



Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Engenharia de Computação
e Automação Industrial

**CONTRIBUIÇÕES AO PROBLEMA DE SEPARAÇÃO
CEGA DE FONTES, COM ÊNFASE NO ESTUDO DE
SINAIS ESPARSOS**

Autor: Everton Zaccaria Nadalin

Orientador: Prof. Dr. Romis Ribeiro de Faissol Attux

Coorientador: Prof. Dr. Ricardo Suyama

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: **Engenharia de Computação.**

Banca Examinadora

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

Prof. Dr. Luiz Wagner Pereira Biscainho — UFRJ

Prof. Dr. Phillip Mark Seymour Burt — USP

Prof. Dr. Rafael Santos Mendes — DCA/FEEC/UNICAMP

Prof. Dr. Renato da Rocha Lopes — DECOM/FEEC/UNICAMP

Campinas – SP

22/08/2011

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

N121c Nadalin, Everton Zaccaria
 Contribuições ao problema de separação cega de
fontes, com ênfase no estudo de sinais esparsos / Everton
Zaccaria Nadalin. --Campinas, SP: [s.n.], 2011.

 Orientadores: Romis Ribeiro de Faissol Attux,
Ricardo Suyama.

 Tese de Doutorado - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

 1. Processamento digital de sinais. 2. Filtros
adptativos. 3. Computação bioinspirada. I. Attux,
Romis Ribeiro de Faissol. II. Suyama, Ricardo. III.
Universidade Estadual de Campinas. Faculdade de
Engenharia Elétrica e de Computação. IV. Título.

Título em Inglês: Contributions to the problem of blind source separation,
with emphasis on the study of sparse signals

Palavras-chave em Inglês: Digital signal processing, Adaptive filters, Bio-
inspired computing

Área de concentração: Engenharia de Computação

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Luiz Wagner Pereira Biscainho, Phillip Mark
Seymour Burt, Rafael Santos Mendes, Renato da
Rocha Lopes

Data da defesa: 22-08-2011

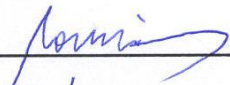
Programa de Pós Graduação: Engenharia Elétrica


COMISSÃO JULGADORA - TESE DE DOUTORADO

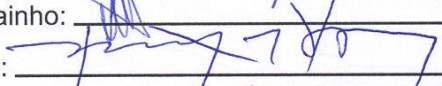
Candidato: Everton Zaccaria Nadalin

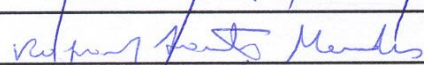
Data da Defesa: 22 de agosto de 2011

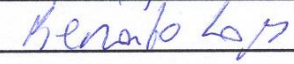
Título da Tese: "Contribuições ao Problema de Separação Cega de Fontes, com Ênfase no Estudo de Sinais Esparsos"

Prof. Dr. Romis Ribeiro de Faissol Attux (Presidente): 

Prof. Dr. Luiz Wagner Pereira Biscaíno: 

Prof. Dr. Phillip Mark Seymour Burt: 

Prof. Dr. Rafael Santos Mendes: 

Prof. Dr. Renato da Rocha Lopes: 

Resumo

Neste trabalho, foi estudado o problema de Separação Cega de Fontes (BSS), com ênfase nos casos chamados de subparametrizados, isto é, em que o número de fontes é maior do que o de misturas. A primeira contribuição proposta foi a de um limitante relacionado ao erro de inversão intrínseco ao problema quando é utilizada uma estrutura linear de separação. As outras contribuições estão relacionadas à hipótese de que as fontes são esparsas: *i*) uma proposta de metodologia híbrida, que se utiliza de conceitos baseados em independência e esparsidade dos sinais de forma simultânea para estimar tanto o sistema misturador quanto o número de fontes existentes em misturas com dois sensores; *ii*) a utilização de ferramentas de otimização baseadas na operação do sistema imunológico para a estimação do sistema misturador em problemas intrinsecamente multimodais; por fim, *iii*) uma proposta de utilização de um critério baseado em esparsidade para separação de fontes, sendo derivado um processo de otimização baseado na norma ℓ_1 para este fim.

Palavras-chave: separação cega de fontes, esparsidade, análise de componentes esparsos, sistemas subparametrizados, otimização bioinspirada.

Abstract

In this work, we studied the problem of Blind Source Separation (BSS), with emphasis on cases referred to as underdetermined, which occur when the number of sources is greater than the number of mixtures. The first contribution was a proposal of a bound to the inversion error that is intrinsic to the problem when a linear structure is used to perform separation. The other contributions are related to the hypothesis that the signals of the sources are sparse: i) the proposal of a hybrid methodology that employs concepts based on signal independence and sparsity to simultaneously estimate both the mixing system and the number of existing sources in mixtures with two sensors; ii) the use of optimization tools based on the modus operandi of the immune system to estimate the mixing system in problems that are inherently multimodal; finally, iii) the use of a criterion based on sparsity for source separation, which is derived from an optimization process based on the ℓ_1 norm.

Keywords: blind source separation, sparsity, sparse component analysis, underdetermined systems, bio-inspired optimization.

Agradecimento

Aos meus pais, por toda sua dedicação, carinho e por me incentivarem a seguir este caminho. À minha querida avó, pelas boas conversas, pelas tardes agradáveis e pelo complemento financeiro.

À minha querida e bela Denise, por ter estado sempre ao meu lado e que certamente continuará participando ativamente dos melhores momentos da minha vida.

Aos amigos Romis e Suyama, por me ajudarem a trilhar este caminho do doutorado.

A Levy, Kazuo e Leonardo, por colaborarem diretamente com alguns dos resultados apresentados nesta tese.

Ao professor Roberto Noritaka Ohashi, por ter aceitado minha proposta de projeto de fim de curso no COTUCA (1998), que era justamente tentar separar fontes de áudio.

Aos professores João Marcos Romano e Renato Lopes, e à Celi, que, juntamente com o Romis, conseguiram proporcionar um ambiente de trabalho tão agradável e proveitoso.

Ao grande amigo Diogo, que tem batalhado comigo desde Circuitos I, no primeiro ano de graduação, até sabe lá quando.

A todas as pessoas do DSPCom, pelos últimos 4 excelentes anos.

Ao pessoal do ondevamo e amigos que me acompanharam de um modo ou de outro durante esta jornada.

A toda minha família que me apoiou.

A todos que, de uma forma ou outra, contribuíram para que este trabalho fosse realizado.

À CAPES, pelo apoio financeiro.

À Unicamp, por toda a estrutura oferecida.

Aos que pagam os impostos. Espero poder retornar de alguma forma todo o investimento feito em minha formação.

A Deus.

Aos meus queridos pais.

À minha querida família.

À minha querida Denise.

*“If at first, the idea is not absurd, then
there is no hope for it”*

Albert Einstein

*“Whoever wants to portray something
that does not exist has to obey certain
rules. Those rules are more or less the
same as for the teller of fairy tales: he
has to apply the function of contrasts;
he has to cause a shock.”*

M. C. Escher

Sumário

CAPÍTULO 1	INTRODUÇÃO	1
1.1	ORGANIZAÇÃO DO CONTEÚDO DO TRABALHO	2
1.2	PUBLICAÇÕES	3
CAPÍTULO 2	SEPARAÇÃO CEGA DE FONTES	5
2.1	ANÁLISE DE COMPONENTES INDEPENDENTES.....	7
2.1.1	<i>ICA via Informação Mútua</i>	10
2.1.2	<i>ICA via Não-Gaussianidade</i>	11
CAPÍTULO 3	ESPARSIDADE E ANÁLISE DE COMPONENTES ESPARSOS	15
3.1	QUANTIFICAÇÃO ESPARSIDADE.....	16
3.2	ANÁLISE DE COMPONENTES ESPARSOS	19
3.2.1	<i>Escolha de uma Base Adequada</i>	20
3.2.2	<i>Estimação do Número de Fontes e da Matriz de Mistura</i>	21
3.2.3	<i>Separação / Reconstrução das Fontes</i>	26
3.3	CONSIDERAÇÕES FINAIS	27
CAPÍTULO 4	LIMITANTE RELACIONADO À ESTIMAÇÃO DE FONTES POR ESTRUTURAS LINEARES	29
4.1	CRITÉRIO DE WIENER (MSE)	30
4.2	CASO DE SEPARAÇÃO DE FONTES	34
4.2.1	<i>Análise da Soma dos MSEs Ótimos</i>	34
4.2.2	<i>Simulações e Resultados</i>	36
4.2.3	<i>Análise para Sinais Correlacionados</i>	39
4.3	CASO DE EQUALIZAÇÃO	40
4.3.1	<i>Simulações e Resultados</i>	41
4.4	CONSIDERAÇÕES FINAIS	42
CAPÍTULO 5	SEPARAÇÃO DE SINAIS BASEADA EM ESPARSIDADE E INDEPENDÊNCIA	43
5.1	A UTILIZAÇÃO DE ICA EM PROBLEMAS SUBPARAMETRIZADOS	44
5.2	IDENTIFICAÇÃO DO SISTEMA MISTURADOR.....	47
5.2.1	<i>Estimação do Número de Clusters</i>	49
5.3	MÉTODO PROPOSTO	50
5.4	RESULTADOS DE SIMULAÇÃO	52
5.4.1	<i>Descrição dos Cenários</i>	52
5.4.2	<i>Resultados da Estimação do Número de Fontes</i>	55
5.4.3	<i>Resultados da Estimação da Matriz de Mistura</i>	57
5.5	CONSIDERAÇÕES FINAIS	60
CAPÍTULO 6	ESTIMAÇÃO DA MATRIZ DE MISTURA ATRAVÉS DE OTIMIZAÇÃO MULTIMODAL BIOINSPIRADA	61
6.1	BREVE DESCRIÇÃO DE SISTEMAS IMUNOLÓGICOS ARTIFICIAIS	62

6.1.1	<i>Funcionamento da Opt-aiNet</i>	62
6.2	MÉTODO PROPOSTO.....	64
6.2.1	<i>Opt-aiNet Aplicada a SCA</i>	65
6.2.2	<i>Identificação das Colunas da Matriz de Mistura</i>	68
6.3	SIMULAÇÕES E RESULTADOS.....	68
6.4	CONSIDERAÇÕES FINAIS.....	76
CAPÍTULO 7 SEPARAÇÃO / EXTRAÇÃO DE FONTES BASEADAS NUM CRITÉRIO DE QUANTIFICAÇÃO DE ESPARSIDADE.....		77
7.1	MEDIDA DE ESPARSIDADE BASEADA NA NORMA ℓ_1	78
7.2	EXTRAÇÃO DE FONTES – ANÁLISE DA FUNÇÃO CUSTO.....	79
7.2.1	<i>Algoritmo de Extração Baseado no Gradiente</i>	84
7.2.2	<i>Extração de Fontes sem Ortogonalidade Disjunta</i>	84
7.2.3	<i>Simulações e Resultados</i>	88
7.3	SEPARAÇÃO DE FONTES UTILIZANDO A NORMA ℓ_1	91
7.3.1	<i>Método de Separação Baseado no Gradiente</i>	94
7.3.2	<i>Separação de Fontes sem Ortogonalidade Disjunta</i>	94
7.3.3	<i>Simulações e Resultados</i>	96
7.4	ESTIMAÇÃO DA MATRIZ DE MISTURA NO CASO SUBPARAMETRIZADO.....	97
7.5	CONSIDERAÇÕES FINAIS.....	98
CAPÍTULO 8 CONCLUSÕES E PERSPECTIVAS.....		99
BIBLIOGRAFIA		101

Lista de figuras

Fig. 2.1 – Cenário clássico de separação de fontes	6
Fig. 2.2 – Exemplo de mistura de duas fontes independentes com distribuição uniforme. Distribuição (a) das fontes, (b) das misturas antes e (c) após o branqueamento.....	9
Fig. 3.1 – Ilustração geométrica do resultado de otimização usando medidas ℓ_0 , $\ell_{1/2}$, ℓ_1 e ℓ_2 . Mais detalhes podem ser obtidos em [COMON e JUTTEN, 2010].....	18
Fig. 3.2 – Gráfico de dispersão de x_1 por x_2 em um caso em que $M = 2$ e $N = 3$. Os sinais se encontram nos domínios (a) do tempo, e (b) tempo-frequencial.....	22
Fig. 3.3 – Superfície da função custo dada por (3.13) para um caso com $N = 5$ e $M = 3$	25
Fig. 4.1 – (a) Modelo de equalizador SISO; (b) o canal do equalizador visto como sendo um problema de BSS.	31
Fig. 4.2 – Intepretação geométrica para um caso em que $N = 3$ e $M = 2$, situações em que o plano criado pelo conjunto atingível de c : (a) não coincide com nenhuma fonte; e (b) coincide com duas fontes.....	34
Fig. 4.3 – Fontes originais (x), e recuperadas (o). É possível ver que somente a primeira fonte conseguiu ser recuperada com sucesso.	37
Fig. 4.4 – $J(W_{MMSE})$ como uma função da correlação r entre as fontes.	40
Fig. 5.1 – Espectrograma de uma mistura com duas vozes.....	44
Fig. 5.2 – Gráficos de dispersão <i>Sensor 1</i> x <i>Sensor 2</i> , com as indicações das direções encontradas pelo algoritmo de ICA para os seguintes casos: (a) Duas fontes e dois sensores; (b) Uma fonte e dois sensores; (c) Quatro fontes e dois sensores; e (d) Nenhuma fonte e dois sensores.	46
Fig. 5.3 – Histograma das direções encontradas numa mistura com 3 fontes de voz	48
Fig. 5.4 – Valor de J_K quando o número de fontes é igual a 3 e a 6, respectivamente.....	49
Fig. 5.5 – Espectrograma de uma mistura com um ruído branco uniforme e um sinal de voz	54
Fig. 5.6 – Histograma das direções encontradas numa mistura com duas fontes de voz e um ruído branco.....	54
Fig. 5.7 – Espectrograma de uma mistura com um sinal de voz e uma multidão batendo palmas e gritando.....	55

Fig. 5.8 – Histograma das direções encontradas numa mistura com duas fontes de voz e uma multidão batendo palmas e gritando	55
Fig. 5.9 – % de estimações corretas para os três diferentes índices	56
Fig. 5.10 – % de estimativas corretas para os três diferentes cenários, utilizando o índice I_{Log}	57
Fig. 5.11 – Valor médio de EAM dos algoritmos para o primeiro cenário (todas as fontes são sinais de voz).....	58
Fig. 5.12 – Valor médio de EAM dos algoritmos para o segundo cenário (com uma das fontes sendo ruído branco)	59
Fig. 5.13 – Valor médio de EAM dos algoritmos para o terceiro cenário (com uma das fontes sendo a de multidão).....	59
Fig. 6.1 – 3 etapas da otimização da função custo, dada por (3.13), para uma situação com 5 fontes e 3 misturas: (a) solução obtida pela opt-aiNet; (b) depois das etapas de refinamento e supressões; e (c) população final, contendo somente as estimativas das direções das fontes. Pode-se ver no detalhe que um dos dois indivíduos que aparecem em (a) já foi suprimido em (b).....	67
Fig. 6.2 – % de estimativas corretas do número de fontes na mistura	73
Fig. 6.3 – Valor médio de EAM para diferentes números de fontes	74
Fig. 6.4 – Número obtido de vetores (fontes).....	75
Fig. 6.5 – Erros das estimações dos vetores de mistura	75
Fig. 7.1 – Valor de $\Psi_y - \Psi_{si}$ por Θ_w . Quando w é ortogonal à fonte mais esparsa, a extração é efetivamente feita.	82
Fig. 7.2 – Curva de nível da função custo Ψ_y , para um caso com 4 fontes e 3 misturas. As setas indicam os mínimos da função.	83
Fig. 7.3 – Superfície 3D da função custo Ψ_y para um caso com 4 fontes e 3 misturas	83
Fig. 7.4 – Valor da norma ℓ_1 de vetores com potência unitária, gerados utilizando (7.10) para $1 < \beta < 4$	86
Fig. 7.5 – (a) O quadrado centrado na origem descreve o espaço amostral de uma variável randômica s , sendo que a área sombreada representa a região onde a integral dupla em (7.12) é calculada. (b) $E\{ w^T x \}$ por Θ	87

Fig. 7.6 – SIR das fontes extraídas por β em um cenário com duas fontes e duas misturas. As fontes foram geradas utilizando (7.10) para $1 < \beta < 4$	92
Fig. 7.7 – Arranjo dos ângulos entre os vetores de mistura e os de separação.....	94
Fig. 7.8 – Função custo da norma ℓ_1 em um cenário subdeterminado	98

Lista de tabelas

Tab. 4.1 – Separação de fontes: Soluções de Wiener e MSEs correspondentes	36
Tab. 4.2 – Caso ruidoso ($\sigma^2 = 0,005$): Soluções de Wiener e MSEs correspondentes	38
Tab. 4.3 – Caso com 50 fontes e o número de misturas variando. São mostrados os valores mínimo, máximo, médio e total de $J(w)$	38
Tab. 4.4 – Primeiro cenário de equalização: Soluções de Wiener e MSEs correspondentes	41
Tab. 4.5 – Segundo cenário de equalização: Soluções de Wiener e MSEs correspondentes	42
Tab. 6.1 – Valores de erros ε , calculados utilizando a Eq. (6.3), da estimação da matriz de mistura	70
Tab. 6.2 – Erros, calculados utilizando a Eq. (6.3), da estimação da matriz de mistura	71
Tab. 7.1 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte, com diferentes graus de esparsidade. Os resultados são uma média de 100 diferentes realizações.....	88
Tab. 7.2 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte, com diferentes graus de ruído presente nos sinais. Os resultados são uma média de 100 diferentes realizações	90
Tab. 7.3 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte. As fontes são geradas com distribuição uniforme. Os resultados são uma média de 100 diferentes realizações	91
Tab. 7.4 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes, com diferentes graus de esparsidade. Os resultados são uma média de 100 diferentes realizações.....	96
Tab. 7.5 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes, com diferentes graus de ruído presente nos sinais. Os resultados são uma média de 100 diferentes realizações	97
Tab. 7.6 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes. As fontes são geradas com distribuição uniforme. Os resultados são uma média de 100 diferentes realizações	97

Capítulo 1

Introdução

Nos últimos anos, tem surgido um grande interesse em problemas de processamento de sinais não-supervisionado, que são caracterizados por formulações baseadas na utilização da menor quantidade de informações *a priori* possível. Um dos exemplos importantes nesta área é o problema de Separação Cega de Fontes (*Blind Source Separation* – BSS) [COMON e JUTTEN, 2010; HYVÄRINEN *et al.*, 2001; ROMANO *et al.*, 2010], cujo objetivo é a recuperação das informações geradas por um conjunto de fontes utilizando somente as misturas destes sinais.

Um dos modos clássicos de tratar o problema de BSS é através da Análise de Componentes Independentes (*Independent Component Analysis* – ICA) [COMON, 1994], uma técnica estatística que busca restaurar uma condição de independência comprometida pela sobreposição inerente aos sinais recebidos. Em condições ideais, i.e., quando o número de fontes não excede o de misturas e o sistema misturador é linear, instantâneo e invariante no tempo, esta suposição de independência é considerada suficiente para a recuperação das fontes.

Entretanto, em diversas aplicações reais, nem todas estas suposições podem ser consideradas válidas, sendo intuitivo acreditar que é mais difícil ainda encontrar situações nas quais todas estas hipóteses ocorram de forma simultânea. Porém, se não for considerada nenhuma hipótese sobre as fontes ou sobre o sistema em que elas se encontram, o problema de BSS se torna insolúvel.

Devido à inviabilidade de se conseguir uma solução geral para qualquer tipo de problema de BSS, a pesquisa nesse tema acaba sendo dividida em duas “frentes”: uma que congrega trabalhos visando construir uma teoria que seja a mais generalista possível, e,

desta forma, leva em consideração o emprego de um mínimo de suposições; e aquela relativa à solução de problemas específicos, utilizando, neste caso, a maior quantidade de informações possível.

Nesta segunda “frente”, destaca-se um problema de BSS em que as técnicas de ICA não são bem sucedidas: o caso em que o número de fontes é superior ao número de misturas. Como explicaremos com mais detalhes durante este trabalho, esta situação, que se associa à ideia de *sistema de mistura subparametrizado*, é caracterizada pela não-existência de um separador linear que consiga recuperar todas as fontes perfeitamente. Neste caso, uma saída mostrada por diversos trabalhos [ABRARD *et al.*, 2001; BOFILL e ZIBULEVSKY, 2000] é considerar que diversos tipos de fontes possuem certa esparsidade, ou seja, momentos relativamente longos sem transmissão de informação. Esta abordagem é comumente chamada de Análise de Componentes Esparsos (*Sparse Component Analysis – SCA*).

É nesta linha de pesquisa que as propostas neste trabalho se inserem: no estudo e no desenvolvimento de novas abordagens relacionadas à utilização de esparsidade nos problemas de separação cega de fontes, com o foco em situações em que o número de fontes é maior do que o de misturas.

1.1 Organização do Conteúdo do Trabalho

A tese se divide em duas partes principais. A primeira delas contém dois capítulos, nos quais é apresentada a teoria necessária para o desenvolvimento das abordagens propostas. No Capítulo 2, é descrito o problema de separação cega de fontes e é feita uma explicação sobre ICA. O Capítulo 3 descreve o conceito de esparsidade, mostrando um breve estudo sobre representação esparsa e sua relação com as técnicas de SCA.

Na segunda parte, que abrange os capítulos restantes, são mostradas as abordagens propostas e seus resultados. No Capítulo 4, é feita uma análise sobre os erros de estimação encontrados em problemas subparametrizados, e é proposto um limite teórico, baseado no erro quadrático médio, para o erro encontrado na recuperação das fontes. No Capítulo 5, é proposta uma abordagem que considera as hipóteses de independência e esparsidade de forma simultânea na estimação do sistema de mistura, bem como na identificação do

número de fontes existentes. O método proposto foi aplicado a diversos sinais reais de vozes, gravadas separadamente.

O Capítulo 6 aborda o estudo da estimação do sistema de mistura no caso subparametrizado. Tomando como base o trabalho proposto por [NAINI *et al.*, 2008], é mostrado que estes tipos de problemas levam a funções custo multimodais, sendo todos os ótimos, ou pelo menos a maior parte, necessários para a resolução do problema. Desta forma, foi proposta a utilização de ferramentas bioinspiradas de otimização multimodal, sendo mostrados resultados de simulação obtidos através do algoritmo opt-aiNet [DE CASTRO e TIMMIS, 2002].

O Capítulo 7 propõe a utilização de uma medida baseada em esparsidade como função custo para a recuperação das fontes. É mostrado que a norma ℓ_1 é capaz de originar uma função custo eficiente, a partir da qual são propostos métodos de extração e separação de fontes, no caso em que o número de fontes e misturas é o mesmo, e também um método para estimação do sistema misturador, no caso subparametrizado.

Por fim, o Capítulo 8 conclui o trabalho com uma análise de todos os resultados, incluindo também as perspectivas para continuações do trabalho desenvolvido. Também são colocadas aqui sugestões para pesquisas posteriores.

1.2 Publicações

Apresentamos a seguir uma lista completa de publicações feitas no período do doutorado.

- Artigos completos publicados em ou submetidos a periódicos:
 - NADALIN, E. Z.; SUYAMA, R.; ATTUX, R.; “An ICA-Based Method for Blind Source Separation in Sparse Domains. Lecture Notes in Computer Science”, v. 5441, p. 597-604, 2009.
 - NADALIN, E. Z.; TAKAHATA, A. K.; DUARTE, L. T. ; SUYAMA, R.; ATTUX, R.; “Blind Extraction of the Sparsest Component”. Lecture Notes in Computer Science, v. 6365, p. 394-401, 2010.

-
- TAKAHATA, A. K.; **NADALIN, E. Z.**; FERRARI, R.; DUARTE, L. T. ; SUYAMA, R.; LOPES, R. R.; ROMANO, J. M. T.; TYGEL, M.; “On Unsupervised Processing of Geophysical Signals”. **Submetido à IEEE Signal Processing Magazine, Special Issue: Geophysical Signal Processing.**

 - Trabalhos completos publicados ou aceitos para publicação em anais de congressos:
 - SORIANO, D. C.; **NADALIN, E. Z.**; WADA, C.; FERRARI, R.; SUYAMA, R.; ATTUX, R.; “Equalização Cega com Realimentação de Decisões Baseada em Redes Imunológicas Artificiais”. XXVI Simpósio Brasileiro de Telecomunicações – SBrT’08, 2008, Rio de Janeiro.
 - PEREIRA, F. O.; **NADALIN, E. Z.**; SUYAMA, R.; ATTUX, R.; “Análise do Emprego de Ferramentas de Computação Natural no Problema de Separação de Misturas com Não-Linearidade Posterior”. XXVII Simpósio Brasileiro de Telecomunicações – SBrT’09, 2009, Blumenau.
 - **NADALIN, E. Z.**; SUYAMA, R.; ATTUX, R.; “Estimating the number of audio sources in a stereophonic instantaneous mixture”. 7o Congresso de Engenharia de Áudio – AES 2009, 2009, São Paulo.
 - ROSA, E. S.; **NADALIN, E. Z.**; ROMANO, J. M. T.; SUYAMA, R.; “Blind Source Separation in Polarization Division Multiplexed Optical Systems”. International Telecommunications Symposium - ITS 2010, 2010, Manaus.
 - **NADALIN, E. Z.**; BOCCATO, L. ; ATTUX, R. ; DUARTE, L. T. ; LOPES, A. ; ROMANO, J. M. T. ; SUYAMA, R.. “Multimodal Optimization in the Context of Sparse Component Analysis”. IEEE Symposium Series in Computational Intelligence – SCCI 2011, 2011, Paris.

Capítulo 2

Separação Cega de Fontes

O problema de separação cega de fontes (*Blind Source Separation* - BSS) é comumente ilustrado por meio do exemplo conhecido como *Cocktail Party Problem* [HYVÄRINEN *et al.*, 2001], que pode ser descrito da seguinte forma: imaginemos uma sala onde se encontram diversas pessoas conversando e vários microfones são posicionados em diferentes locais do ambiente; tendo em mãos somente os sinais captados pelos microfones, o problema é conseguir obter separadamente os sinais de voz de cada pessoa conversando.

A Fig. 2.1 mostra um esquema clássico de separação de fontes. No exemplo dado, pode-se dizer que cada voz é uma *fonte* representada por $s_i(n)$, sendo i um índice que varia entre 1 e N (número total de fontes), e n é o índice da amostra do sinal. Os sinais recebidos pelos sensores (que, neste caso, são os microfones) são chamados de *misturas*, e são representados por $x_j(n)$, sendo o j um índice que varia entre 1 e M (número total de misturas). O ambiente em que as pessoas se encontram corresponde ao sistema misturador, e é representado por \mathcal{A} . Reinterpretando o problema de BSS a partir da Fig. 2.1, se tivermos à disposição somente os valores x_j , pretende-se descobrir se existe um sistema separador, designado por \mathcal{W} , que gere um conjunto de valores $y_i(n)$ que sejam estimativas precisas das fontes.

Como não temos informações sobre o sistema misturador, e também não existe uma forma de gerar sinais de treinamento, de forma a ajustar os parâmetros ou adaptar o sistema misturador, este tipo processo é chamado na literatura de processamento de sinais *não-supervisionado*.

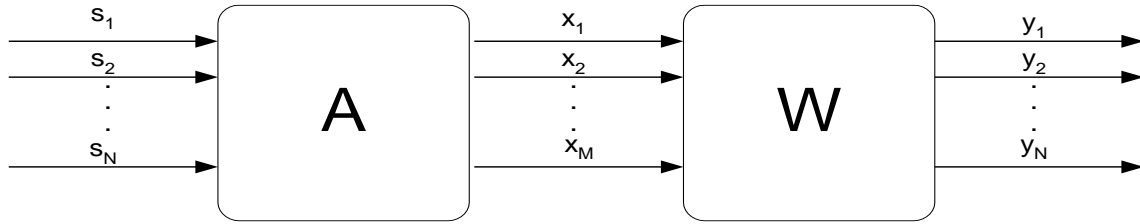


Fig. 2.1 – Cenário clássico de separação de fontes

Porém, na ausência completa de informações adicionais acerca dos sinais e/ou do sistema misturador, o problema de BSS não admite solução única. Dessa forma, consideremos, então, uma situação em que o sistema de mistura é linear, instantâneo (sem memória), invariante no tempo e sem ruído. A partir da Fig. 2.1, podemos descrever matematicamente o modelo de mistura considerando que $\mathbf{s}(n)$ é o n -ésimo instante do vetor N -dimensional contendo as fontes, o que permite expressar o vetor $\mathbf{x}(n)$ contendo as misturas como

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \quad (2.1)$$

sendo que, agora, o sistema de mistura \mathbf{A} passa a ser definitivamente descrito como uma matriz $M \times N$.

Faremos aqui uma divisão do problema geral em dois processos correlatos: o de recuperação de uma só fonte, processo este que será chamado de extração; e o de separação de todas as fontes ao mesmo tempo, que será chamado simplesmente de processo de separação.

Extração Cega de Fontes

O processo de extração cega de fontes (*Blind Source Extraction – BSE*) é um caso caracterizado pela recuperação de apenas uma das fontes. Ele é formulado em termos da busca não-supervisionada por um vetor \mathbf{w} , tal que $y(n)$, dado por

$$y(n) = \mathbf{w}^T \mathbf{x}(n), \quad (2.2)$$

seja uma boa estimativa de uma das fontes contidas em $\mathbf{s}(n)$.

Separação Cega de Fontes

Como dito, o problema de BSS se refere à recuperação de todas as fontes de forma simultânea. Para cumprir esta meta, se o número de fontes não for superior ao de sensores

($N \leq M$), adota-se, tipicamente, um sistema separador com estrutura análoga à da mistura, ou seja

$$\mathbf{y}(n) = \mathbf{W}^T \mathbf{x}(n), \quad (2.3)$$

onde \mathbf{W} é chamado de matriz de separação, ou seja, é a matriz que inverte o sistema de mistura, recuperando as fontes. Note que isso é equivalente a realizar, de maneira simultânea, a extração das N fontes. Dessa forma, podemos encarar as colunas da matriz \mathbf{W} como sendo os vetores de extração das fontes, i.e., $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_N]$.

Quando o número de fontes é maior do que o de sensores ($N > M$), o sistema de mistura é comumente chamado de *sistema subparametrizado*, sendo que não existe matriz \mathbf{W} (tamanho $M \times N$) que o inverta perfeitamente. Este caso, como discutido no Capítulo 1, será bastante relevante na sequência deste trabalho.

2.1 Análise de Componentes Independentes

As suposições de que o sistema misturador é linear, instantâneo, invariante no tempo e inversível são necessárias para garantir a existência de pelo menos uma matriz \mathbf{W} que inverta o sistema de mistura, mas não são suficientes para que esta matriz seja encontrada de forma não-supervisionada. Se desejamos criar um critério de otimização para encontrar o sistema separador sem ter conhecimento das fontes e do sistema de mistura, é preciso fazer uso de outras suposições. Um exemplo é encontrado nos métodos que se baseiam na hipótese de independência mútua entre as fontes, o que significa que a sua função de densidade de probabilidade (fdp) conjunta pode ser fatorada da seguinte maneira:

$$p(s_1, s_2, \dots, s_N) = \prod_{i=1}^N p(s_i) \quad (2.4)$$

Em outras palavras, tem-se que a densidade de probabilidade conjunta é igual ao produto das correspondentes densidades marginais. A abordagem que se utiliza deste conceito é comumente chamada na literatura de *análise de componentes independentes* (*Independent Component Analysis – ICA*) [COMON, 1994]. Vale frisar que a hipótese de independência mútua entre as fontes é bem plausível em diversos problemas reais, especialmente naqueles em que os sinais das fontes são gerados de forma separada.

É importante ressaltar que o conceito de independência mútua entre as fontes é mais restritivo do que assumir a simples decorrelação entre elas. De fato, independência implica decorrelação, mas o contrário não é necessariamente verdade [HYVÄRINEN *et al.*, 2001]. Quando se lida com variáveis aleatórias mutuamente independentes, temos que, para quaisquer funções f e g , vale:

$$E\{f(s_1)g(s_2)\} = E\{f(s_1)\}E\{g(s_2)\}, \quad (2.5)$$

uma condição que abrange a decorrelação como um caso especial (no qual f e g são funções identidade). Por outro lado, a decorrelação só implica a independência para variáveis aleatórias gaussianas [PAPOULIS, 1993].

Porém, embora tornar os dados decorrelacionados não seja equivalente a torná-los independentes, em muitos casos, o processo de decorrelação pode ser encarado como uma etapa de pré-processamento dos sinais, auxiliando na tarefa de separação das fontes. Este processo é chamado de branqueamento dos sinais.

Branqueamento

Dado um vetor contendo amostras de um conjunto de sinais de média nula $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$, se seus elementos x_i são decorrelacionados e possuem variância unitária, este sinal é chamado de espacialmente branco¹ [HYVÄRINEN *et al.*, 2001]. Matematicamente, isto significa que $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{I}$, onde \mathbf{I} é a matriz identidade. Caso um vetor \mathbf{x} não seja decorrelacionado, pode-se branqueá-lo por meio de uma transformação linear dada por

$$\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}, \quad (2.6)$$

onde \mathbf{T} é uma matriz, chamada de *matriz de branqueamento*, calculada da seguinte forma²:

$$\mathbf{T} = \mathbf{D}^{-1/2}\mathbf{E}^T, \quad (2.7)$$

onde $\mathbf{E} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n]$ é a matriz contendo os autovetores com norma unitária da matriz de covariância $\mathbf{C}_x = E\{\mathbf{x}\mathbf{x}^T\}$, e $\mathbf{D} = \text{diag}(d_1 \ d_2 \ \dots \ d_n)$ é a matriz diagonal contendo os autovalores de \mathbf{C}_x .

A Fig. 2.2 ilustra a limitação do processo de branqueamento no problema de separação dos sinais. Na Fig. 2.2 (a), temos um gráfico que representa conjuntamente as amostras das misturas, normalmente chamado de gráfico de dispersão, em um caso com

¹ Para fins de comodidade, trataremos os sinais espacialmente brancos simplesmente de brancos.

² Neste texto, optou-se pela matriz de branqueamento descrita na Eq. (2.7), porém se multiplicarmos esta por qualquer matriz ortogonal, o resultado também será uma matriz de branqueamento.

duas fontes independentes com distribuição uniforme: a independência, aliás, pode ser inferida pelo fato de que as “laterais” do quadrado formado estão ortogonais aos eixos. Quando os sinais passam por um sistema misturador, o gráfico de dispersão passa a “distorcer” o quadrado para o desenho de um paralelogramo sem lados ortogonais entre si, como pode ser visto na Fig. 2.2 (b). Se os sinais das misturas são branqueados, os pontos voltam a formar um quadrado ligeiramente rotacionado em relação aos eixos, como pode ser visto na Fig. 2.2 (c). Neste caso, os sinais estão descorrelacionados, mas não são independentes, e, conseqüentemente, a separação não foi realizada, e, desta forma, o problema passaria a ser o de encontrar um critério de otimização para encontrar uma matriz de rotação que levaria a distribuição observada na Fig. 2.2 (c) ao gráfico de dispersão original das fontes.

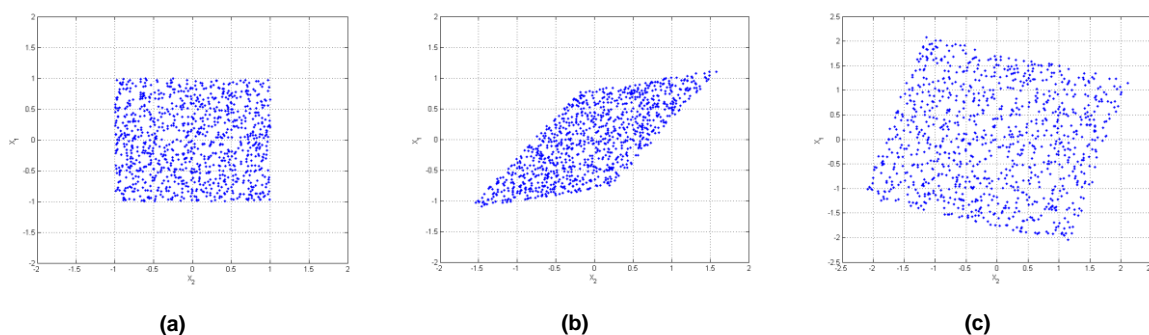


Fig. 2.2 – Exemplo de mistura de duas fontes independentes com distribuição uniforme. Distribuição (a) das fontes, (b) das misturas antes e (c) após o branqueamento.

Tendo em vista a limitação inerente ao processo de branqueamento, é imperativo lidar com a hipótese de independência mútua das fontes. Tendo em mãos somente os dados provenientes do vetor de mistura $\mathbf{x}(n)$, descrito pela equação (2.1), pode-se buscar recuperar as fontes ajustando a matriz de separação \mathbf{W} de forma que as estimativas $\mathbf{y}(n)$ sejam tão mutuamente independentes quanto possível. Se $M = N$ e \mathbf{A} for de posto completo, é possível inverter o mapeamento que esta matriz de mistura realiza a menos de fatores de escala, visto que a condição de independência é invariante à amplitude, e de uma permutação, já que a ordem das colunas da matriz \mathbf{W} também não interfere na independência dos sinais. Portanto, a matriz de separação ideal seria

$$\mathbf{W}^T = \mathbf{P}\mathbf{D}\mathbf{A}^{-1}, \quad (2.8)$$

onde \mathbf{P} é uma matriz de permutação e \mathbf{D} uma matriz diagonal.

Em um clássico trabalho [COMON, 1994], Comon provou que ajustar a matriz \mathbf{W} de modo a aproximar tanto quanto possível a condição de independência inerente às fontes é suficiente para garantir a separação de acordo com (2.8), desde que duas condições sejam garantidas: que haja no máximo uma fonte com distribuição gaussiana e que a matriz \mathbf{A} seja inversível. Nesse contexto, resta apenas encontrar formas de quantificar a independência, de modo a definir uma função custo³ capaz de ser aplicada para fornecer estimativas apropriadas para o sistema separador. Essa ideia deu origem à teoria de BSS baseada em ICA.

Entre os principais conceitos utilizados em ICA com a finalidade de construir critérios para medir independência entre os sinais envolvidos estão a informação mútua e a não-gaussianidade. Será feita uma breve explicação das duas abordagens, levando-se em conta principalmente os conceitos que serão utilizados posteriormente neste trabalho. Caso o leitor queira realizar um estudo mais detalhado, sugerem-se referências como [HYVÄRINEN *et al.*, 2001; COMON e JUTTEN, 2010; ROMANO *et al.*, 2010]. Uma outra abordagem para separação de fontes, diretamente relacionada à ICA, se baseia no critério de máxima verossimilhança. Como esta abordagem só será utilizada na Seção 7.3.2, optamos por deixar sua explicação para aquela seção.

2.1.1 ICA via Informação Mútua

A informação mútua, um dos conceitos fundamentais da teoria de informação [COMON e JUTTEN, 2010; HYVÄRINEN *et al.*, 2001], dá origem à mais direta abordagem de ICA, já que traduz matematicamente o grau de independência entre variáveis aleatórias. Dado um vetor de variáveis aleatórias \mathbf{y} , pode-se definir a informação mútua $I(\cdot)$ como sendo a diferença entre a soma de suas entropias marginais e sua entropia conjunta:

$$I(y_1, y_2, \dots, y_N) = \sum_{j=1}^N H(y_j) - H(\mathbf{y}), \quad (2.9)$$

onde $H(\cdot)$ corresponde à entropia diferencial de uma variável aleatória. A entropia diferencial é uma forma de quantificar a incerteza acerca dos valores de uma variável contínua, sendo definida da seguinte forma

³ No caso de separação de fontes, esta função custo também pode ser denominada de função contraste [COMON, 1994]

$$H(\mathbf{y}) = - \int_{-\infty}^{\infty} p_y(\mathbf{v}) \ln p_y(\mathbf{v}) d\mathbf{v}. \quad (2.10)$$

O ponto crucial, que justifica o uso da informação mútua em critérios que reflitam a condição de independência entre os sinais, é que $I(\cdot)$ sempre retorna um valor não-negativo, fornecendo um valor nulo apenas no caso em que as variáveis aleatórias são mutuamente independentes [COVER e THOMAS, 2006].

2.1.2 ICA via Não-Gaussianidade

O uso de não-gaussianidade como critério para ICA se associa ao teorema central do limite, o qual afirma, em termos simples, que, sob algumas condições, uma soma de variáveis aleatórias independentes aproxima-se de uma variável com distribuição de probabilidade normal conforme o número de variáveis envolvidas tende ao infinito [PAPOULIS, 1993]. De fato, conclui-se, intuitivamente, que, ao se tornar os dados “o menos gaussiano possível”, estaríamos contrabalançando o processo de mistura – visto que este é, em essência, uma soma ponderada dos sinais das fontes, que são, por hipótese, independentes entre si –, ou seja, recuperando as fontes.

Devido ao fato de os métodos de otimização visarem maximizar a não-gaussianidade de um sinal, suas derivações são feitas somente para o caso de BSE, havendo tipicamente um processo de normalização de potência para evitar soluções triviais. O processo de separação através destes métodos é normalmente feito por uma técnica chamada deflação [DELFOSSÉ e LOUBATON, 1995], em que a fonte estimada é subtraída dos sinais de mistura para que se estime a próxima fonte.

Duas funções de contraste são comumente utilizadas quando se lida com o conceito de não-gaussianidade, conforme exporemos a seguir.

Curtose

O cumulante de k -ésima ordem (κ_k) [MENDEL, 1991; HYVÄRINEN *et al.*, 2001], de uma variável aleatória é definido como

$$\kappa_k = (-j)^k \left. \frac{\partial^k \ln[\Phi_y(\omega)]}{\partial \omega^k} \right|_{\omega=0}, \quad (2.11)$$

onde $\Phi_y(\omega)$ é a função geradora de momentos associada a variável aleatória y , conhecida como. Quando a variável aleatória possui média nula, o cumulante de quarta ordem é chamado de curtose, que é definida no caso real por

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y\})^2. \quad (2.12)$$

Um ponto interessante é que a curtose pode ser interpretada como sendo uma forma de quantificar o “alargamento” da cauda da fdp do sinal. Ela se torna nula caso a distribuição de y seja gaussiana (é importante notar que o contrário não é necessariamente válido, ou seja, nem toda variável aleatória com curtose nula é gaussiana). Caso o valor da curtose de um sinal seja negativo, temos os chamados sinais sub-Gaussianos; caso contrário, fala-se em sinais super-Gaussianos. Deste modo, uma forma de se extrair um sinal de uma mistura linear é através de um processo de otimização em \mathbf{w} (no contexto da Eq. (2.2)) que maximize o módulo da curtose do sinal estimado.

Negentropia

Como visto anteriormente, a entropia pode ser entendida como o grau de incerteza associado a uma variável aleatória. Um fato interessante associado a essa grandeza é que, caso consideremos todos os possíveis sinais com mesma variância, os de maior entropia são aqueles que possuem distribuição gaussiana [HYVÄRINEN *et al.*, 2001]. Tendo isto em mente, se considerarmos a função custo $J(y)$, dada por

$$J(y) = H(y_{\text{Gauss}}) - H(y), \quad (2.13)$$

onde y_{Gauss} denota uma variável aleatória com a mesma variância de y . Desta forma, teremos indiretamente uma medida de não-gaussianidade, visto que a negentropia nos dá uma ideia de quão diferente é a distribuição de y em relação a uma distribuição gaussiana. Neste caso, quanto maior for o valor de $J(y)$, menos gaussiana será a fdp de y .

Visto que a procura é feita de forma que a variância de y seja mantida constante, o primeiro termo da Eq. (2.13) pode ser ignorado e, conseqüentemente, maximizar $J(y)$ é equivalente a minimizar a entropia do sinal y , o que significa que, feita uma normalização de potência, o vetor de extração \mathbf{w} pode ser obtido como solução do seguinte problema

$$\mathbf{w} = \arg \min E\{-\log p(y)\}. \quad (2.14)$$

Tendo apresentado os fundamentos do problema de separação de fontes, da abordagem de análise de componentes independentes e dos paradigmas de ICA que serão particularmente importantes para o restante da exposição, passaremos ao estudo da ideia de esparsidade – um dos temas centrais ao presente trabalho – e de sua aplicação em BSS.

Capítulo 3

Esparsidade e Análise de Componentes Esparsos

Em diferentes cenários práticos, as hipóteses consideradas para a separação dos sinais com ICA são plausíveis. No entanto, conforme comentado no capítulo anterior, a recuperação perfeita dos sinais depende (entre outras suposições) de que o número de fontes for igual ao de misturas, o que implica em uma matriz de mistura de posto completo. Porém, em muitos cenários, pode não ser possível recuperar as fontes por meio de ICA, como em certos casos em que o número de fontes é maior do que os sensores, ou seja, no caso de um sistema de misturas subparametrizado.

Uma vez que a hipótese de independência pode não ser suficiente para solucionar os problemas de BSS, devemos considerar outras informações a respeito dos sinais e do sistema misturador. Uma possibilidade é considerar que os sinais envolvidos apresentem um determinado grau de esparsidade. Em termos simples, um sinal esparso pode ser definido como aquele cujos coeficientes, em sua maioria, possuem valores nulos ou próximos de zero. Em outras palavras, a maior parte da energia de um sinal desse tipo está concentrada em uma pequena quantidade de seus coeficientes. Visando compreender melhor como a esparsidade pode ser utilizada no processo de separação de fontes, será realizada neste capítulo, primeiramente, uma discussão teórica mais ampla sobre o conceito de esparsidade, e, em seguida, será mostrado como ele é utilizado nos processos de BSE e BSS.

3.1 *Quantificação Esparsidade*

Embora grande parte dos pesquisadores concorde com a noção geral de que um sinal esparso possui a maior parte de seus valores próximos de zero, não existe um consenso acerca de uma métrica ideal para quantificar esparsidade. Em [HURLEY e RICKARD, 2009], foi feita uma discussão interessante sobre maneiras de comparar e mensurar esparsidade, na qual são expostos seis critérios a que um quantificador de esparsidade deve obedecer. Os quatro primeiros são baseados nas Leis de Dalton [DALTON, 1920], propostas em 1920, que tratam da análise, do ponto de vista matemático, da desigualdade da distribuição de renda, visto que esta possui um cunho intrinsecamente esparso: uma ínfima quantidade da população concentra a maior parte da renda mundial. Os outros dois critérios, os últimos da lista abaixo, foram adicionados por [HURLEY e RICKARD, 2009]:

- Visto que a maior parte da energia está concentrada em apenas alguns coeficientes, se esta energia for distribuída, a esparsidade irá diminuir;
- A esparsidade é invariante ao escalamento dos sinais, ou seja, se multiplicarmos um sinal por uma constante, sua esparsidade não deve ser alterada;
- Se adicionarmos uma constante a todos os coeficientes, a esparsidade irá diminuir (isto se os sinais forem considerados definidos semi-positivos, o que é o caso para a distribuição de renda);
- Se dois sinais, ou populações, idênticos se unirem, a esparsidade total deve se manter a mesma;
- Tendo o coeficiente de maior energia, se este valor aumentar sem alterar a energia dos outros coeficientes, a esparsidade irá aumentar;
- Por fim, se forem adicionados coeficientes de valor nulo ao conjunto total, a esparsidade irá aumentar.

Apesar de a lista acima trazer requisitos objetivos para se escolher um índice para a medição de esparsidade, cabe ressaltar que, em várias situações práticas, nem todos os requisitos precisam ser satisfeitos.

Um dos modos mais simples de quantificar a esparsidade de um vetor s é por meio da medida ℓ_0 (também descrita como $\|s\|_0$), que corresponde ao número de componentes

não-nulos do sinal. É usualmente dito que, se $\|\mathbf{s}\|_0 = k$, então o sinal é k -esparso. Apesar de esta medida ter levado a importantes resultados teóricos [CANDÈS e TAO, 2005], ela não é, em termos estritos, adequada a aplicações práticas. Isto se deve a, pelo menos, dois fatores.

O primeiro deles é a existência de ruído causa pequenas flutuações em torno de zero que pode acarretar uma grande variação na medida ℓ_0 do sinal. Esta deficiência pode, em certo grau, ser sanada através da transformação da medida ℓ_0 em $\ell_{0\epsilon}$, que é definida como sendo o número de componentes que têm valores absolutos maiores que um limiar⁴ ϵ . O segundo fator é que ambas as medidas, ℓ_0 e $\ell_{0\epsilon}$, são difíceis de tratar quando a solução mais esparsa é a desejada, como o caso de uma otimização baseada na minimização destas medidas, uma vez que não é possível derivar um método de otimização baseado em gradiente e o resultado ótimo deve ser obtido a partir da solução de um problema combinatório NP-difícil [HURLEY e RICKARD, 2009; NATARAJAN, 1995]. Além disso, se voltarmos para os seis critérios citados anteriormente, verificamos que a medida ℓ_0 só obedece ao segundo e ao sexto. Quando o critério é alterado para $\ell_{0\epsilon}$, o segundo critério deixa de ser satisfeito.

Por conta destas desvantagens, as medidas baseadas em ℓ_0 são normalmente substituídas por medidas de normas⁵ ℓ_p :

$$\|\mathbf{s}\|_p = \left(\sum_n |s(n)|^p \right)^{1/p}. \quad (3.1)$$

Considerando que, no modelo de mistura dado pela Eq. (2.1), tanto a matriz de mistura \mathbf{A} , quanto sinais de mistura \mathbf{x} são conhecidos, e supondo que as fontes são esparsas, a estimação dos sinais pode ser expressa por meio de um problema de otimização com restrições, descrito na Eq. (3.2):

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}\|_p \text{ s.a } \mathbf{x} = \mathbf{A}\mathbf{s} \quad (3.2)$$

Uma interpretação geométrica desta otimização, no contexto de um problema bidimensional, pode ser vista na Fig. 3.1. Como a linha da figura representa a restrição

⁴ Na prática, uma definição precisa do valor apropriado do limiar pode ser uma tarefa bem complexa.

⁵ Apesar de (3.1) ser rigorosamente uma norma somente para $p \geq 1$, neste trabalho ela será informalmente chamada de norma também para $0 \leq p < 1$. Note que para $p = 0$, é usada a convenção de que $|s(n)|^0 = 1$, se $|s(n)| > 0$, e $0^0 = 0$.

$\mathbf{x} = \mathbf{A}\mathbf{s}$, o ótimo corresponde à intersecção entre esta linha e a menor bola de norma ℓ_p , definida por $B_p(r) = \{\mathbf{s} \in \mathbb{R}^2 : \|\mathbf{s}\|_p \leq r\}$ [JACQUES e VANDERGHEYNST, 2010]. Na Fig. 3.1, os pontos ótimos para $p = 1/2$, $p = 1$ e $p = 2$ são apresentados. É importante notar que, para a medida ℓ_0 , as soluções ótimas são as intersecções com os eixos, admitindo que \mathbf{s} é 1-esparso, ou seja $\|\mathbf{s}\|_0 = 1$. Pode-se ver também que, para $p = 2$, a solução de mínimos quadrados tem dois elementos não-nulos, enquanto, para $p = 1$ e $p = 1/2$, as soluções se encontram sobre os eixos e são 1-esparsas. Este fato de não haver garantia de existirem elementos nulos na solução quando $p = 2$ pode ser relacionado com a ideia de decorrelação, visto que, como mostrado na Seção 2.1, a estrutura de momentos de ordem dois não é suficiente para a recuperação perfeita das fontes. Intuitivamente, no contexto da otimização de (3.2), as normas ℓ_p , com $0 < p \leq 1$, podem ser aproximadas para a medida ℓ_0 porque as respectivas bolas das normas ℓ_p têm vértice nos eixos, enquanto, para $p > 1$, as bolas das normas ℓ_p são suaves e convexas.

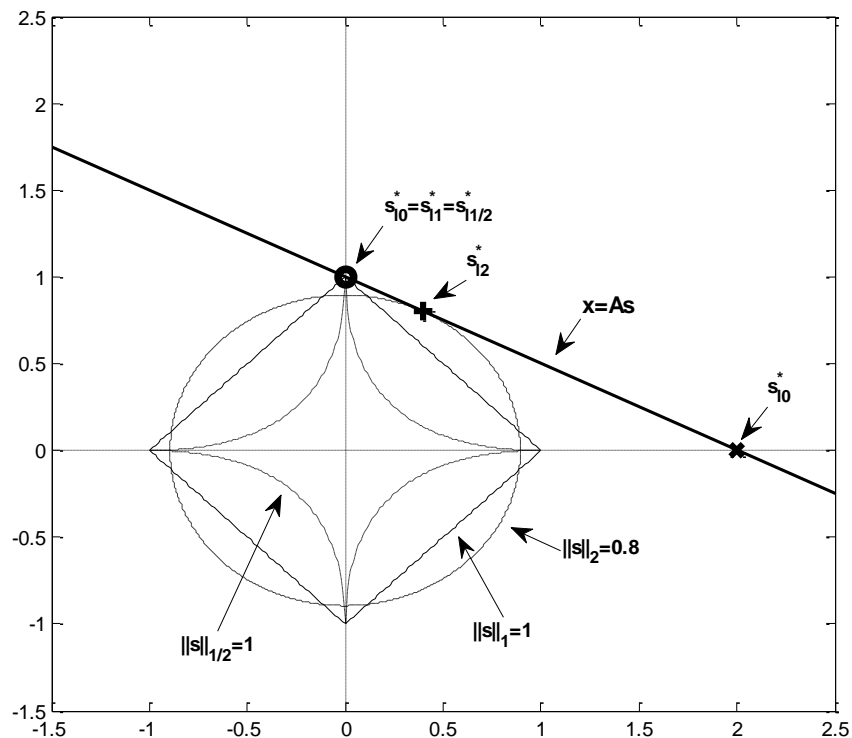


Fig. 3.1 – Ilustração geométrica do resultado de otimização usando medidas ℓ_0 , $\ell_{1/2}$, ℓ_1 e ℓ_2 . Mais detalhes podem ser obtidos em [COMON e JUTTEN, 2010]

É interessante ressaltar que a utilização da medida ℓ_p também tem motivações estatísticas [COMON e JUTTEN, 2010]. Os elementos de \mathbf{s} são considerados esparsos ou super-Gaussianos se suas funções de densidade de probabilidade estiverem concentradas em volta da origem, e com caudas espessas, um padrão que é comumente modelado com a ajuda da distribuição gaussiana generalizada:

$$p_{\mathbf{s}}(\mathbf{s}) \propto \exp(-\alpha \|\mathbf{s}\|_p^p) \quad (3.3)$$

com $0 < p \leq 1$ ($p = 1$ dá origem à distribuição laplaciana), quando os elementos de \mathbf{s} são considerados estatisticamente independentes. Neste caso, a verossimilhança entre \mathbf{s} e $\hat{\mathbf{s}}$, estimado a partir da otimização dada por (3.2), é maximizada quando a medida ℓ_p é minimizada, o que motiva a sua utilização como medida de esparsidade.

Como a escolha $0 < p < 1$ leva a um problema de otimização não-convexo, para o qual só existem algoritmos que asseguram convergência para um mínimo local [COMON e JUTTEN, 2010], $p = 1$ é normalmente preferido, já que, na maioria dos cenários, o problema de otimização resultante pode ser eficientemente resolvido com o uso de técnicas de otimização convexa ou de programação linear [TROPP, 2006]. Além disto, foi mostrado que a utilização da norma ℓ_1 é equivalente ao uso da medida ℓ_0 se o sinal for considerado “suficientemente esparsos” [DONOHO e HUO, 2001], além de ela ser mais robusta a cenários ruidosos, já que uma variação pequena do sinal causa pouco efeito na norma ℓ_1 , em contraste com o comportamento da medida ℓ_0 nesta mesma situação. Por conta destas vantagens, a norma ℓ_1 tem sido uma escolha popular para medição de esparsidade em diversas situações, incluindo separação de fontes [COMON e JUTTEN, 2010; BOFILL e ZIBULEVSKY, 2000], o que motiva seu uso em contribuições apresentadas neste trabalho, como será mostrado no Capítulo 7.

3.2 Análise de Componentes Esparsos

Conforme comentado anteriormente, a motivação para a hipótese de sinais esparsos está ligada à separação dos sinais em cenários em que há um número de fontes superior ao de sensores, i.e., uma mistura subparametrizada. O fato de haver momentos em que os sinais estão em “silêncio”, ou seja, regiões nulas de informação, faz com que exista a possibilidade de que nem todas as fontes estejam ativas ao mesmo instante. Desta maneira,

um sistema subdeterminado poderia ser considerado localmente determinado, visto que, em um determinado instante, pode existir uma quantidade menor de fontes ativas do que o número de misturas. Às técnicas que utilizam este tipo de abordagem damos o nome de Análise de Componentes Esparsos (*Sparse Component Analysis* – SCA).

Os métodos de SCA são usualmente divididos em três estágios de operação, que serão discutidos individualmente a seguir.

3.2.1 Escolha de uma Base Adequada

Como exposto na Seção 3.1, não existe um consenso sobre a definição matemática que melhor captura o conceito de esparsidade, ao contrário do que ocorre quando se lida com independência, caso em que é possível expressar matematicamente, e de maneira inequívoca, a condição ideal. Porém, um ponto interessante é que um sinal não precisa, necessariamente, ser esparsos no domínio do tempo: um exemplo disso seria um sinal senoidal, claramente não esparsos no domínio temporal, mas representado por apenas um coeficiente no domínio frequencial. Dessa forma, podemos procurar transformações lineares, a serem aplicadas aos sinais observados, de maneira que a representação dos sinais na base definida por tal transformação seja esparsa, sem, no entanto, alterar a característica de independência mútua das fontes.

Apesar de o exemplo apresentado anteriormente considerar o sinal no domínio frequencial, em geral, a esparsidade de um sinal é obtida por meio de sua representação em uma base sobrecompleta, ou seja, em que existe um número de vetores-base maior do que o necessário para se representar o sinal. Desta forma, é provável que esta nova representação possua menos coeficientes do que a original, deixando, desta forma, o sinal mais esparsos. Dentre as abordagens mais comuns para se obter uma representação esparsa do sinal estão aquelas baseadas em transformadas como *wavelet* e *short-time Fourier transform* (STFT). Um exemplo ilustrativo é fornecido por sinais de áudio, que, apesar de possuírem esparsidade temporal [ABRARD *et al.*, 2001], são mais eficientemente tratados no domínio tempo-frequencial.

3.2.2 Estimação do Número de Fontes e da Matriz de Mistura

Depois de ser garantida a obtenção de uma representação na qual os sinais são considerados esparsos, é possível seguir para o próximo passo, que é tipicamente a estimação do número de fontes e também da matriz de mistura. A maioria dos métodos relacionados a SCA considera que, além de as fontes serem esparsas, é necessário que exista uma certa esparsidade conjunta entre as mesmas, de forma a haver regiões em que não haja sobreposição entre os sinais. Em outras palavras, isto significa que, para que determinado método funcione, é esperado que o número de fontes que estejam ativas ao mesmo tempo deva ser limitado. Uma possibilidade usual é a de haver ortogonalidade disjunta, o que ocorre quando no máximo uma fonte está ativa em cada instante. Matematicamente, tem-se que

$$s_i(n)s_j(n) = 0, \forall i \neq j \quad (3.4)$$

Se voltarmos para o sistema de mistura dado por (2.1), teremos que

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) = \mathbf{a}_i s_i(n), \quad (3.5)$$

onde \mathbf{a}_i é a coluna de \mathbf{A} , que define a *direção* em que será projetada a i -ésima fonte. Desta forma, é possível reconstruir a matriz de mistura através da estimação das diversas direções \mathbf{a}_i das fontes.

Com o intuito de exemplificar a relação entre ortogonalidade disjunta e estimação da matriz de mistura, a Fig. 3.2 mostra dois gráficos de dispersão $x_1 \times x_2$, em uma situação com duas misturas e quatro fontes. Se compararmos a Fig. 3.2 (a) com a Fig. 3.2 (b), perceberemos que a primeira possui uma nuvem densa de pontos enquanto, na segunda figura, os pontos se encontram bem alinhados, sendo que cada direção de alinhamento se associa a uma das colunas da matriz \mathbf{A} .

É interessante notar que a ortogonalidade disjunta não é essencial para a estimação da matriz de mistura: o importante é que as fontes sejam esparsas o suficiente para o que algoritmo proposto consiga estimar todas as direções. Desta forma, a identificação da matriz de mistura pode ser feita através de um processo de clusterização, que pode também ser útil na identificação do número de fontes na mistura [ARBERET *et al.*, 2006; NADALIN *et al.*, 2009; HULLE, 1999].

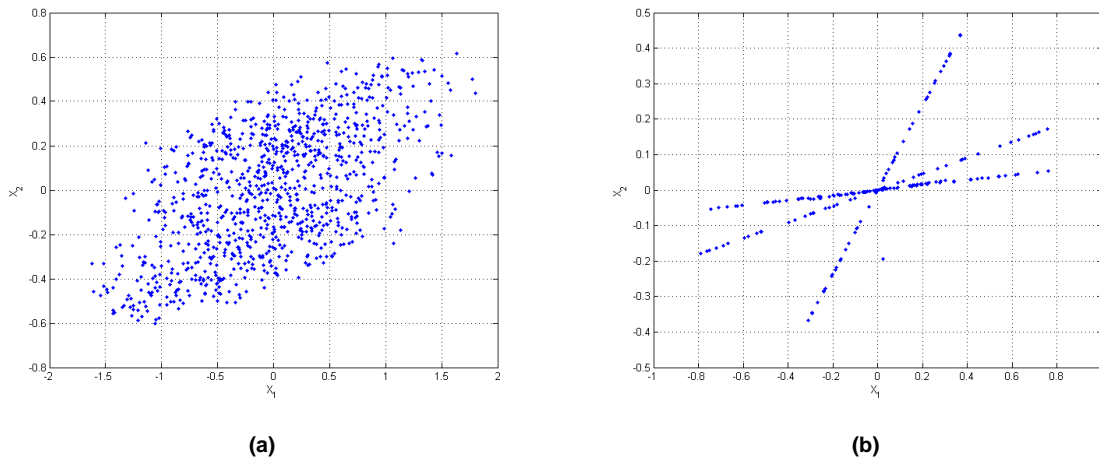


Fig. 3.2 – Gráfico de dispersão de x_1 por x_2 em um caso em que $M = 2$ e $N = 3$. Os sinais se encontram nos domínios (a) do tempo, e (b) tempo-frequencial.

Dentre as técnicas existentes para estimação da matriz de mistura, apresentar-se-ão duas que serão utilizadas como base para métodos propostos neste trabalho.

Através do Espectrograma

O método primeiramente proposto por [ABRARD *et al.*, 2001] se baseia na suposição de que existe uma representação esparsa o suficiente para que, ao longo dos sinais da mistura, só exista, tipicamente, uma fonte ativa por intervalo considerado. A princípio, o método foi desenvolvido para trabalhar com sinais na representação tempo-frequencial (normalmente obtida por meio de uma STFT, dada por

$$X_i(t, f) = \int e^{-jft} x_i(\tau) h(\tau - t) d\tau, \quad (3.6)$$

onde $h(\tau - t)$ é a janela em que a transformada é realizada), o que se justifica pelo pano de fundo de aplicação a sinais de áudio. No entanto, é possível adaptar a proposta para uso no âmbito de qualquer tipo de representação esparsa do sinal.

Em termos gerais, pretende-se encontrar regiões em que só exista uma fonte ativa, e, desta forma, utilizar a informação dada pela diferença entre as amplitudes indicadas em cada sinal de mistura com o intuito de estimar a direção do vetor representado por uma das colunas da matriz de mistura. Será mostrada aqui uma versão adaptada do método original, com alterações influenciadas pelo artigo [HULLE, 1999], para um caso genérico

subparametrizado contendo duas misturas⁶ e N fontes. Desta forma, se a matriz de mistura for definida como sendo

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \end{bmatrix}, \quad (3.7)$$

a Eq. (2.1) pode ser reescrita da seguinte forma

$$\begin{cases} x_1(n) = \sum_{i=1}^N a_{1i} s_i(n) \\ x_2(n) = \sum_{i=1}^N a_{2i} s_i(n) \end{cases} \quad (3.8)$$

Se considerarmos agora uma transformada tempo-frequencial do sinal, dada por

$$\begin{cases} X_1(t, f) = \sum_{i=1}^N a_{1i} S_i(t, f) \\ X_2(t, f) = \sum_{i=1}^N a_{2i} S_i(t, f) \end{cases}, \quad (3.9)$$

onde $X_j(t, f)$ e $S_i(t, f)$ representam, respectivamente, as transformadas tempo-frequencial da j -ésima mistura e da i -ésima fonte, podemos gerar uma razão complexa $\alpha(t, f)$ dada por

$$\alpha(t, f) = \frac{X_1(t, f)}{X_2(t, f)} = \frac{a_{11}S_1(t, f) + a_{12}S_2(t, f) + \cdots + a_{1N}S_N(t, f)}{a_{21}S_1(t, f) + a_{22}S_2(t, f) + \cdots + a_{2N}S_N(t, f)}. \quad (3.10)$$

Se, em um determinado ponto ou região na representação tempo-frequencial (t_c, f_c) , somente uma fonte S_j estiver ativa, todas as outras fontes $S_i \forall i \neq j$ possuirão valor nulo, desta forma, teremos:

$$\alpha_j(t_c, f_c) = \frac{a_{1j}S_j(t_c, f_c)}{a_{2j}S_j(t_c, f_c)} = \frac{a_{1j}}{a_{2j}}, \quad (3.11)$$

que é justamente a razão entre os coeficientes de uma das colunas da matriz de mistura, e, conseqüentemente, o valor do arco tangente desta razão indicará um ângulo θ_j , referente à direção daquela coluna. Como não é possível saber quantas fontes estão ativas, podemos dizer que, para cada instante, existe um θ_c candidato a ser direção de uma das colunas da matriz de mistura, o qual é calculado diretamente a partir dos sinais das misturas, de acordo com

$$\theta_c = \tan^{-1} \frac{X_1(t_c, f_c)}{X_2(t_c, f_c)}. \quad (3.12)$$

⁶ Apesar de o método ser proposto para duas misturas, é possível expandi-lo para uma situação mais genérica com $M > 2$ misturas.

É importante notar que o subscrito c pode ser tanto relativo a um único instante quanto ao valor médio de uma determinada janela de dados, sendo somente necessário adequar a Eq. (3.12) ao caso em questão. Depois de o cálculo acima ter sido feito para todos os dados, uma das formas de se estimar a matriz de mistura é através da clusterização sobre todos os valores de θ calculados, visto que, se na maior parte dos dados houver no máximo uma fonte ativa em cada janela, existirão muitos valores de θ corretos, ou seja, que indicam a direção de uma das colunas de \mathbf{A} . Deste modo, é possível determinar a matriz de mistura a menos da permutação das colunas e de ganhos em amplitude dados por cada coluna. É importante notar que, se o processo for bem sucedido, o número de clusters indicará a quantidade de fontes existentes na mistura.

Através da Procura por Subespaços

Considerando que os dados sejam esparsos em alguma representação, o método proposto em [NAINI *et al.*, 2008] tenta estimar a matriz de mistura \mathbf{A} assumindo que existe uma média de k fontes ativas na mistura. Apesar de o espaço gerado pelos sensores ser menor do que o das fontes, visto que o sistema é subparametrizado, o fato de sinais das fontes serem esparsos implica que os dados observados estarão concentrados em torno de subespaços k -dimensionais, que são gerados por conjuntos de colunas da matriz de mistura. Desta forma, a ideia principal desta abordagem é, primeiro, encontrar esses *subespaços de concentração* das fontes, e, depois, utilizá-los para determinar os vetores de mistura.

A estimação dos subespaços de concentração é feita através da maximização de uma função custo baseada na medida denominada ℓ_0 suave (*Smoothed ℓ_0*) [MOHIMANI *et al.*, 2007], cuja ideia principal é suavizar a medida ℓ_0 a partir da utilização de gaussianas em torno dos valores não-nulos. A função custo é descrita na Eq. (3.13)

$$f_{\sigma}(\mathbf{B}) = \sum_n \exp\left(\frac{-d^2(\mathbf{x}(n), \mathbf{B})}{2\sigma_a^2}\right), \quad (3.13)$$

onde $d(\mathbf{x}(n), \mathbf{B})$ é a distância entre $\mathbf{x}(n)$ e um subespaço k -dimensional arbitrário representado pela matriz $\mathbf{B} \in \mathbb{R}^{M \times k}$. O parâmetro σ_a tem um papel crítico na definição do formato da função custo, pois controla a dispersão das gaussianas que suavizam a norma ℓ_0 . Como discutido pelos autores, quando o valor de σ_a for grande, os picos correspondentes às soluções terão uma base larga, de forma que é possível que dois ou mais picos se

sobreponham formando um só, não sendo, desta forma, possível encontrar todas as direções. Conforme o valor de σ_a vai diminuindo, os picos correspondentes às soluções são mais salientados. Entretanto, se σ_a for muito pequeno, máximos locais que não correspondem aos subespaços de concentração começarão a surgir, o que pode comprometer o processo de identificação. A Fig. 3.3 mostra a superfície da função custo $f_a(\mathbf{B})$ em função dos ângulos que definem as direções de \mathbf{B} , para um caso em que $N = 5$ e $M = 3$.

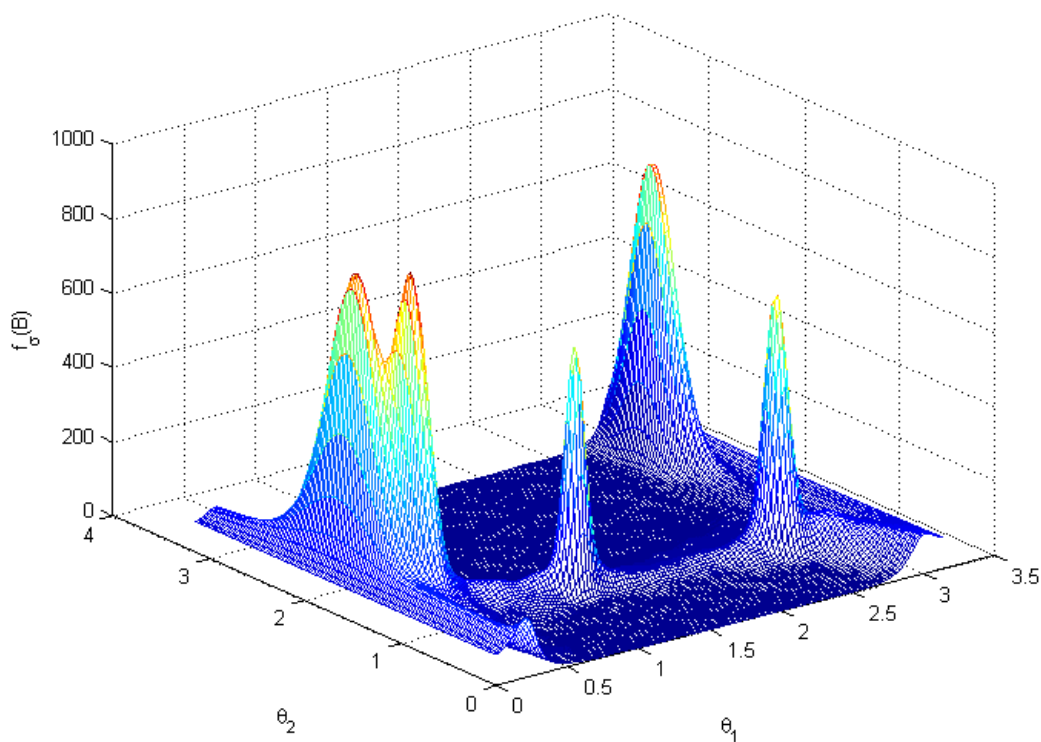


Fig. 3.3 – Superfície da função custo dada por (3.13) para um caso com $N = 5$ e $M = 3$.

Depois de determinar os subespaços de concentração dados pelas matrizes B_i , sendo $i = 1, \dots, N_p$, onde N_p é o número total de subespaços, ainda resta a tarefa de estimar os vetores de mistura, que, de certo modo, aparecem na “intersecção” desses subespaços de concentração. Isto significa que cada vetor de mistura se encontra em q subespaços, onde q depende dos valores de N e k . Por conseguinte, o algoritmo proposto em [NAINI *et al.*, 2008] busca vetores que melhor representem estas intersecções, os quais são candidatos a serem as direções que definem a matriz de mistura, sendo que esta otimização é feita com

algoritmos baseados na informação trazida pelo vetor gradiente. Uma vez que cada vetor associado a uma fonte deve estar contido / próximo de q subespaços gerados, vetores estimados que estejam próximos a um número menor de subespaços serão descartados. Desta forma, idealmente, o número final de vetores informa a quantidade de fontes na mistura, e a composição de suas direções forma a matriz de mistura.

3.2.3 Separação / Reconstrução das Fontes

Uma vez que a estimação da matriz de mistura tenha sido feita, o último passo é realizar a separação ou reconstrução das fontes. A diferença entre estes dois termos (separação e reconstrução) vem do fato de que, se existirem regiões em que ocorra algum tipo de sobreposição dos sinais das fontes de forma a criar uma ambiguidade na solução final, não será possível separar as fontes de forma integral. Nestes casos, são propostos métodos que tentam reconstruir os dados nestas regiões.

É mostrado em [RICKARD, 2006] que é teoricamente possível separar as fontes no caso de ortogonalidade disjunta entre as fontes, mesmo com $M < N$. Esta hipótese é utilizada por [BOFILL e ZIBULEVSKY, 2000] para o algoritmo chamado de *shortest path*, que consiste em realizar uma otimização da norma ℓ_1 da fonte s_i sujeita à Eq. (3.8). Desta forma, como não existe sobreposição entre as fontes, cada amostra da mistura será simplesmente o valor não-nulo da fonte ativa naquele instante, a menos de um ganho dado pelo sistema misturador.

Similarmente ao que vem ocorrendo nos trabalhos relacionados à estimação da matriz de mistura, nas abordagens que propõem separar sinais esparsos, hipóteses menos restritivas do que a ortogonalidade disjunta vêm sendo utilizadas. Dentre os trabalhos que estenderam suas metodologias para situações de fontes não disjuntas, podemos citar, por exemplo, o algoritmo proposto em [AISSA-EL-BEY *et al.*, 2004], na qual é apenas requerido que todos os pares formados pelas colunas das matrizes sejam linearmente independentes, e que o número de fontes ativas seja menor do que o de misturas em todas as janelas de dados escolhidas.

3.3 Considerações Finais

O conceito de esparsidade e a abordagem de SCA dele derivada têm sido empregados com sucesso na descrição de informação *a priori* em separação de fontes [BOFILL e ZIBULEVSKY, 2000] e em outras importantes áreas de processamento de sinais, como *denoising* [CHEN e DONOHO, 1998] e *compressed sensing* [CANDÈS *et al.*, 2006; DONOHO, 2006]. Não obstante, o uso de esparsidade em processamento de sinais ainda é relativamente recente, de modo que há aspectos a explorar tanto do ponto de vista de relações com conceitos mais consolidados como o de independência, quanto em termos da adequação de métricas e procedimentos de otimização empregados. Antes, porém, de passarmos a uma análise desses aspectos, exporemos uma contribuição relativa ao entendimento das limitações estruturais inerentes à inversão de sistemas subparametrizados.

Capítulo 4

Limitante Relacionado à Estimação de Fontes por Estruturas Lineares

Classicamente os problemas de BSE e BSS são resolvidos de modo exato em um contexto teórico que inclui duas hipóteses: o processo de mistura subjacente é linear e também é inversível. Estas hipóteses, aceitáveis em determinadas aplicações, possibilitam uma recuperação perfeita das fontes via um sistema linear de separação. Entretanto, quando o número de fontes é maior que o de misturas, a hipótese de extração / separação perfeita não se mantém. Isto também ocorre no âmbito da tarefa de equalização cega monocal [ROMANO *et al.*, 2010], um problema tipicamente subparametrizado em consequência das particularidades associadas à tarefa de inversão de modelos de canal de resposta ao impulso finita (*Finite Impulse Response* – FIR) através da utilização de um equalizador também FIR.

Uma abordagem para o problema de se estabelecer um valor de referência para o caso subparametrizado pode ser feita considerando-se uma situação supervisionada, visto que existe aí uma vantagem teórica com relação aos cenários cegos. Em outras palavras, a maior informação disponível no caso supervisionado o credencia como base para uma análise de limitantes relativos à estrutura de processamento. Faremos isto, conforme será mostrado a seguir, utilizando o critério de erro quadrático médio.

4.1 Critério de Wiener (MSE)

Neste capítulo, busca-se obter um limitante para a qualidade da estimação de múltiplas fontes linearmente misturadas (em ambos os contextos de BSE / BSS e de equalização). Este limite será baseado no critério de Wiener – i.e. no erro quadrático médio (*Mean-Squared Error* – MSE) – escolhido por pelo menos dois motivos. O primeiro é justamente o fato de o critério de MSE ser supervisionado, o que significa que ele utiliza efetivamente informações sobre os sinais a serem recuperados, estando, assim, em uma posição *a priori* vantajosa em comparação com os métodos cegos. O segundo motivo é o fato de que existe uma relação direta entre o critério MSE e o uso de projeções euclidianas em um espaço combinado (modelo de mistura + sistema de separação / canal + equalizador). Essa relação leva a uma interpretação geométrica do problema de recuperação de fontes que tem sido útil em diversos trabalhos importantes [GU e TONG, 1999; DING e KENNEDY, 1992; REGALIA e MBOUP, 1999], e que será relevante para ilustrar o processo de “perda de informação” que é característico dos casos subparametrizados.

Do ponto de vista matemático, podemos partir do modelo de mistura dado pela Eq. (2.1). Se o problema for de BSS ou BSE, o vetor de fontes é dado por $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$, onde $s_i(n)$ é a i -ésima fonte. Assume-se ainda que $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_M(n)]^T$ é um vetor composto por M misturas. Se definirmos o problema de equalização – mais especificamente no caso SISO (*single-input / single-output*) – como mostrado na Fig. 4.1 (a), $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T$ corresponderá ao vetor de versões atrasadas do mesmo sinal. É interessante notar que é possível reescrever este sistema misturador como sendo um modelo MIMO (*multiple-input / multiple-output*), como mostra a Fig. 4.1 (b). Neste caso, as diversas fontes serão dadas pelas versões atrasadas de um único sinal de informação, i.e., $\mathbf{s}(n) = [s(n), s(n-1), \dots, s(n-N+1)]^T$. Desta forma, o problema de equalização pode ser descrito como um caso particular de BSS.

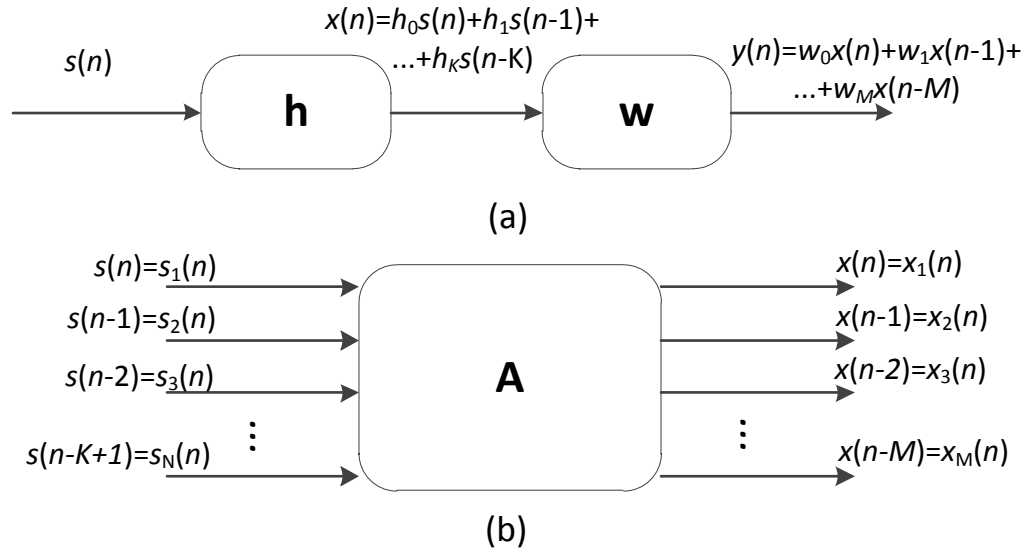


Fig. 4.1 – (a) Modelo de equalizador SISO; (b) o canal do equalizador visto como sendo um problema de BSS.

A distinção em relação a $\mathbf{s}(n)$ não é de grande relevância para este trabalho, visto que assumimos que nos casos de BSS e BSE, as fontes são mutuamente independentes, e que no caso de equalização, o sinal transmitido é composto por amostras independentes e identicamente distribuídas (i.i.d.). Em ambos os casos, a matriz \mathbf{A} é $M \times N$, onde, no caso de equalização, N é igual a $M+K-1$, sendo K o comprimento do canal FIR. No contexto de BSS/BSE, como discutido, é possível, em certas circunstâncias, assumir que $M = N$, o que leva à possibilidade de inversão perfeita, ao passo que, no caso de equalização de um canal SISO, uma hipótese deste tipo seria extremamente restritiva. Note que o cenário subparametrizado surge quando $M < N$.

Caso se deseje estimar $s_i(n)$, o i -ésimo elemento de $\mathbf{s}(n)$, o critério de Wiener (MSE) pode ser descrito em termos da seguinte função custo:

$$J(\mathbf{w}) = E\{[s_i(n) - y(n)]^2\} \quad (4.1)$$

onde $y(n)$, a versão estimada de $s_i(n)$, é encontrada a partir da Eq. (2.2), sendo que \mathbf{w} é o vetor de parâmetros com respeito ao qual a função custo deve ser minimizada. Se a função for devidamente expandida, é possível escrever [HAYKIN, 1995]:

$$J(\mathbf{w}) = \mathbf{w}^T \mathbf{R}_{xx} \mathbf{w} - \mathbf{w}^T \mathbf{p} - \mathbf{p}^T \mathbf{w} + \sigma_{s_i}^2, \quad (4.2)$$

sendo \mathbf{R}_{xx} a matriz de correlação, definida como:

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \quad (4.3)$$

e \mathbf{p} o vetor de correlação cruzada, definido como:

$$\mathbf{p} = E\{s_i(n)\mathbf{x}(n)\}, \quad (4.4)$$

e $\sigma_{s_i}^2$ é a variância do sinal desejado $s_i(n)$. Assume-se que todo sinal envolvido possui média zero e é estacionário no sentido amplo. Também, por uma questão de simplicidade, considera-se que todos os elementos de $\mathbf{s}(n)$ possuem potência unitária. Note que não existe perda de generalidade nesta última suposição, visto que qualquer valor distinto pode ser gerado a partir de um ajuste adequado dos coeficientes da matriz \mathbf{A} .

O mínimo de (4.1), a chamada solução de Wiener, é dado por:

$$\mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1}\mathbf{p}. \quad (4.5)$$

As suposições levantadas, juntamente com a hipótese de independência previamente exposta, permitem-nos escrever, com a ajuda de (2.1):

$$\mathbf{R}_{xx} = E\{\mathbf{A}\mathbf{s}(n)\mathbf{s}^T(n)\mathbf{A}^T\} = \mathbf{A}E\{\mathbf{s}(n)\mathbf{s}^T(n)\}\mathbf{A}^T = \mathbf{A}\mathbf{I}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T \quad (4.6)$$

e

$$\mathbf{p} = E\{s_i(n)\mathbf{A}\mathbf{s}(n)\} = \mathbf{A}E\{s_i(n)\mathbf{s}(n)\} = \mathbf{A}\mathbf{u}_i = \mathbf{a}_i, \quad (4.7)$$

sendo $\mathbf{u}_i = [0 \ 0 \ \dots \ 1(i\text{-ésimo}) \ \dots \ 0 \ 0]^T$ e \mathbf{a}_i a i -ésima coluna de \mathbf{A} . Portanto, (4.5) pode ser reescrita como:

$$\mathbf{w}_{opt} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{u}_i = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}_i. \quad (4.8)$$

Seguiremos agora uma linha de raciocínio distinta. Usando (2.2), podemos escrever:

$$y(n) = \mathbf{w}^T\mathbf{x}(n) = \mathbf{w}^T\mathbf{A}\mathbf{s}(n) = \mathbf{c}^T\mathbf{s}(n), \quad (4.9)$$

sendo $\mathbf{c} = \mathbf{A}^T\mathbf{w}$ um vetor combinado que estabelece uma relação direta entre $y(n)$ e o sinal a ser estimado. Portanto, de acordo com o espírito do problema supervisionado definido em (4.1), idealmente, gostaríamos de ter a distância entre \mathbf{c} e a configuração ideal \mathbf{u}_i como sendo a menor possível. Em outras palavras, podemos definir a seguinte função custo:

$$J^*(\mathbf{w}) = \|\mathbf{u}_i - \mathbf{A}^T\mathbf{w}\|^2 = (\mathbf{u}_i - \mathbf{A}^T\mathbf{w})^T(\mathbf{u}_i - \mathbf{A}^T\mathbf{w}). \quad (4.10)$$

Interessantemente, uma expansão de (4.10) leva a:

$$J^*(\mathbf{w}) = \mathbf{u}_i^T\mathbf{u}_i - \mathbf{w}^T\mathbf{A}\mathbf{u}_i - \mathbf{u}_i^T\mathbf{A}^T\mathbf{w} + \mathbf{w}^T\mathbf{A}\mathbf{A}^T\mathbf{w}. \quad (4.11)$$

Levando em consideração as suposições destacadas – fontes i.i.d. e sistema de mistura linear e instantâneo, é possível verificar que $J^*(\mathbf{w})$ é idêntica a $J(\mathbf{w})$, e, desta forma, sua solução ótima também é dada por (4.5). Esta equivalência mostra um ponto interessante: é possível olhar o problema da estimação de uma fonte no sentido do mínimo erro quadrático médio (*minimum* MSE – MMSE) como sendo o problema de encontrar o

valor de \mathbf{w} que gera uma resposta combinada \mathbf{c} tão próxima quanto possível, em termos do quadrado da distância euclidiana, de uma resposta combinada ideal \mathbf{u}_i .

É interessante ressaltar que a solução MMSE corresponde à projeção ortogonal do vetor de resposta ideal no espaço gerado pelas misturas, o que é feito por meio do produto $\mathbf{O} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$, que representa o operador de projeção [TREFETHEN e BAU, 1997]. Note que, dependendo de \mathbf{A} , pode não ser possível obter todos os possíveis vetores \mathbf{c} . Desta forma, o conjunto de vetores \mathbf{c} gerados pelos vetores \mathbf{w} para um dado problema corresponde a um *conjunto atingível* [LI e DING, 1995]. É exatamente o fato de que, para um problema subparametrizado genérico, vetores ideais \mathbf{u}_i não estão contidos no conjunto atingível que explica a impossibilidade de uma perfeita recuperação das fontes através de uma estrutura linear (definida por \mathbf{w}) e do critério MSE. Este erro existente de estimação pode ser expressado por meio da distância euclidiana entre o conjunto atingível e vetores que não pertencem a este conjunto.

Fazendo uma interpretação geométrica, o processo de mistura irá corresponder a projeção das informações presentes em um espaço N -dimensional em outro espaço M -dimensional de menor dimensão. Desse ponto de vista, os sistemas de extração e de equalização lineares possuem um papel complexo: operando em um espaço de menor dimensão, eles têm como objetivo encontrar a projeção que seja mais adequada à estimação de uma fonte, i.e., de um sinal associado a um espaço de maior dimensão. Se estivermos lidando com BSS, a tarefa se torna ainda mais ambiciosa: a ideia é encontrar N projeções que levem às estimativas associadas a todas as coordenadas do espaço original a partir da informação transmitida pelas projeções num espaço de menor dimensão (as misturas). Esta incompatibilidade está por trás do fato de que, exceto por alguns casos especiais, todas as fontes são recuperadas com um MSE residual.

As Figs. 4.2 (a) e (b) mostram a interpretação geométrica para um caso de três fontes independentes e duas misturas. Como o sistema é subparametrizado, a dimensão do subespaço de vetores atingíveis \mathbf{c} é menor do que a de \mathbf{u}_i ; desta forma, não é possível uma recuperação perfeita das fontes. No caso da Fig. 4.2 (a), o plano que define o conjunto atingível por \mathbf{c} não coincide com nenhuma fonte, diferentemente da Fig. 4.2 (b), onde duas fontes são recuperadas perfeitamente e uma das fontes se encontra ortogonal ao subespaço.

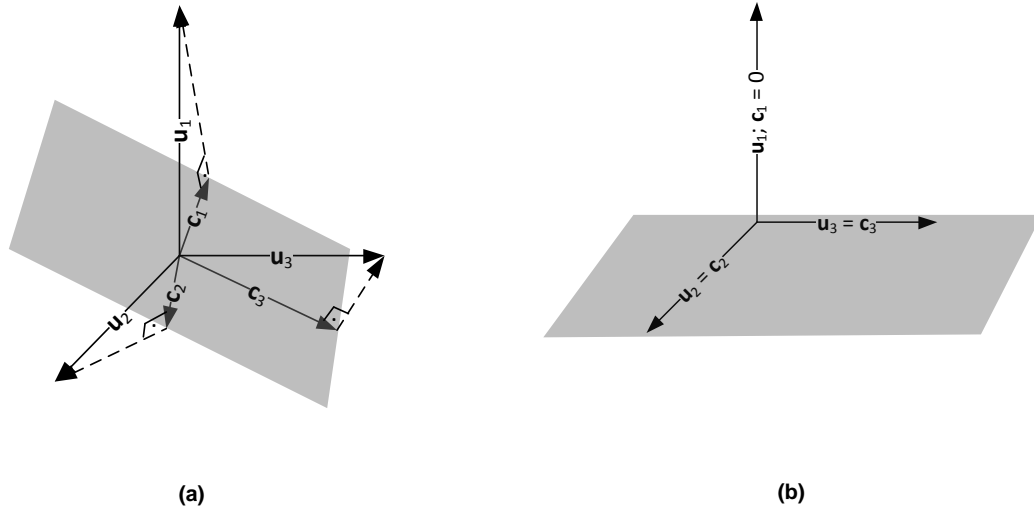


Fig. 4.2 – Interpretação geométrica para um caso em que $N = 3$ e $M = 2$, situações em que o plano criado pelo conjunto atingível de c : (a) não coincide com nenhuma fonte; e (b) coincide com duas fontes.

4.2 Caso de Separação de Fontes

4.2.1 Análise da Soma dos MSEs Ótimos

A fim de analisar o erro de estimação das fontes em sistemas de misturas subparametrizados, vejamos qual seria o mínimo erro de estimação obtido quando a recuperação das fontes é feita por meio de uma estrutura linear. Deste modo, definimos um critério erro quadrático médio global, que corresponde à soma dos erros quadráticos médios associados à estimação de cada uma das fontes, e que é matematicamente definido pela seguinte função custo:

$$J(\mathbf{W}) = \|\mathbf{e}(n)\|^2 = \text{Tr}\{E\{\mathbf{e}(n)\mathbf{e}^T(n)\}\} \quad (4.12)$$

onde $\mathbf{e}(n) = \mathbf{W}\mathbf{x}(n) - \mathbf{s}(n)$. Depois de algumas manipulações, é possível mostrar que (4.12) pode ser escrita como se segue

$$J(\mathbf{W}) = \text{Tr}(\mathbf{W}\mathbf{R}_{\mathbf{xx}}\mathbf{W}^T - \mathbf{W}\mathbf{P}_{\mathbf{xs}} - \mathbf{P}_{\mathbf{xs}}^T\mathbf{W}^T + \mathbf{R}_{\mathbf{ss}}) \quad (4.13)$$

onde

$$\mathbf{P}_{\mathbf{xs}} = E\{\mathbf{x}(n)\mathbf{s}^T(n)\} = \mathbf{A}\mathbf{R}_{\mathbf{ss}} = \mathbf{A}. \quad (4.14)$$

Note que estamos usando o fato de $\mathbf{R}_{\mathbf{ss}}$ ser a identidade, o que decorre da hipótese de que as fontes são mutuamente independentes e que a suas variâncias são unitárias.

O mínimo valor da função custo dada (4.13) é obtido para \mathbf{W} tal que o gradiente da função custo seja nulo, i.e.

$$\frac{\partial}{\partial \mathbf{W}} J(\mathbf{W}) = 2\mathbf{W}\mathbf{R}_{\mathbf{xx}} - 2\mathbf{P}_{\mathbf{xs}}^T = 0 \quad (4.15)$$

Assim, essa solução pode ser escrita como sendo

$$\mathbf{W}_{\text{MMSE}} = \mathbf{P}_{\mathbf{xs}}^T \mathbf{R}_{\mathbf{xx}}^{-1}, \quad (4.16)$$

que corresponde à solução de Wiener.

O mínimo valor de (4.13) pode ser encontrado através da substituição de (4.16) em (4.13):

$$J(\mathbf{W}_{\text{MMSE}}) = \text{Tr}(\mathbf{R}_{\mathbf{ss}} - \mathbf{P}_{\mathbf{xs}}^T \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{P}_{\mathbf{xs}}). \quad (4.17)$$

Tendo por base (4.14), o fato de que, por suposição, $\mathbf{R}_{\mathbf{xx}} = \mathbf{A}\mathbf{A}^T$ e $\mathbf{R}_{\mathbf{ss}} = \mathbf{I}$, e a observação de que o traço é invariante a permutações cíclicas, i.e., $\text{Tr}(\mathbf{XYZ}) = \text{Tr}(\mathbf{YZX})$, pode-se reescrever (4.17) como:

$$J(\mathbf{W}_{\text{MMSE}}) = \text{Tr}(\mathbf{R}_{\mathbf{ss}}) - \text{Tr}(\mathbf{P}_{\mathbf{xs}}^T \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{P}_{\mathbf{xs}}) \quad (4.18)$$

$$= \text{Tr}(\mathbf{I}) - \text{Tr}(\mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{A}) \quad (4.19)$$

$$= N - \text{Tr}((\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{A}\mathbf{A}^T) \quad (4.20)$$

$$= N - M. \quad (4.21)$$

É importante observar que a matriz $\mathbf{A}\mathbf{A}^T$ é considerada inversível.

A Eq. (4.21), principal contribuição apresentada neste capítulo, revela que, para a solução de Wiener em BSS, o MSE total é dado pela diferença entre o número de fontes e o de misturas. Por exemplo, se $N = M$, o valor de MSE é nulo, o que é esperado, visto que é possível a inversão perfeita da matriz de mistura. Conseqüentemente, se a diferença entre N e M aumenta, o mesmo ocorre com o erro resultante. É interessante notar que se houver ruído aditivo, este será tratado como sendo uma nova fonte, como será mostrado na seção de simulações e resultados.

Outro ponto interessante a ser observado é que, como temos N fontes contribuindo para o MSE total, a partir de (4.21), podemos dizer que o erro médio associado à recuperação de uma fonte é dado por

$$\text{MSE}_{\text{médio}} = \frac{N - M}{N} = 1 - \frac{M}{N}. \quad (4.22)$$

Como esperado, pelo menos uma fonte será recuperada com $\text{MSE} \leq \text{MSE}_{\text{médio}}$ e uma fonte será recuperada com $\text{MSE} \geq \text{MSE}_{\text{médio}}$. Portanto, em um caso subparametrizado, conforme

o número de misturas aumenta, o “melhor caso” de MSE vai se tornando cada vez mais favorável.

4.2.2 Simulações e Resultados

Com o intuito de ilustrar o resultado mostrado pela Eq. (4.21), foram realizadas simulações em três cenários. O primeiro deles se vincula à separação simultânea de fontes no caso de matrizes subdeterminadas. Depois, mostra-se que a linha de análise proposta também é útil em situações para as quais as observações são corrompidas por ruído aditivo. Por fim, o último cenário mostra os resultados relativos ao aumento do número de misturas para um mesmo valor de N .

Primeiro Cenário

Primeiramente, será considerado um caso simples de sistema subdeterminado, para o qual $N = 3$, $M = 2$ e a matriz de mistura é dada por

$$\mathbf{A} = \begin{bmatrix} 1 & 0,6 & -0,3 \\ 0,3 & 1 & -0,6 \end{bmatrix} \quad (4.23)$$

Na Tab. 4.1, são mostradas as soluções de Wiener associadas às estimações de todas as três fontes e seus respectivos valores de MSE. É importante notar que a soma dos elementos da última coluna é – desconsiderando os erros de aproximação – igual a um, como esperado a partir de (4.21). Outro ponto interessante é que, devido à estrutura da matriz (e das projeções realizadas por ela), a primeira fonte pôde ser recuperada com um grau significativo de precisão. Desta forma, pode-se dizer que a maior parte do erro total foi distribuída entre as duas outras estimativas, sendo que a estimativa da terceira fonte foi a mais prejudicada.

Número da fonte	Solução de Wiener	MSE
1	[1,20286 -0,68903]	0,0038457
2	[-0,22434 0,85675]	0,2778549
3	[0,22754 -0,58327]	0,7182993

Tab. 4.1 – Separação de fontes: Soluções de Wiener e MSEs correspondentes

Para exemplificar este caso, são mostradas na Fig. 4.3 amostras das fontes geradas de forma aleatória utilizando uma distribuição uniforme (indicadas na figura por ‘x’) e os valores de \mathbf{y} (indicados por ‘o’), encontrados a partir da extração utilizando as soluções de

Wiener encontradas. Como esperado, a primeira fonte é estimada com precisão significativa, ao contrário das demais.

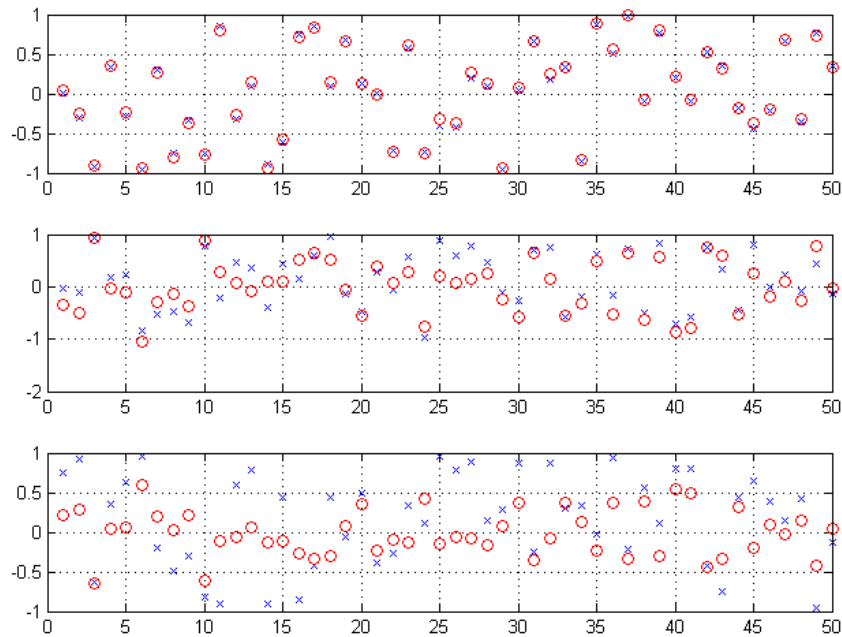


Fig. 4.3 – Fontes originais (x), e recuperadas (o). É possível ver que somente a primeira fonte conseguiu ser recuperada com sucesso.

Segundo Cenário

Consideremos agora o mesmo cenário, exceto pela inclusão de um ruído aditivo e branco de variância σ^2 . Visto que cada uma das duas misturas possui um termo de ruído independente, esta situação pode ser modelada a partir de uma nova matriz em que o número de fontes é aumentado de dois (i.e. $N = 5$). Desta forma, a nova matriz de mistura é dada por

$$\mathbf{A} = \begin{bmatrix} 1 & 0,6 & -0,3 & \sigma & 0 \\ 0,3 & 1 & -0,6 & 0 & \sigma \end{bmatrix} \quad (4.24)$$

Sabemos, por meio da aplicação de (4.21), que a soma dos MSEs será igual a $N - M = 3$. Entretanto, um aspecto chave é a questão de como este erro será dividido. Vamos começar a análise por um caso em que o ruído possui potência relativamente baixa em relação às fontes: $\sigma^2 = 0,005$. A Tab. 4.2 mostra as soluções de Wiener, bem como os respectivos valores de MSE. A partir destes resultados, pode-se notar que, apesar de as duas “fontes de ruído” terem causado um aumento de 2 no valor da soma dos MSEs, elas

sozinhas “absorveram” um MSE total de 1,9847, o que significa que apenas uma pequena parcela deste ruído foi distribuída entre as outras estimativas.

Número da fonte	Solução de Wiener	MSE
1	[1,189743 -0,676923]	0,013333
2	[-0,217751 0,848915]	0,281736
3	[0,222485 -0,577515]	0,720237
4 (ruído)	[0,108228 -0,080334]	0,992347
5 (ruído)	[-0,080334 0,108228]	0,992347

Tab. 4.2 – Caso ruidoso ($\sigma^2 = 0,005$): Soluções de Wiener e MSEs correspondentes

Se for aumentado o valor de σ^2 , o MSE total permanece o mesmo, porém as duas estimativas das “fontes” adicionais não irão incorporar uma parcela tão significativa do erro. Por exemplo, se $\sigma^2 = 0,05$, a soma dos dois últimos MSEs será 1,8616, o que significa que as três fontes de informação terão que dividir um valor adicional de MSE igual a 0,13843. Quando o valor de σ^2 se aproxima da potência unitária, as duas “fontes de ruído” começam a “exigir” uma recuperação mais justa, degradando significativamente os sinais desejados.

Terceiro Cenário

Para mostrar o potencial das Eqs. (4.21) e (4.22), é proposto um cenário com 50 fontes e o número de misturas variando entre 10 e 49. Dentre os resultados de $J(\mathbf{w})$ encontrados através das soluções de Wiener, são mostrados os valores mínimo, máximo e médio da função, além do valor total $J(\mathbf{W}_{\text{MMSE}})$. Como esperado, conforme M aumenta, todos os valores de erro diminuem, sendo que (4.21) é sempre válida. Além disso, a Eq. (4.22) se mostra uma boa referência sobre o que ocorre com os valores de $J(\mathbf{w})$ relativos a cada fonte.

Quantidade de Misturas	MSE min	MSE max	MSE médio	MSE total
10	0,659356	0,916151	0,800000	40,000000
20	0,431645	0,758461	0,600000	30,000000
30	0,209049	0,646127	0,400000	20,000000
40	0,081797	0,367136	0,200000	10,000000
49	5×10^{-7}	0,125032	0,020000	1,000000

Tab. 4.3 – Caso com 50 fontes e o número de misturas variando. São mostrados os valores mínimo, máximo, médio e total de $J(\mathbf{w})$.

4.2.3 Análise para Sinais Correlacionados

Voltaremos ao modelo indicado pela Eq. (2.1), página 6, mas agora considerando uma situação em que as fontes são correlacionadas. O intuito é investigar como o grau de correlação entre as fontes influencia a soma – dada pela Eq. (4.17) – dos valores de MSE. Para isto, vamos considerar uma situação simples em que $M = 1$ e $N = 2$. Além disto, suponhamos que a matriz de correlação das fontes seja dada por:

$$\mathbf{R}_{ss} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \quad (4.25)$$

e que a matriz de mistura seja dada por $\mathbf{A} = [a_1 \ a_2]$. Neste cenário, a Eq. (4.17) pode ser escrita como:

$$J(\mathbf{W}_{MMSE}) = 2 - \frac{(a_1 + ra_2)^2 + (ra_1 + a_2)^2}{[(a_1 + ra_2)a_1 + (ra_1 + a_2)a_2]} \quad (4.26)$$

Levando em conta que $-1 \leq r \leq 1$, pode-se mostrar que, se $a_1 > a_2$, a maximização de (4.26) com respeito a r ocorre quando $r = -a_2 / a_1$. Reciprocamente, se $a_1 < a_2$, então (4.26) tem seu máximo valor em $r = -a_1 / a_2$. Para ilustrar este resultado, consideraremos um exemplo numérico, com $\mathbf{A} = [1 \ 0,5]$. Na Fig. 4.4, é mostrado o valor de $J(\mathbf{W}_{MMSE})$ em função do parâmetro r , isto é, a correlação entre as fontes. Antes de tudo, notemos que, quando $r = -1$ ou $r = 1$, a soma de MSEs se torna nula. Isto é esperado, visto que, em tal situação, as fontes são perfeitamente correlacionadas, o que, na prática, leva a um caso com apenas uma fonte e uma mistura. Em segundo lugar, pode-se observar que o pior caso – dado pelo máximo de $J(\mathbf{W}_{MMSE}) = 1,25$ – é atingido com $r = -a_2 / a_1 = -0,5$. Finalmente, quando $r = 0$, obtém-se $J(\mathbf{W}_{MMSE}) = 1$, que é, como esperado, o resultado previsto pela Eq. (4.21). Este exemplo indica que fontes correlacionadas podem prover tanto valores maiores quanto menores de MSE total, dependendo da estrutura de correlação subjacente, o que pode ser relevante, por exemplo, para a concepção / análise de pré-codificadores baseados em MSE.

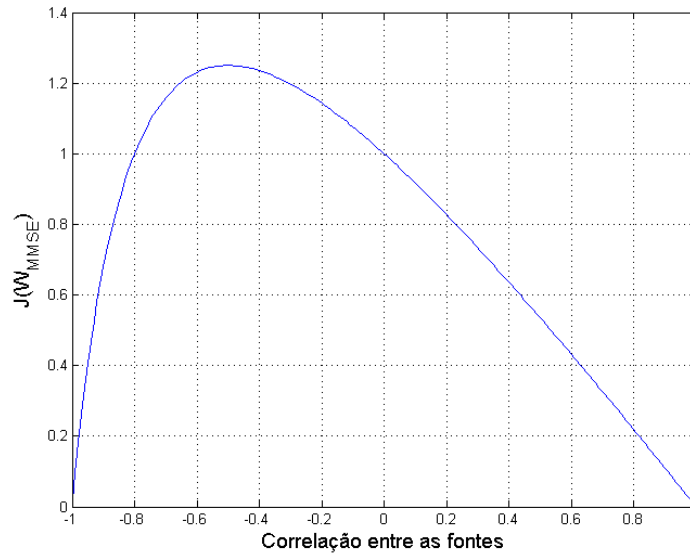


Fig. 4.4 – $J(\mathbf{W}_{MMSE})$ como uma função da correlação r entre as fontes.

4.3 Caso de Equalização

Como exposto na Seção 4.1, o problema de equalização pode ser formulado como um problema de BSS. Deste modo, o resultado expresso na Eq. (4.21) pode ser estendido para o problema de equalização SISO de um canal FIR usando um equalizador FIR. Neste caso, se utilizarmos a formulação descrita na referida seção, a função custo expressa em (4.13) corresponderá à soma dos MSEs associados à recuperação de todas as versões atrasadas que compõem $\mathbf{s}(n)$, dada por $J(\mathbf{W}_{MMSE}) = K - 1$. De forma análoga ao caso de separação de fontes, temos $K + M - 1$ estimativas possíveis, sendo cada uma obtida para um atraso diferente de equalização. Analogamente à Eq. (4.22), o valor médio de MSE para cada atraso é dado por

$$\text{MSE}_{\text{médio}} = \frac{(K - 1)}{(K + M - 1)}, \quad (4.27)$$

de forma que existe pelo menos um atraso de equalização que dá origem a estimativa com $\text{MSE} \leq \text{MSE}_{\text{médio}}$ e um com $\text{MSE} \geq \text{MSE}_{\text{médio}}$. Este valor diminui com o aumento do comprimento M do equalizador, o que justifica a bem estabelecida e intuitiva ideia de que o acréscimo de coeficientes a esse dispositivo tem o potencial de levar a uma melhoria de desempenho.

4.3.1 Simulações e Resultados

Para analisar o limite proposto no caso de equalização, são propostos dois cenários, cada um caracterizado por um canal diferente.

Primeiro Cenário

Pretende-se equalizar um canal FIR com parâmetros $\mathbf{h} = [1 \ 0,4 \ 0,2]$ com um equalizador FIR de dois coeficientes. A matriz \mathbf{A} pode ser descrita como sendo uma matriz de convolução associada ao canal \mathbf{h} :

$$\mathbf{A} = \begin{bmatrix} 1 & 0,4 & 0,2 & 0 \\ 0 & 1 & 0,4 & 0,2 \end{bmatrix} \quad (4.28)$$

Como mencionado anteriormente, no contexto de equalização, as diferentes fontes são as versões atrasadas do mesmo sinal. Neste caso, a partir de (4.28), podemos dizer que temos $N = 4$ e $M = 2$.

As soluções de Wiener obtidas e os valores de MSE são mostrados na Tab. 4.4. Nota-se que a soma dos valores de MSE novamente confirma a validade de (4.21), sendo o valor de $J(\mathbf{W}_{\text{MMSE}})$, neste caso, igual a 2. Uma vez mais, a distribuição dos MSEs é desigual, sendo que, claramente, as duas primeiras fontes (atrasos) são favorecidas.

Número do atraso	Solução de Wiener	MSE
1	[0.99206 -0.39683]	0.0079365
2	[0.00000 0.83333]	0.1666667
3	[0.03968 0.31746]	0.8650794
4	[-0.07937 0.19841]	0.9603175

Tab. 4.4 – Primeiro cenário de equalização: Soluções de Wiener e MSEs correspondentes

No caso de equalização, isto pode ser interpretado em termos da característica de fase mínima do canal em questão, que tende a favorecer os atrasos de equalização mais próximos de zero. Nota-se ainda que a Tab. 4.4 confirma nossa previsão de que pelo menos uma fonte deveria ser recuperada com $\text{MSE} \leq 0,5$ e pelo menos uma com $\text{MSE} \geq 0,5$.

Segundo Cenário

Consideremos agora um problema discutido em [MONTALVÃO *et al.*, 1999; CHAVES *et al.*, 2007]: a equalização de um canal com um zero em +1 ou -1. Considerando o caso mais simples possível, com $\mathbf{h} = [1 \ 1]$ e o equalizador sendo um filtro de dois coeficientes, temos a matriz \mathbf{A} da seguinte forma

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad (4.29)$$

A Tab. 4.5 traz tanto as soluções de Wiener quanto seus respectivos valores de MSE.

Número do atraso	Solução de Wiener	MSE
1	[0.66667 -0.33333]	0.33333
2	[0.33333 0.33333]	0.33333
3	[-0.33333 0.66667]	0.33333

Tab. 4.5 – Segundo cenário de equalização: Soluções de Wiener e MSEs correspondentes

Nota-se que, neste caso, o MSE total (igual a $N - M = 1$) é dividido igualmente entre as fontes (que são recuperadas exatamente com o valor do $\text{MSE}_{\text{médio}}$): isto faz com que seja particularmente difícil lidar com este canal. Uma possibilidade interessante neste caso é fazer uso da fronteira dada por $\text{MSE}_{\text{médio}}$ e aumentar o número de coeficientes do equalizador até que o valor desejado de MSE seja atingido. Por exemplo, se $M = 2$, como $K = 2$, $N = 6$ e teríamos $\text{MSE}_{\text{médio}} = 1/6$, que seria exatamente o valor de cada MSE. Se o nível ainda não é baixo o suficiente, um novo aumento pode ser considerado.

4.4 Considerações Finais

Tendo em vista o interesse em melhor compreender o problema de BSS nos casos em que o sistema misturador é subparametrizado, assim como o problema de equalização temporal, foi proposto neste capítulo um limitante do erro de estimação das fontes baseado no critério de Wiener. Como resultado mais importante, foi verificado que a soma dos valores de MSE entre as fontes originais e as recuperadas é sempre igual a $N - M$. Além disso, foi possível investigar o efeito do aumento do número de misturas e do comprimento do equalizador sobre o desempenho do processo de estimação e também interpretar sob uma nova perspectiva o modo pelo qual um ruído aditivo prejudica a recuperação de sinais de informação.

Capítulo 5

Separação de Sinais Baseada em Esparsidade e Independência

No Capítulo 2, foi descrito o problema de separação cega de fontes, tendo sido mostrado que, quando as misturas são lineares, instantâneas e invariantes no tempo, e o número de fontes N não ultrapassa o de sensores M , é possível separar as fontes utilizando a hipótese de independência dos sinais. Por outro lado, como mostrado no Capítulo 4, quando $N > M$, não é possível separar perfeitamente as fontes usando um sistema linear de separação. Foi visto no Capítulo 3 que esse é um fator que motiva a substituição da hipótese de independência pela suposição de esparsidade. Porém, cabe ressaltar que as hipóteses não são disjuntas: de fato, em algumas situações, é razoável supor que as fontes, além de serem mutuamente independentes, possuam um certo grau de esparsidade.

O problema abordado no presente capítulo é como utilizar os conceitos de independência e esparsidade de forma simultânea. Como explicado, diversas técnicas de SCA procuram regiões em que somente uma das fontes esteja ativa, e, desta forma, estimam a direção da coluna da matriz de mistura referente àquela fonte. Porém, se considerarmos que existe mais de um sensor, é possível imaginar que mais de uma direção possa ser estimada de forma simultânea.

Uma forma de caracterizar esparsidade é dizer que existe uma concentração de energia em poucos coeficientes do sinal, sendo que os outros possuem valores nulos ou próximos de zero. É interessante notar que esses coeficientes não-nulos podem se encontrar bem espalhados, ou concentrados em algumas regiões. Como exemplo deste último caso, podemos citar os sinais de áudio. Apesar de esparsos no domínio do tempo,

sua esparsidade é mais acentuada no domínio tempo-frequencial [ABRARD e DEVILLE, 2003]. Quando analisamos um sinal de voz neste domínio, percebemos regiões contínuas de valores não-nulos intercaladas com outras regiões de ausência de sinal. Na Fig. 5.1, é mostrado o espectrograma de uma mistura realizada com duas fontes, após ser realizada uma STFT nos sinais. Note a presença de contraste entre as regiões mais escuras (onde o sinal é mais forte) e as regiões mais claras (que caracterizam ausência de sinal), o que indica a presença de um grau significativo de esparsidade.

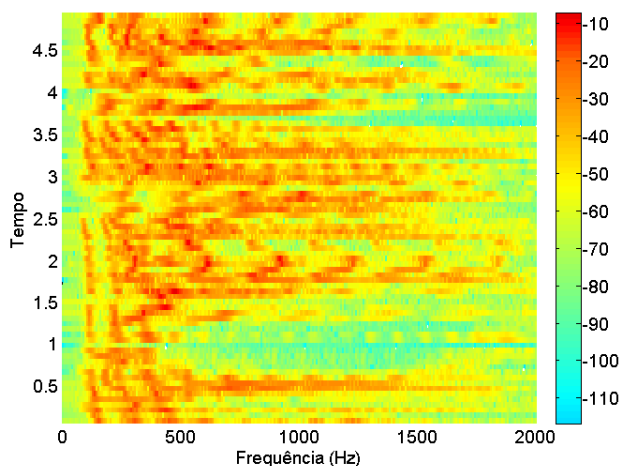


Fig. 5.1 – Espectrograma de uma mistura com duas vozes

Para estes tipos de dados, podemos considerar que o sistema misturador, que é subparametrizado, passa a ser localmente determinado se, em uma janela de dados, o número de fontes que estão ativas não ultrapassar o número de sensores. Sabendo que os sinais das fontes são mutuamente independentes, é possível considerar a aplicação de ICA a cada uma dessas janelas de dados do sinal.

Neste capítulo, mostraremos como usar esta proposta de abordagem híbrida ICA+SCA tanto para identificar a matriz de mistura quanto para estimar o número de fontes.

5.1 A Utilização de ICA em Problemas Subparametrizados

De acordo com as ideias expostas nos Capítulos 2 e 3, se as fontes são esparsas e mutuamente independentes, faz sentido pensarmos que, para um modelo de mistura subparametrizado, podemos aplicar, pelo menos localmente (isto é, em pequenas janelas de

dados dos sinais), alguma técnica de ICA com o intuito de estimar as direções dos vetores formados pelas colunas da matriz de mistura. Neste contexto, esta técnica funcionaria adequadamente se o número de fontes em cada janela de sinais fosse menor que ou igual ao de misturas existentes. Intuitivamente, se na maioria dos casos isto for verdade, será possível estimar a matriz de mistura. Não obstante, é preciso analisar qual será o significado da resposta gerada pelo processo de otimização quando $M \neq N$. Com essa finalidade, veremos qual seria o resultado obtido em quatro cenários:

1 – **O número de fontes ativas é igual ao número de sensores:** este é o caso mais simples, no qual o algoritmo encontrará uma matriz cujas direções corresponderão às colunas da matriz de mistura. Na Fig. 5.2 (a), pode-se ver o gráfico de dispersão das misturas em um caso em que $M = N = 2$. As retas indicam as direções das colunas da matriz estimada.

2 – **O número de fontes ativas é menor que o número de sensores:** quando isto ocorre, os algoritmos baseados em ICA tendem a encontrar as direções associadas a todas as fontes ativas, sendo que os demais graus de liberdade serão influenciados por outros fatores, como ruído aditivo ou erros de aproximação de cálculo, por exemplo. É de se esperar, entretanto, que as direções não associadas às fontes ativas tendam a variar consideravelmente de uma região tempo-frequencial para outra. A Fig. 5.2 (b) ilustra esta situação, onde existem duas misturas e apenas uma fonte ativa. Como esperado, uma das direções encontradas indica exatamente a direção da fonte existente.

3 – **O número de fontes ativas é maior do que o número de sensores:** isto gera um problema localmente indeterminado. Desta forma, é possível que algumas direções sejam encontradas com razoável precisão, enquanto outras não serão encontradas, no espírito da discussão sobre limitações estruturais apresentada no Capítulo 4. A saída do sistema de separação provavelmente não será composta por sinais independentes, e, em geral, nenhuma condição de decorrelação perfeita será alcançada. Aliás, é interessante notar que, como as estimativas dos sinais separados serão necessariamente correlacionadas, é possível avaliar estes dados e classificar suas estimações como inválidas. Na Fig. 5.2 (c), temos um exemplo com quatro fontes e dois sensores, no qual nenhuma das direções da matriz de mistura foi estimada com precisão.

4 – **Não existem fontes ativas no intervalo escolhido:** neste caso, teremos, novamente, graus de liberdade que não poderão ser utilizados, o que significa que a aplicação de um método de ICA não trará informação relevante. Analogamente ao que foi visto para o segundo cenário, as direções obtidas nestes blocos de dados tendem a variar consideravelmente. A Fig. 5.2 (d) ilustra uma situação desta natureza, com dois sensores e nenhuma fonte ativa.

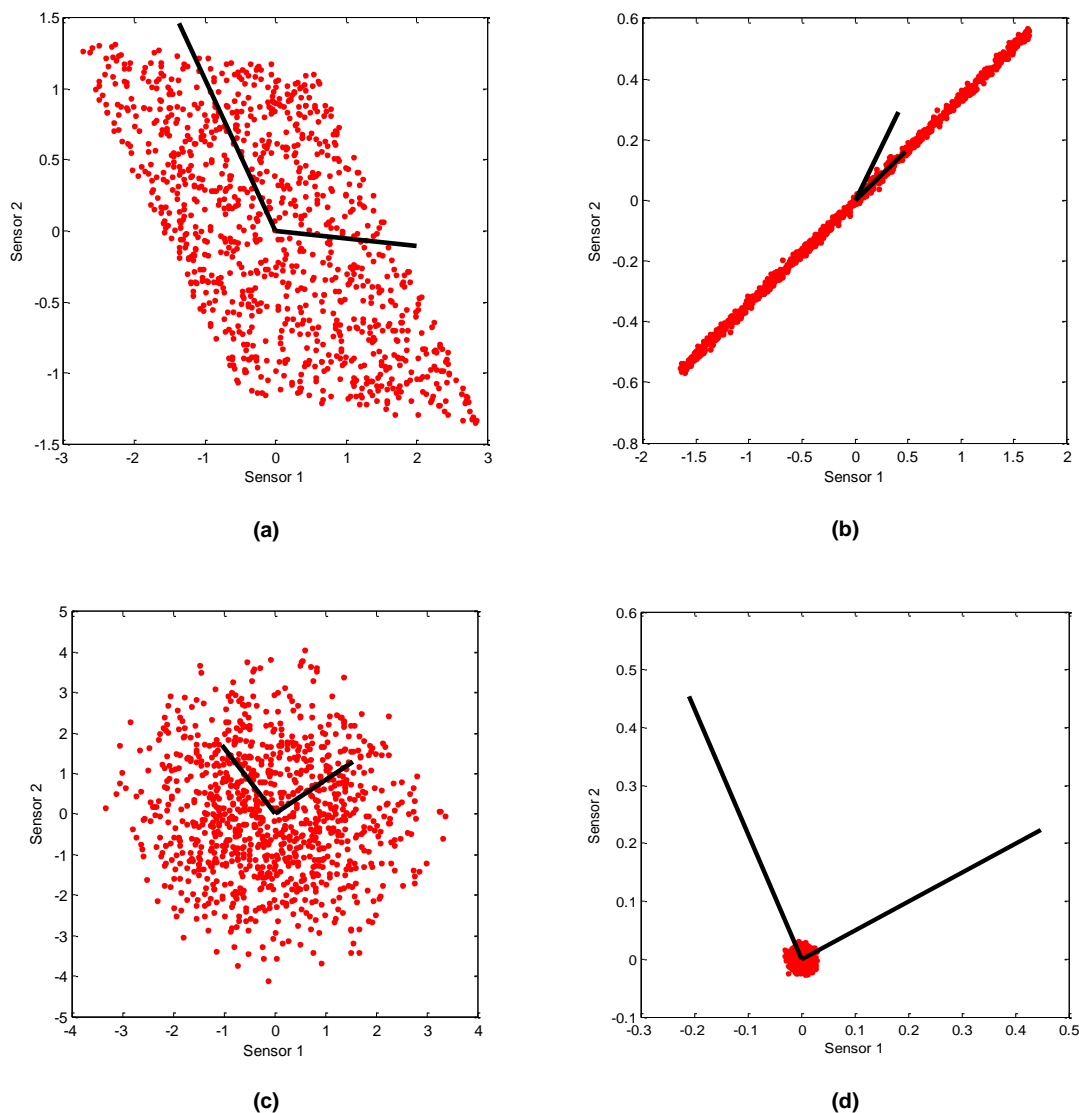


Fig. 5.2 – Gráficos de dispersão *Sensor 1* x *Sensor 2*, com as indicações das direções encontradas pelo algoritmo de ICA para os seguintes casos: (a) Duas fontes e dois sensores; (b) Uma fonte e dois sensores; (c) Quatro fontes e dois sensores; e (d) Nenhuma fonte e dois sensores.

Desta forma, depois de aplicar ICA a diversas janelas de dados, e descartando as soluções que geram sinais altamente correlacionados (como descrito no terceiro cenário), é lícito crer que os valores das estimativas estejam mais concentrados em torno das direções das colunas da matriz de mistura, visto que as estimativas que não estão associadas a nenhuma fonte tendem a ficar igualmente distribuídas entre todas as direções possíveis.

5.2 Identificação do Sistema Misturador

O problema de identificação do sistema de mistura em um caso em que este é subparametrizado é dividido em duas partes: descobrir o número de fontes existentes e estimar a matriz de mistura (assumindo que o sistema misturador é linear e instantâneo). Visto que foi mostrado na Seção 3.2.2 um método para identificar o sistema misturador a partir de um espectrograma utilizando técnicas baseadas em SCA, queremos ver agora o quanto técnicas baseadas em ICA podem ser úteis na estimação do sistema. Para isto, iremos propor uma abordagem híbrida, que utiliza tanto o conceito de esparsidade quanto o de independência.

Se, para uma janela de dados da mistura, aplicarmos ICA, obteremos uma matriz de separação \mathbf{W} de tamanho $M \times M$. Invertendo esta matriz, obteremos M conjuntos de vetores-coluna, cujas direções corresponderão a um dos quatro cenários descritos na Seção 5.1. Caso o sistema possua somente dois sensores, estas direções podem ser escritas na forma polar, através de um ângulo θ para cada vetor. Depois de realizada a ICA em todas as janelas de dados, teremos uma massa de valores de θ .

Em seguida, o método opera definindo primeiramente o número de clusters mais adequado para representar estes dados, o que provê informação sobre o número de fontes existentes na mistura. A partir daí, é utilizada uma estratégia para definir qual é a posição mais representativa de cada cluster, a qual será justamente o valor de θ mais representativo de cada cluster, que é, conseqüentemente, um forte candidato a representar a direção de uma das colunas da matriz de mistura.

Um exemplo ilustrativo dos valores de θ encontrados pelo algoritmo pode ser visto na Fig. 5.3, que mostra o histograma feito com todos os valores obtidos em um caso com duas misturas e três fontes: os traços verticais indicam as direções verdadeiras. É

importante notar que, conforme esperado, as direções encontradas pelo algoritmo proposto se concentram em torno das verdadeiras direções da matriz de mistura, e o número de picos indica o número de fontes.

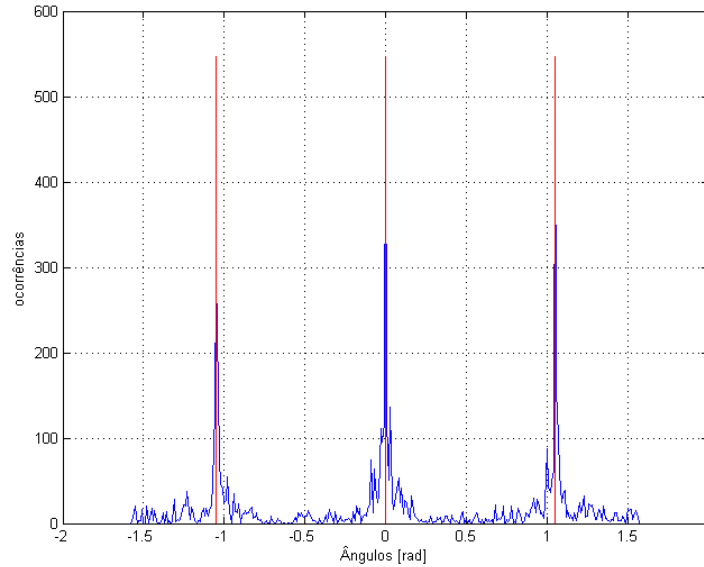


Fig. 5.3 – Histograma das direções encontradas numa mistura com 3 fontes de voz

Um método tradicional e bem estabelecido para realizar a tarefa de clusterização requerida é o algoritmo *k-means* [MAULIK e BANDYOPADHYAY, 2002; CHINRUNGRUENG e SEQUIN, 1995], que pode ser entendido como um método de busca estocástica associado à busca por centros de clusters que minimizem a seguinte função custo:

$$J_K^{(\alpha)} = \sum_{j=1}^l \sum_{k=1}^K u_{kj} \|\theta_j - C_k\|^\alpha \quad (5.1)$$

onde C_k representa o centro do k -ésimo cluster e K é o número total de clusters; $u_{kj} = 1$ se $\theta_j \in C_k$, em caso contrário, $u_{kj} = 0$, e α vale tipicamente 2. Em nosso caso, no entanto, verificamos que $\alpha = 1$ trouxe melhores resultados para localizar os centros dos clusters. É provável que isto se deva ao fato de não utilizarmos um sistema circular de parametrização do valor de θ , ou seja, em nossa descrição das colunas da matriz de mistura, $-\pi/2$ e $\pi/2$ representam a mesma direção, e isto não é levado em consideração pelo método de clusterização.

Uma forma de se empregar o *k-means* de forma adaptativa [CHINRUNGRUENG e SEQUIN, 1995] pode ser descrita resumidamente do seguinte modo: primeiramente são gerados os valores iniciais de todos os centros dos clusters, e, em seguida, para todo valor de θ_j , é procurado o C_k mais próximo, que será adaptado da seguinte forma:

$$C_k = C_k + \mu(\theta_j - C_k), \quad (5.2)$$

onde μ , sendo $\mu < 1$, é o passo de adaptação. Esta segunda etapa é repetida diversas vezes utilizando a mesma massa de dados, até que um critério de parada seja atingido.

5.2.1 Estimação do Número de Clusters

O método tradicional de aplicação do *k-means* assume que o valor de K é dado *a priori*. Porém, se o número de clusters não é conhecido, é possível tentar utilizar J_K para estimar este valor. Esta abordagem considera o fato de que o mínimo valor de J_K (que idealmente corresponde à solução dada pelo algoritmo *k-means*) irá, geralmente, apresentar pequena variação quando o número proposto de clusters K for maior do que o necessário, isto é, maior do que o número correto de clusters. A Fig. 5.4 mostra a variação de J_K quando o número de clusters vai de 1 até 15, para cenários com 3 e 6 fontes, respectivamente, ambos com duas misturas e fontes esparsas.

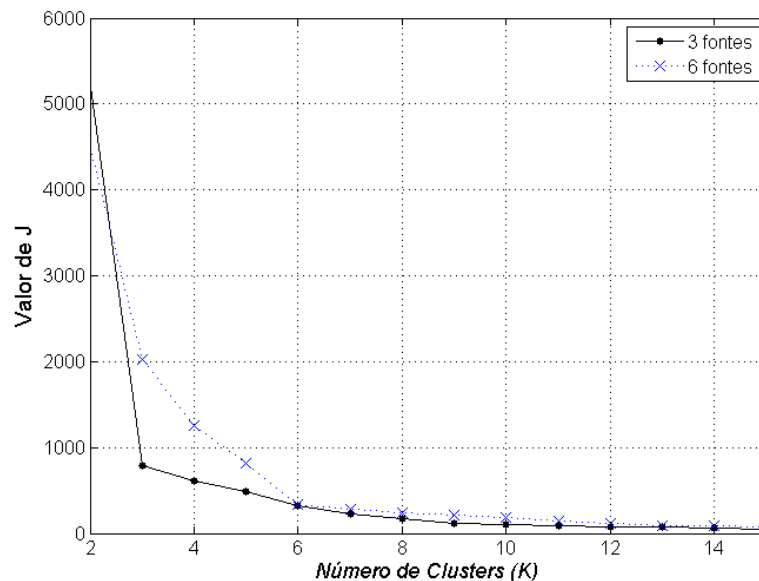


Fig. 5.4 – Valor de J_K quando o número de fontes é igual a 3 e a 6, respectivamente

O problema passa agora a ser como conseguir encontrar o último acréscimo de cluster para o qual a função custo J_K ainda seja reduzida de forma significativa. Para

encontrar este valor, proporemos três índices, mostrados abaixo. Posteriormente, buscaremos analisar qual deles é o mais adequado ao problema.

$$I_{Normal}(K) = \frac{J_{K+1} - J_K}{J_K - J_{K-1}} \quad (5.3)$$

$$I_{Log}(K) = \frac{\log(J_{K+1}) - \log(J_K)}{\log(J_K) - \log(J_{K-1})} \quad (5.4)$$

$$I_{Sum}(K) = I_{Normal}(K) + I_{Log}(K), \quad (5.5)$$

e o número estimado de fontes é dado pelo índice i para o qual o valor em que cada I_u ($u \in \{Normal, Log, Sum\}$) é máximo, ou seja,

$$i_{max} = \arg \max_{i \in \{1, \dots, K_{max}\}} \{I_u(i)\}. \quad (5.6)$$

5.3 Método Proposto

Para derivar o método proposto, foi obtida uma representação esparsa dos sinais a partir da aplicação de uma STFT às misturas. Esta opção não é a única possibilidade, mas é uma solução eficiente e largamente utilizada [RIOUL e VETTERLI, 1991; ABRARD e DEVILLE, 2003]. Abordaremos aqui uma situação com apenas duas misturas, embora os resultados possam ser estendidos para situações com mais sensores.

O método é baseado naquele apresentado na Seção 3.2.2, e pode ser descrito de acordo com o seguinte algoritmo:

Passo 1 - Obter a representação tempo-frequencial de cada sinal observado $x_i(t, f)$ usando a STFT[$x_i(t)$] para $i=1,2$, escolhendo adequadamente o tamanho da janela de amostras e a razão de sobreposição entre elas.

Passo 2 – Utilizando uma técnica de ICA, estimar as matrizes de mistura para a vizinhança⁷ $\Omega_{t,f}$ de cada ponto tempo-frequencial (t, f) . O algoritmo de ICA irá prover uma “matriz de mistura local” de tamanho $M \times M$. A partir desta matriz, é possível estimar as direções associadas a todas as fontes usando a seguinte equação:

⁷ É importante notar que a quantidade de amostras existentes nas vizinhanças $\Omega_{t,f}$ pode afetar o resultado. Se o número for baixo, o algoritmo baseado em ICA terá dificuldades em convergir para os valores exatos. Por outro lado, caso o número de amostras seja elevado, a probabilidade de existirem várias fontes sobrepostas é maior.

$$\theta_i(t, f) = \tan^{-1} \frac{[\mathbf{W}]_{2,i}}{[\mathbf{W}]_{1,i}}, \quad i = 1, 2 \quad (5.7)$$

A validade da estimação obtida depende, como discutido, da existência ou ausência de mais de duas fontes ativas. Este fato justifica a inclusão do passo 2.1.

Passo 2.1 – Com o intuito de evitar estimativas não confiáveis, soluções que gerem saídas com coeficiente de correlação maior do que certo limiar (tipicamente 0,08 em nossos ensaios) são descartadas. Esta política pretende rejeitar as direções obtidas a partir de cenários não modeláveis, como explicado na Seção 5.1.

Passo 3 – Depois que todas as direções tiverem sido obtidas, a estimativa final das direções da mistura decorre da aplicação de um algoritmo de clusterização à massa de dados gerada pelo conjunto de todos os valores de $\theta(t, f)$ estimados que não foram descartados no passo 2.1. Para K variando⁸ entre 1 e $(K_{\max} + 1)$, onde K_{\max} é dado pelo máximo número de fontes que se acredita existir na mistura, aplicar o algoritmo do k -means em θ , guardando todos os valores de $J_K^{(2)}$ encontrados.

Passo 4 – Para se obter o número correto de clusters, e, conseqüentemente, o número de fontes, calcular $I(K)$, para $K=2, \dots, K_{\max}$, a partir de um dos três índices sugeridos (Eqs. (5.3), (5.4) e (5.5)). O número de fontes estimado será dado pela Eq. (5.6).

Passo 5 – Para se estimar as direções da matriz de mistura, aplicar novamente o algoritmo k -means, utilizando agora $K = i_{\max}$ (dado pelo Passo 4) e $\alpha = 1$. As direções serão dadas pelos valores de C_k encontrados no final da otimização.

Como dito, o método proposto possui certo grau de similaridade com as abordagens propostas em [ABRARD e DEVILLE, 2003], estando a principal diferença exatamente na utilização de ICA para auxiliar na estimação da matriz de mistura. Esta diferença pode ter um efeito relevante, como veremos a seguir com a ajuda de uma análise baseada num conjunto de simulações de cenários representativos.

⁸ Apesar de K_{\max} ser o máximo número de fontes que se acredita existir, é preciso calcular também o valor de J_K para $(K_{\max} + 1)$, pois este valor é necessário para se calcular o valor de $I_u(K_{\max})$.

5.4 Resultados de Simulação

Para averiguar o desempenho da metodologia proposta, foi aplicada a STFT às misturas, sendo considerada uma janela de 512 amostras, com meia janela de sobreposição. O algoritmo de ICA utilizado foi o proposto por Comon [COMON e MOREAU, 1997], o qual se baseia na maximização de uma medida de curtose, exposta na Seção 2.1.2. A principal razão para a utilização deste método é o fato de ele atingir uma situação de convergência com um número reduzido de amostras. Este fator é importante, visto que a janela de dados não pode ser muito grande, já que isto aumenta a chance de existirem mais fontes ativas que sensores no intervalo escolhido.

Como os dados de entrada do algoritmo possuem valores complexos, optou-se por tratar suas partes reais e imaginárias como elementos separados de um vetor, i.e., o vetor de dados será ordenado da seguinte forma: $[real(.) \text{ imag}(.)]$. Isto faz com que o algoritmo de ICA seja favorecido, já que o número de amostras utilizadas dobra, enquanto o tamanho da vizinhança $\Omega_{t,f}$ permanece o mesmo.

Em todos os casos, o método proposto foi comparado com os algoritmos DEMIX (que estima tanto o número de fontes quanto a matriz de mistura) e DEMIXN (que considera conhecido o número de fontes) [ARBERET *et al.*, 2006], sendo o último, em certo sentido, semelhante sob alguns aspectos metodológicos ao método proposto, particularmente no uso de um processo de clusterização. Devido ao fato de os algoritmos (DEMIX e DEMIXN) terem mostrado melhores resultados, quando comparados pelos autores com outras abordagens existentes, este trabalho se restringe a fazer uma comparação entre nossa proposta e esses métodos.

Com o intuito de facilitar a exposição dos resultados, primeiramente serão apresentados os cenários propostos, e, em seguida, os resultados obtidos.

5.4.1 Descrição dos Cenários

O cenário experimental utilizado foi similar ao descrito em [ARBERET *et al.*, 2006]. As fontes foram escolhidas a partir de um conjunto de 200 gravações de 5 segundos de vozes polonesas, todas com taxa de amostragem de 4 kHz, estando a base disponível em [IEEE MLSP, 2005].

Primeiro Cenário

O primeiro cenário foi construído de acordo com as condições descritas em [ARBERET *et al.*, 2006]: as misturas foram criadas a partir de um número de fontes variando entre 2 e 15, sendo as direções da mistura sempre uniformemente distribuídas. Para um exemplo ilustrativo de um mistura com duas fontes, podemos voltar à Fig. 5.1.

Segundo Cenário

Teoricamente, a vantagem esperada do algoritmo proposto ocorre nas situações em que o número de fontes ativas em cada instante deixa de ser unitário, e passa a ser equivalente ao número de misturas. Desta forma, é sugerido um segundo cenário, no qual é incluído como uma das fontes um ruído branco uniforme. Esta inclusão pode ser justificada como teste de robustez, visto que este tipo de fonte é não esparsa no domínio tempo-frequencial. Particularmente, nesta situação em que temos mais fontes do que sensores e uma das fontes é não esparsa, não é possível aplicar de forma separada, pelo menos teoricamente, nem as técnicas clássicas de ICA, visto que o sistema é subparametrizado, nem as abordagens típicas de SCA, já que existe um sinal não esparsa na mistura. A Fig. 5.5 mostra o espectrograma de uma mistura que contém uma fonte de voz e um sinal de distribuição branca e uniforme. É importante ressaltar que os métodos propostos em [ARBERET *et al.*, 2006] são fortemente baseados na suposição de que a maioria das regiões tem no máximo uma fonte ativa, o que não é válido nesta situação, visto que o ruído branco sempre está ativo.

Para ilustrar as dificuldades que este tipo de fonte traz, é mostrado o histograma obtido a partir de uma mistura com duas fontes de voz e uma correspondente a um ruído branco (Fig. 5.6). Comparando estes resultados com o da Fig. 5.3, pode-se notar que duas das concentrações de dados do histograma decrescem significativamente, aumentando a dificuldade associada ao estágio de clusterização do método.

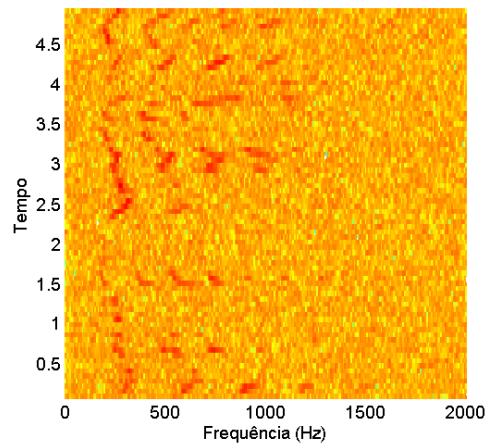


Fig. 5.5 – Espectrograma de uma mistura com um ruído branco uniforme e um sinal de voz

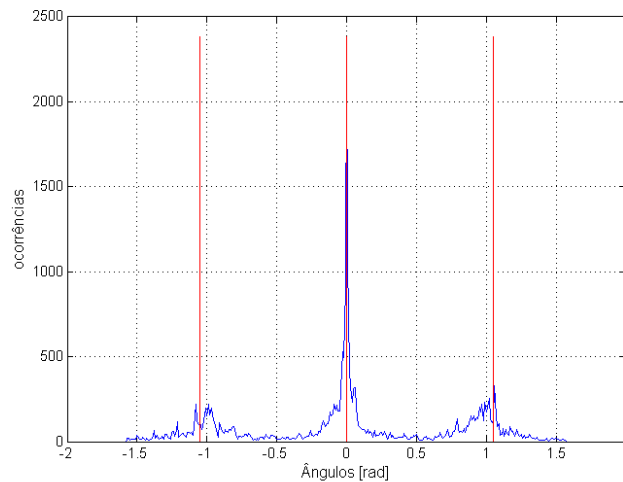


Fig. 5.6 – Histograma das direções encontradas numa mistura com duas fontes de voz e um ruído branco

Terceiro Cenário

O terceiro cenário é, em certo sentido, uma variação do segundo. Porém, em vez de um ruído branco, que não é um modelo particularmente representativo de um sinal de áudio, a fonte adicional foi definida de modo mais próximo a uma situação real. Nesse espírito, foi escolhido um sinal que tem amplo espectro em frequência e possui uma duração relativamente longa: o som de uma multidão gritando e batendo palmas. É importante notar que o nosso sistema misturador é instantâneo, e, deste modo, todos os diferentes sons desta multidão acabam sendo resumidos automaticamente a uma única fonte. A Fig. 5.7 mostra o espectrograma de uma mistura entre um sinal de voz e uma

multidão batendo palmas e gritando, e a Fig. 5.8 mostra o histograma de uma mistura contendo esta fonte misturada a duas fontes de voz.

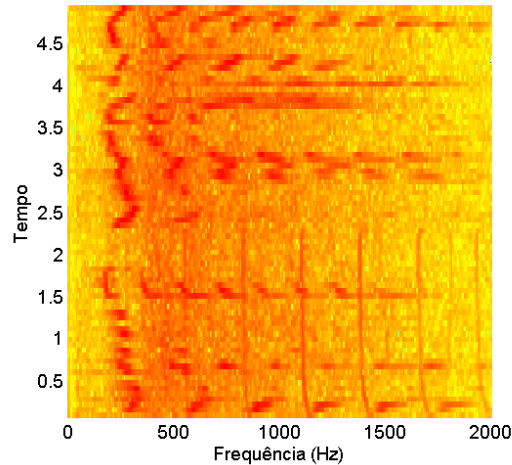


Fig. 5.7 – Espectrograma de uma mistura com um sinal de voz e uma multidão batendo palmas e gritando

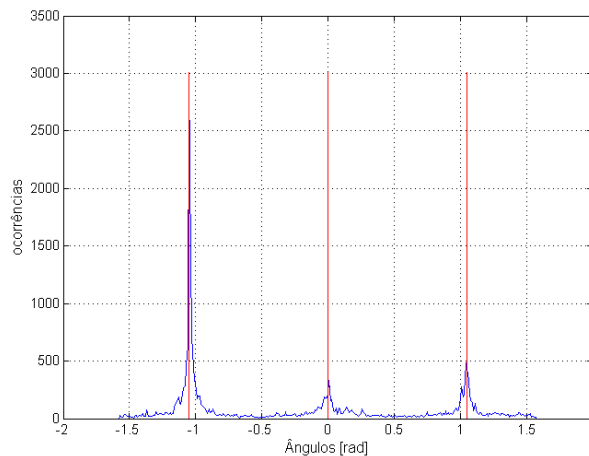


Fig. 5.8 – Histograma das direções encontradas numa mistura com duas fontes de voz e uma multidão batendo palmas e gritando

5.4.2 Resultados da Estimação do Número de Fontes

Primeiramente foi rodado novamente o experimento do primeiro cenário explicado na Seção 5.4.1, sendo que, para cada número de fontes, foram feitas 50 realizações. Para cada situação, aplicamos os três índices propostos em 5.2.1 com o intuito de estabelecer elementos de comparação entre os mesmos. A porcentagem de estimações corretas do número de fontes é mostrada na Fig. 5.9.

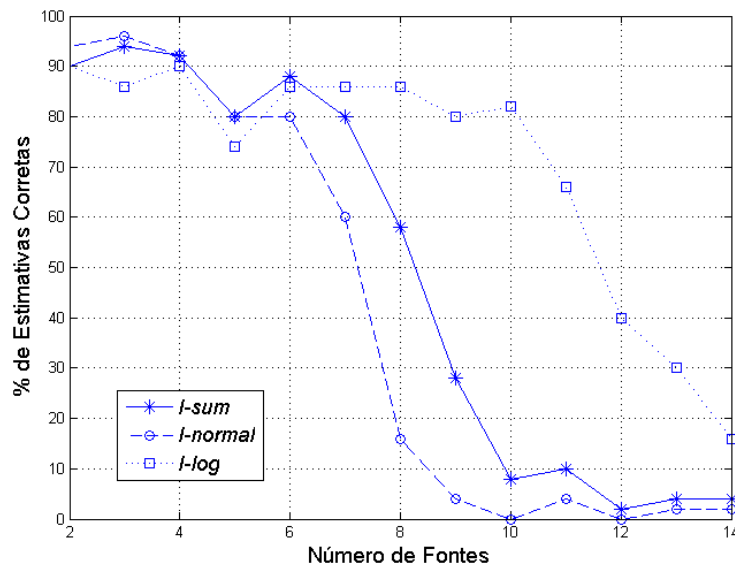


Fig. 5.9 – % de estimações corretas para os três diferentes índices

Pode-se perceber que os melhores resultados foram encontrados utilizando o índice I_{log} , mas, se considerarmos apenas os cenários com número de fontes menor do que seis, percebe-se um resultado melhor com a utilização do índice I_{Normal} . O índice I_{sum} não se mostrou tão bom compromisso relativamente aos dois outros índices como esperado. Outro resultado importante a ser considerado é o fato de que o índice I_{log} encontra o valor correto da estimativa do número de fontes em mais de 80% dos casos apresentados até 11 fontes, exceto pelo caso de 5 fontes (75%). Nesta mesma situação, o método DEMIX proposto por [ARBERET *et al.*, 2006] consegue estimar corretamente com 80% de acertos somente até 8 fontes, falhando sempre quando $N > 10$.

Como mencionado anteriormente, uma das vantagens da utilização de um método baseado em ICA é a possibilidade de lidar com cenários em que a suposição de esparsidade não é necessariamente válida para todas as fontes. Com o intuito de corroborar esta proposição, foram feitos testes equivalentes aos descritos nos segundo e terceiro cenários da Seção 5.4, com uma das fontes da mistura sendo substituída por um sinal branco com distribuição uniforme, e uma gravação de uma multidão gritando e batendo palmas, respectivamente.

Os mesmos testes foram aplicados aos dois cenários, agora usando somente o índice I_{log} . Os resultados são mostrados na Fig. 5.10. Até seis fontes, os resultados são similares

aos obtidos quando temos somente fontes esparsas. Depois disto, os erros começam a crescer por conta de as direções associadas às fontes não esparsas não serem tão evidentes, formando baixas concentrações de dados no histograma.

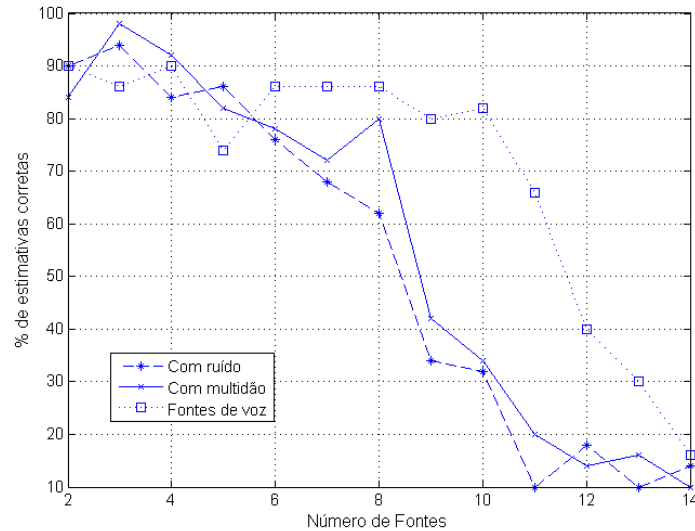


Fig. 5.10 – % de estimativas corretas para os três diferentes cenários, utilizando o índice I_{Log}

5.4.3 Resultados da Estimação da Matriz de Mistura

Primeiro cenário

Para cada número de fontes N , os algoritmos foram rodados 20 vezes e foi calculado o valor do erro angular médio (EAM), definido como a média das diferenças entre os ângulos estimados e os das direções da matriz de mistura original, para cada um deles. Na Fig. 5.11, encontramos a evolução do EAM com respeito ao número de fontes (para uma média de 20 diferentes realizações, cada uma utilizando um número diferente de vozes, que foram escolhidas aleatoriamente a partir do conjunto total). A linha contínua representa os resultados encontrados pelo algoritmo proposto, enquanto a linha tracejada é referente aos resultados do DEMIXN. Os dois métodos podem ser considerados como tendo desempenhos aproximadamente equivalentes, com ligeira vantagem para o DEMIXN.

Vale, no entanto, ressaltar que o DEMIXN não foi capaz de encontrar o conjunto completo das direções de mistura em 4% dos casos, o que foi causado pelo fato de o algoritmo ter criado dois clusters relacionados à mesma direção. Estas situações não foram levadas em consideração no processo do cálculo do EAM mostrado na Fig. 5.11. O método

proposto baseado em ICA, por outro lado, não produziu nenhuma estimativa errônea nos testes realizados. Como explicado em [ARBERET *et al.*, 2006], o DEMIXN converge até 13 fontes, mas, para números maiores, ele começa a divergir, pois não encontra todos os clusters. Por outro lado, o algoritmo proposto não teve problemas nos testes realizados até 15 fontes. Não foram realizados testes para um número maior de fontes devido ao fato de que este tipo de teste não traria informações significativas adicionais, visto que em casos reais as direções raramente estão igualmente distribuídas.

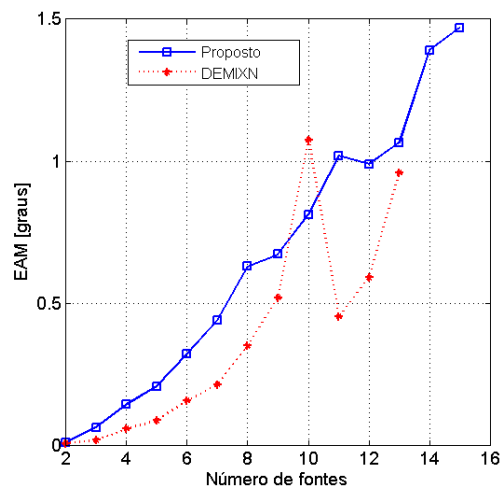


Fig. 5.11 – Valor médio de EAM dos algoritmos para o primeiro cenário (todas as fontes são sinais de voz)

Segundo Cenário

Nesta situação, o método proposto não experimenta uma perda significativa de desempenho em nenhum dos testes. Por outro lado, o DEMIXN não encontra todas as direções em 40% dos casos, e, mais ainda, deixa de convergir com 10 ou mais fontes. A Fig. 5.12 mostra que, até 9 fontes e nos casos em que o DEMIXN encontra valores aceitáveis, os dois algoritmos têm desempenho próximos, sendo que o proposto possui desempenho ligeiramente superior na maioria dos casos. O aumento dos valores de EAM em relação aos encontrados no primeiro cenário já era esperado, e pode ser explicado pelo fato de que a direção exata relativa à fonte com ruído branco é mais difícil de ser estimada. Se descartarmos esta fonte, os valores de EAM caem a valores comparáveis aos do primeiro cenário.

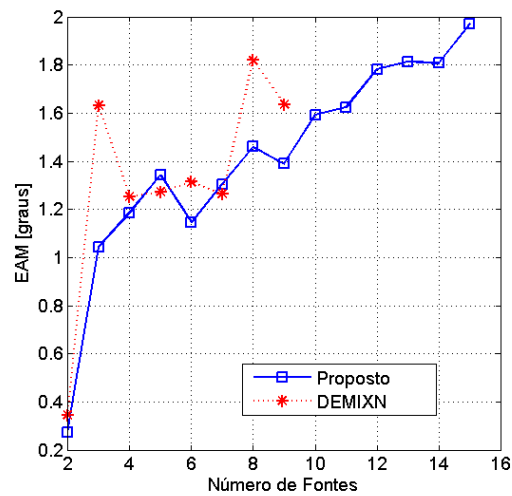


Fig. 5.12 – Valor médio de EAM dos algoritmos para o segundo cenário (com uma das fontes sendo ruído branco)

Terceiro Cenário

Como esperado, os resultados encontrados foram intermediários em relação aos dois primeiros cenários. Nesta situação, o DEMIXN falhou em encontrar as direções da matriz de mistura em mais de 10% das tentativas. Como mostra a Fig. 5.13, ambos os resultados são levemente melhores do que os obtidos no segundo cenário. Mesmo existindo regiões com apenas uma fonte ativa, o método proposto leva vantagem em relação ao DEMIXN por conseguir estimar corretamente as direções para casos cujo número de fontes é maior do que 10.

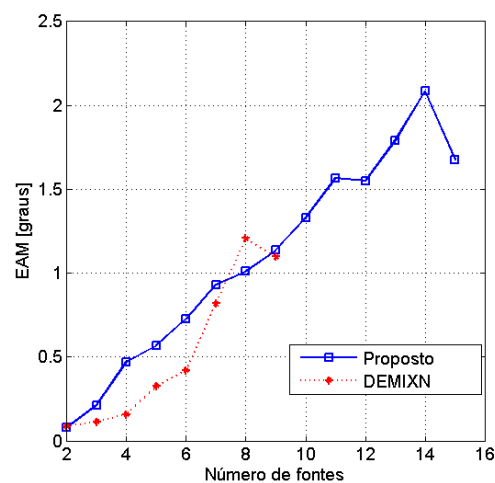


Fig. 5.13 – Valor médio de EAM dos algoritmos para o terceiro cenário (com uma das fontes sendo a de multidão)

5.5 Considerações Finais

Neste capítulo, foi proposta uma metodologia híbrida, que emprega conjuntamente os conceitos de independência e esparsidade para lidar com problemas subparametrizados de BSS. O método derivado apresentou bons resultados em cenários estereofônicos, ou seja, com duas misturas, tanto na estimação do sistema misturador quanto na determinação do número de fontes existentes. Uma vantagem importante do método proposto relativamente a abordagens que utilizam apenas esparsidade é a sua capacidade de lidar com a presença de fontes não esparsas na mistura, o que foi atestado pelos resultados obtidos.

Capítulo 6

Estimação da Matriz de Mistura

Através de Otimização Multimodal

Bioinspirada

Na Seção 3.2.2, foram explicados dois métodos de estimação do sistema misturador, sendo que um deles é baseado na busca por subespaços de concentração [NAINI *et al.*, 2008]. No que diz respeito a esta técnica, dois pontos interessantes podem ser levantados. O primeiro é que a formulação apresentada pode também ser utilizada na procura de subespaços unidimensionais, similarmente ao que é feito em outras abordagens. Se este for o caso, os subespaços de concentração já permitem uma estimação direta dos vetores de mistura, e, desta forma, uma etapa subsequente não é estritamente necessária. O segundo aspecto é que, ao contrário de diversos problemas de otimização em que queremos encontrar o ótimo global, este problema é inerentemente multimodal, ou seja, todos os máximos são igualmente importantes. Isto exige a utilização de técnicas de otimização capazes tanto de explorar adequadamente o espaço de busca quanto de realizar refinamentos locais. Neste sentido, meta-heurísticas bioinspiradas se tornam naturalmente opções bem interessantes, como, por exemplo, as ferramentas baseadas em sistemas imunológicos artificiais (*Artificial Immune Systems – AIS*) [DE CASTRO, 2006].

6.1 Breve Descrição de Sistemas Imunológicos Artificiais

O campo de pesquisa de AIS tem recebido uma atenção crescente nos últimos anos, o que se explica, ao menos em parte, pela sua aplicabilidade a uma ampla gama de tarefas: reconhecimento de padrões, navegação autônoma, controle, entre outros [DE CASTRO, 2006]. Merece destaque, neste trabalho, o uso de AIS's em problemas de otimização, mais especificamente, de otimização multimodal. Isto se justifica pelas seguintes características desejáveis:

- Há AIS's que permitem um controle dinâmico do número de indivíduos (células) em uma determinada população;
- AIS's possuem uma capacidade inerente de manter a diversidade populacional;
- AIS's tendem a preservar as soluções encontradas correspondentes a ótimos locais.

Uma proposta que reúne estas características, e, além disto, vem sendo aplicada com sucesso no contexto de processamento de sinais [ATTUX, 2005; ATTUX *et al.*, 2003], é a rede imunológica conhecida como opt-aiNet [DE CASTRO e TIMMIS, 2002], a qual é potencialmente capaz de encontrar simultaneamente os máximos de uma função custo como a dada pela Eq. (3.13). A seguir, será descrito o funcionamento da opt-aiNet, e, na próxima seção, será proposta adaptação desta ferramenta de otimização para a sua aplicação ao problema de BSS no contexto de SCA.

6.1.1 Funcionamento da Opt-aiNet

Um sistema imunológico [DE CASTRO, 2006] pode ser descrito como sendo aquele que possui capacidade de produzir anticorpos com objetivo de se combinarem com um grande número de microrganismos invasores, denominados antígenos, de forma a imunizar o organismo a ser protegido. Esta capacidade está relacionada à adaptação o sistema, que pode ser compreendida com o auxílio de dois fatores: *i*- as noções de *seleção clonal* [BURNET, 1959] e *maturação de afinidade* [ADA e NOSSAL, 1987]; *ii*- a ideia de rede imunológica [JERNE, 1974].

De maneira simples, a seleção clonal pode ser explicada em termos de um processo de clonagem de células que obtiveram um certo grau de reconhecimento positivo com o antígeno. Este reconhecimento positivo está ligado à afinidade estrutural que os receptores

presentes em cada célula possuem com as características superficiais do antígeno. Nesse contexto, a adaptação do sistema é feita através de mutações nos receptores das células durante o processo de clonagem. Como este processo ocorre com intensidade inversamente proporcional ao grau de afinidade entre a célula e o antígeno, é entendido que este é um processo de *maturação de afinidade* entre os dois. Por fim, existem ainda mecanismos que podem ser modelados por meio da introdução de novos *indivíduos* na *população* de células de defesa. Por outro lado, a noção de rede estabelecida por [JERNE, 1974] fornece uma visão distinta, mais voltada às interações entre as próprias células de defesa, as quais dão origem a uma gama de comportamentos próprios (*eigenbehaviors*).

Desta forma, tendo em vista a combinação entre as ideias de seleção clonal / maturação de afinidade e a noção de rede imunológica, podemos dizer que a adaptação existente em um sistema imunológico pode ser vista como um “processo de otimização das células”. Fazendo este paralelo, podemos dizer que cada possível solução do problema a tratar corresponde a um certo indivíduo da população de células. A função custo a ser otimizada é representada pela medida de afinidade (ou *fitness*) entre as células e os antígenos.

Assim, o objetivo da opt-aiNet é maximizar uma determinada função custo (também chamada de função de *fitness*) com respeito a um conjunto de parâmetros que, de certa forma, representam a estrutura de um anticorpo que deve identificar um dado antígeno. A opt-aiNet pode ser implementada a partir da seguinte sequência de passos:

Passo 1 – *Inicialização com valores aleatórios de uma população P de células contendo $N_{inicial}$ vetores de solução.*

Passo 2 – *Cada uma das células existentes na população P gera N_c clones (cópias exatas), sendo que, para todos menos um desses clones, será aplicada uma mutação, que segue uma distribuição gaussiana, conforme as seguintes equações:*

$$P_i' = P_i + \alpha N(0,1) ,$$

$$\alpha = \frac{1}{\beta} e^{-Fit^*} .$$
(6.1)

Em (6.1), $N(0,1)$ é uma variável aleatória de distribuição gaussiana, com média zero e desvio padrão unitário, β é o parâmetro que controla o decaimento da função exponencial

inversa e Fit^ é o valor do fitness (função custo) da célula original normalizado para estar no intervalo $[0,1]$. É importante observar que o grau de maturação α é controlado pelo fitness de cada indivíduo: indivíduos menos adaptados experimentam uma mutação mais acentuada, enquanto indivíduos mais adaptados sofrem mutação em menor grau.*

Passo 3 – *Todo o repertório de células é avaliado, e, para cada grupo de clones, a célula com maior valor de fitness é selecionada. Como critério de parada, pode-se calcular o valor médio do fitness de todas as células: caso este valor seja relativamente maior⁹ do que o anterior, voltar para o **Passo 2**; caso contrário ir para o **Passo 4**.*

Passo 4 – *É determinada a similaridade entre pares de células (normalmente por meio de uma distância euclidiana): se esta for menor que o limiar de supressão, dado por σ_s , a célula com menor fitness do par em questão é suprimida. Finalmente, a cada quantidade determinada de iterações, um número de células (normalmente definido a partir de uma porcentagem $d\%$ do tamanho da população antes da supressão) é aleatoriamente gerado e introduzido na população. Caso um critério de parada pré-definido não seja atingido, voltar para o **Passo 2**.*

É importante notar que o limiar de supressão σ_s é um parâmetro crítico da opt-aiNet, visto que a sua escolha idealmente requer um certo grau de conhecimento *a priori* da função a ser otimizada.

6.2 Método proposto

Como explicado na Seção 3.2.2, é proposto em [NAINI *et al.*, 2008] um método de estimação da matriz de mistura **A** baseado na procura por intersecções de subespaços gerados pelos sinais das fontes. Como mostra a Fig. 3.3 (reproduzida abaixo), a função custo proposta pelos autores gera uma superfície nitidamente multimodal, sendo necessário encontrar todos os máximos, ou sua maioria, para se conseguir estimar a matriz de mistura. Além disto, o número de máximos é um indicativo de quantas fontes existem no sistema.

⁹ Na prática, o limiar subjacente pode ser difícil escolha. Por isto, em diversas situações, o critério de parada pode ser trocado para uma quantidade máxima de iterações.

Deste modo, como delineado, é proposta a utilização de um método de otimização global baseado em sistemas imunológicos à abordagem implementada por [NAINI *et al.*, 2008] tanto para se obter a matriz de mistura quanto para se estimar o número de fontes, em um sistema de mistura subparametrizado, sabendo-se que as fontes são esparsas.

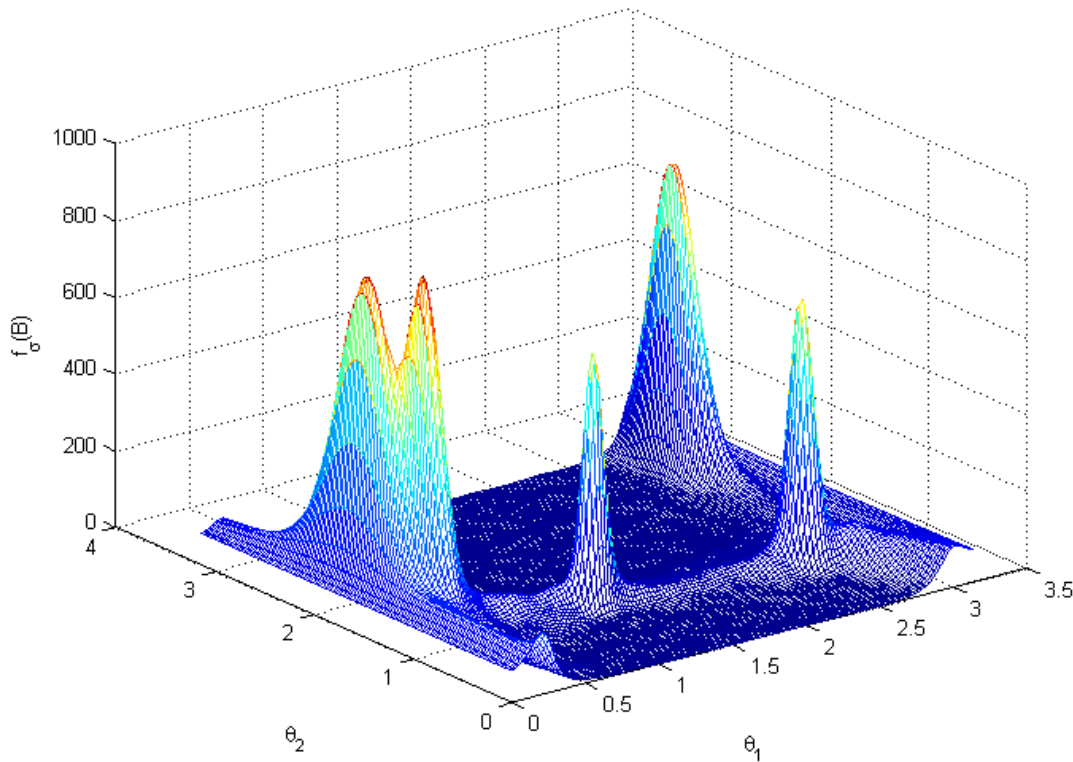


Fig. 3.3 – Superfície da função custo dada por (3.13) para um caso com $N = 5$ e $M = 3$.

6.2.1 Opt-aiNet Aplicada a SCA

Essencialmente, a opt-aiNet é aplicada para realizar uma busca multimodal na função custo $f_\sigma(\mathbf{B})$, proposta por [NAINI *et al.*, 2008] e descrita em (3.13) (replicada abaixo), e, conseqüentemente, estimar a matriz de mistura \mathbf{A} .

$$f_\sigma(\mathbf{B}) = \sum_n \exp\left(\frac{-d^2(\mathbf{x}(n), \mathbf{B})}{2\sigma_a^2}\right) \quad (3.13)$$

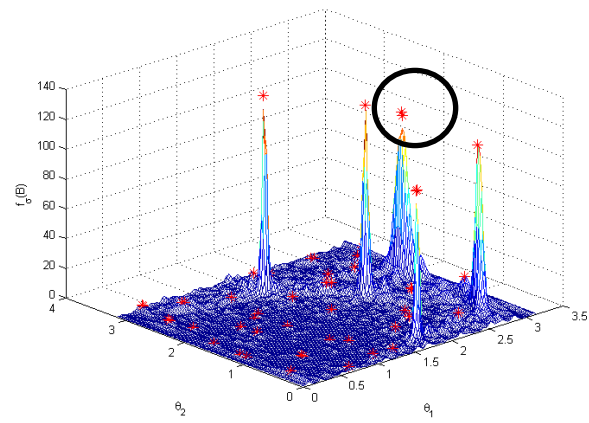
Visto que as colunas de \mathbf{A} são consideradas como sendo de norma unitária, analogamente ao que é feito em [NAINI *et al.*, 2008], cada célula pode ser representada por

um vetor real contendo $M - 1$ componentes que definem a representação polar de um vetor no \mathbb{R}^M . A população inicial é gerada com ângulos uniformemente distribuídos no intervalo $(0, \pi)$, sendo que, no passo de mutação, é introduzida uma restrição: se o ângulo mutado não pertencer ao intervalo $(0, \pi)$, ele volta a possuir o seu valor original. Isto assegura que somente células válidas são geradas.

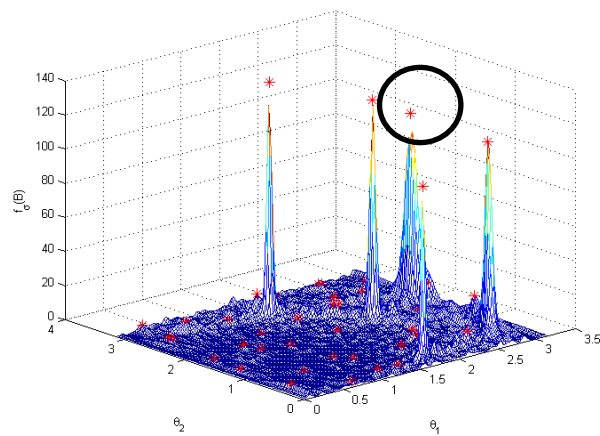
Neste trabalho, em vez de utilizar uma métrica de distância no sistema de coordenadas polares, optou-se que as supressões fossem feitas utilizando a correspondente representação cartesiana da célula. Neste caso, visto que não existe, por hipótese, nenhum par de colunas \mathbf{A} separado de menos de 6 graus, foi escolhido o limiar $\sigma_s = 0.1047$ radianos (6 graus).

Já que se optou por um critério de parada dado por um limite pré-definido de iterações, é possível que, ao final da operação do algoritmo, haja espaço para algum refinamento local. Além disso, é possível que a distância entre algumas células esteja abaixo do limiar de supressão. Por conta destes fatores, foi aplicado um passo adicional de supressão, e, em seguida, as células restantes foram sujeitas a hipermutação, ou seja, as etapas de clonagem, mutação e seleção foram repetidas algumas vezes. Pelo mesmo motivo, é necessário executar novamente o passo de supressão para obter a população final de células que serão utilizadas na identificação dos vetores de mistura.

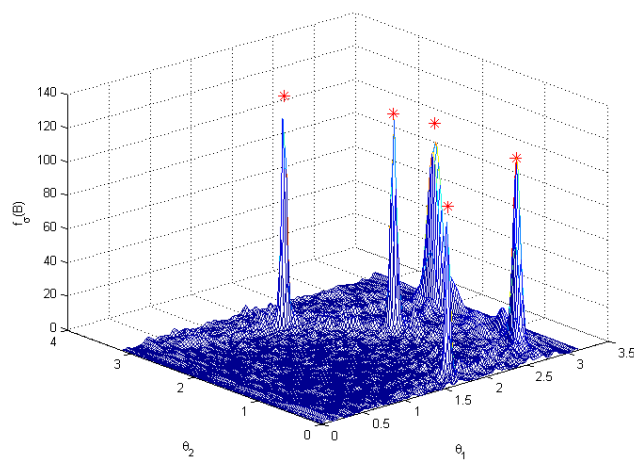
Na Fig. 6.1, é possível verificar as três etapas da otimização da função custo, dada por (3.13), para uma situação com 5 fontes e 3 misturas. Os asteriscos indicam os indivíduos da população. Em (a), temos a solução obtida pela opt-aiNet, dado um número fixo de iterações. Em (b), são mostrados os resultados depois de feitas as etapas de refinamento e supressão. É possível observar dois fatores: alguns indivíduos que se encontravam próximos a um ótimo local acabaram por atingi-lo; e, quando dois ou mais indivíduos se encontraram próximos em (a), em (b) todos exceto o melhor indivíduo foram suprimidos. Por fim, é feita a supressão de todos os indivíduos com valor de custo inferior a um limiar pré-definido: desta forma (c) apresenta a população final, contendo somente as estimativas das direções das fontes.



(a)



(b)



(c)

Fig. 6.1 – 3 etapas da otimização da função custo, dada por (3.13), para uma situação com 5 fontes e 3 misturas: (a) solução obtida pela opt-aiNet; (b) depois das etapas de refinamento e supressões; e (c) população final, contendo somente as estimativas das direções das fontes. Pode-se ver no detalhe que um dos dois indivíduos que aparecem em (a) já foi suprimido em (b).

6.2.2 Identificação das Colunas da Matriz de Mistura

A última parte do método proposto é dedicada à extração das estimativas dos vetores de mistura a partir da população. Como mencionado, os valores mais altos de $f_\sigma(\mathbf{B})$, descrita em (3.13), estão associados aos valores reais dos vetores de mistura, e suas amplitudes deveriam ser bem superiores às associadas aos máximos locais restantes. Entretanto, a validação deste cenário depende fortemente do parâmetro σ da função custo. Isto significa que, se σ for adequadamente ajustado, a diferença entre as amplitudes dos picos associados com as colunas de \mathbf{A} e dos máximos espúrios poderia ser utilizada para determinar automaticamente o número de fontes.

Um procedimento simples explorando esta ideia pode ser estabelecido adotando a seguinte abordagem. Primeiramente, o vetor de *fitness* é ordenado, e, em seguida, é computada a diferença entre os valores adjacentes. Finalmente, é localizado o maior valor deste vetor, que indica onde ocorre o maior intervalo entre os valores de *fitness* na população. Desta forma, é possível estimar as direções das fontes, que são relativas aos maiores picos, as quais compõem a solução do problema. Pode-se ainda realizar um refinamento local nos picos maiores, aumentando, conseqüentemente, a acurácia da estimação.

É importante mencionar que, em vez de se usarem os valores brutos de *fitness*, propõe-se o emprego de uma transformação logarítmica antes do passo de ordenação: desta forma, é enfatizada a lacuna existente entre as melhores soluções e o resto da população.

6.3 Simulações e Resultados

Com intuito de avaliar a abordagem proposta, foram realizados três experimentos diferentes. No primeiro cenário, é comparado o desempenho da opt-aiNet com o de uma busca exaustiva. Este resultado é importante para verificar a capacidade de busca multimodal do algoritmo. No segundo cenário, é avaliado o processo de estimação do número de fontes presente na mistura. Por fim, é sugerido um cenário em cujo contexto aplicamos a metodologia a um ambiente parecido com o do primeiro cenário, embora mais severo.

Em todas as simulações, a geração das fontes foi realizada de acordo com o modelo proposto em [NAINI *et al.*, 2008], segundo o qual os sinais esparsos correspondem a uma combinação de fontes gaussianas:

$$s_i \sim pN(0, \sigma_{on}) + (1 - p)N(0, \sigma_{off}) \quad (6.2)$$

onde p (dada por k/N , sendo o valor de k definido em cada cenário) é a probabilidade de uma fonte estar ativa e σ_{on} e σ_{off} são os desvios padrão nos modos ativo e inativo, respectivamente. Desta forma, com o intuito de garantir esparsidade, é necessário que $\sigma_{off} \ll \sigma_{on}$ e $p \ll 1$. Nos experimentos, foram considerados os valores $\sigma_{on} = 1$ e $\sigma_{off} = 0,01$.

Como modelo de mistura para simulações, foi utilizada a Eq. (2.1). A matriz de mistura foi gerada de forma aleatória e cada uma das colunas foi normalizada. Além disto, forçou-se que o mínimo ângulo de distância entre quaisquer duas colunas de \mathbf{A} valesse pelo menos 6 graus. Finalmente, para cada n , $\mathbf{x}(n)$ também foram normalizados.

Primeiro Cenário

Visto que a opt-aiNet pretende localizar os maiores picos da função $f_\sigma(\mathbf{B})$, e, em teoria, nenhum valor encontrado neste processo de otimização deve ser melhor do que o encontrado em uma busca exaustiva bem construída. Desta forma, com objetivo de validar o método proposto e verificar a sua capacidade de estimar corretamente a matriz de mistura, é proposto um cenário para comparar os resultados obtidos nos dois procedimentos. Visto que este tipo de teste só é computacionalmente factível para um espaço de busca de baixa dimensão, foram escolhidos os seguintes parâmetros para gerar as misturas: $N = 5$, $M = 3$, $k = 2$ e $T = 4000$.

Tendo por base simulações preliminares, foi adotado o valor $\sigma = 0.02$ e os parâmetros da opt-aiNet foram definidos como: $N_{inicial} = 40$, $N_c = 10$, $\beta = 10$, $d = 30\%$. O número de máximo de iterações escolhido foi 400. Duas fases de hipermutação foram aplicadas, sendo os valores de β iguais a 10 e 20, respectivamente, e, em ambas as situações, $N_c = 10$. Entretanto, realizaram-se 10 iterações para a primeira fase de hipermutação e 100 iterações para a segunda.

Para fins de análise de desempenho, um valor de erro ε foi calculado utilizando a expressão:

$$\varepsilon = \min_{\mathbf{P} \in \Lambda} \|\mathbf{A} - \hat{\mathbf{A}}\mathbf{P}\|_2 \quad (6.3)$$

onde $\hat{\mathbf{A}}$ é a matriz de mistura estimada e Λ é o conjunto de todas as matrizes de permutação [NAINI *et al.*, 2008]. Para uma mesma matriz de mistura, foram realizadas 10 realizações independentes do algoritmo. A Tab. 6.1 mostra estes resultados, o erro médio das realizações e também o erro associado à busca exaustiva.

Experimento	Opt-aiNet
1	0,0043
2	0,0034
3	0,0048
4	0,0049
5	0,0035
6	0,0028
7	0,0030
8	0,0031
9	0,0029
10	0,0052
Média	0,0038
Variância	8,45 x 10⁻⁷
Busca exaustiva	0,0020

Tab. 6.1 – Valores de erros ϵ , calculados utilizando a Eq. (6.3), da estimação da matriz de mistura

O intuito desta simulação foi o de verificar se a convergência do algoritmo ocorre para estimativas similares, visto que as inicializações são feitas de forma aleatória. Dada a proximidade entre os resultados obtidos pela opt-aiNet e pelo método exaustivo, é possível concluir que a abordagem proposta foi capaz de identificar corretamente a matriz de mistura em todas as situações, e que o algoritmo consegue convergir para soluções similares, mesmo com inicializações diferentes, dado o baixo valor da variância. Caso fosse permitido que a opt-aiNet realizasse mais iterações, é provável que o erro ficasse ainda mais próximo do ideal.

Consideremos agora uma situação em que a matriz de mistura é gerada de forma aleatória em cada experimento, não permanecendo mais constante como no cenário anterior. Neste caso, para cada situação, são calculados os valores de ϵ tanto para o método proposto quanto para o exaustivo, permitindo uma análise da robustez do método no que diz respeito à estimação correta da matriz de mistura. Na Tab. 6.2, são mostrados os erros das estimativas obtidas com a opt-aiNet e com a busca exaustiva, desta vez considerando dez diferentes matrizes de misturas. Novamente, é possível observar a proximidade dos dois resultados, visto que, exceto pelo experimento 5, os valores obtidos pela opt-aiNet e pelo método exaustivo ($\Delta\epsilon$) não diferem mais do que 15%. O motivo de existirem algumas estimativas da opt-aiNet com valores de erro menores do que os relativos à busca exaustiva se deve ao fato de que a implementação desta última não é ideal, visto que a sua grade de busca possui precisão finita.

Experimento	Opt-aiNet	Exaustivo	$\Delta\epsilon(\%)$
1	0,0064	0,0064	0,0
2	0,0043	0,0040	7,5
3	0,0077	0,0068	13,2
4	0,0066	0,0061	8,2
5	0,0096	0,0066	45,4
6	0,0049	0,0054	-9,3
7	0,0030	0,0035	-14,3
8	0,0034	0,0035	-2,8
9	0,0032	0,0035	-8,6
10	0,0035	0,0034	2,9

Tab. 6.2 – Erros, calculados utilizando a Eq. (6.3), da estimação da matriz de mistura

É também interessante mencionar que, usando a mesma configuração de simulação, o método descrito em [NAINI *et al.*, 2008] gerou um erro estimado de 0,0161. Comparando este valor com os apresentados na Tab. 6.2, é possível concluir que o método proposto tem um desempenho ligeiramente melhor, apesar de esta diferença poder ser explicada por uma escolha sub-ótima dos parâmetros da abordagem canônica. De qualquer forma, isto,

juntamente com a comparação com o procedimento de busca exaustiva, revela que a opt-aiNet é de fato adequada a este tipo de problema de SCA.

Segundo Cenário

A intenção agora é verificar a acurácia do método em estimar corretamente o número de fontes. Para atingir este objetivo, é sugerido um cenário com $k = 1$, $M = 2$, N variando entre 2 e 15 e $T = 3000$. Foi adotado o valor $\sigma = 0,02$, e foram definidos os seguintes parâmetros para a opt-aiNet: $N_{\text{inicial}} = 30$, $N_c = 10$, $\beta = 50$ and $d = 30\%$. O número máximo de iterações foi 400, e duas fases de hipermutação foram aplicadas, ambas com $\beta = 100$ e $N_c = 5$, sendo realizadas 10 iterações para a primeira fase e 5 iterações para a segunda. Todos os parâmetros foram ajustados em uma etapa prévia, no contexto de uma situação com seis fontes presentes na mistura. Foram feitas no total 50 realizações independentes, nas quais ambas fontes e matrizes de mistura foram geradas de forma aleatória.

Os resultados mostrados na Fig. 6.2 indicam que o método consegue estimar corretamente o número de fontes em 100% dos casos quando M é igual a 6 e 7, indicando também 88% de acerto nos outros casos em que M vale entre 2 e 9, o que pode ser caracterizado como um bom desempenho. Por outro lado, o fato de os melhores resultados terem sido obtidos na vizinhança do caso com $M = 6$ mostra que existe uma sensibilidade do método ao ajuste dos parâmetros.

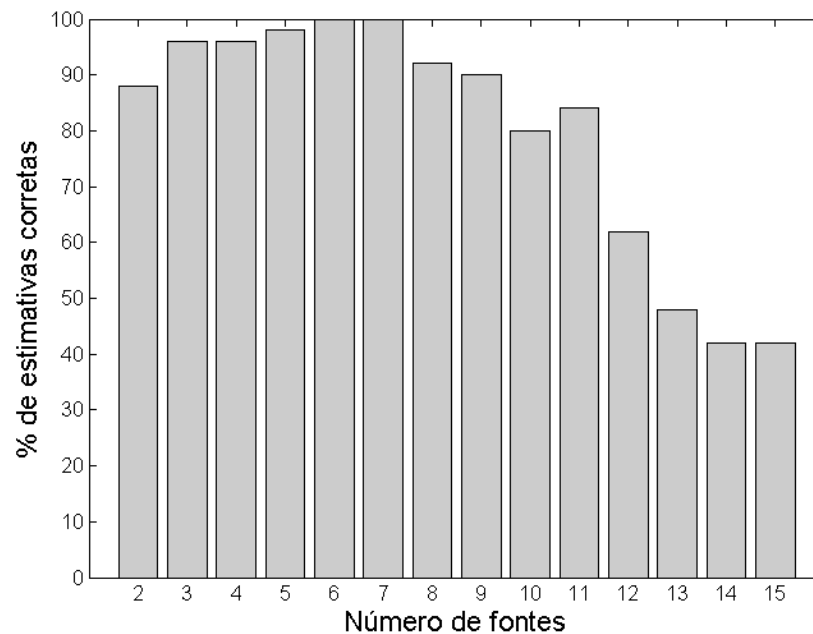


Fig. 6.2 – % de estimativas corretas do número de fontes na mistura

Outro fato importante a se verificar é a qualidade da estimação de mistura nestas situações. Isto pode ser feito com a ajuda do erro angular médio (EAM) entre as matrizes de mistura originais e suas respectivas estimativas. A Fig. 6.3 mostra o valor médio de EAM considerando todas as situações em que o método proposto estimou corretamente o número de fontes. Dados os baixos valores encontrados (menores do que 3% dos valores das distâncias originais) para os casos até 14 fontes, conclui-se que as matrizes foram estimadas corretamente dentro de uma precisão razoável. Outro fato relevante é o de existir um acréscimo dos valores de EAM com o aumento do número de fontes, o que era esperado, tendo em vista que a quantidade de subespaços existentes cresce em progressão geométrica.

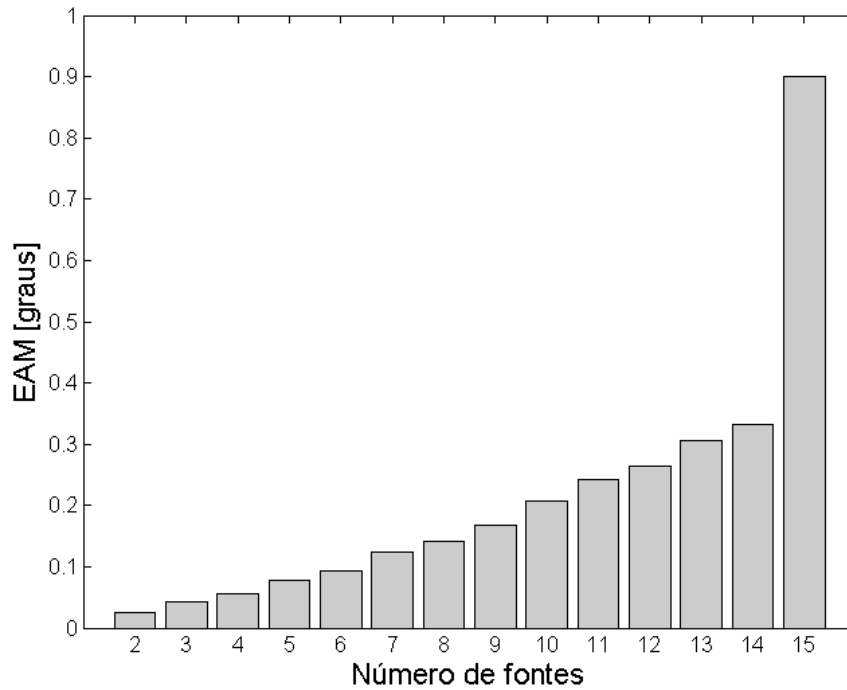


Fig. 6.3 – Valor médio de EAM para diferentes números de fontes

Terceiro Cenário

Por fim, é proposto um último experimento, em que é considerado um cenário com $N = 8$, $M = 4$, $k = 2$ e $T = 4000$. Neste caso, foi adotado o valor $\sigma = 0,03$, e os parâmetros da opt-aiNet são: $N_{\text{initial}} = 50$, $N_c = 10$, $\beta = 10$, $d = 30\%$. Foram realizadas 700 iterações do algoritmo, e, novamente, foram aplicadas duas fases adicionais de hipermutação: a primeira com $\beta = 10$, e a segunda com $\beta = 25$, ambas com $N_c = 10$, sendo rodadas 100 iterações de cada fase.

Foram realizadas 100 simulações com este cenário. A Fig. 6.4 mostra o número de fontes obtido pelo algoritmo em todas as simulações, enquanto a Fig. 6.5 mostra os correspondentes erros de estimação, segundo a Eq.(6.3). Quando o algoritmo estimou menos de 8 fontes, o erro foi calculado a partir somente das colunas obtidas.

Como ilustrado na Fig. 6.4, o método foi capaz de identificar corretamente o número de fontes em 97% dos casos. Além disto, os erros na estimação da matriz de mistura foram bastante satisfatórios: na maioria das simulações, eles se mantiveram abaixo

de 0,02. Ambos os resultados novamente indicam o grande potencial do método na identificação da matriz de mistura.

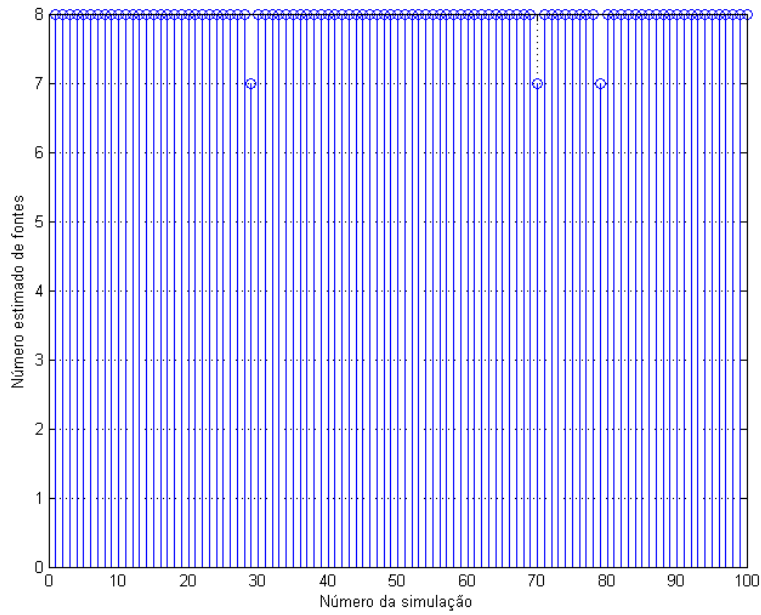


Fig. 6.4 – Número obtido de vetores (fontes)

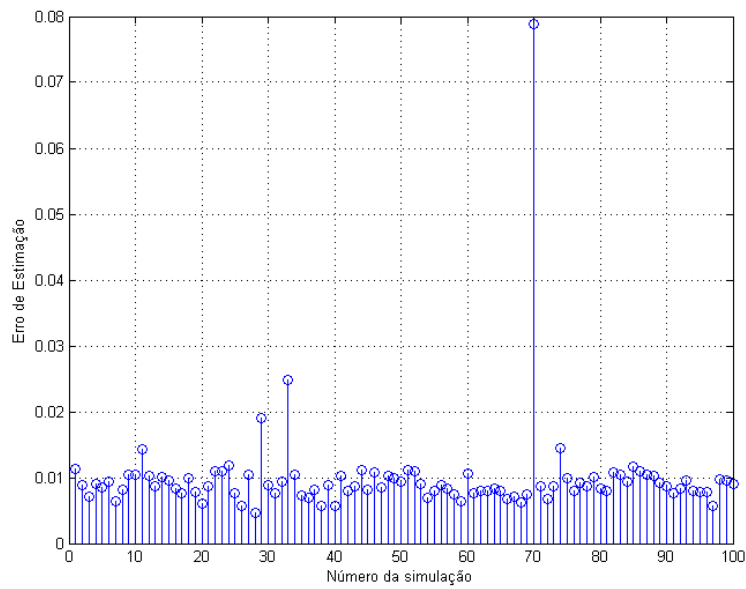


Fig. 6.5 – Erros das estimações dos vetores de mistura

6.4 Considerações Finais

As técnicas de SCA baseadas em busca por subespaços são ferramentas importantes para lidar problemas subparametrizados de BSS. Interessantemente, a aplicação dessas técnicas leva a funções custo intrinsecamente multimodais e a um problema de otimização em que a obtenção de múltiplas soluções é essencial, inclusive para que se possa estimar adequadamente o número de fontes presentes nas misturas.

Para lidar com esse problema, foi proposto o uso de um algoritmo bioinspirado de comprovada eficiência em tarefas desse tipo, a rede imunológica denominada opt-aiNet. Os resultados obtidos comprovam a relevância do uso desse algoritmo, abrindo a possibilidade de emprego de uma gama mais ampla de meta-heurísticas.

Capítulo 7

Separação / Extração de Fontes

Baseadas num Critério de

Quantificação de Esparsidade

A breve revisão mostrada no Capítulo 3 indica uma leve diferença entre os processos de estimação da matriz de mistura e os métodos de separação (ou reconstrução, quando a separação estrita não for possível) de sinais, visto que estes últimos são usualmente derivados a partir de formulações baseadas em problemas de otimização convexa, como descrito na Seção 3.2.3, enquanto os primeiros ainda se utilizam de métodos empíricos, normalmente baseados em clusterização (como é caso do método proposto no Capítulo 5).

Levamos então uma questão: seria possível criar uma função custo otimizável baseada numa quantificação numérica do conceito abstrato de esparsidade para resolver o problema de separação de fontes como um todo (tanto na estimação da matriz de mistura, quanto na separação/recuperação de fontes)?

Uma resposta parcial a este questionamento pode ser encontrada em [RICKARD, 2006], que afirma que é possível separar fontes esparsas se elas possuírem ortogonalidade disjunta, condição esta que, apesar de suficiente, não é necessária. Isto nos leva à adaptação da questão proposta para a seguinte: assumindo que as fontes são esparsas, sob quais condições um critério baseado diretamente em esparsidade conduzirá à separação das fontes? Em outras palavras, quando é possível separar as fontes decompondo as misturas em sinais que sejam tão esparsos quanto possível? É importante notar que esta questão é

bem próxima da levantada por Comon em seu clássico trabalho [COMON, 1994], no qual se analisa a questão de separação de sinais baseando-se na independência estatística, como explicado na Seção 2.1.

Respondendo estas perguntas para o caso de SCA, pretendemos obter uma visão mais quantitativa sobre o papel da esparsidade na separação de fontes, abrindo caminho, desta forma, para novos desenvolvimentos junto à teoria de BSS. Como explicado na Seção 3.1, o trabalho [HURLEY e RICKARD, 2009] faz um estudo detalhado sobre formas de se mensurar a esparsidade, mas a norma ℓ_1 não é a apontada como a medida mais adequada para esse fim. No entanto, em diversos estudos práticos, esta medida tem trazido bons resultados no problema de separação/reconstrução das fontes quando a matriz de mistura é conhecida.

Tendo isso em vista, nossa proposta é verificar a utilização da norma ℓ_1 também para o caso de separação totalmente cega, no sentido de que nem os sinais das fontes nem o sistema de mistura sejam conhecidos. Naturalmente, para ser bem sucedida, esta otimização baseada na norma ℓ_1 precisa ser capaz de abranger as situações já solucionadas atualmente por outras abordagens, como os casos de extração e separação de fontes na situação em que $M = N$. No caso subparametrizado, além de lidar com a separação/reconstrução de sinais quando o sistema misturador é conhecido, é preciso verificar se também é possível estimar tanto o número de fontes quanto a matriz de mistura.

7.1 Medida de Esparsidade Baseada na Norma ℓ_1

Para um dado sinal $g(n)$, a “medida de esparsidade” proposta Ψ_g , baseada na norma ℓ_1 , é definida como sendo:

$$\Psi_g = \sum_n |g(n)| \quad (7.1)$$

Em [HURLEY e RICKARD, 2009], é mostrado que a norma ℓ_1 não respeita todos os seis critérios mostrados na Seção 3.1, os quais são considerados pelos autores como sendo importantes para que se estabeleça um bom quantificador de esparsidade. Porém, em diversas aplicações, não é necessário cumprir todos estes critérios, motivo pelo qual a norma ℓ_1 é largamente utilizada. Um exemplo é o fato de que o número de amostras dos

sinais não é alterado nos processos de recuperação das fontes, o que significa que o último critério – que afirma que se forem adicionados coeficientes de valor nulo ao conjunto total, a esparsidade irá aumentar – não precisa ser levado em consideração.

Por outro lado, pelo fato de que, nos problemas que pretendemos abordar, é necessário comparar a esparsidade de dois sinais distintos, alguns problemas podem surgir na utilização da norma ℓ_1 . Se voltarmos para os critérios sugeridos na Seção 3.1, um deles diz que a esparsidade é invariante ao escalamento dos sinais, ou seja, se multiplicarmos um sinal por uma constante, sua esparsidade não deve ser alterada. Isto não ocorre com a norma ℓ_1 , que é susceptível a variações de amplitude. Ou seja, se compararmos um sinal $g(n)$ com o mesmo sinal multiplicado por uma constante k qualquer, a esparsidade será a mesma, mas, se fizermos o cálculo da norma ℓ_1 , teremos

$$q(n) = k \cdot g(n) \rightarrow \Psi_q = \sum_n |q(n)| = \sum_n |k \cdot g(n)| = |k| \sum_n |g(n)| = |k| \Psi_g \quad (7.2)$$

Outro problema que ocorre é que há situações em que, nitidamente, um sinal é mais esparso que o outro, e o valor da norma ℓ_1 ainda assim é o mesmo. Isto é ilustrado pelo exemplo a seguir

$$\begin{aligned} g(n) = \{0, 0, 5, 0, 0\} \rightarrow \Psi_g &= \sum_n |g(n)| = 5 \\ q(n) = \{1, 1, 1, 1, 1\} \rightarrow \Psi_q &= \sum_n |q(n)| = 5 \end{aligned} \quad (7.3)$$

Felizmente, se a potência dos sinais for normalizada, os dois problemas automaticamente deixam de existir. Porém, esta necessidade de normalização de potência dos sinais faz com que não seja possível a recuperação da amplitude original dos sinais, problema este também encontrado nas técnicas baseadas em ICA, conforme exposto na Seção 2.1. Desta forma, para todas as análises feitas neste capítulo, será necessário sempre algum procedimento que garanta que não existe variação de potência nas transformações lineares que serão realizadas.

7.2 Extração de Fontes – Análise da Função Custo

No Capítulo 2, foi explicado o problema de BSE, cujo objetivo é extrair uma única fonte da mistura. Considerando que a função Ψ_g quantifica corretamente o grau de

esparsidade de um sinal, podemos utilizá-la como critério a ser seguido na adaptação de um vetor de extração de sinais. Em outras palavras, o objetivo seria obter um vetor \mathbf{w} tal que o sinal estimado $y(n) = \mathbf{w}^T \mathbf{x}(n)$ seja o mais esparso possível, i.e., Ψ_y seja o menor possível. Com intuito de entender melhor esta minimização, Ψ_y será descrito em função de $\mathbf{s}(n)$, de forma a analisarmos os mínimos da função. Desta forma, obtemos:

$$\Psi_y = \sum_n |y(n)| = \sum_n |\mathbf{w}^T \mathbf{x}(n)| = \sum_n |\mathbf{w}^T \mathbf{A} \mathbf{s}(n)| = \sum_n \left| \sum_j \mathbf{w}^T \mathbf{a}_j s_j(n) \right|. \quad (7.4)$$

A fim de empreender uma análise inicial, admitamos que as fontes possuam ortogonalidade disjunta. Além disto, para que seja garantida que a potência de y não seja alterada com a variação de \mathbf{w} , consideremos que este vetor será normalizado na forma $\|\mathbf{w}\|_2 = 1$. Assim, teremos que a norma ℓ_1 de y pode ser escrita como sendo:

$$\begin{aligned} \Psi_y &= \sum_n \left| \sum_j \mathbf{w}^T \mathbf{a}_j s_j(n) \right| = \sum_n \sum_j |\mathbf{w}^T \mathbf{a}_j s_j(n)| \\ &= \|\mathbf{w}\|_2 \sum_j \|\mathbf{a}_j\|_2 |\cos(\theta_j)| \Psi_{s_j} = \sum_j \|\mathbf{a}_j\|_2 |\cos(\theta_j)| \Psi_{s_j}, \end{aligned} \quad (7.5)$$

onde \mathbf{a}_j denota a j -ésima coluna de \mathbf{A} , θ_j é o ângulo entre \mathbf{a}_j e \mathbf{w} , e Ψ_{s_j} é o valor de Ψ da j -ésima fonte.

O primeiro passo da análise da função custo será mostrar que Ψ_y alcança seu mínimo quando \mathbf{w} é ortogonal a $M - 1$ fontes: desta forma, se $N = M$, teremos que \mathbf{w} anulará a contribuição de todas as fontes menos uma. Para que se mostre isto, consideremos um vetor \mathbf{w}_V que é ortogonal a um conjunto V de vetores \mathbf{a}_j , cuja cardinalidade é $0 < |V| \leq M - 2$ (assumindo $N \geq M$), de forma que temos $|\cos(\theta_j)| = 0 \forall j \in V$ e $|\cos(\theta_j)| > 0 \forall j \notin V$. Sabe-se que $|x|$ não é diferenciável em $x = 0$; isto implica que Ψ_y não é diferenciável em nenhum dos vetores \mathbf{w}_V . Entretanto, se considerarmos

$$\Psi_j^V = \sum_{j \notin V} \|\mathbf{a}_j\|_2 |\cos(\theta_j)| \Psi_{s_j}, \quad (7.6)$$

observamos que esta função é diferenciável se restringirmos a análise ao hiperplano α_V , de dimensão $M - |V|$, que é ortogonal aos vetores $\mathbf{a}_j \in V$. Visto que Ψ_j^V é dado por uma soma de funções não-negativas que são côncavas na vizinhança de \mathbf{w}_V , $|\cos(\theta_j)| > 0$, e

concluimos que Ψ_j^V também é côncava na vizinhança de \mathbf{w}_V [BOYD e VANDENBERGHE, 2004]. Consequentemente, nenhum \mathbf{w}_V satisfaz a condição de segunda ordem requerida para ser considerada um mínimo de Ψ_j^V . Visto também que a Eq. (7.6) é exatamente a Eq. (7.5) se restringirmos \mathbf{w} ao hiperplano α_V , concluimos que \mathbf{w}_V também não pode ser mínimo de Ψ_j . Se considerarmos os vetores \mathbf{w}_V que não são ortogonais a nenhum vetor \mathbf{a}_j , i.e., $V = \{\}$, teremos $|\cos(\theta_j)| > 0 \forall j$ na Eq. (7.5), tal que Ψ_y seja diferenciável e côncavo nestes \mathbf{w}_V . Desta forma, concluimos que nenhum mínimo de Ψ_y pode existir entre estes pontos.

Caso se assuma que, nos casos não-degenerados, não é possível ter um conjunto V tal que $|V| \geq M$, podemos finalizar nossa análise estudando o caso em que $|V| = M - 1$. Neste caso, não é possível estabelecer um hiperplano do tipo α_V com o intuito de obter uma função diferenciável em termos da Eq. (7.6). Portanto, Ψ_j não é diferenciável nos vetores \mathbf{w} que sejam ortogonais a $M - 1$ vetores \mathbf{a}_j . Desta forma, estes pontos de Ψ_j não diferenciáveis são qualificados como pontos críticos. Observando que Ψ_y é delimitado, visto que é definido para todo \mathbf{w} , tal que $\|\mathbf{w}\|_2 = 1$, que é um conjunto compacto de pontos, concluimos que estes pontos críticos são os únicos candidatos a serem mínimos locais de Ψ_j . Além disto, se notarmos o fato de que, em todos estes pontos críticos, (2.1) fornece as projeções de $M - 1$ fontes, observamos que, quando $M = N$, é possível extrair uma fonte em cada um dos mínimos locais de Ψ_y .

Com o intuito de ilustrar o resultado dado por (7.5), consideremos um cenário com duas misturas e duas fontes, sendo a matriz de mistura \mathbf{A} ortonormal. O vetor de separação \mathbf{w} , neste caso, pode ser representado por um único ângulo Θ_w . Neste exemplo, o ponto ótimo de Ψ_j é alcançado quando \mathbf{w} é ortogonal à coluna de \mathbf{A} que não é associada ao sinal mais esparso.

Na Fig. 7.1, são mostrados os valores de $\Psi_y - \Psi_{si}$ por Θ_w . Neste exemplo, a direção da coluna de \mathbf{A} associada à fonte mais esparsa é $-\pi/6$, enquanto a direção da segunda coluna é dada por $\pi/3$. É importante notar que a diferença $\Psi_y - \Psi_{si}$ é nula quando o vetor de separação tem uma direção dada por $\Theta_w = \pi/3$, e, desta forma, ele é de fato o vetor ortogonal à fonte mais esparsa. É possível notar também a existência de outro mínimo local na função, que corresponde à solução \mathbf{w} que extrai a fonte restante. Este fato já era esperado, visto que nossa análise concluiu que cada um dos mínimos de Ψ_y é relativo a uma das colunas da matriz de mistura.

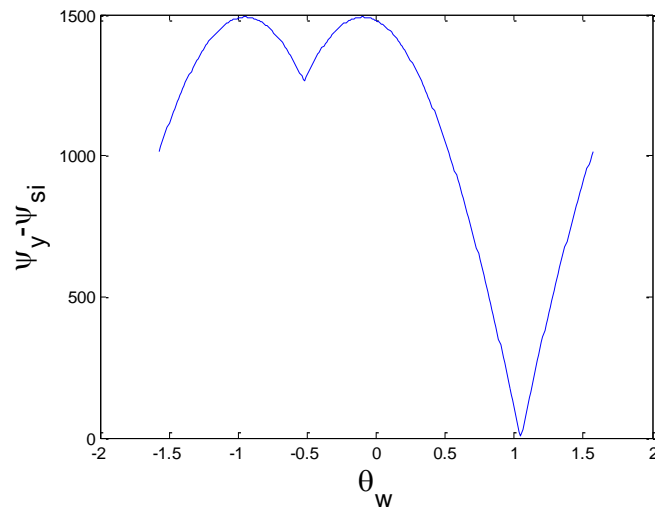


Fig. 7.1 – Valor de $\Psi_y - \Psi_{si}$ por θ_w . Quando w é ortogonal à fonte mais esparsa, a extração é efetivamente feita.

Analisaremos agora o caso de um sistema subparametrizado, em que $M = 3$ e $N = 4$, para verificar o significado de cada um dos mínimos da função. Nas Figs. 7.2 e 7.3, são mostradas as curvas de nível e de superfície 3D dos valores de Ψ_y . Nelas, podem ser observados alguns “morros” e “vincos”. Os morros são formados por partes côncavas da função Ψ_y , onde w não é ortogonal a nenhuma fonte. Por outro lado, os vincos correspondem à intersecção entre as bolas $\|w\|_2 = 1$ e os hiperplanos que são ortogonais a uma das fontes. Teremos, então, mínimos locais onde dois desses hiperplanos se encontram, sendo, neste caso, w perpendicular a duas das fontes existentes na mistura.

O tratamento matemático apresentado mostra que a norma ℓ_1 é capaz de originar um contraste efetivo para a extração de uma fonte quando existe ortogonalidade disjunta, o que dá uma resposta parcial à questão, mencionada anteriormente, que guia o desenvolvimento deste trabalho. Agora, será mostrado que é possível desenvolver, a partir deste resultado teórico, uma metodologia aplicável em um processo de otimização. O primeiro passo nesta direção é construir um algoritmo adaptativo para extração de fontes; em seguida, será estudado o que ocorre quando as fontes não possuem ortogonalidade disjunta.

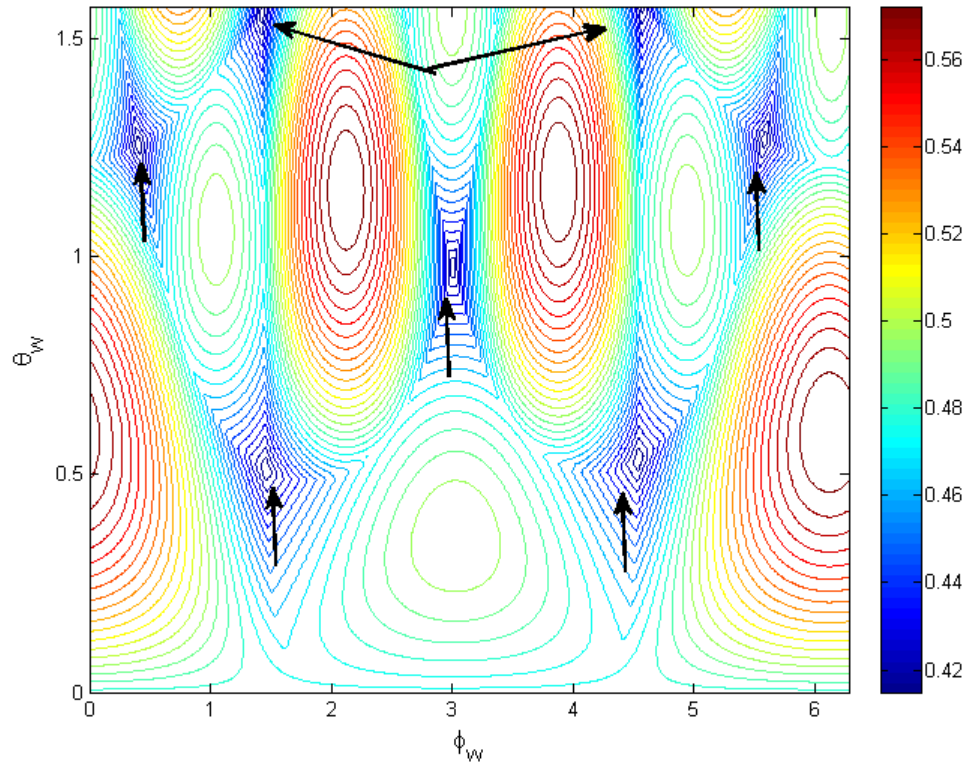


Fig. 7.2 – Curva de nível da função custo Ψ_y , para um caso com 4 fontes e 3 misturas. As setas indicam os mínimos da função.

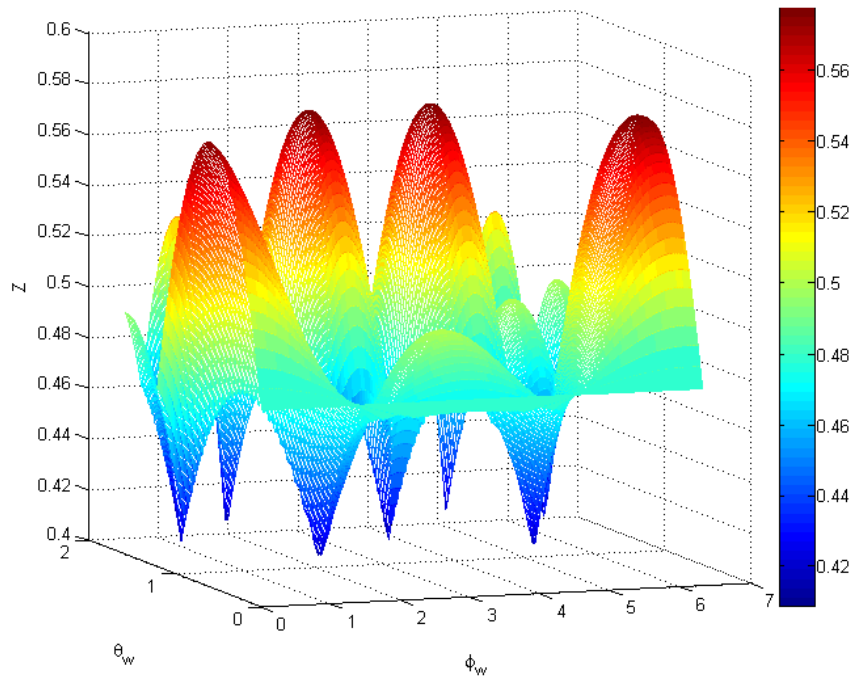


Fig. 7.3 – Superfície 3D da função custo Ψ_y para um caso com 4 fontes e 3 misturas

7.2.1 Algoritmo de Extração Baseado no Gradiente

Como mostrado no começo deste capítulo, se o número de fontes é igual ao de misturas, é possível extrair uma fonte procurando um mínimo local na função custo Ψ_y . Neste caso, um método de otimização que pode ser utilizado é aquele baseado na direção do gradiente da função custo [LUENBERGER, 2003]. É importante salientar que, apesar de a função módulo não ser derivável em todos os pontos, para a análise é possível utilizar a função sinal como derivada da função módulo.

Visto que a norma ℓ_1 é sensível a uma variação de amplitude (como explicado na Seção 7.1), o processo de branqueamento acaba sendo uma peça importante para garantir que os sinais terão a mesma potência, e, conseqüentemente, que será possível fazer uma comparação justa entre as esparsidades dos sinais utilizando a norma ℓ_1 .

Depois do branqueamento, é possível minimizar a função custo em (7.4). Para isto, primeiramente será calculado o gradiente de Ψ_y em função de \mathbf{w} . Desta forma, a partir de (7.4), teremos que

$$\nabla_{\mathbf{w}} \Psi_y = \sum_n \text{sign}[y(n)] \tilde{\mathbf{x}}^T(n), \quad (7.7)$$

onde $\tilde{\mathbf{x}}$ representa a mistura branqueada e $\text{sign}[\cdot]$ denota a função sinal. Deste modo, a adaptação será dada por

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}} \Psi_y(\mathbf{w}(k)), \quad (7.8)$$

onde $\mu > 0$ é o fator multiplicativo que indica o tamanho do passo a ser dado em cada iteração. O valor inicial de \mathbf{w} é gerado de forma aleatória, e, com o intuito de garantir que $\|\mathbf{w}\|_2 = 1$, ao final de cada iteração é necessário realizar a seguinte normalização

$$\mathbf{w}(k+1) = \frac{\mathbf{w}(k+1)}{\|\mathbf{w}(k+1)\|_2}. \quad (7.9)$$

Depois desta etapa, uma nova iteração é iniciada, a partir da Eq. (7.7), até que um critério de parada seja obedecido.

7.2.2 Extração de Fontes sem Ortogonalidade Disjunta

Até agora, fizemos uma análise somente para o caso específico em que as fontes possuem ortogonalidade disjunta. Dando continuidade ao estudo, iremos realizar uma interpretação do que ocorre com a otimização baseada na norma ℓ_1 para outros tipos de

fontes. No caso de ortogonalidade disjunta, foi possível fazer uma análise completamente determinística dos sinais, visto que a otimização foi feita com base em uma norma, cuja definição não se baseia em conceitos estatísticos. Entretanto, o algoritmo descrito pelas Eqs. (7.7) e (7.8) pode também ser obtido se considerarmos uma abordagem estatística baseada na maximização da negentropia (descrita na Seção 2.1.2), como iremos mostrar a seguir.

Tipicamente, no processo de minimização da entropia, em vez de estimar $p(y)$ a cada iteração, é realizada uma aproximação da fdp, baseada em funções não-lineares fixas [HYVÄRINEN *et al.*, 2001]. Um aspecto interessante é que, se for utilizada a distribuição laplaciana (considerada uma distribuição esparsa [FEVOTTE e GODSILL, 2006]) como aproximação de $p(y)$, é possível encontrar uma relação entre a maximização da negentropia e o método proposto em 7.2.1. Para isto, consideremos a distribuição laplaciana, um caso particular da distribuição gaussiana generalizada, dada por

$$p(y) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|y-\mu|}{\alpha}\right)^\beta}, \quad (7.10)$$

quando $\beta = 1$. Neste caso, iremos considerar também $\alpha = 1$ e $\mu = 0$. Desta forma, teremos que (2.14), $\mathbf{w} = \arg \min E\{-\log p(y)\}$, pode ser reescrita como sendo

$$\begin{aligned} \mathbf{w} &= \arg \min E\left\{-\log \frac{1}{2} e^{-|y|}\right\} \\ \mathbf{w} &= \arg \min \left(-\log \frac{1}{2} + E\{|y|\}\right) \\ \mathbf{w} &= \arg \min E\{\|y\|_1\} \end{aligned} \quad (7.11)$$

Quando a esperança em (7.11) é substituída por um estimador amostral, o critério resultante se torna equivalente à otimização baseada na norma ℓ_1 . Isto mostra que, nos casos em que as fontes são esparsas, mas não possuem ortogonalidade disjunta, ainda é possível extrair uma das fontes utilizando a norma ℓ_1 , visto que se estará, de certo modo lidando com a maximização de uma medida de não-gaussianidade.

Com o intuito de compreender melhor este fato, será feito agora um estudo acerca de casos intermediários. Desta forma, as fontes serão geradas utilizando-se a distribuição gaussiana generalizada, descrita em (7.10). Como dito, quando $\beta = 1$, $p(y)$ descreve uma distribuição laplaciana, e o caso de uma distribuição uniforme é dado quando $\beta \rightarrow \infty$. A Fig. 7.4 mostra os valores da norma ℓ_1 de vetores com potência unitária, gerados utilizando

(7.10) para $1 < \beta < 4$, $\alpha = 1$ e $\mu = 0$. Conforme o valor de β aumenta, o sinal gerado vai se tornando menos esparsos. Nesta situação, como era esperado, a norma ℓ_1 aumenta, sendo que o menor e o maior valor são encontrados quando $\beta = 1$ e $\beta = 4$, respectivamente.

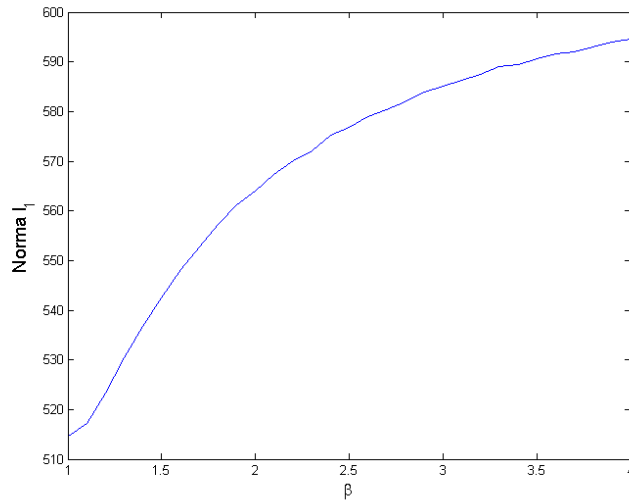


Fig. 7.4 – Valor da norma ℓ_1 de vetores com potência unitária, gerados utilizando (7.10) para $1 < \beta < 4$.

Este exemplo mostra que uma otimização baseada na minimização da norma ℓ_1 levará ao sinal mais super-gaussiano, e, conseqüentemente, ao sinal mais esparsos.

Caso Específico de Sinais Sub-Gaussianos

Apesar do fato de a abordagem proposta e a baseada em negentropia usarem conceitos bem diferentes, mostramos que a primeira pode ser vista como um caso especial da última. Visto que a negentropia também pode ser utilizada com sinais sub-Gaussianos, é esperado que a abordagem proposta também consiga bons resultados com estes tipos de fontes. Porém, em contraste com o que foi feito até agora, o processo de otimização deve ser feito através da *maximização* da norma ℓ_1 .

Com o intuito de confirmar esta suposição, consideremos como exemplo um sistema com duas fontes e duas misturas que, neste caso, serão geradas a partir de distribuições uniformes. A escolha da distribuição uniforme não se deve apenas ao fato de, como dito, ela ser sub-gaussiana, mas também porque quando usamos a gaussiana generalizada, esta distribuição é atingida quando $\beta \rightarrow \infty$, ou seja, exatamente no extremo

oposto das fontes esparsas. A Fig. 7.5 (a) representa a distribuição conjunta das amostras das fontes, que, neste caso, é delimitada por um quadrado centrado na origem.

Primeiramente, podemos notar, a partir da Eq. (7.4), que $\Psi_y \rightarrow nE\{|\mathbf{w}^T \mathbf{x}|\}$ para um valor grande de amostras n . Sem perda de generalidade, vamos considerar uma matriz de mistura ortonormal. Com o auxílio da Fig. 7.5 (a), $E\{|\mathbf{w}^T \mathbf{x}|\}$ pode ser calculada pela Eq. (7.12), no intervalo $0 \leq \theta \leq \pi/4$, onde Ω é a área da distribuição das amostras, sendo que a integração será feita sobre a área sombreada.

$$E\{|\mathbf{w}^T \mathbf{x}|\} = \frac{2}{\Omega} \iint x_1 \cos \theta + x_2 \sin \theta \, d\Omega = \frac{2}{3} \sin \theta \tan \theta + \frac{1}{3} \cos \theta (1 - \tan^2 \theta) \quad (7.12)$$

A partir da Fig. 7.5 (b), que mostra o resultado da Eq. (7.12), encontramos o máximo da função quando $\theta = 0$, ou seja, a direção de uma das colunas da matriz de mistura, que é relativa a uma das fontes existentes. Por outro lado, o mínimo ocorre quando $\theta = \pi/4$. Como já era esperado, a maximização da função custo Ψ_y implica a extração da fonte menos esparsa. Neste ponto, é interessante notar mais uma vez que tanto a interpretação de maximização da não-gaussianidade (dada pelo conceito de negentropia) quanto a de minimização da esparsidade (interpretação da abordagem proposta) dos sinais são totalmente aceitáveis.

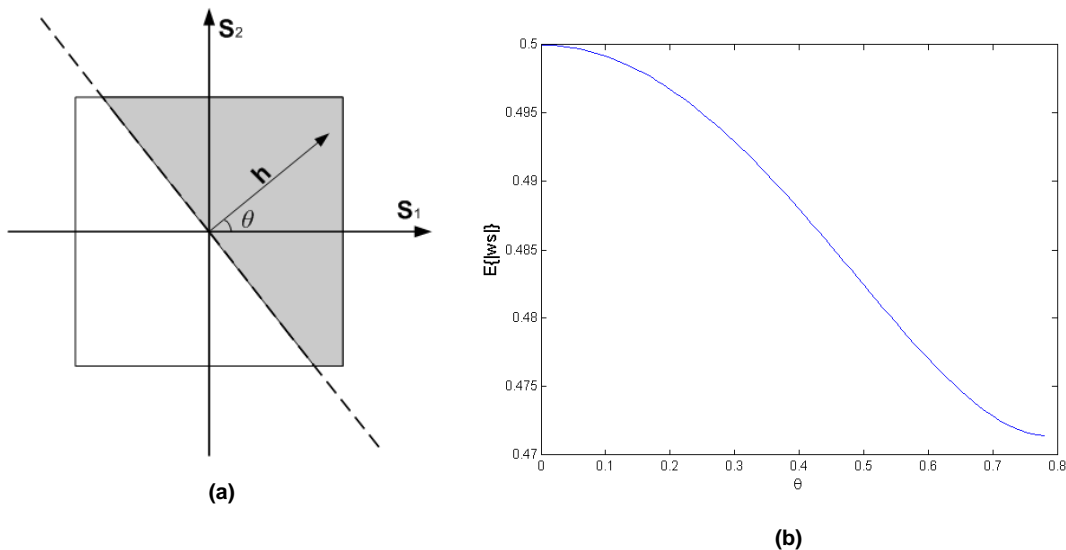


Fig. 7.5 – (a) O quadrado centrado na origem descreve o espaço amostral de uma variável randômica s , sendo que a área sombreada representa a região onde a integral dupla em (7.12) é calculada. (b) $E\{|\mathbf{w}^T \mathbf{x}|\}$ por θ .

7.2.3 Simulações e Resultados

Com o intuito de melhor compreender a abordagem proposta, são sugeridos três diferentes cenários.

Primeiro Cenário

Primeiramente, foram realizados testes em um cenário para o qual o número de fontes esparsas varia entre 5 e 100, a matriz de mistura \mathbf{A} é gerada de forma aleatória, e $M = N$. As fontes, cada uma possuindo 4000 amostras, foram geradas a partir de uma distribuição bernoulli-gaussiana [KORMYLO e MENDEL, 1982], com probabilidade de se ter um valor não-nulo variando entre 10 e 75%, de acordo com a Tab. 7.1. É importante notar que o aumento do número de fontes e/ou da probabilidade de uma amostra da fonte estar ativa faz com que a suposição de ortogonalidade disjunta seja cada vez menos plausível.

Foi realizado um conjunto de simulações com o objetivo de comparar a abordagem proposta em (7.8) com um método que minimiza um valor estimado de negentropia. Em ambos os casos, foi utilizado um procedimento de otimização baseado no gradiente. Para a aproximação da fdp na negentropia, foi utilizada como função não-quadrática [HYVÄRINEN *et al.*, 2001] a tangente hiperbólica. Foi calculado o valor da relação sinal-interferente (*signal-to-interference ratio* – SIR), dado em dB, entre cada sinal recuperado e a fonte original correspondente. Os resultados mostrados na Tab. 7.1 são uma média de 100 diferentes realizações.

% de prob.	5 fontes		10 fontes		50 fontes		100 fontes	
	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1
10%	45,81	54,40	40,93	51,20	33,50	46,62	30,17	45,56
25%	40,46	50,96	36,24	47,89	28,45	47,29	25,19	45,14
50%	34,67	49,22	30,36	47,32	22,48	45,42	18,82	43,69
75%	27,94	48,23	23,18	46,02	12,11	31,99	1,78	1,00

Tab. 7.1 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte, com diferentes graus de esparsidade. Os resultados são uma média de 100 diferentes realizações

A Tab. 7.1 mostra a presença de valores relativamente altos de SIR, o que indica que a otimização baseada na norma ℓ_1 pôde extrair fontes adequadamente na maioria dos casos, sendo estes resultados melhores quando os sinais são mais esparsos, como já era esperado. Visto que as fontes são independentes entre si, a otimização baseada em negentropia também obteve bons resultados, porém ligeiramente inferiores.

Nas simulações com 50 e 100 fontes, em que existia 75% de probabilidade de cada amostra estar ativa, o algoritmo baseado em negentropia não conseguiu extrair adequadamente nenhuma fonte. Entretanto, se o número de amostras utilizado for aumentando para 10.000, este problema não ocorre mais. Por outro lado, o algoritmo proposto continua a falhar no caso de 100 fontes e 75% de probabilidade, mesmo com o aumento da quantidade de amostras, o que pode ser atribuído a um grau insuficiente de esparsidade conjunta entre as fontes.

Dois motivos podem nos ajudar a compreender o fato de o algoritmo baseado na norma ℓ_1 gerar melhores resultados. O primeiro deles é justamente a relação entre negentropia e o método proposto: o algoritmo baseado na norma ℓ_1 pode ser visto como um caso especial do método baseado em negentropia, assumindo que as fontes possuem distribuição laplaciana, como explicado na Seção 7.2.1. Como as fontes são super-gaussianas, é esperado que o modelo de nossa proposta seja mais adequada aos dados testados.

Outra explicação é o fato de os mínimos da função custo baseados na norma ℓ_1 possuírem uma vizinhança caracterizada por um formato pontiagudo, como pode ser visto na Fig. 7.1 (para um caso sem ruído), o que potencialmente gera ótimos mais bem definidos. Desta forma, para um mesmo número de amostras, o aumento da quantidade de fontes acaba afetando menos a abordagem proposta do que a baseada em ICA.

Segundo Cenário

O segundo cenário é similar ao primeiro, porém, agora a probabilidade de cada amostra estar ativa permanece em 25% em todas as situações, e os instantes inativos não são mais associados a valores nulos, mas sim preenchidos com amostras geradas a partir de uma distribuição gaussiana cujo desvio padrão varia entre 5 e 30% do valor do desvio das amostras ativas. Para facilitar nossa explicação, a partir de agora, as amostras “inativas”

com valores não nulos serão chamadas de ruído. É importante notar que, nesta situação, a hipótese de ortogonalidade disjunta entre as fontes não é mais válida. Os resultados deste cenário se encontram na Tab. 7.2.

% de ruído	5 fontes		10 fontes		50 fontes		100 fontes	
	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1
5%	38,68	40,90	34,21	36,51	26,22	28,98	22,88	25,66
10%	34,52	33,65	30,50	29,60	22,43	21,86	19,20	18,80
15%	31,55	28,42	26,62	24,41	18,81	16,89	15,45	13,87
20%	25,94	24,31	22,91	19,98	15,02	12,49	11,48	9,56
25%	23,35	20,68	18,88	15,61	10,69	6,94	4,03	1,78
30%	17,94	14,34	14,07	9,83	3,18	1,25	0,53	0,45

Tab. 7.2 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte, com diferentes graus de ruído presente nos sinais. Os resultados são uma média de 100 diferentes realizações

Visto que o ruído adicionado às amostras inativas da bernoulli-gaussiana possui distribuição gaussiana, quanto maior for a potência deste ruído, maior será o grau de gaussianidade das fontes, causando dificuldades no processo de extração. Este fato pode levar à conclusão de que a abordagem proposta obteria resultados melhores do que os do algoritmo baseado em negentropia, visto que este último lida diretamente com a não-gaussianidade do sinal, enquanto o primeiro pode ser visto como uma medida da distância relativamente à laplaciana. Contudo, isto não foi observado, o que significa que a diminuição da esparsidade dos sinais foi o fator predominante.

Outro fato a observar é que a otimização baseada na norma ℓ_1 consegue realizar a extração em situações com 100 fontes e 10% de ruído nas amostras inativas. Isto reforça a ideia que não é necessário que exista ortogonalidade disjunta entre as fontes. Se interpretarmos este resultado de outra forma, podemos dizer que, nesta situação, a norma ℓ_1 faz antes o papel de medida de não-gaussianidade do que de aproximação da norma ℓ_0 , explicada na Seção 3.2.1.

Terceiro Cenário

Na Seção 7.2.2, são mostradas análises teóricas que sugerem a utilização da norma ℓ_1 para fontes sub-gaussianas. Para verificar esta possibilidade, é proposto um terceiro

cenário, utilizando agora fontes com distribuição uniforme. Todas as outras definições são mantidas idênticas às adotadas no primeiro cenário. Os resultados das simulações para este caso são mostrados na Tab. 7.3.

	5 fontes		10 fontes		50 fontes		100 fontes	
	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1	Neg.	Norma ℓ_1
Distribuição uniforme	32,24	26,13	28,48	22,39	18,91	2,05	0,99	0,42

Tab. 7.3 – Resultados de simulação da SIR (em dB) associada com a extração de uma fonte. As fontes são geradas com distribuição uniforme. Os resultados são uma média de 100 diferentes realizações

Como esperado, a estratégia de maximização da norma ℓ_1 é capaz de lidar com fontes geradas com distribuição uniforme. Também era esperado que os resultados não fossem tão bons quanto os encontrados pelo método baseado em negentropia, visto que este último foi desenvolvido para abranger a extração de fontes sub-gaussianas. Este resultado parece corroborar a idéia de que a norma ℓ_1 pode ser usada também como medida de sub-gaussianidade, mas sua aplicação é nitidamente mais adequada a fontes esparsas.

Por último, as fontes são geradas através da Eq. (7.10), com $1 < \beta < 4$. O cenário possui duas fontes e duas misturas, e a matriz de mistura é ortonormal. A Fig. 7.6 mostra o resultado do valor de SIR para os casos de minimização (linha contínua) e maximização (linha tracejada) da norma ℓ_1 . A extração é feita corretamente utilizando um processo de minimização para $\beta < 1.9$ e de maximização quando $\beta > 2.1$. Entre estes dois valores, a extração não é feita de forma correta. É importante salientar que as duas curvas se encontram exatamente quando $\beta = 2$, ou seja, quando a Eq. (7.10) representa uma distribuição gaussiana, o que dá suporte aos resultados mostrados na Seção 7.2.2.

7.3 Separação de Fontes Utilizando a Norma ℓ_1

Como mostrado na Seção 7.2, se todas as fontes são esparsas e $M = N$, cada mínimo da função Ψ_y , Eq. (7.4), corresponde a um vetor \mathbf{w} que extrai uma das fontes. Também foi mostrado que é possível extrair uma das fontes através da minimização da norma ℓ_1 dos sinais. Dando sequência à nossa análise da utilização da norma ℓ_1 como função custo de esparsidade, será proposta agora uma estratégia para separar simultaneamente todas as fontes, em vez de realizar a extração de um único sinal.

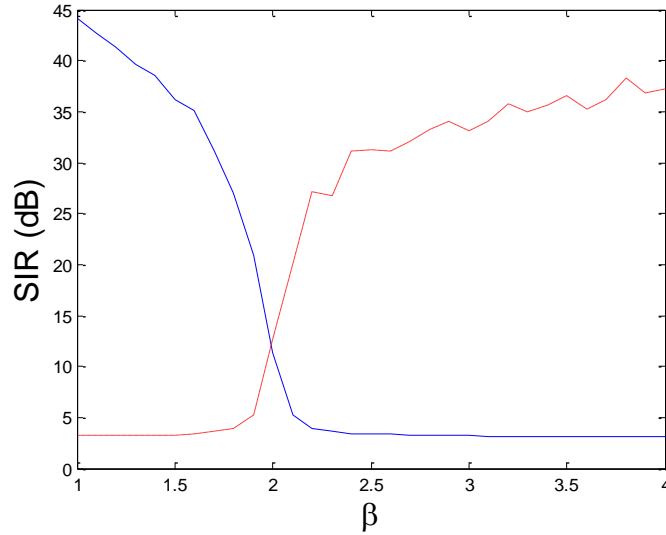


Fig. 7.6 – SIR das fontes extraídas por β em um cenário com duas fontes e duas misturas. As fontes foram geradas utilizando (7.10) para $1 < \beta < 4$.

Tendo por base as Eqs. (2.3) e (7.4), é possível determinar uma Ψ_{y_i} para cada vetor y_i .

$$\Psi_{y_i} = \sum_n |y_i(n)| = \sum_n |\mathbf{w}_i^T \mathbf{x}(n)| = \sum_n |\mathbf{w}_i^T \mathbf{A} \mathbf{s}(n)| = \sum_n \left| \sum_j \mathbf{w}_i^T \mathbf{a}_j s_j(n) \right| \quad (7.13)$$

Aqui, faremos novamente uma análise para o caso de fontes com ortogonalidade disjunta, e, em seguida, faremos um estudo sobre os outros casos. Sendo considerada a ortogonalidade disjunta entre as fontes, o mínimo valor de Ψ_{y_i} é dado quando \mathbf{w}_i é ortogonal a um dos vetores \mathbf{a}_i , como mostrado na Seção 7.2. Neste caso, todos os mínimos de Ψ_{y_i} são encontrados quando \mathbf{W} tem o formato dado por (2.8), $\mathbf{W}^T = \mathbf{PDA}^T$. Proporemos, então, uma função custo J_{SEP} para separar todas as fontes simultaneamente. Esta função será dada pela soma de todos os Ψ_{y_i} , ou seja:

$$\begin{aligned} J_{SEP} &= \sum_i \Psi_{y_i} = \sum_i \sum_n \left| \sum_j \mathbf{w}_i^T \mathbf{a}_j s_j(n) \right| = \sum_i \sum_j |\mathbf{w}_i^T \mathbf{a}_j| \sum_n |s_j(n)| \\ &= \sum_i \sum_j |\mathbf{w}_i^T \mathbf{a}_j| \Psi_{s_j}. \end{aligned} \quad (7.14)$$

Para realizar a análise, vamos assumir primeiramente que somente duas das fontes não são recuperadas. Deste modo, existem dois sinais de $\mathbf{y}(n)$ que serão combinações lineares dos sinais das fontes. Como estamos considerando a hipótese de ortogonalidade disjunta entre as fontes, isto garante que não existe sobreposição entre suas amostras, e, conseqüentemente, não existe interferência destrutiva quando os sinais das fontes passam pelo sistema misturador. Visto que os valores de Ψ_{sj} permanecem constantes, qualquer combinação de $|\mathbf{w}_i^T \mathbf{a}_j| \Psi_{sj}$ gerará um valor de J_{SEP} maior que o mínimo valor de Ψ_{sj} , a menos do ganho dado por \mathbf{A} . Considerando que os sinais das misturas são branqueados e \mathbf{W} é ortonormal, ao minimizarmos J_{SEP} , estaremos estimando simultaneamente todas as fontes. Isto pode ser provado da seguinte forma.

Afirmção: Se \mathbf{A} é uma matriz ortogonal, \mathbf{W} é uma matriz ortonormal e \mathbf{s} possui ortogonalidade disjunta, J_{SEP} , descrita pela Eq. (7.14), possui seus mínimos locais quando \mathbf{W} separa as fontes.

Prova: podemos reescrever (7.14) como sendo

$$J_{SEP} = \sum_i \sum_j |\mathbf{w}_i^T \mathbf{a}_j| \Psi_{sj} = \sum_i \sum_j \Psi_{sj} \|\mathbf{a}_j\|_2 |\cos \theta_{ij}|, \quad (7.15)$$

onde θ_{ij} é o ângulo entre \mathbf{w}_i e \mathbf{a}_i , e $0 < \theta_{ij} < \pi/2$ sem perda de generalidade. No caso bidimensional, temos:

$$J_{SEP} = \Psi_{s_1} \|\mathbf{a}_1\|_2 (|\cos \theta_{11}| + |\cos \theta_{21}|) + \Psi_{s_2} \|\mathbf{a}_2\|_2 (|\cos \theta_{12}| + |\cos \theta_{22}|). \quad (7.16)$$

De acordo com a Fig. 7.7, podemos definir $\theta_{11} = \theta_{22} = \theta$. Desta forma, teremos $\theta_{12} = \pi/2 - \theta$ e $\theta_{21} = \pi/2 + \theta$. Poderemos, portanto, reescrever (7.19) como sendo:

$$J_{SEP} = \Psi_{s_1} \|\mathbf{a}_1\|_2 (|\cos \theta| + |\cos(\pi/2 + \theta)|) + \Psi_{s_2} \|\mathbf{a}_2\|_2 (|\cos \theta| + |\cos(\pi/2 - \theta)|). \quad (7.17)$$

Pode ser observado que J_{SEP} não pode atingir seu mínimo se $\theta < \pi/2$ porque, neste caso, $|\cos \theta| > 0$ e $|\cos(\pi/2 \pm \theta)| > 0$, e, portanto, J_{SEP} se torna uma soma não-negativa de funções côncavas. Visto que J_{SEP} é bem definida no intervalo especificado, o mínimo deve estar em $\pi/2$, e a afirmação é válida para o caso bidimensional.

Para o caso N -dimensional, é possível observar que um mínimo local não pode ser atingido se $\theta_{ij} \neq \pi/2$ para todo i, j , porque, neste caso, $|\cos \theta_{ij}| > 0$ e J_{SEP} se tornar uma soma não-negativa de funções côncavas. Se $\theta_{ij} = \pi/2$ para um par i, j , \mathbf{w}_i é ortogonal a todos menos um vetor \mathbf{a}_i , visto que \mathbf{W} e \mathbf{A} são matrizes ortogonais. Se fixarmos \mathbf{w}_i para este

valor, o problema deve ser reiniciado em $N - 1$ dimensões. Admitindo que esta propriedade também é válida para $N - 1$ dimensões, provamos nossa afirmação por indução.

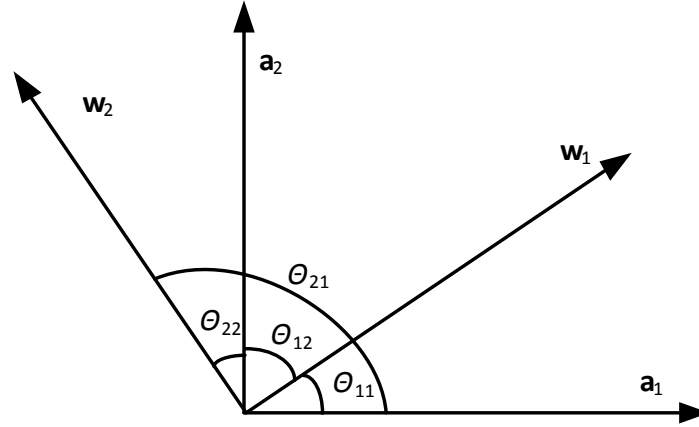


Fig. 7.7 – Arranjo dos ângulos entre os vetores de mistura e os de separação.

7.3.1 Método de Separação Baseado no Gradiente

Analogamente à derivação proposta na Seção 7.2.1, iremos propor um método de otimização baseado no gradiente de J_{SEP} . Esse gradiente é dado por

$$\nabla J_{SEP}(\mathbf{W}) = \sum_n \text{sign}[\mathbf{y}(n)] \tilde{\mathbf{x}}^T(n), \quad (7.18)$$

onde, novamente, $\tilde{\mathbf{x}}(n)$ representa a mistura branqueada. A adaptação de \mathbf{W} é feita por¹⁰

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu \nabla J_{SEP}(\mathbf{W}(k)) \quad (7.19)$$

$$\mathbf{W}(k+1) = \mathbf{V} \mathbf{D}^{-1/2} \mathbf{V}^T \mathbf{W}(k+1), \quad (7.20)$$

onde $\mu > 0$ é o passo de adaptação, \mathbf{V} é a matriz dos autovetores de $\mathbf{W}(k) \mathbf{W}^T(k)$, e \mathbf{D} é uma matriz diagonal contendo seus autovalores.

7.3.2 Separação de Fontes sem Ortogonalidade Disjunta

Se o processo de extração baseado na norma ℓ_1 possui relações com o método de ICA baseado em negentropia, é esperado que a abordagem proposta para separação de todas as fontes também possua uma relação com algum outro método de ICA. De fato, é interessante notar que o procedimento de minimização análogo ao mostrado em (7.18) e

¹⁰ Uma outra forma de se fazer esta adaptação é através da utilização do gradiente natural [CICHOCKI e AMARI., 2002], que obtém melhores resultados quando a matriz precisa ser ortogonalizada a cada iteração. Nesse caso, a Eq. (7.19) poderia ser reescrita como sendo $\mathbf{W}(k+1) = \mathbf{W}(k) - \mu \nabla J_{SEP}(\mathbf{W}(k)) \mathbf{W}(k) \mathbf{W}^T(k)$.

(7.19) pode ser obtido no contexto da metodologia de máxima verossimilhança considerando um modelo de mistura sem ruído.

Consideremos a função logarítmica da verossimilhança dada por [CARDOSO, 1998]

$$\frac{1}{T} \log p(\mathbf{X}_{1:T}; \mathbf{A}, q) = \frac{1}{T} \sum_{t=1}^T \log q(\mathbf{A}^{-1} \mathbf{x}(t)) - \log |\det \mathbf{A}|, \quad (7.21)$$

onde $q(\cdot)$ corresponde à distribuição conjunta das fontes. Em um cenário verdadeiramente cego, nenhuma suposição sobre a forma detalhada de $q(\cdot)$ poderia ser feita. Não obstante, ainda é possível realizar a separação das fontes quando a distribuição de $q(\cdot)$ adotada difere da distribuição real das fontes: esta é uma ideia importante para a aplicação do princípio Infomax [BELL e SEJNOWSKI, 1995].

Quando o modelo de mistura se restringe ao grupo de matrizes ortogonais, o último termo de (7.21) desaparece. Além disto, assumindo que $q(\cdot)$ possa ser fatorada, i.e., que as fontes são estatisticamente independentes, então a verossimilhança logarítmica em (7.21) pode ser reescrita como

$$\log L(\mathbf{W}) = \sum_{t=1}^T \sum_{i=1}^n \log p_i(\mathbf{w}_i^T \mathbf{x}(t)) \quad (7.22)$$

Novamente, podemos assumir uma distribuição laplaciana de média nula como aproximação para a fdp de um sinal esparso. Desta forma, se calcularmos o gradiente desta expressão, ele será dado por

$$\nabla L(\mathbf{W}) = - \sum_n \text{sign}[\mathbf{y}(n)] \mathbf{x}^T(n) \quad (7.23)$$

e, deste modo, o algoritmo baseado em gradiente para maximizar a verossimilhança será dado por

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu \nabla L(\mathbf{W}(k)) = \mathbf{W}(k) - \mu \sum_n \text{sign}[\mathbf{y}(n)] \mathbf{x}^T(n) \quad (7.24)$$

A cada iteração, o passo de ortogonalização, dado pela Eq. (7.20), deve ser aplicado para manter \mathbf{W} coerente com as hipóteses feitas.

Observando as Eqs. (7.18) e (7.23), torna-se claro que o desenvolvimento da norma ℓ_1 para separação de fontes é equivalente à formulação de máxima verossimilhança em que o modelo das fontes é dado por distribuições laplacianas. Apesar de a relação entre o uso da

norma ℓ_1 e a estimação de máxima verossimilhança com distribuição laplaciana ser bem conhecida (pode-se citar [BOFILL e ZIBULEVSKY, 2000] como exemplo para o caso subdeterminado), o presente desenvolvimento foi feito a partir de uma abordagem determinística baseada na minimização da norma ℓ_1 como medida de esparsidade, a qual interessantemente levou a um caso especial da abordagem estocástica baseada em máxima verossimilhança.

Uma consequência é que, assim como ocorre no processo de extração, quando é aplicado o algoritmo de separação a fontes esparsas, é possível encontrar bons resultados, mesmo que as fontes não possuam ortogonalidade disjunta.

7.3.3 Simulações e Resultados

Os três cenários descritos na Seção 7.2.3 foram novamente empregados, com a diferença que, agora, o problema consiste da separação de todas as fontes de forma simultânea. A abordagem proposta foi comparada com a clássica metodologia Infomax – que, como mostrado por Cardoso [CARDOSO, 1997], é equivalente à formulação de máxima verossimilhança – e os resultados podem ser vistos nas Tabs. 7.4, 7.5 e 7.6.

Os resultados confirmam a viabilidade da utilização da norma ℓ_1 como função custo para separação de fontes. É interessante notar que, quando a probabilidade de a amostra ser ativa aumenta, os valores de SIR parecem indicar maior robustez do método de separação, em comparação com o caso de extração (pode-se verificar isto comparando as Tabs. 7.1 e 7.4). Isto pode ser uma consequência do fato de que, em certo sentido, os múltiplos problemas de extração são resolvidos conjuntamente, sendo que qualquer melhora na solução atingida por uma das estimativas tem o potencial de afetar as soluções restantes.

% de prob.	5 fontes		10 fontes		50 fontes		100 fontes	
	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1
10%	36,94	38,32	32,72	33,81	24,89	25,34	21,79	22,16
25%	36,20	38,17	32,00	33,69	24,28	25,37	21,13	22,12
50%	34,41	38,47	30,15	33,64	22,32	25,31	19,07	22,00
75%	29,70	38,03	25,39	33,50	0,62	25,02	0,34	0,51

Tab. 7.4 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes, com diferentes graus de esparsidade. Os resultados são uma média de 100 diferentes realizações

% de ruído	5 fontes		10 fontes		50 fontes		100 fontes	
	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1
5%	35,86	36,44	32,18	31,57	23,73	24,30	20,61	21,18
10%	34,49	34,27	30,02	29,65	22,32	21,98	19,15	18,82
15%	32,56	30,90	27,99	26,49	20,12	18,71	16,93	15,56
20%	29,40	26,77	25,20	22,57	17,20	14,87	3,90	11,72
25%	25,91	22,48	21,23	17,98	13,37	10,21	0,34	0,68
30%	20,87	17,06	16,66	12,54	0,62	1,01	0,33	0,41

Tab. 7.5 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes, com diferentes graus de ruído presente nos sinais. Os resultados são uma média de 100 diferentes realizações

	5 fontes		10 fontes		50 fontes		100 fontes	
	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	Norma ℓ_1	Infomax	ℓ_1 norm
Distribuição uniforme	33,58	28,99	29,16	24,49	0,65	1,23	0,34	0,39

Tab. 7.6 – Resultados de simulação da SIR (em dB) associada com a separação de todas as fontes. As fontes são geradas com distribuição uniforme. Os resultados são uma média de 100 diferentes realizações

7.4 Estimação da Matriz de Mistura no Caso Subparametrizado

Para concluir nossa análise sobre a utilização da norma ℓ_1 nos problemas de BSE e BSS, iremos agora verificar sua utilização como função custo para a estimação da matriz de mistura no caso em que o número de fontes é maior do que o de sensores. Neste tipo de problema subparametrizado, a obtenção de um vetor que minimiza a norma ℓ_1 dos sinais estimados não implica a recuperação perfeita das fontes. Entretanto, é interessante notar que a informação trazida por todos os vetores encontrados pode ser ainda utilizada para a estimação da matriz de mistura. O ponto principal é que os vetores que minimizam a norma ℓ_1 são ortogonais ao subespaço gerado pelas colunas da matriz de mistura. Com o intuito de verificar este fato, consideremos um cenário simples com três sensores e no máximo duas fontes simultaneamente ativas.

Na Fig. 7.8, é apresentada a norma ℓ_1 da saída do sinal \mathbf{y} com respeito aos parâmetros de \mathbf{w} , considerando que existem, na média, duas fontes ativas em cada instante de tempo. É possível notar certos vales na superfície de custo, que correspondem aos vetores \mathbf{w} que são ortogonais a uma das colunas da matriz de mistura. Portanto, na intersecção de tais vales, teremos vetores \mathbf{w} que serão ortogonais a subespaços gerados por pares de colunas da matriz de mistura \mathbf{A} , como explicado anteriormente.

É importante notar que encontrar estes vetores \mathbf{w} é equivalente a encontrar os subespaços de concentração, como definido em [NAINI *et al.*, 2008], que fornecem as informações necessárias para a estimação de \mathbf{A} , processo este descrito na Seção 3.2.2. Desta forma, a diferença frente ao método proposto em [NAINI *et al.*, 2008] seria por conta da troca da função custo descrita pela Eq. (3.13) pela proposta em (7.4). O método de otimização pode ser tanto o descrito em [NAINI *et al.*, 2008] (também discutido na Seção 3.2.2), quanto a abordagem proposta no Capítulo 6.

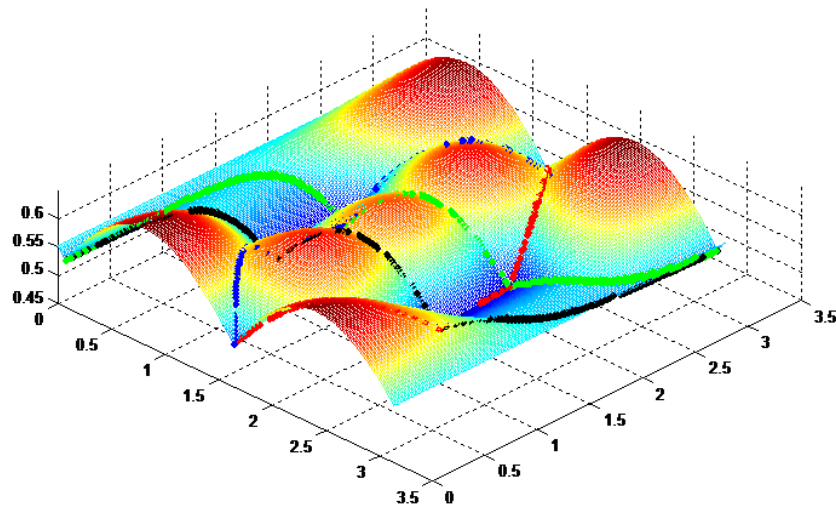


Fig. 7.8 – Função custo da norma ℓ_1 em um cenário subdeterminado

7.5 Considerações Finais

Neste capítulo, foi proposta a utilização da norma ℓ_1 como base para um critério de recuperação de fontes esparsas. Foi mostrado que, no caso de haver ortogonalidade disjunta entre as fontes, essa norma origina uma função custo tanto para extração quanto para separação de fontes no caso em que $N = M$. Quando a ortogonalidade disjunta não ocorre, foi mostrado que a otimização baseada na norma ℓ_1 pode ser considerada um caso particular de uma estratégia baseada em entropia, desde que se assuma uma fdp laplaciana e independência entre as fontes. Ademais, caso o problema seja subparametrizado, é possível estimar o número de fontes e a matriz de mistura, em uma abordagem similar à proposta no Capítulo 6.

Capítulo 8

Conclusões e Perspectivas

Em processamento não-supervisionado de sinais, uma das áreas que mais tem crescido é o estudo do problema de separação cega de fontes. Devido a sua grande complexidade, esse problema vem sendo tratado no âmbito de diversas formulações teóricas, sendo uma delas baseada em sistemas de mistura subparametrizados. É justamente nesta linha de pesquisa que se encaixa o trabalho apresentado.

A primeira contribuição desta tese, relatada no Capítulo 4, está relacionada à quantificação do erro intrínseco a soluções baseadas em sistemas lineares, o que abre perspectivas analíticas particularmente relevantes para o estudo de problemas subparametrizados. Essa quantificação toma a forma de um limitante inferior calculado a partir do erro quadrático médio (MSE) entre as fontes originais e as recuperadas: fundamentalmente, a soma dos valores de MSE associados à recuperação das fontes será sempre dada pela diferença entre o número de fontes e o de misturas, $N - M$. Desse limitante são obtidos resultados que ilustram o efeito do aumento do número de misturas / coeficientes do equalizador sobre a qualidade da recuperação dos sinais de interesse e também é apresentada uma nova perspectiva para o entendimento do efeito do ruído aditivo.

No Capítulo 5, foi estudada uma abordagem híbrida, utilizando ambas as suposições de esparsidade e independência, do problema de BSS no caso de sistemas de mistura subparametrizados. Foi proposto também um algoritmo para estimação do número de fontes e do sistema misturador no caso em que $N = 2$. Os resultados obtidos são bastante promissores, principalmente quando há fontes não esparsas na mistura, situação que não é contemplada por técnicas tradicionais de SCA.

A contribuição dada no Capítulo 6 é baseada na aplicação da *opt-aiNet*, ferramenta bioinspirada de otimização multimodal, ao problema de estimação da matriz de mistura e determinação do número de fontes existentes no âmbito de busca por subespaços baseada em uma métrica ℓ_0 suave. Uma vez que a solução do problema requer a busca simultânea de vários ótimos, há uma adequação conceitual significativa no que se refere ao uso da referida rede imunológica, adequação esta que é suportada pelos resultados obtidos.

O Capítulo 7 propõe a utilização de uma função baseada na norma ℓ_1 como contraste para o uso de esparsidade no problema de BSS. Diferentemente de outras abordagens existentes, a otimização é feita a partir da minimização da norma ℓ_1 do sinal estimado pelo sistema de extração / separação, considerando que tanto as fontes quanto as misturas são desconhecidas. A partir daí, foram derivados métodos aplicáveis tanto quando o sistema possui um mesmo número de fontes e sensores, quanto para a estimação da matriz de mistura em casos subparametrizados.

Como perspectivas de trabalhos futuros, vislumbramos, de imediato, as seguintes possibilidades:

- Estender a aplicação da metodologia híbrida exposta no Capítulo 5 a cenários com um maior número de misturas e com misturas convolutivas.
- Utilização da metodologia de busca multimodal bioinspirada discutida no Capítulo 6 com funções custo baseadas em métricas distintas da ℓ_0 suave.
- Aprofundar a análise das relações teóricas entre o contraste baseado na norma ℓ_1 e grandezas derivadas da teoria da informação, especialmente a entropia. Consideramos que esse estudo pode levar à construção de métricas ainda mais adequadas ao tratamento de sinais esparsos.
- Relacionar de maneira direta as abordagens propostas à recuperação / reconstrução efetiva das fontes no caso de sistemas subparametrizados.

Bibliografia

- ABRARD, F.; DEVILLE, Y. **Blind Separation of Dependent Sources Using the “Time-frequency Ratio of Mixtures” Approach**. Proceedings of ISSPA 2003. França: [s.n.], 2003.
- ABRARD, F.; DEVILLE, Y.; WHITE, P. From Blind Source Separation to Blind Source Cancellation in the Underdetermined Case: A New Approach Based on Time-frequency Analysis. **Proceedings of ICA 2001**, EUA, 2001. 734-739.
- ADA, G. L.; NOSSAL, G. J. V. The Clonal Selection Theory. **Scientific American**, v. 257, n. 2, p. 50-57, 1987.
- AISSA-EL-BEY, A.; LINH-TRUNG, K.; ABED-MERAIM, K.; BELOUHRANI, A.; GRENIER, Y. Underdetermined Blind Separation of Nnon-disjoint Sources in the Time-frequency Domain. **IEEE Transactions on Signal Processing**, v. 52, n. 7, p. 1830-1847, 2004.
- ARBERET, S.; GRIBONVAL, R.; BIMBOT, F. **A Robust Method to Count and Locate Audio Sources in a Stereophonic Linear Instantaneous Mixture**. Proceedings of ICA2006. EUA: [s.n.], 2006. p. 536-543.
- ATTUX, R. R. F. **Novos Paradigmas para Equalização e Identificação de Canais Baseados em Estruturas Não-lineares e Algoritmos evolutivos**. Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas. Campinas. 2005.
- ATTUX, R.; LOIOLA, M. B.; SUYAMA, R.; DE CASTRO, L. N.; VON ZUBEN, F. J.; ROMANO, J. M. T. Blind Search for Optimal Wiener Equalizers Using an Artificial Immune Network Model. **EURASIP Journal on Applied Signal Processing**, v. 2003, n. 8, p. 740-747, 2003.
- BELL, A.; SEJNOWSKI, T. An Information-Maximization Approach to Blind Separation. **Neural Computation**, v. 7, p. 1004-1034, 1995.

- BOFILL, P.; ZIBULEVSKY, M. **Blind Separation of More Sources than Mixtures Using Sparsity of their Short-time Fourier Transform**. Proceedings of ICA2000. Helsinki, Finlândia: [s.n.]. 2000. p. 87-92.
- BOYD, S.; VANDENBERGHE, L. **Convex Optimization**. Cambridge: Cambridge University Press, 2004.
- BURNET, F. M. **The Clonal Selection Theory of Acquired Immunity**. EUA: Vanderbilt University Press, 1959.
- CANDÈS, E. J.; ROMBERG, T.; TAO, T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. **IEEE Transactions on Information Theory**, v. 52, n. 2, p. 489-509, 2006.
- CANDÈS, E. J.; TAO, T. Decoding by Linear Programming. **IEEE Transactions on Information Theory**, v. 51, n. 12, p. 4203-4215, 2005.
- CARDOSO, J. Blind Signal Separation: Statistical Principles. **Proceedings of the IEEE**, v. 86, p. 2009-2025, 1998.
- CARDOSO, J. F. Infomax and Maximum Likelihood for Source Separation. **IEEE Signal Processing Letters**, v. 4, p. 112-114, 1997.
- CHAVES, F. D. S.; FERNANDES, F. G.; ATTUX, R.; SUYAMA, R.; LOPES, R. R.; ROMANO, J. M. T. **Une Analyse du Problème de l'Égalisation Basée sur la Théorie des Jeux**. Actes du Colloque GRETSI. Troyes: [s.n.]. 2007. p. 941-944.
- CHEN, S. S.; DONOHO, D. L. Atomic Decomposition by Basis Pursuit. **SIAM Journal on Scientific Computing**, v. 20, n. 1, p. 33-61, 1998.
- CHINRUNGRUENG, C.; SEQUIN, C. Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate. **IEEE Transactions on Neural Networks**, v. 6, n. 1, p. 157-169, 1995.
- CICHOCKI, A.; AMARI, S.-I. **Adaptive Blind Signal and Image Processing - Learning Algorithms and Applications**. Chichester: John Wiley & Sons, 2002.
- COMON, P. Independent Component Analysis, a New Concept? **Signal Processing**, v. 36, p. 287-314, 1994.
- COMON, P.; JUTTEN, C. (Eds.). **Handbook of Blind Source Separation - Independent Component Analysis and Applications**. Burlington: Academic Press, 2010.

-
- COMON, P.; MOREAU, E. **Improved Contrast Dedicated to Blind Separation in Communications**. Proc. ICASSP 1997. Munique: [s.n.]. 1997. p. 20-24.
- COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2^a. ed. Nova Jersey: John Wiley & Sons, 2006.
- DALTON, H. The Measurement of the Inequity of Incomes. **Econom. J.**, v. 30, p. 348, 1920.
- DE CASTRO, L. N. **Fundamentals of Natural Computing - Basic Concepts, Algorithms, and Applications**. Boca Raton: Chapman & Hall/CRC, 2006.
- DE CASTRO, L. N.; TIMMIS, J. **An Artificial Immune Network for Multimodal Function Optimization**. IEEE International Conference on Evolutionary Computation. EUA: [s.n.]. 2002. p. 674-699.
- DELFOSSÉ, N.; LOUBATON, P. Adaptive Blind Separation of Independent Sources: A Deflation Approach. **Signal Processing**, v. 45, p. 59-83, 1995.
- DING, Z.; KENNEDY, R. A. On the Whereabouts of Local Minima for Blind Adaptive Equalizers. **IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing**, v. 39, n. 2, p. 119-123, 1992.
- DONOHU, D. L. Compressed Sensing. **IEEE Transactions on Information Theory**, v. 52, n. 2, p. 1289-1306, 2006.
- DONOHU, D. L.; HUO, X. Uncertainty Principles and Ideal Atomic Decomposition. **IEEE Transactions on Information Theory**, v. 47, n. 7, p. 2845-2862, 2001.
- FEVOTTE, C.; GODSILL, S. J. A Bayesian Approach for Blind Separation of Sparse Sources. **IEEE Transactions on Audio, Speech and Language Processing**, v. 14, n. 6, p. 2174-2188, 2006.
- GU, M.; TONG, L. Geometrical Characterization of Constant Modulus Receivers. **IEEE Transactions on Signal Processing**, v. 47, n. 10, p. 2745-2756, 1999.
- HAYKIN, S. **Adaptive Filter Theory**. EUA: Prentice-Hall, 1995.
- HULLE, M. V. **Clustering Approach to Square and Non-square Blind Source Separation**. Proceedings of NNSP99. EUA: [s.n.]. 1999. p. 315-323.
- HURLEY, N.; RICKARD, S. Comparing Measures of Sparsity. **IEEE Transactions on Information Theory**, v. 55, n. 10, p. 4723-4721, 2009.

- HYVÄRINEN, A.; KARHUNEN, J.; OJA, E. **Independent Component Analysis**. Nova York: John Wiley & Sons, 2001.
- IEEE MLSP. **Data Analysis and Signal Processing Competition, 2005 IEEE International Workshop on Machine Learning for Signal Processing**, 2005. Disponível em: <<http://mlsp2005.conwiz.dk/index.php?id=30.html>>. Acesso em: 06 maio 2011.
- JACQUES, L.; VANDERGHEYNST, P. Compressed Sensing: 'When Sparsity Meets Sampling'. In: CRISTBAL, G.; SCHELKENS, P.; THIENPONT, H. **Optical and Digital Image Processing - Fundamental and Applications**. [S.l.]: Wiley-Blackwell, 2010.
- JERNE, N. K. Towards a Network Theory of the Immune System. **Ann. Immunol. (Inst. Pasteur)**, v. 125C, p. 373-389, 1974.
- KORMYLO, J. J.; MENDEL, J. M. Maximumlikelihood Detection and Estimation of Bernoulli-Gaussian Processes. **IEEE Transactions on Information Theory**, v. 28, p. 482-488, 1982.
- LI, Y.; DING, Z. Convergence Analysis of Finite Length Blind Adaptive Equalizers. **IEEE Transactions on Signal Processing**, v. 43, n. 9, p. 2120-2129, 1995.
- LUENBERGER, D. G. **Linear and Nonlinear Programming**. 2ª. ed. Norwell: Kluwer Academic Publishers, 2003.
- MAULIK, U.; BANDYOPADHYAY, S. Performance Evaluation of Some Clustering Algorithms and Validity Indices. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 12, p. 1650-1654, 2002.
- MENDEL, J. M. Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications. **Proceedings of the IEEE**, v. 79, n. 3, p. 278-305, Março 1991.
- MOHIMANI, H.; BADAIE-ZADEH, M.; JUTTEN, C. **Fast Sparse Representation based on Smoothed L0 norm**. Proceedings of ICA2007. Inglaterra: [s.n.]. 2007. p. 389-396.
- MONTALVÃO, J. R.; MOTA, J. C. M.; DORIZZI, B. Some Theoretical Limits of Efficiency of Linear and Nonlinear Equalizers. **Journal of the Brazilian Telecommunications Society**, v. 14, n. 2, p. 85-92, 1999.

-
- NADALIN, E. Z.; SUYAMA, R.; ATTUX, R. **An ICA-based Method for Blind Source Separation in Sparse Domains**. Proceedings of ICA2009. Paraty, Brasil: [s.n.]. 2009. p. 597-604.
- NAINI, F. M.; MOHIMANI, G. H.; BABAIE-ZADEH, M.; JUTTEN, C. Estimating the Mixing Matrix in Sparse Component Analysis (SCA) Based on Partial k-dimensional Subspace Clustering. **Neurocomputing**, v. 71, p. 2330-2343, 2008.
- NATARAJAN, B. K. Sparse Approximate Solutions to Linear Systems. **SIAM Journal on Computing**, v. 24, n. 2, p. 227-234, 1995.
- PAPOULIS, A. **Probability, Random Variables, and Stochastic Processes**. 3^a. ed. EUA: McGraw Hill, 1993.
- REGALIA, P. A.; MBOUP, M. Undermodeled Equalization: A Characterization of Stationary Points for a Family of Blind Criteria. **IEEE Transactions on Signal Processing**, v. 47, p. 760-770, 1999.
- RICKARD, S. **Sparse Sources are Separated Sources**. Proceedings of the 16th Annual European Signal Processing Conference. Itália: [s.n.]. 2006.
- RIOUL, O.; VETTERLI, M. Wavelets and Signal Processing. **IEEE Signal Processing Magazine**, v. 8, p. 14-38, 1991.
- ROMANO, J. M. T.; ATTUX, R.; CAVALCANTE, C. C.; SUYAMA, R. **Unsupervised Signal Processing - Channel Equalization and Source Separation**. Boca Raton: CRC Press, 2010.
- TREFETHEN, L. N.; BAU, D. **Numerical Linear Algebra**. Philadelphia: Siam, 1997.
- TROPP, J. A. Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise. **IEEE Transactions on Information Theory**, v. 52, n. 3, p. 1030-1051, 2006.