWADA, H., et al. “Reducing Musical Noise by a Fine-Shift Overlap-Add Method Applied to Source Separation Using a Time-Frequency Mask”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006)*, v. 3, pp. 81–84, Apr. 2005.
- [142] WANG, Z. “Comparison of Four Kinds of Fuzzy C-Means Clustering Methods”, *International Symposium on Information Processing (ISIP 2010)*, pp. 563–566, Oct. 2010.
- [143] AVARGEL, Y., COHEN, I. “System Identification in the Short-Time Fourier Transform Domain with Crossband Filtering”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 4, pp. 1305–1319, May 2007.
- [144] AVARGEL, Y., COHEN, I. “Adaptive System Identification in the Short-Time Fourier Transform Domain Using Cross-Multiplicative Transfer Function Approximation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 1, pp. 162–173, Jan. 2008.

- [145] AVARGEL, Y., COHEN, I. “On Multiplicative Transfer Function Approximation in the Short-Time Fourier Transform Domain”, *IEEE Signal Processing Letters*, v. 14, n. 5, pp. 337–340, May 2007.
- [146] MAZUR, R., MERTINS, A. “On Separation Performance Enhancement in Convolutional Blind Source Separation”, *Proc. of the 32nd Asilomar Conference on Signals, Systems and Computers (ACSSC’08)*, pp. 1718–1721, Oct. 2008.
- [147] DE M. PREGO, T., DE LIMA, A. A., NETTO, S. L., et al. “A Blind Algorithm for Reverberation-Time Estimation Using Subband Decomposition of Speech Signals”, *J. Acoust. Soc. Am.*, v. 131, pp. 2811–2816, Apr. 2012.
- [148] WEN, J. Y. C., HABETS, E. A. P., NAYLOR, P. A. “Blind Estimation of Reverberation Time Based on the Distribution of Signal Decay Rates”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2008)*, pp. 329–332, Apr. 2008.
- [149] LÖLLMAN, H. W., VARY, P. “Estimation of the Reverberation Time in Noisy Environments”, *Proceedings of the IEEE International Workshop Acoustic Echo and Noise Control*, pp. 1–4, Sept. 2008.
- [150] WHYTE, K. R. L., SHEIKHZADEH, H., BRENNAN, R. L., et al. “Complexity Reduction of Partial Update Oversampled Subband Adaptive Algorithms by Selective Pruning of Polyphase Components”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2005)*, v. 5, pp. 673–676, Mar. 2005.
- [151] PETRAGLIA, M. R., ALVES, R. G., DINIZ, P. S. R. “New Structures for Adaptive Filtering in Subbands with Critical Sampling”, *IEEE Transactions on Signal Processing*, v. 48, n. 12, pp. 3316–3327, Dec. 2000.
- [152] GILLOIRE, A., VETTERLI, M. “Adaptive Filtering in Subbands with Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancellation”, *IEEE Transactions on Signal Processing*, v. 40, n. 8, pp. 1862–1875, Aug. 1992.
- [153] PETRAGLIA, M. R., ALVES, R. G. “A New Adaptive Subband Structure with Critical Sampling”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 1999)*, v. 4, pp. 1861–1864, Aug. 1999.
- [154] BUCHNER, H., AICHNER, R., KELLERMAN, W. “TRINICON-based Blind System Identification with Application to Multiple-Source Localization

- and Separation”. In: *Blind Speech Separation*, 1 ed., cap. 4, Springer, 2007.
- [155] MUKAI, R., ARAKI, S., SAWADA, H., et al. “Removal of Residual Crosstalk Components in Blind Source Separation Using Time-Delayed Spectral Subtraction”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2002)*, v. 2, pp. 1789–1792, May 2002.
- [156] MUKAI, R., ARAKI, S., SAWADA, H., et al. “Removal of Residual Crosstalk Components in Blind Source Separation Using LMS filters”, *Neural Networks for Signal Processing*, pp. 435–444, Sept. 2002.
- [157] PARK, K. S., PARK, J. S., SON, K. S., et al. “Postprocessing with Wiener Filtering Technique for Reducing Residual Crosstalk in Blind Source Separation”, *IEEE Signal Processing Letters*, v. 13, n. 12, pp. 749–751, Dec. 2006.
- [158] AICHNER, R., ZOURUB, M., BUCHNER, H., et al. “Post-Processing for Convolutional Blind Source Separation”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006)*, v. 5, pp. 37–40, May 2006.
- [159] NAKAGAWA, K., TAKAHASHI, K. “Analysis and Improvement of Post-Processing for BSS Aiming to Suppress Recognizable Crosstalk”, *SICE*, pp. 1406–1413, Sept. 2007.
- [160] NGUYEN, T. Q. “Digital Filter Bank Design Quadratic-constrained Formulation”, *IEEE Transactions on Speech and Audio Processing*, v. 43, n. 9, pp. 2103–2108, Sept. 1995.
- [161] PETRAGLIA, M. R., BATALHEIRO, P. B. “Filter Bank Design for a Subband Adaptive Filtering Structure with Critical Sampling”, *IEEE Transactions on Circuits and Systems - I: Regular Papers*, v. 51, n. 6, pp. 1194–1202, Jun. 2004.
- [162] VLEESCHOUWER, C. D., MACQ, B. “Subband Dictionaries for Low-Cost Matching Pursuit of Video Residues”, *IEEE Transactions on Circuits and Systems for Video Technology*, v. 9, n. 7, pp. 984–993, Oct. 1999.
- [163] TROPP, J. A. “Greed is Good: Algorithmic Results for Sparse Approximation”, *IEEE Transactions on Information Theory*, v. 50, n. 10, pp. 2231–2242, Oct. 2004.

- [164] ELAD, M., BRUCKSTEIN, A. M. “A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases”, *IEEE Transactions on Information Theory*, v. 48, n. 9, pp. 2558–2567, Sept. 2002.
- [165] IERODIACONOU, S., BYRNE, J., BULL, D. R., et al. “Unsupervised Image Compression Using Graphcut Texture Synthesis”, *16th IEEE International Conference on Image Processing (ICIP 2009)*, pp. 2289–2292, Nov. 2009.
- [166] JOST, P., VANDERGHEYNST, P., FROSSARD, P. “Tree-Based Pursuit: Algorithm and Properties”, *IEEE Transactions on Signal Processing*, v. 54, n. 12, pp. 4685–4697, Dec. 2006.
- [167] TOSIC, I., FROSSARD, P. “Dictionary Learning”, *IEEE Signal Processing Magazine*, v. 28, n. 2, pp. 27–38, Mar. 2011.
- [168] LOVISOLO, L., TCHEOU, M. P., DA SILVA, E. A. B., et al. “Modeling of Electric Disturbance Signals Using Damped Sinusoids Via Atomic Decompositions and Its Applications”, *EURASIP Journal on Applied Signal Processing*, pp. 1–16, Mar. 2007.
- [169] MALLAT, S. G., ZHANG, Z. “Matching Pursuits with Time-Frequency Dictionaries”, *IEEE Transactions on Signal Processing*, v. 41, n. 12, pp. 3397 – 3415, Dec. 1993.
- [170] PATI, Y. C., REZAIIFAR, R., KRISHNAPRASAD, P. S. “Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decompositions”, *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers (ACSSC’93)*, v. 1, pp. 40–44, Nov. 1993.
- [171] REBOLLO-NEIRA, L., LOWE, D. “Optimized Orthogonal Matching Pursuit Approach”, *IEEE Signal Processing Letters*, v. 9, n. 4, pp. 137–140, Apr. 2002.
- [172] CHEN, S. S., DONOHO, D. L., SAUNDERS, M. A. “Atomic Decomposition by Basis Pursuit”, *SIAM Journal on Scientific Computing*, v. 20, n. 1, pp. 33–61, Aug. 1998.
- [173] FRIEDMAN, J. H., STUETZLE, W. “Projection Pursuit Regression”, *Journal of American Statistical Association*, v. 13, n. 376, pp. 435–475, Dec. 1981.
- [174] SCHMID-SAUGEON, P., ZAKHOR, A. “Dictionary Design for Matching Pursuit and Application to Motion-compensated Video Coding”, *IEEE*

Transactions on Circuits and Systems for Video Technology, v. 14, n. 6, pp. 880–886, Jun. 2004.

- [175] FERRANDO, S. E., KOLASA, L. A., KOVACEVIC, N. “Algorithm 820: A Flexible Implementation of Matching Pursuit for Gabor Functions on the Interval”, *ACM Transactions on Mathematical Software*, v. 28, n. 3, pp. 337–353, Sept. 2002.
- [176] LOVISOLO, L., DA SILVA, E. A. B., RODRIGUES, M. A. M., et al. “Efficient Coherent Adaptive Representations of Monitored Electric Signals in Power Systems using Damped Sinusoids”, *IEEE Transactions on Signal Processing*, v. 53, n. 10, pp. 3831–3846, Oct. 2005.
- [177] LOVISOLO, L. *Representações de Sinais Usando Dicionários Redundantes*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Ago. 2006.
- [178] TCHEOU, M. P. *Compressão de Sinais Usando Decomposições Atômicas com Base em Dicionários Redundantes*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Mar. 2011.
- [179] SUGDEN, P., CANAGARAJAH, N. “Underdetermined Noisy Blind Separation Using Dual Matching Pursuits”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2004)*, v. V, pp. 557–560, May 2004.
- [180] GRIBONVAL, R. “Sparse Decomposition of Stereo Signals with Matching Pursuit and Application to Blind Separation of More Than Two Sources from a Stereo Mixture”, *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2002)*, v. 3, pp. 3057–3060, May 2002.
- [181] STURM, B. L., SHYNK, J. J., DAUDET, L., et al. “Dark Energy in Sparse Atomic Estimations”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 3, pp. 671–676, Mar. 2008.
- [182] STURM, B. L., SHYNK, J. J. “Sparse Approximation and the Pursuit of Meaningful Signal Models With Interference Adaptation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 3, pp. 461–472, Mar. 2010.
- [183] LOVISOLO, L., DA SILVA, E. A. B., DINIZ, P. S. R. “On the Statistics of Matching Pursuit Angles”, *Signal Processing*, v. 90, pp. 3164–3184, Dec. 2010.

- [184] JAFARI, M. G., PLUMBLEY, M. D. “Fast Dictionary Learning for Sparse Representations of Speech Signals”, *IEEE Journal of Selected Topics in Signal Processing*, v. 5, n. 5, pp. 1025–1031, Sept. 2011.
- [185] R. RUBINSTEIN, T. P., ELAD, M. “Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model”, *IEEE Transactions on Signal Processing*, v. 61, n. 3, pp. 661–677, Feb. 2013.
- [186] ITOYAMA, K., GOTO, M., KOMATANI, K., et al. “Simultaneous Processing of Sound Source Separation and Musical Instrument Identification Using Bayesian Spectral Modeling”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2011)*, pp. 3816–3819, May 2011.
- [187] ENGAN, K., AASE, S. O., HUSOY, J. H. “Multi-frame Compression: Theory and Design”, *EURASIP Signal Processing*, v. 80, n. 10, pp. 2121–2140, Oct. 2000.
- [188] KREUTZ-DELGADO, K., MURRAY, J., RAO, D., et al. “Dictionary Learning Algorithms for Sparse Representations”, *Neural Comput.*, v. 15, n. 2, pp. 349–396, Feb. 2003.
- [189] GERSHO, A., GRAY, R. M. *Vector Quantization and Signal Compression*. 2 ed. Boston, Kluwer Academic Publishers, 1992.
- [190] AHARON, M., ELAD, M., BRUCKSTEIN, A. “K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”, *IEEE Transactions on Signal Processing*, v. 54, n. 11, pp. 4311–4322, Nov. 2006.
- [191] SMITH, L. N., ELAD, M. “Improving Dictionary Learning: Multiple Dictionary Updates and Coefficient Reuse”, *IEEE Signal Processing Letters*, v. 20, n. 1, pp. 79–82, Jan. 2013.
- [192] ENGAN, K., AASE, S. O., HAKON-HUSOY, J. H. “Method of Optimal Directions for Frame Design”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1999)*, v. 5, pp. 2443–2446, Mar. 1999.
- [193] BENVENISTE, A., GOURSAT, M., RUGET, G. “Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications”, *IEEE Transactions on Automatic Control*, v. 25, n. 3, pp. 385–399, Jun. 1980.

- [194] SATO, Y. “Two Extensional Applications of the Zero-Forcing Equalization Method”, *IEEE Transactions on Communications*, v. 23, n. 6, pp. 684–687, Jun. 1975.
- [195] GODARD, D. “Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems”, *IEEE Transactions on Communications*, v. 28, n. 11, pp. 1867–1875, Nov. 1980.
- [196] TREICHLER, J., AGEE, B. “A New Approach to Multipath Correction of Constant Modulus Signals”, *IEEE Transactions on Acoustic, Speech and Signal Processing*, v. 31, n. 2, pp. 459–472, Apr. 1983.
- [197] AMARI, S., CICHOCKI, A. “Adaptive Blind Signal Processing - Neural Network Approaches”, *Proceedings of the IEEE*, v. 86, n. 10, pp. 2026–2048, Oct. 1998.
- [198] SUN, X., DOUGLAS, S. C. “Mean Square Error Analyses of Multichannel Blind Deconvolution Algorithms”, *Proceedings of the 35th. Annual Asilomar Conference on Signals, Systems, and Computers*, v. 1, pp. 648–652, Nov. 2001.
- [199] CHEVALIER, P., ALBERA, L., COMON, P., et al. “Comparative Performance Analysis of Eight Blind Source Separation Methods on Radio-communications Signals”, *IEEE International Joint Conference on Neural Networks (IJCNN 2004)*, pp. 273–278, Jul. 2004.
- [200] SINGH, Y., RAI, C. S. “A Comparison of BSS Algorithms”, *IEEE International Joint Conference on Neural Networks (IJCNN 2001)*, v. 1, pp. 932–935, Jul. 2001.
- [201] LEBLANC, J. P., JR., C. R. J. “Global CMA Error Surface Characteristics, Source Statistic Effects: Polytopes and Manifolds”, *13th International Conference on Digital Signal Processing Proceedings*, pp. 131–134, Jul. 1997.
- [202] CHUNG, W., JR., C. R. J. “Characterization of the Regions of Convergence of CMA Adapted Blind Fractionally Spaced Equalizer”, *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, v. 1, pp. 493–497, Nov. 1998.
- [203] CUSANI, R., LAURENTI, A. “Convergence Analysis of the CMA Blind Equalizer”, *IEEE Transactions on Communications*, v. 43, n. 2, pp. 1304–1307, Feb. 1995.

- [204] NEVES, A., PANAZIO, C. “A Class of Channels Resulting in Ill-Convergence for CMA in Decision Feedback Equalizers”, *IEEE Transactions on Signal Processing*, v. 58, n. 11, pp. 5736–5746, Nov. 2010.
- [205] MAI, J., SAYED, A. H. “A Feedback Approach to the Steady-State Performance of Fractionally Spaced Blind Adaptive Equalizers”, *IEEE Transactions on Signal Processing*, v. 48, n. 1, pp. 80–91, Jan. 2000.
- [206] AMARI, S., DOUGLAS, S. C., CICHOCKI, A., et al. “Multichannel Blind Deconvolution and Equalization using the Natural Gradient”, *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, v. 13, n. 1, pp. 101–104, Apr. 1997.
- [207] ZHANG, L., CICHOCKI, A., AMARI, S. “Multichannel Blind Deconvolution of Nonminimum-Phase Systems Using Filter Decomposition”, *IEEE Transactions on Signal Processing*, v. 52, n. 5, pp. 1430 – 1442, May 2004.
- [208] LEBLANC, J. P., DOGANCAI, K., KENNEDY, R. A., et al. “Effects of Input Data Correlation on the Convergence of Blind Adaptive Equalizers”, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1994)*, v. 3, pp. 313–316, Apr. 1994.
- [209] LJUNG, L. “On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes”, *IEEE Transactions on Automatic Control*, v. 22, n. 4, pp. 539–551, Aug. 1977.
- [210] LJUNG, L. “Analysis of Recursive Stochastic Algorithms”, *IEEE Transactions on Automatic Control*, v. 22, n. 4, pp. 551–575, Aug. 1977.
- [211] KUSHNER, H. J. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*. 1 ed. Boston, Birkhäuser, 1990.
- [212] POWELL, T. D., WIBERG, D. M., LJUNQUIST, D. “A Counterexample of Proof of Convergence Using Asymptotic Averaging”, *Proceedings of the 32th Conference on Decision and Control*, pp. 2060–2061, Dec. 1993.
- [213] KHALIL, H. K. *Nonlinear Systems*. 3 ed. New Jersey, Prentice Hall, 2002.
- [214] ZUFIRIA, P. J. “On the Discrete-Time Dynamics of the Basic Hebbian Neural-Network Node”, *IEEE Transactions on Neural Networks*, v. 13, n. 6, pp. 1342–1352, Nov. 2002.
- [215] REGALIA, P. A. *Adaptive IIR Filtering in Signal Processing and Control*. 1 ed. New York, Marcel Dekker, 1995.