

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE COMUNICAÇÕES

Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário

Charles Casimiro Cavalcante

Tese de Doutorado

Orientador: Prof. Dr. João Marcos Travassos Romano

Co-Orientador: Prof. Dr. Francisco Rodrigo Porto Cavalcanti

Banca Examinadora:

Prof. Dr. João Cesar Moura Mota, (DETI, UFC)
Prof. Dr. Paulo Sérgio Ramirez Diniz, (COPPE, UFRJ)
Prof. Dr. Fernando José Von Zuben, (DCA, UNICAMP)
Prof. Dr. Paulo Cardieri, (DECOM, UNICAMP)
Prof. Dr. Reginaldo Palazzo Júnior, (DT, UNICAMP)
Prof. Dr. Sérgio Santos Mühlen, (DEB, UNICAMP)

*Tese apresentada à Faculdade de Engenharia Elétrica e de
Computação da Universidade Estadual de Campinas como
parte dos requisitos para obtenção do título de **Doutor em
Engenharia Elétrica***

Campinas, Abril de 2004

*Aos meus pais,
Casimiro e Ivoneide*

*A Erika
pelo amor*

*Ao meu filho Matheus
pela alegria*

“Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim.”

Chico Xavier

Agradecimentos

Muitas pessoas participaram ativamente, no campo profissional e pessoal, para tornar esta tese possível. Mais que justo é dedicar-lhes uma pequena parte do resultado de três anos em forma de agradecimentos.

Gostaria de agradecer profundamente a meus pais, **Casimiro e Ivoneide**, pelo apoio e amor incondicional durante toda a minha vida, me permitindo chegar bem mais longe do que um dia acreditei que poderia. Painho e Mainha, este trabalho é para vocês, por tudo que representam para mim.

Quero também agradecer muito a minha esposa, **Erika**, que compartilha comigo a paixão pela ciência, e que soube suportar a distância e as dificuldades e permanecer comigo para dividir este momento, compreendendo a importância dele nas nossas vidas. Agradeço a ela também, o milagre da vida, de nós dois agora sermos três, já que temos agora nosso **Matheus**, a quem também agradeço pela imensa alegria que trouxe às nossas vidas e de nossas famílias. Amo vocês e este amor tornou isso possível.

Ao meu orientador, amigo, colega de desafios e projetos, conselheiro e um pouco pai, **João Marcos T. Romano**. Quaisquer adjetivos usados seriam pouco para expressar minha profunda gratidão pela maneira muito segura que me orientou durante os três anos do meu doutorado, deixando-me muito à vontade para seguir minhas intuições e sabendo me trazer a portos mais seguros em momentos decisivos. Além disso, me proporcionou inúmeras oportunidades, as quais definiram com bastante precisão os caminhos que hoje escolho trilhar. Sentirei enorme saudade dos nossos cafés regados com ciência, política, arte, literatura e outros assuntos que sempre permearam nossos encontros.

Ao meu co-orientador **Rodrigo Cavalcanti**, pela amizade, interesse no trabalho e discussões técnicas sempre proveitosas que me permitiram enxergar aspectos práticos de implementação de técnicas de processamento de sinais. Também quero agradecer sua sempre calorosa acolhida no GTEL (Grupo de Pesquisa em Telecomunicações Sem Fio) permitindo que durante as temporadas em Fortaleza meu trabalho não fosse prejudicado.

Ao amigo e colega **João Cesar M. Mota**, pelo imenso interesse que sempre demonstrou com meu doutorado e pelas discussões técnicas muito proveitosas e agradáveis permitindo vislumbrar muitos pontos de continuidade do trabalho e ajudando a sanar dúvidas. Quero também lhe agradecer pela cuidadosa leitura das primeiras versões da tese.

Às excelentes sugestões e o rigor com notações do professor **Paulo Diniz**. Este texto não teria a mesma qualidade sem eles. Quero agradecê-lo profundamente pelo enorme esforço que dedicou à leitura da tese e por todos os seus comentários extremamente pertinentes que enriqueceram bastante o trabalho.

Ao professor **Fernando Von Zuben**, pelas valiosas sugestões que permitiram melhorar bastante a qualidade e o teor didático do texto.

Melhoria significativa do texto em relação às aplicações biomédicas e em comunicações são graças as intervenções precisas dos professores **Sérgio Mühlen** e **Paulo Cardieri**. Os aspectos de teoria da informação e o rigor matemático de algumas demonstrações encontraram a sua adequada expressão graças ao professor **Reginaldo Pallazo**. A todos eles quero agradecer fortemente por toda a atenção que tiveram com o trabalho desde seu início.

Agradecimento muito especial a todos os colegas do DSPCom (Laboratório de Processamento de Sinais para Comunicações) que durante toda a duração do trabalho foram sempre presença mais que marcante. A Aline, Cynthia, Laís, Eloísa, Celi (damas primeiro!), Romis, Danilo, Cristiano Panazio, Maurício Sol, Suyama, Rafael, Murilo, Renato, Moisés, Dayan, Gustavo, Tiago, Leonardo, Leandro Castro, Leandro Rangel, Tarciano, quero agradecer pelos rodízios de pizza, pelos jogos de *WAR* nas madrugadas, pela festa junina na casa da prima da Aline, pela viagem a Ubatuba, pelos almoços na Biologia e outras cantinas, pelas reuniões na casa da Cynthia, pelas noites tocando violão, pelos churrascos, pelas sessões de cinema e de discussão pós-filmes, pelos seminários e por tantos outros momentos bastante agradáveis que tornaram a época de minha estada em Campinas um período da minha vida que não será esquecido jamais. Além disso, agradeço também pelas discussões técnicas muito agradáveis que resultaram em várias partes contempladas nesta tese, particularmente ao Romis e Suyama pela ajuda e leitura do texto sobre equivalências entre critérios. A Eloísa e Celi quero agradecer particularmente pelo auxílio mais que necessário com questões administrativas bastante importantes. A vocês meu muito obrigado de coração!

A **Daniel** e **Fábio**, companheiros de república, que dividiram momentos alegres e tristes, faxinas e almoços de domingos bem como piadas e gozações sobre minha habilidade em futebol. Mais que amigos, ganhei dois irmãos muito queridos que tornaram as coisas “menos Alcione”. Quero também agradecer àqueles que estiveram presentes na vida cotidiana da vida estudantil tornando-a um pouco menos penosa em alguns aspectos: Divanilson, Raul, Yusef, Gustavo, Fabrício, Carlos Henrique, Romeu, André dentre os vários que com certeza não lembro agora para citar.

A **Josué**, meu amigo-irmão (ou irmão-amigo) desde sempre. Pela atenção com que ouviu minhas angústias, pelo apoio moral sempre presente, pelas novas piadas, pelos passeios em São Paulo, pelo carnaval em Socorro, por ter tirado algumas de minhas (muitas) dúvidas em estatística e probabilidade e por estar presente em todos os momentos meu muito obrigado.

A todos que compõem o GTEL pela amizade e apoio nos momentos em que compartilhamos o ambiente de trabalho. Ao Walter e ao André, em particular, quero agradecer pela ajuda e colaboração nos projetos em parceria com a Ericsson.

Aos meus irmãos **Kleber**, **Rogério** e **Marcelo**, pelo amor e apoio à distância.

A CAPES que financiou esta pesquisa, e ao povo brasileiro, que por meio de seus impostos, permite que pesquisa de qualidade seja desenvolvida no nosso país.

Terminar de escrever uma tese e tornar-se pai de primeira viagem não foi tarefa fácil, quero então também agradecer às famílias da minha esposa e minha, especialmente à **Xênia** e **D. Socorro**, pelas noites de vigília ao lado do berço do Matheus, permitindo-me um pouco mais de tranqüilidade numa etapa bastante crítica de finalização do documento.

A Deus, a quem tudo devo e que tem me proporcionado tantos momentos felizes na vida.

Charles Casimiro Cavalcante

Índice

Agradecimentos	i
Resumo	ix
Abstract	xi
Lista de Figuras	xiii
Lista de Tabelas	xvii
Notações e Abreviações	xix
1 Introdução	1
1.1 Recuperação da informação: motivação	2
1.1.1 Estatísticas de ordem superior	3
1.1.2 Teoria da informação	3
1.2 O problema de separação cega de fontes	4
1.2.1 <i>Cocktail party</i>	4
1.2.2 Processamento de sinais biomédicos	5
1.3 Processamento multi-usuário	8
1.4 Organização, estrutura e contribuições	10
I Separação Cega de Fontes e Processamento Multiusuário	13
2 Separação Cega de Fontes: Princípios e Conceitos	15
2.1 Um pouco de história...	16
2.2 Modelagem matemática: um pouco de formalismo	20
2.2.1 Quais estatísticas são necessárias para resolver o problema?	23
2.3 <i>Independent Component Analysis</i>	27
2.3.1 Independência estatística	28

2.3.2	Entropia e informação mútua	31
2.3.3	Divergência de Kullback-Leibler	32
2.3.4	Negentropia	34
2.3.5	<i>Kurtosis</i>	35
2.3.6	Funções de contraste	37
2.4	Conclusões e síntese	42
3	Separação Cega de Fontes: Estratégias e Aplicações	43
3.1	Principais estratégias de separação de fontes	44
3.1.1	<i>MaxEnt</i> e <i>InfoMax</i>	44
3.1.2	Máxima verossimilhança e abordagem bayesiana	46
3.1.3	Critério “universal”	51
3.2	Algoritmos	53
3.2.1	Algoritmo Jutten-Hérault e variantes	53
3.2.2	Equivariância e gradiente relativo/natural	55
3.2.3	<i>Deflation</i>	57
3.2.4	<i>FastICA</i>	58
3.2.5	Aprendizado anti-Hebbiano não-linear	60
3.2.6	Análise não-linear por componentes principais - <i>Nonlinear PCA</i>	62
3.3	Aplicações	65
3.3.1	Processamento de sinais de voz	65
3.3.2	Sinais de comunicação digital	68
3.3.3	Processamento de sinais biomédicos	71
3.3.4	Aquisição de sinais eletroencefalográficos	74
3.3.5	Extração de imagens	77
3.4	Conclusões e síntese	79
4	Processamento Multiusuário	81
4.1	Esquemas de múltiplo acesso	82
4.2	Antenas adaptativas	85
4.3	Caracterização da resposta do canal e do sistema de separação dos usuários	87
4.3.1	Processamento espacial	88
4.3.2	Processamento espaço-temporal	91
4.4	Caracterização das fontes de um sistema multiusuário	96
4.4.1	Fontes discretas	96
4.4.2	Alfabeto finito	97
4.4.3	Mesma distribuição estatística	97
4.5	Estratégias de separação de fontes em sistemas multiusuário	98
4.5.1	Solução ótima e algoritmos supervisionados	99
4.5.2	<i>Multiuser Constant Modulus Algorithm</i>	101
4.5.3	<i>Fast Multiuser Constant Modulus Algorithm</i>	103
4.5.4	<i>Multiuser Kurtosis Maximization</i>	106

4.5.5	Discussão e comentários	111
4.6	Conclusões e síntese	114
II Proposições e Análises		115
5	Métodos Multiusuário Baseados na Estimação de Densidade de Probabilidade	117
5.1	Revisão sobre método para equalização cega	118
5.2	Análise do método FP	123
5.2.1	Propriedades de convergência	123
5.2.2	Equivalência entre o critério FP e outros critérios	131
5.2.3	Alguns outros aspectos relevantes	137
5.3	Estimação de densidade de probabilidade	140
5.3.1	Estimação pelo método de Parzen	142
5.3.2	Critérios baseados em métodos de estimação de fdp	144
5.4	Família de critérios multiusuário baseados na estimação da fdp	148
5.4.1	Utilização da decorrelação explícita	149
5.4.2	Critério com restrição	153
5.5	Critério para minimização da probabilidade de erro	159
5.6	Conclusões e síntese	162
6	Análise da Influência das HOS em Algoritmos Adaptativos para Separação de Fontes	165
6.1	Motivação	166
6.2	Aproximações de funções de densidade de probabilidade	168
6.2.1	Expansão de Gram-Charlier	169
6.2.2	Expansão de Edgeworth	174
6.2.3	Comentários e considerações	174
6.2.4	Expansão em série em torno de distribuições de misturas gaussianas	177
6.3	Evolução temporal da estimativa da densidade de probabilidade	180
6.3.1	Comparação MUK \times MU-CFPA	182
6.4	Separação cega de fontes \times desconvolução cega no contexto de HOS	187
6.5	Conclusões e síntese	189
7	Aplicações em Sistemas de Mistura Instantânea e Convolutiva	191
7.1	Sistemas de mistura instantânea: processamento espacial	192
7.1.1	Receptor com sensores simples	192
7.1.2	Receptores com arranjos de antenas	196
7.1.3	Estimação do número de usuários	201
7.2	Resolução de colisão em redes de pacotes <i>Slotted ALOHA</i>	205
7.3	Sistemas de mistura convolutiva: processamento espaço-temporal	210

7.4	Conclusões e síntese	216
8	Conclusões, Sugestões e Perspectivas	217
8.1	Conclusões e contribuições	218
8.2	Sugestões e perspectivas para trabalhos futuros	221
III	Apêndices	225
A	Definições Matemáticas	227
A.1	Cumulantes e momentos	227
A.1.1	História	227
A.1.2	Cumulantes e momentos de distribuições de probabilidade	228
A.1.3	Algumas propriedades	229
A.1.4	Cumulantes e momentos	230
A.1.5	Cumulantes conjuntos	231
A.1.6	Cumulantes condicionais	231
A.2	Entropia de variáveis aleatórias	232
A.2.1	Definição de entropia	232
A.2.2	Distribuições com máxima entropia	232
A.2.3	Entropia de uma variável gaussiana: abordagem alternativa	234
B	Demonstrações e Derivações	237
B.1	Recuperação de sinais baseado no Teorema de Shalvi-Weinstein	237
B.2	Teorema de Benveniste-Goursat-Ruget	239
B.3	Ortogonalização de Gram-Schmidt	241
	Referências	243
	Índice Remissivo	261

Resumo

Esta tese é dedicada ao estudo de técnicas de separação cega de fontes aplicadas ao contexto de processamento multiusuário em comunicações digitais. Utilizando estratégias de estimação da função de densidade de probabilidade (fdp), são propostos dois métodos de processamento multiusuário que permitem recuperar os sinais transmitidos pela medida de similaridade de Kullback-Leibler entre a fdp dos sinais à saída do dispositivo de separação e um modelo paramétrico que contém as características dos sinais transmitidos. Além desta medida de similaridade, são empregados diferentes métodos que garantem a decorrelação entre as estimativas das fontes de tal forma que os sinais recuperados sejam provenientes de diferentes fontes. É ainda realizada a análise de convergência dos métodos e suas equivalências com técnicas clássicas resultando em algumas importantes relações entre critérios cegos e supervisionados, tais como o critério proposto e o critério de máxima *a posteriori*. Estes novos métodos aliam à capacidade de recuperação da informação uma baixa complexidade computacional. A proposição de métodos baseados na estimativa da fdp permitiu a realização de um estudo sobre o impacto das estatísticas de ordem superior em algoritmos adaptativos para separação cega de fontes. A utilização da expansão da fdp em séries ortonormais permite avaliar através dos cumulantes a dinâmica de um processo de separação de fontes. Para tratar com problemas de comunicação digital é proposta uma nova série ortonormal, desenvolvida em torno de uma função de densidade de probabilidade dada por um somatório de gaussianas. Esta série é utilizada para evidenciar as diferenças em relação ao desempenho em tempo real ao se reter mais estatísticas de ordem superior. Simulações computacionais são realizadas para evidenciar o desempenho das propostas frente a técnicas conhecidas da literatura em várias situações de necessidade de alguma estratégia de recuperação de sinais.

Palavras-chave: Separação cega de fontes, processamento multiusuário, estimação da fdp, estatísticas de ordem superior, expansão em séries ortonormais.

Abstract

This thesis is devoted to study blind source separation techniques applied to multiuser processing in digital communications. Using probability density function (pdf) estimation strategies, two multiuser processing methods are proposed. They aim for recovering transmitted signal by using the Kullback-Leibler similarity measure between the signals pdf and a parametric model that contains the signals characteristics. Besides the similarity measure, different methods are employed to guarantee the decorrelation of the sources estimates, providing that the recovered signals origin from different sources. The convergence analysis of the methods as well as their equivalences with classical techniques are presented, resulting on important relationships between blind and supervised criteria such as the proposal and the maximum a posteriori one. Those new methods have a good trade-off between the recovering ability and computational complexity. The proposal of pdf estimation-based methods had allowed the investigation on the impact of higher order statistics on adaptive algorithms for blind source separation. Using pdf orthonormal series expansion we are able to evaluate through cumulants the dynamics of a source separation process. To be able to deal with digital communication signals, a new orthonormal series expansion is proposed. Such expansion is developed in terms of a Gaussian mixture pdf. This new expansion is used to evaluate the differences in real time processing when we retain more higher order statistics. Computational simulations are carried out to stress the performance of the proposals, faced to well known techniques reported in the literature, under the situations where a recovering signal strategy is required.

Keywords: Blind source separation, multiuser processing, pdf estimation, higher order statistics, orthonormal series expansion.

Lista de Figuras

1.1	Ilustração do problema de <i>cocktail party</i>	5
1.2	Aquisição de sinais a partir de sensores na superfície corporal.	5
1.3	Ilustração do método de medida do eletrocardiograma fetal/materno a partir de sensores na superfície corporal.	7
1.4	Aquisição de (eletro/magneto)encefalograma.	8
1.5	Cenário típico de comunicações sem fio.	9
2.1	Esquema do processamento de separação cega de fontes.	21
2.2	Estrutura da matriz de autocorrelação de dados esferatizados.	25
2.3	Visão geométrica da diferença entre ICA e PCA.	29
2.4	Funções de distribuição de probabilidade conjuntas de variáveis gaussianas e uniformes em um sistema de mistura instantânea.	30
2.5	Exemplos de fdps gaussianas, sub-gaussianas e super-gaussianas.	36
2.6	Interpretação geométrica no espaço das distribuições entre a informação e a negentropia.	38
3.1	Rede neural artificial de dois neurônios para solução do problema de separação de mistura de duas variáveis.	54
3.2	Esquema de separação de fontes através do procedimento <i>deflation</i>	57
3.3	Rede neural artificial com dois neurônios e inibição lateral para implementação da regra anti-Hebbiana.	61
3.4	Fontes de sinais de voz.	66
3.5	Misturas disponíveis nos sensores provenientes de sinais de voz.	67
3.6	Estimativas das fontes de voz.	67
3.7	Sinais de mistura de um sistema de comunicação digital com modulação QPSK.	69
3.8	Fontes estimadas de sinais QPSK.	70
3.9	Evolução da matriz de separação global para um sistema de fontes de comunicação com modulação QPSK.	70
3.10	Alguns sinais biomédicos típicos.	72
3.11	Sinais disponíveis nos sensores de um sistema de processamento de sinais biomédicos.	72

3.12	Estimativas das fontes de sinais biomédicos.	73
3.13	Fontes de um sistema de aquisição de sinais eletroencefalográficos.	75
3.14	Misturas de sinais eletroencefalográficos.	75
3.15	Estimativas das fontes de sinais eletroencefalográficos.	76
3.16	Imagens fonte.	78
3.17	Misturas bidimensionais resultantes da composição de imagens em um sistema de mistura instantânea.	78
3.18	Estimativas dos sinais imagem fontes.	79
4.1	Recursos compartilhados em estratégias de múltiplo acesso.	84
4.2	Ilustração do plano de onda incidente em um arranjo linear de antenas.	86
4.3	Arranjo de antenas e processamento espacial.	90
4.4	Arranjo de antenas e processamento espaço-temporal a partir do k -ésimo usuário.	93
4.5	Alguns tipos de constelações de modulações digitais.	97
5.1	Modelo em banda básica equivalente de um sistema de comunicação digital.	119
5.2	Exemplo de dispersão intraclasse para um sistema digital contendo duas classes.	121
5.3	Comportamento da função custo de FP para vários valores de σ_r^2 utilizando canal $H_{AR}(z)$	127
5.4	Comportamento da função custo de FP para vários valores de σ_r^2 utilizando canal $H_{MA}(z)$	128
5.5	Trajетórias de convergência para diferentes inicializações e valores de fator de passo utilizando canal $H_{AR}(z)$ e modulação BPSK.	129
5.6	Trajетórias de convergência para diferentes inicializações e valores de fator de passo utilizando canal $H_{MA}(z)$ e modulação BPSK.	130
5.7	Comparação das funções de custo dos critérios DD e FP.	134
5.8	Comparação das curvas de nível dos critérios DD e FP.	134
5.9	Comparação entre funções de referência para os critérios de Sato, CM e FP.	139
5.10	Desempenho do MU-FPA em relação ao fator de regularização γ e número de usuários perdidos.	152
5.11	Comparação entre os algoritmos MU-FPA e MU-CFPA.	156
5.12	Comparação em termos de CME entre os algoritmos MUK e MU-CFPA.	157
5.13	Análise das funções de custo e curvas de nível dos critérios MP e FP modificado.	158
6.1	Evolução temporal do erro do módulo constante para os algoritmos MUK e MU-CFPA.	184
6.2	Evolução dinâmica da estimativa da função de densidade de probabilidade para o algoritmo MUK em diferentes intervalos temporais.	185
6.3	Evolução dinâmica da estimativa da função de densidade de probabilidade para o algoritmo MU-CFPA em diferentes intervalos temporais.	186
7.1	Evolução temporal dos erro do módulo constante para processamento espacial com sensores simples em um sistema de 4 usuários e 6 sensores.	194

7.2	Constelações dos 10% símbolos finais para cada um dos algoritmos.	195
7.3	Evolução do erro do módulo constante para um sistema de 4 usuários e um arranjo de antenas de 8 elementos.	197
7.4	Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o MU-CMA.	199
7.5	Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o FMU-CMA.	200
7.6	Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o MU-CFPA.	201
7.7	Evolução da estimativa do número de usuários ativos no sistema em função do número de iterações utilizadas para estimativa da matriz de autocorrelação. . .	204
7.8	Número mínimo de iterações necessárias para a correta estimação de $K = 8$ usuários em função do número de sensores utilizados.	205
7.9	Probabilidade de atingir transmissão com 100% de eficiência $P(K K, M)$ para $M = 8$ antenas.	209
7.10	Resultados de vazão em função da carga oferecida para os cenários de diversidade (DIV) e linha de visada (LOS).	210
7.11	Evolução temporal da interferência residual média para os diversos algoritmos utilizados no processamento espaço-temporal.	212
7.12	Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o LMS.	213
7.13	Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o MU-FPA.	214
7.14	Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o MU-CMA.	215

Lista de Tabelas

3.1	Resumo das características de alguns algoritmos de separação cega de fontes. . .	64
3.2	Parâmetros do sistema de mistura/separação de processamento de sinais de voz. . .	66
3.3	Parâmetros do sistema de mistura/separação de sinais de comunicação digital com modulação QPSK.	68
3.4	Parâmetros do sistema de mistura/separação de sinais biomédicos.	71
3.5	Parâmetros do sistema de mistura/separação de sinais eletroencefalográficos. . .	74
3.6	Parâmetros do sistema de mistura/separação de extração de imagens.	77
4.1	Algoritmo FMU-CMA.	105
4.2	Algoritmo MUK.	108
5.1	Resumo do <i>Multiuser Fitting pdf Algorithm</i> para os casos espacial e espaço-temporal.	151
5.2	Algoritmo MU-CFPA.	155
6.1	Parâmetros de simulação dos algoritmos MUK e MU-CFPA para comparação das estatísticas de ordem superior.	183
7.1	Parâmetros de simulação para processamento multiusuário espacial para 100 canais independentes num sistema de 4 usuários e 6 sensores.	193
7.2	Configuração dos usuários e suas respectivas direções de chegada num sistema que emprega um receptor com um arranjo de antenas.	196
7.3	Parâmetros de simulação para processamento multiusuário em um sistema com 4 usuários empregando um arranjo de antenas com 8 no receptor.	197
7.4	Parâmetros de simulação para resolução de pacotes em uma rede S-ALOHA. . .	208
7.5	Configuração do sistema para processamento espaço-temporal considerando 2 usuários com 2 percursos cada.	211
7.6	Parâmetros de simulação para processamento espaço-temporal.	211

Notações e Abreviações

NESTA seção são listados os símbolos, abreviações e acrônimos que ocorrem com frequência no decorrer do trabalho. Na tabela, a primeira coluna refere-se ao conceito/acrônimo, a segunda sua definição e a terceira a página do texto no qual o mesmo foi definido pela primeira vez.

Uma notação geral é que letras e símbolos em itálico representam valores escalares, vetores são representados por letras romanas minúsculas em negrito, por exemplo \mathbf{x} e matrizes são representadas por letras romanas maiúsculas em negrito, como \mathbf{X} . Entretanto, devido ao grande número de símbolos, para evitar notações repetidas são utilizados letras em formato e fontes diferentes que são definidas propriamente no decorrer do texto, por exemplo, \mathbf{f} e \mathbf{f} representam vetores diferentes.

Outro tipo de notação é relativo a listagem de hipóteses, condições ou propriedades. A lista é dada com a inicial do que é descrito (H, C ou P) com subscritos indicando o assunto da listagem.

Lista de símbolos

$E\{\cdot\}$: operador esperança	23
$p_a(\mathbf{a})$: função de densidade de probabilidade da variável \mathbf{a}	28
$p(\mathbf{a}, \mathbf{b})$: função de densidade de probabilidade (fdp) conjunta de \mathbf{a} e \mathbf{b}	31
$\mathcal{H}(\cdot)$: entropia	31
$\mathcal{H}(\mathbf{a} \mathbf{b})$: entropia condicional de \mathbf{a} dada a observação de \mathbf{b}	31
$I(\mathbf{a}, \mathbf{b})$: informação mútua entre \mathbf{a} e \mathbf{b}	31
$D(\bullet \circ)$: Divergência de Kullback-Leibler	32
$p(\mathbf{a} \mathbf{b})$: função de densidade de probabilidade (fdp) condicional de \mathbf{a} dada a observação de \mathbf{b}	31
\mathbf{H}	: matriz de mistura convolutiva (pode corresponder ao canal de um sistema de múltiplos usuários e múltiplos receptores)	21
\mathbf{x}	: vetor de amostras no receptor espaço-temporal provenientes do processamento da matriz de mistura convolutiva	21

\mathbf{W}	: matriz de separação	22
\mathbf{y}	: vetor contendo as saídas y_i do sistema de separação	22
\mathbf{a}	: vetor contendo os sinais a_i das fontes ou os sinais temporais $a(n)$ de uma única fonte	20
\mathbf{A}^{-1}	: inversa da matriz \mathbf{A}	40
\mathbf{a}^H	: hermitiano (transposto conjugado) de \mathbf{a} (<i>vetoroumatriz</i>)	90
\mathbf{a}^*	: complexo conjugado de \mathbf{a}	99
$N_G(\mathbf{a})$: negentropia de \mathbf{a}	34
$N(\mathbf{a}, \mathbf{b})$: distribuição gaussiana (normal) com média \mathbf{a} e variância \mathbf{b}	34
$U(\mathbf{a}, \mathbf{b})$: distribuição uniforme no intervalo de \mathbf{a} a \mathbf{b}	45
$\mathcal{K}\{\cdot\}$: kurtosis	35
$ \bullet $: módulo (valor absoluto)	36
ℓ	: atraso	25
$\Psi(\cdot)$: função de contraste	37
$I(\mathbf{a})$: informação mútua de \mathbf{a}	34
$I(\mathbf{a}^G)$: informação mútua de uma versão gaussiana, com a mesma média e variância, de \mathbf{a}	39
$\Psi_{\text{ICA}}(\cdot)$: função de contraste que utiliza análise por componentes independentes	39
$\det(\mathbf{A})$: determinante da matriz \mathbf{A}	40
M	: número de sensores (antenas) no receptor	20
K	: número de fontes (usuários)	20
f'	: primeira derivada de uma função f	59
\propto	: proporcional à	45
$[h_{ij}]$: elemento da i -ésima linha e j -ésima coluna da matriz \mathbf{H}	50
\mathbf{I}	: matriz identidade	22
$\nabla J(\mathbf{A})$: gradiente de J em relação ao parâmetro \mathbf{A} (matriz ou vetor)	53
\mathbb{R}	: conjunto dos números reais	57
c_i	: cumulante de i -ésima ordem	51
$\langle \mathbf{A} \rangle$: operação sobre a matriz (vetor) \mathbf{A}	55
$\hat{\mathbf{E}}$: estimador equivariante	55
$\text{Re}(\mathbf{a})$: parte real de \mathbf{a}	69
$\text{Im}(\mathbf{a})$: parte imaginária de \mathbf{a}	69
n	: índice temporal discreto	20
$\text{diag}(\mathbf{a})$: matriz diagonal com os elementos do vetor \mathbf{a}	92
\mathfrak{A}	: alfabeto da modulação	97
$\text{card}(\mathbf{a})$: cardinalidade de \mathbf{a}	97
\mathfrak{C}	: cardinalidade do alfabeto da fonte	97
\mathcal{H}_k	: matriz de convolução do canal espaço-temporal do k -ésimo usuário	94
γ	: fator de regularização do termo de decorrelação de algoritmos de processamento multi-usuário	101
\mathbf{W}^e	: matriz de separação na etapa de equalização de algoritmos de processamento multi-usuário	107

$\text{sign}(\cdot)$:	operador que extrai o sinal de uma variável, $\text{sign}(\mathbf{a}) = \begin{cases} 1, & \mathbf{a} \geq 0 \\ -1, & \mathbf{a} < 0 \end{cases}$	107
$\ \mathbf{A}\ $:	norma de \mathbf{A} (vetor ou matriz)	108
$\mathcal{K}_{\mathbf{a}}$:	<i>kurtosis</i> de \mathbf{a}	106
ρ_2 :	dispersão da constelação	101
$\mathcal{L}_{\mathbf{B}}(\mathbf{A})$:	log-verossimilhança do modelo \mathbf{A} , a partir de \mathbf{B} amostras	47
\tilde{x} :	componente do sinal recebido sem ruído	109
$\mathbf{A}^{\#}$:	pseudo-inversa da matriz \mathbf{A}	109
$\delta(\cdot)$:	função de Dirac	50
δ_{ij} :	função de Kronecker	59
\mathcal{H} :	matriz de convolução do canal mono-usuário	119
\mathbf{g} :	resposta global de um sistema mono-usuário	119
\mathbf{G} :	matriz de separação global de um sistema de separação cega de fontes	68
\mathbf{h}_i :	polinômio de Hermite de i -ésima ordem	51
\mathbf{a}_i :	i -ésimo elemento (escalar) do alfabeto \mathfrak{A}	120
\mathbb{L}_p :	distância estatística de ordem p	144
$K_{\sigma}(\cdot)$:	<i>kernel</i> gaussiano de desvio padrão σ	145
$\text{dec}(\cdot)$:	função de decisão	133
j :	número complexo ($j = \sqrt{-1}$)	86
κ_k :	momento centrado de k -ésima ordem	169
$\Omega_Y(\omega)$:	função geradora de momentos (ou função característica) da variável aleatória y	169
$p_Y^{(k)}(y)$:	k -ésimo termo da expansão em série de $p_Y(y)$	182
$\text{rank}(\mathbf{A})$:	número de colunas da matriz \mathbf{A}	202
$\mathcal{O}(\mathbf{a})$:	complexidade computacional da ordem de \mathbf{a} operações	203
\mathbb{Z} :	conjunto dos números inteiros	194
$\text{tr}(\mathbf{A})$:	traço da matriz \mathbf{A}	235
$\Upsilon_Y(\omega)$:	função geradora de cumulantes	176

Abreviações e acrônimos

8-PSK :	<i>8-Phase Shift Keying</i>	96
fdc :	função de distribuição cumulativa	45
fdp :	função densidade de probabilidade	17
i.i.d. :	independentes e identicamente distribuídas	26
nPCA :	<i>Nonlinear PCA</i>	62
v.a. :	variável aleatória	29
AMUSE :	<i>Algorithm for Multiple Unknown Signals Extraction</i>	26

AR	: <i>Auto Regressive</i>	126
BER	: <i>Bit Error Rate</i>	112
BGR	: <i>Benveniste-Goursat-Rouget</i>	140
BPSK	: <i>Binary Phase Shift Keying</i>	96
BSS	: <i>Blind Source Separation</i>	4
CAG	: <i>Controle Automático de Ganho</i>	23
CDMA	: <i>Code Division Multiple Access</i>	83
CM	: <i>Constant Modulus</i>	60
CMA	: <i>Constant Modulus Algorithm</i>	60
CME	: <i>Constant Modulus Error</i>	152
DD	: <i>Decisão Dirigida</i>	133
DMI	: <i>Direct Matrix Inversion</i>	100
DOA	: <i>Direction Of Arrival</i>	85
DQ	: <i>Distância Quadrática</i>	145
DS-CDMA	: <i>Direct Sequence Code Division Multiplex Access</i>	112
EASI	: <i>Equivariant Adaptive Separation via Independence</i>	56
ECG	: <i>EletroCardioGrama</i>	6
ECGF	: <i>EletroCardioGrama Fetal</i>	6
ECGM	: <i>EletroCardioGrama Materno</i>	7
EEG	: <i>EletroEncefaloGrama</i>	7
EQM	: <i>Erro Quadrático Médio</i>	99
FDMA	: <i>Frequency Division Multiple Access</i>	82
FIR	: <i>Finite Impulse Response</i>	92
FMU-CMA	: <i>Fast Multiuser Constant Modulus Algorithm</i>	103
FP	: <i>Fitting pdf</i>	123
FT	: <i>Função de Transferência</i>	118
FPA	: <i>Fitting pdf Algorithm</i>	123
FPC	: <i>Fitting pdf Criterion</i>	122
FT	: <i>Função de Transferência</i>	118
GCMA	: <i>Generalized Constant Modulus Algorithm</i>	113
HOS	: <i>Higher Order Statistics</i>	17
ICA	: <i>Independent Component Analysis</i>	17
InfoMax	: <i>Information Maximization</i>	45
ISI	: <i>InterSymbol Interference</i>	84
ICALAB	: <i>Independent Component Analysis Laboratory</i>	65
KLD	: <i>Kullback-Leibler Divergence</i>	32
LMS	: <i>Least Mean Squares</i>	100
LOS	: <i>Line-Of-Sight</i>	208
LSAD-CMA	: <i>Least-Squares with Adaptive Decorrelation Constant Modulus Algorithm</i>	103
MaxEnt	: <i>Maximization of Entropy</i>	45
MA	: <i>Moving Average</i>	128

MAC	: <i>Medium Access layer</i>	205
MAI	: <i>Multiple Access Interference</i>	83
MAP	: <i>Maximum a Posteriori</i>	160
MEG	: <i>MagnetoEncefaloGrama</i>	7
MIMO	: <i>Multiple-Input Multiple-Output</i>	112
MMSE	: <i>Minimum Mean Square Error</i>	99
MP	: <i>Matched pdf</i>	147
MUD	: <i>Multiuser Detection</i>	84
MUK	: <i>Multiuser Kurtosis Maximization</i>	106
MV	: <i>Máxima Verossimilhança</i>	46
MU-CFPA	: <i>Multiuser Constrained Fitting pdf Algorithm</i>	154
MU-CFPC	: <i>Multiuser Constrained Fitting pdf Criterion</i>	153
MU-CMA	: <i>Multiuser Constant Modulus Algorithm</i>	101
MU-FPA	: <i>Multiuser Fitting pdf Algorithm</i>	150
MU-FPC	: <i>Multiuser Fitting pdf Criterion</i>	149
NG-FICA	: <i>Natural Gradient Flexible ICA</i>	72
NMI	: <i>Negative Mutual Information</i>	40
PCA	: <i>Principal Component Analysis</i>	24
SOS	: <i>Second Order Statistics</i>	16
SOM	: <i>Self-Organizing Maps</i>	17
SISO	: <i>Single-Input Single-Output</i>	17
RNA	: <i>Redes Neurais Artificiais</i>	16
QPSK	: <i>Quadrature Phase Shift Keying</i>	68
RI	: <i>Residual Interference</i>	211
SDMA	: <i>Spatial Division Multiple Access</i>	83
SEA	: <i>Super-Exponential Algorithm</i>	60
SNR	: <i>Signal-to-Noise Ratio</i>	65
SPF	: <i>Sampled pdf Fitting</i>	146
SVD	: <i>Singular Value Decomposition</i>	23
SW	: <i>Shalvi-Weinstein</i>	106
S-ALOHA	: <i>Slotted ALOHA</i>	206
TDMA	: <i>Time Division Multiple Access</i>	83
TICA	: <i>Thin algorithm for Independent Component Analysis</i>	66
ZF	: <i>Zero Forcing</i>	118

– “O começo é a metade de tudo.”

Pythagore

1

Introdução

GRANDES avanços em sistemas de recuperação de informações têm sido registrados nos últimos anos. Muito disso se deve a técnicas avançadas de processamento de sinais, que permitem um incremento no tratamento das informações no receptor de tal forma a beneficiar-se ao máximo de suas características originais.

Os objetivos e a estrutura da presente tese, que versa sobre algumas destas técnicas avançadas de processamento de sinais, são apresentados neste capítulo introdutório. Além disto, para situar o leitor nos principais aspectos teóricos do trabalho, discute-se também alguns conceitos e idéias fundamentais que serão aprofundados no decorrer da tese.

O capítulo é então organizado da seguinte forma. Na Seção 1.1 é introduzida a motivação principal do trabalho relativo à recuperação dos sinais destacando ferramentas clássicas utilizadas na concepção e análise de métodos para remoção de interferência. O problema genérico de interferência em sinais é apresentado na Seção 1.2 e alguns problemas que se enquadram no modelo são também descritos. Um caso particular e de grande interesse prático é brevemente discutido na Seção 1.3. Finalmente, a estrutura, organização e as contribuições da tese são listadas na Seção 1.4.

1.1 Recuperação da informação: motivação

Um dos objetivos básicos de qualquer sistema de transmissão/recepção de informação é prover capacidade de recuperação do sinal transmitido da maneira mais fiel possível, para que a informação obtida no receptor seja confiável. Neste sentido, algumas técnicas devem ser concebidas, uma vez que o meio de transmissão corrompe o sinal transmitido. Por exemplo, a utilização de codificação para inserir redundância é uma destas técnicas, assim como a utilização de protocolos que permitam assegurar a correta recepção.

Do ponto de vista de processamento de sinais, a inserção de um dispositivo que atue no sinal recebido para mitigar os efeitos do canal é uma das mais freqüentes estratégias. Quando vários sinais provenientes de diferentes fontes estão presentes simultaneamente no sistema, tal dispositivo, além de retirar a interferência, deve ser capaz de separar os sinais dos diversos usuários.

Duas linhas de atuação são possíveis para a otimização do dispositivo de recuperação:

- processamento supervisionado ou treinado;
- processamento cego ou não-supervisionado¹.

Para a utilização de estratégias supervisionadas são utilizadas seqüências de treinamento conhecidas no transmissor e no receptor. Embora sejam bastante robustas, há situações em que não é possível, ou pelo menos desejado, o emprego de tais técnicas. Nestas situações, a utilização de estratégias cegas é uma alternativa, apesar de diversas dificuldades, tanto práticas como de análise matemática destes métodos.

Desde a década de 1970, com o trabalho pioneiro de Sato [Sato, 1975], estratégias cegas têm sido estudadas no contexto de recuperação de sinais digitais quando os mesmos são submetidos a interferência inter-simbólica. A maioria delas, entretanto, apresentava-se como soluções de engenharia, isto é, sem uma metodologia de caráter mais geral, aproveitando-se de características bastante específicas dos sinais em questão.

Somente na década seguinte surgiram os primeiros trabalhos que geraram um maior entendimento sobre as condições necessárias para que a recuperação de sinais, sem conhecimento do sinal transmitido, fosse possível.

Tais análises evidenciaram a importância das estatísticas de ordem superior, assunto abordado brevemente no tópico a seguir.

¹Nesta tese serão utilizados os termos cego, autodidata e não-supervisionado como sinônimos.

1.1.1 Estatísticas de ordem superior

Um dos primeiros trabalhos que tratou sobre a necessidade do uso de estatísticas de ordem superior para recuperação de sinais foi proposto por Benveniste, Goursat e Rouget em 1980. Naquele trabalho foi mostrado que, para recuperação de um sinal não-gaussiano através de estruturas lineares, seria necessário igualar as densidades de probabilidade dos sinais de entrada e saída do sistema, sendo com isso necessário igualar todas as estatísticas de ordem superior que caracterizam o sinal.

Posteriormente, em 1982, Lii e Rosenblatt [Lii & Rosenblatt, 1982] mostraram que seria possível tal recuperação cega utilizando somente o momento de ordem dois e um momento de ordem superior. Uma particularização de tal critério foi proposta por Shalvi-Weinstein, em 1990 [Shalvi & Weinstein, 1990], indicando que a utilização específica do cumulante de quarta ordem (*kurtosis*) tornava o processamento possível e com solução única. Estas condições foram colocadas no contexto de recuperação de uma única fonte e para os casos nos quais os sinais são discretos e sua distribuição de probabilidade é conhecida.

Entretanto, nenhum trabalho até o momento se propôs a avaliar as diferenças de resultados obtidos ao se considerar estatísticas de ordem superior distintas e o impacto disto nos métodos de recuperação de sinais. Este é um objetivo fundamental desta tese que se propõe, também, a avaliar tal aspecto em algoritmos adaptativos para recuperação de sinais.

Quando são consideradas múltiplas fontes e estas não são discretas, algumas ferramentas adicionais às estatísticas de ordem superior são necessárias para permitir a recuperação dos sinais de todas as fontes. Uma poderosa ferramenta de análise é discutida no tópico a seguir.

1.1.2 Teoria da informação

A formalização do conceito de informação dado por Claude Shannon em [Shannon, 1948] iniciou uma nova maneira de avaliar fenômenos de transmissão de dados. A chamada *teoria da informação* possibilitou a avaliação do desempenho de sistemas a partir de medidas de quantidade de informação em diferentes estágios de um processo qualquer de aquisição de dados.

As medidas de *entropia* e *informação mútua* permitem avaliar e conceber critérios para a recuperação de sinais quando as fontes não possuem distribuição de probabilidade conhecida, nem mesmo são necessariamente discretas. Isto é de fundamental importância para tratar problemas de natureza bem mais genérica que o caso de recuperação de sinais em sistemas de uma única fonte com distribuição discreta. Um problema de maior envergadura, que congrega várias áreas do conhecimento, é a necessidade de recuperar diferentes sinais desconhecidos,

combinados de forma também não conhecida. Este é o tema tratado ao longo desta tese e está apresentado de forma preliminar na seção seguinte.

1.2 O problema de separação cega de fontes

Durante muitos anos foi tido como impossível a idéia de recuperação de sinais, combinados no processo de transmissão, a partir somente de diferentes combinações obtidas e observadas num certo número limitado de sensores.

Apenas em 1985, através do trabalho de Hérault, Jutten e Ans [Hérault et al., 1985], começou-se a delinear a área denominada na época de *separação cega de fontes* (BSS, *Blind Source Separation*), mostrando a possibilidade de resolução de tal problema através do uso de estruturas não-lineares, particularmente utilizando redes neurais artificiais.

A partir deste momento, muito se avançou nas ferramentas utilizadas e, na comunidade científica de processamento de sinais, as técnicas de separação cega de fontes passaram a receber forte atenção, pela sua generalidade e pelo potencial de aplicação em várias áreas, assim como pelos instigantes trabalhos teóricos que delas decorreram. Algumas de suas aplicações práticas são assunto dos tópicos a seguir.

1.2.1 *Cocktail party*

Um dos problemas típicos investigados pelas técnicas de separação cega de fontes é o de separação das diversas fontes existentes em uma situação de *cocktail party*. A Figura 1.1 ilustra um esquema geral do tipo de informação proveniente de diferentes fontes que são captadas por sensores.

No caso de uma reunião ou festa, vários tipos de fontes, algumas delas inclusive correlacionadas, estão presentes e perturbam a compreensão e identificação de uma determinada fonte. Situações como estas são apresentadas ao cérebro humano quase que diariamente e não se nota nenhuma dificuldade para compreender um determinado interlocutor, salvo em meio a ruído excessivo, graças à grande capacidade de processamento do cérebro humano.

No caso de utilizar-se sensores para execução da mesma tarefa, este requer um esforço considerável. Aqui técnicas de separação cega de fontes são necessárias para retirar, da informação que se deseja obter, qualquer interferência das outras fontes, sendo isto feito sem o conhecimento de nenhuma das fontes.

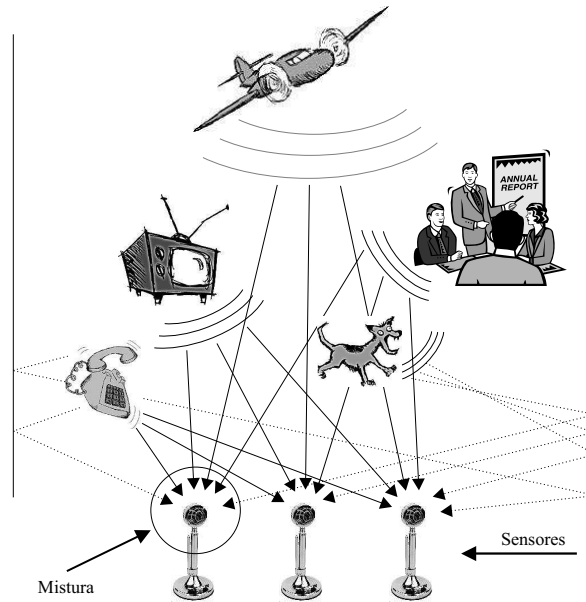


Figura 1.1: Ilustração do problema de *cocktail party*.

1.2.2 Processamento de sinais biomédicos

Uma outra classe de problemas de separação cega de fontes é representada na Figura 1.2.

Quando se deseja saber o máximo possível sobre a situação da saúde de um paciente, os métodos de aquisição de informações que possibilitem tais análises devem ser, idealmente, o mais indolores, não invasivos e confiáveis possível.

Do ponto de vista de menos invasivo (e conseqüentemente indolor), as técnicas que utilizam sensores na superfície corporal são bastante atrativas. Quanto a aspectos de robustez e confiabilidade, sem um tratamento adequado estas não fornecem bons resultados, dado que os sensores na superfície captam sinais provenientes de várias atividades fisiológicas.

A utilização de métodos de separação cega de fontes é empregada para separar as diversas fontes

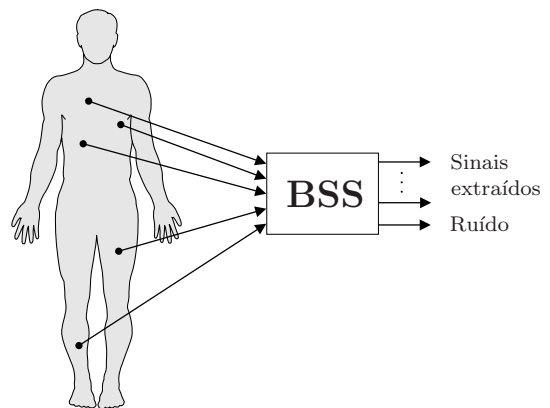


Figura 1.2: Aquisição de sinais a partir de sensores na superfície corporal.

e isolar aquela que fornece a informação buscada.

Alguns problemas específicos de sinais biomédicos apresentam as mesmas características do modelo genérico da Figura 1.2. Dois casos particulares e de bastante interesse em áreas de engenharia biomédica são descritos a seguir.

Eletrocardiograma materno/fetal

O eletrocardiograma (ECG) é um procedimento comum e eficiente, além de indolor e barato, para diagnóstico do estado de saúde de um indivíduo. Uma vez que a atividade cardíaca é regida pela atividade elétrica no músculo cardíaco, o ECG revela importantes informações médicas que não podem ser observadas a partir da auscultação.

Da mesma maneira que para adultos, é possível observar a atividade elétrica do coração de um feto através do *eletrocardiograma fetal* (ECGF) que contém importantes indicações das condições de saúde do feto. A avaliação da taxa cardíaca fetal é prática comum em qualquer exame pré-natal para indicar se o feto está sob estresse (grandes variações de frequência cardíaca) ou com arritmia (frequências cardíacas anormais).

Normalmente, a frequência cardíaca fetal é medida através da utilização de técnicas de ultra-sonografia baseadas na medida do deslocamento Doppler ultrassônico. Entretanto, este método requer que um feixe ultrassônico seja direcionado para o coração do feto, o que pode ser difícil de conseguir devido a movimentações do feto e da mãe.

Além disso, o método apresenta uma baixa precisão devido a alguns fatores, dentre eles [Lathauwer, 1997]:

- os movimentos mecânicos do coração podem introduzir variações espúrias na medida da frequência cardíaca fetal;
- a maioria dos sistemas baseados em Doppler calculam um número médio de variações da frequência cardíaca, não sendo possível medir a variação a cada batimento;
- as unidades de monitoramento são geralmente disponíveis até uma frequência cardíaca de 220 batimentos por minuto; no caso de frequência mais altas o número é dividido por dois resultando numa aparente frequência cardíaca normal quando, na verdade, o feto apresenta taquicardia.

Estas desvantagens não são observadas quando o ECGF é medido a partir de técnicas de aquisição do ECG usando sensores na superfície corporal da mãe. A Figura 1.3 ilustra a aquisição através de tal estratégia.

Neste caso, é necessário a utilização de técnicas de separação cega de fontes para

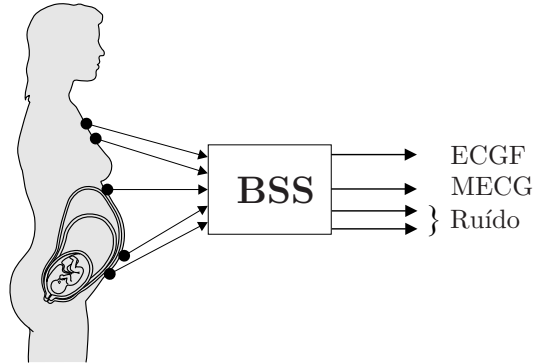


Figura 1.3: Ilustração do método de medição do eletrocardiograma fetal/materno a partir de sensores na superfície corporal.

separar o ECGF dos sinais do ECG materno (ECGM). Quando tais técnicas são empregadas corretamente, elas constituem uma ferramenta bastante robusta para a detecção dos sinais desejados, importantes para a biometria materna e fetal.

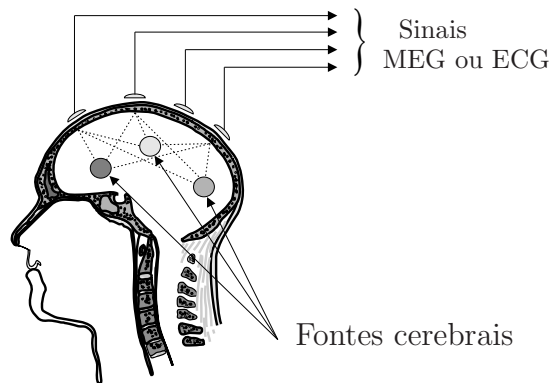
(Eletro/magneto)encefalograma

Devido às características bastante complexas do cérebro humano, a utilização de técnicas não invasivas é mais que desejada para a obtenção de informações sobre os padrões de ativação cerebral.

De uma maneira geral, quando uma região de tecido neuronal, composta aproximadamente de 100 mil células, é sincronamente ativada, correntes elétricas e campos magnéticos extra-celulares são gerados. Estas regiões podem ser modeladas por dipolos de corrente, já que fornecem campos elétricos de corrente dipolar na superfície do crânio. Estes fluxos de corrente através do crânio criam diferenças de potencial em sua superfície, que podem ser detectadas através de eletrodos de superfície num procedimento chamado *eletroencefalograma* (EEG). No caso de se dispor de núcleos condutores na superfície craniana, pode-se então detectar os campos magnéticos gerados pela atividade cerebral pelo procedimento de *magnetoencefalograma* (MEG) [Cichocki & Amari, 2002].

Quando se conhece a posição das fontes, pode-se determinar os padrões elétricos e magnéticos obtidos na superfície craniana. Quando esta informação não está disponível, caso mais comum, determinar a partir dos dados coletados nos sensores elétricos ou magnéticos

a posição das fontes cerebrais (regiões ativadas) é um problema que exige um tratamento mais elaborado. É neste ponto que separação cega de fontes encontra seu ambiente natural de aplicação. A Figura 1.4 ilustra a obtenção dos dados elétricos e/ou magnéticos para determinação das regiões ativadas do cérebro.



Novamente, este é um caso típico de situação prática em que a necessidade de utilização de avançadas técnicas de processamento de sinal é de extrema relevância para permitir que as estimativas das fontes sejam confiáveis o suficiente para prover um diagnóstico preciso para identificação e tratamento de patologias, numa área tão complexa e sensível quanto o cérebro humano.

Figura 1.4: Aquisição de (eletro/magneto)encefalograma.

1.3 Processamento multi-usuário

Num sistema de comunicações digitais sem fio, quando vários usuários em diferentes posições espaciais transmitem ao mesmo tempo e na mesma frequência sinais para o mesmo receptor, o sinal resultante é uma combinação dos sinais das fontes ponderados pelos efeitos dos canais relativos a cada um dos usuários [Cavalcanti, 1999].

Devido à diversidade espacial, um número determinado de sensores é utilizado para permitir a completa separação dos diferentes sinais dos usuários. A Figura 1.5 ilustra o problema de uma maneira genérica.

Também neste tipo de processamento a meta é separar, a partir das misturas obtidas, cada um dos sinais provenientes dos diversos usuários existentes no sistema. Desta maneira, as mesmas técnicas que permeiam o campo da separação cega de fontes podem ser utilizadas para solucionar a questão da remoção da interferência de múltiplo acesso, gerada quando vários usuários compartilham os mesmos recursos do canal, e da interferência inter-simbólica devida às dispersões temporais que o sinal sofre [Proakis, 1995].

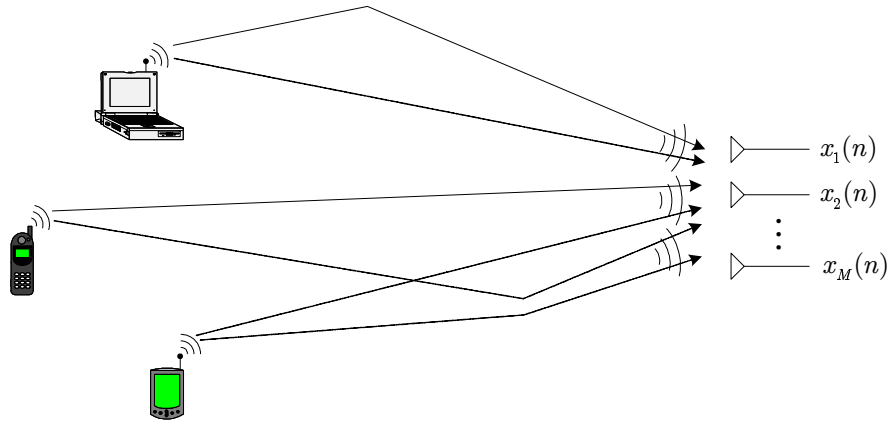


Figura 1.5: Cenário típico de comunicações sem fio.

Entretanto, neste caso, características especiais das fontes e do sistema de mistura permitem a concepção de estratégias baseadas em equalização cega para a detecção dos sinais dos diversos usuários, diminuindo assim a complexidade do método de separação de fontes. Estas características das fontes são relativas às considerações de, por exemplo, alfabeto finito, módulo constante ou cicloestacionariedade. Quanto ao canal, alguns dos parâmetros que permitem uma modelagem mais simplificada dizem respeito à faixa de frequência utilizada, à arquitetura do receptor e à relação do tempo de coerência do canal com o período de amostragem dos sinais.

Embora sejam conhecidos na literatura vários trabalhos que tratam do problema de remoção de interferência e identificação dos sinais em sistemas de múltiplos usuários (ver por exemplo [Verdú, 1998; Cavalcanti, 1999; Giannakis et al., 2001] e suas referências para uma lista bastante extensa), o aspecto geral de um sistema de processamento multi-usuário é raramente abordado e vários métodos de grande potencial não são sequer considerados na área de comunicações móveis.

Uma das metas desta tese é também colocar sob uma abordagem mais unificada os problemas de separação cega de fontes e o de processamento multi-usuário cego, este último também chamado na literatura de *conformação cega de feixes*. Desta maneira, permite-se a aplicação de novas técnicas para um problema de grande interesse no campo de comunicações móveis.

Outra grande meta é o interesse por novos métodos aplicáveis ao contexto de processamento multi-usuário baseados em técnicas de separação cega de fontes, bem como sua análise teórica e comparação com outros métodos disponíveis na literatura, contribuindo para um novo aporte de conhecimentos a uma área de pesquisa tão abrangente.

1.4 Organização, estrutura e contribuições

A organização bem como a estrutura e as contribuições desta tese estão apresentadas da seguinte maneira.

- **Capítulo 2** - *Separação Cega de Fontes: Princípios e Conceitos*
Apresentação dos princípios e conceitos básicos relacionados à separação cega de fontes, bem como uma apresentação das principais ferramentas que são empregadas para que a solução do problema seja factível.
- **Capítulo 3** - *Separação Cega de Fontes: Estratégias e Aplicações*
Neste capítulo são explorados alguns dos principais métodos de separação cega de fontes bem como uma abordagem de critérios capazes de sintetizar e verificar as diferentes estratégias existentes. Além disso, um conjunto de aplicações, dentre as várias possíveis, são apresentadas para ilustrar a potencialidade dos métodos de separação cega de fontes.
- **Capítulo 4** - *Processamento Multiusuário*
Características particulares dos sistemas de comunicação móvel que utilizam técnicas de múltiplo acesso são descritas e avaliadas. Também pelo fato da proposta de novos métodos para este tipo de problema ser um dos focos principais do trabalho, é dado algum destaque para as estratégias com a mesma finalidade e já existentes na literatura.
- **Capítulo 5** - *Métodos Multiusuário Baseados na Estimação de Densidade de Probabilidade*
A proposta original de novos métodos para processamento multi-usuário, baseados na estimativa da função de densidade de probabilidade do sinal na saída do dispositivo de separação, por meio de um modelo paramétrico e utilizando a medida de Kullback-Leibler, é um dos tópicos deste capítulo. Outra contribuição é a análise e comparação da proposta com métodos clássicos, gerando novas relações matemáticas entre diferentes estratégias.
- **Capítulo 6** - *Análise da Influência das Estatísticas de Ordem Superior em Algoritmos Adaptativos para Separação de Fontes*
Motivado pelos resultados obtidos a partir dos métodos propostos, neste capítulo é apresentada uma análise da influência das estatísticas de ordem superior na concepção de algoritmos adaptativos de separação cega de fontes. Para permitir a avaliação em sistemas de comunicação digital, é derivada uma nova expansão em séries ortonormais para a função de densidade de probabilidade utilizando como função de referência uma soma de distribuições gaussianas.

- **Capítulo 7** - *Aplicações em Sistemas de Mistura Instantânea e Convolutiva*
Para ilustrar a capacidade de aplicação das propostas e avaliá-las comparativamente com outras estratégias, são realizados experimentos computacionais que permitem observar o comportamento de tais métodos em diversas situações, tais como sistemas com diversidade de recepção, utilização de arranjos de antenas, resolução de colisão de pacotes em redes S-ALOHA, estimação do número de usuários ativos e sistemas com desvanecimento espaço-temporal. No contexto da estimação do número de usuários, é ainda proposto um algoritmo baseado na decomposição de valores singulares da matriz de autocorrelação dos sinais recebidos.
- **Capítulo 8** - *Conclusões, Sugestões e Perspectivas*
Além das conclusões, este capítulo é dedicado a algumas sugestões de perspectivas e trabalhos futuros decorrentes do conteúdo da tese.
- **Apêndice A** - *Definições Matemáticas*
Algumas importantes definições matemáticas utilizadas no decorrer da tese são destacadas e discutidas em maiores detalhes neste apêndice.
- **Apêndice B** - *Demonstrações e Derivações*
Neste apêndice, derivações e demonstrações de teoremas e métodos de relevância para as propostas da tese são apresentados visando uma maior profundidade dos conceitos transmitidos. Particularmente, uma demonstração do Teorema de Benveniste-Goursat-Rouget é dada num formato original nesta tese.

PARTE I

*Separação Cega de
Fontes e
Processamento
Multiusuário*

– “A principal função de um sistema de comunicação é reproduzir, exatamente ou de forma aproximada, uma informação proveniente de outro ponto diferente.”

Claude Shannon, 1948

2

Separação Cega de Fontes: Princípios e Conceitos

OS ÚLTIMOS ANOS da comunidade de processamento de sinais têm sido bastante agitados pelo campo da recuperação de informação. Verifica-se um intenso número de aplicações decorrentes de outras áreas que, mais recentemente, vieram a integrar o cenário de tratamento de informação. Assim, problemas de desconvolução e separação ganharam uma roupagem bastante interessante ao serem abordados de forma conjunta, possibilitando grandes avanços na comunidade de processamento de sinais.

Este capítulo faz uma revisão do estado da arte e dos principais progressos nas técnicas que tratam o problema de separação cega de fontes, situando o leitor sob um ponto de vista histórico. Outra meta é a discussão, feita talvez de forma mais concisa da que faria jus um assunto tão vasto, de algumas das principais estratégias de BSS, visando fornecer um conhecimento do estado da arte.

O restante do capítulo é organizado da seguinte forma: a Seção 2.1 discorre sobre o estado da arte das técnicas que tratam de separação de fontes sem uma preocupação com o formalismo matemático, abordados de forma mais rigorosa na Seção 2.2. A Seção 2.3 trata do principal método de separação de fontes,

incluindo as ferramentas estatísticas e de teoria da informação envolvidas nesta técnica. Finalmente, na Seção 2.4 são apresentadas as conclusões e também uma breve discussão sobre as técnicas de separação cega de fontes.

2.1 Um pouco de história...

A história do processamento autodidata (ou cego) de fontes tem início nos anos 80 com o trabalho de um grupo francês da *Université de Grenoble*, liderado por Christian Jutten. A motivação inicial surgiu a partir do estudo de trabalhos de processamento de sinais neurofisiológicos, em um modelo simplificado de movimento muscular codificado.

O problema consistia em tentar modelar o sinal coletado no nervo eferente¹ a partir de estímulos que são processados pelo cérebro. No caso da ativação muscular tem-se, dentre outros, a posição angular de um ponto no espaço e a velocidade com a qual o membro (ou membros) deve se mover para alcançar o objeto. Os pesquisadores utilizaram como modelo do sinal incidente no nervo muscular uma combinação linear dos estímulos não disponíveis individualmente para medição [Hérault & Jutten, 1994; Hyvärinen et al., 2001].

Naquele trabalho, a linha principal da estratégia de separação dos impulsos originais foi a utilização de redes neurais artificiais (RNA). Além disto, o modelo simplificado adotado para sistemas reais abriu margem para várias outras vertentes, nas quais a abordagem de um sinal que resulta de uma mistura (não necessariamente linear) de outros faz-se necessária.

O trabalho [Hérault et al., 1985] é considerado *o nascimento* da área de separação de fontes, nem tanto pela contribuição da solução proposta, que originou posteriormente o conhecido algoritmo Jutten-Hérault [Jutten & Héroult, 1991], mas principalmente pela modelagem através de equações de álgebra linear.

A *infância* das técnicas de separação cega de fontes transcorreu com o apoio das já então consolidadas técnicas baseadas em estatísticas de segunda ordem (SOS, *Second Order Statistics*), que possibilitavam um número de problemas solucionáveis bastante limitado devido às restrições impostas às fontes.

Desta forma, o estágio da pesquisa em BSS permaneceu restrito até o final da década de 80 quando surgiram outros grupos de pesquisa franceses que investiram numa maior elaboração matemática dos modelos até então existentes. Neste contexto, dois pesquisadores de extrema importância podem ser citados, Jean-François Cardoso e Pierre Comon.

¹Os nervos eferentes são os responsáveis, dentre outras funções, movimentos de contração e relaxamento muscular.

Estes dois nomes contribuíram fortemente com o desenvolvimento e avanço matemático do tema através da exploração de duas linhas [Cardoso, 1989; Comon, 1989]:

- (1) teoria da informação – a utilização da teoria elaborada por Claude Shannon foi crucial para permitir uma manipulação matemática e mostrar a factibilidade da solução do problema de BSS;
- (2) estatísticas de ordem superior (HOS, *Higher Order Statistics*) – a generalização de várias estratégias decorrentes do problema de desconvolução cega, que é uma simplificação do problema de BSS, possibilitou também um avanço das técnicas utilizadas para a separação de sinais.

Ainda assim, durante alguns anos, o problema foi tido como insolúvel dadas certas características das fontes ou do sistema de combinação. Na busca de técnicas que se mostrassem adequadas, as redes neurais auto-organizáveis (SOM, *Self-Organizing Maps*²) tiveram (e ainda têm, vide [Haykin, 2000a]) um papel fundamental. Apenas em 1994, Pierre Comon propôs uma nova abordagem de encarar o problema de BSS. Ele mostrou em seu trabalho [Comon, 1994] que, se as fontes são independentes entre si, o sistema de separação deveria forçar as saídas a serem, também, independentes, segundo um critério adequadamente definido. Esta técnica consagrou-se com o nome de análise por componentes independentes (ICA, *Independent Component Analysis*) e, conforme será visto na seqüência deste capítulo, tornou-se uma poderosa ferramenta para a confecção de estratégias de separação autodidata de fontes.

Um ponto a ser destacado é que a aplicação da técnica de ICA tem a limitação de necessitar de uma estimativa da função densidade de probabilidade (fdp) dos sinais das fontes. Uma vez que as fontes não estão disponíveis e que nem sempre as fdps podem ser facilmente estimadas a partir dos dados, uma grande contribuição, no sentido de alavancar a aplicação de ICA, foi dada pelo grupo da Universidade Tecnológica de Helsinki na Finlândia. Eles propuseram fórmulas simples de estimação aproximada da fdp a partir de amostras dos dados no receptor. O algoritmo derivado a partir destas aproximações foi nomeado *FastICA* e tem sido amplamente investigado no contexto de diversas aplicações [Hyvärinen & Oja, 2000; Hyvärinen et al., 2001].

A partir deste marco, muito se avançou na busca e desenvolvimento de algoritmos e soluções para o problema de separação de fontes baseado apenas em suas amostras recebidas e com alguma informação estatística das fontes, sem conhecimento prévio do sistema de combinação. Pode-se citar que algumas das técnicas geradas no período posterior a esta época decorreram de estratégias de desconvolução em sistemas SISO (*Single-Input Single-Output*), ou o conhecido problema de equalização de canal, estendidas para o caso multilinear, aproveitando, com mais efeito, as informações contidas nas HOS [Lathauwer, 1997].

²As redes auto-organizáveis surgiram com o advento dos mapas de Kohonen, por isso são freqüentemente referidas por mapas de Kohonen.

Assim, com todas as técnicas desenvolvidas é possível separar fontes das mais diversas características³: reais ou complexas, contínuas ou discretas, de diferentes distribuições de probabilidade, etc. Tal generalidade, entretanto, requer o uso de estruturas não-lineares para a solução do problema. No caso de estruturas lineares, algumas restrições sobre as fontes, tais como fontes discretas, mesma distribuição de probabilidade, sub-gaussianas, independentes e estatisticamente independentes, permitem a separação.

Outro fator de destaque acerca do problema é que as soluções que vêm sendo propostas atualmente vão no sentido de generalizar ao máximo o sistema de combinação dos sinais das fontes, permitindo inclusive combinações não-lineares, criando a possibilidade de promover soluções para os mais diversos tipos de aplicações nas áreas de [Kofidis, 2001]:

- comunicações;
- análise de sinais biomédicos;
- restauração de imagens adquiridas de cenas inacessíveis (ultra-sonografia, imagens astronômicas, imagens de reatores nucleares, ...);
- extração de padrões;
- exploração geofísica;
- ...

A história contemporânea de separação não-supervisionada de fontes se confunde com a dos grupos que se dedicam à investigação e pesquisa deste tema, e também se diversifica dadas as preferências de abordagens e de estruturas que cada um deles domina. Pode-se citar então alguns nomes, observando que a lista a seguir não é de forma alguma, exaustiva; serve apenas como ponto de partida para aqueles que buscam iniciar-se no assunto:

1. **França** - tendo sido o berço da área, é natural que um grande número de pesquisadores deste país se dedique ao mesmo assunto. Várias das importantes contribuições em modelos e critérios/algoritmos foram criadas neste país. Pode-se citar alguns nomes, dentre os vários, de destaque na área:
 - ▷ Jean-François Cardoso, *ENST Paris*;
 - ▷ Pierre Comon, *I3S Sophia-Antipolis*;
 - ▷ Christian Jutten, *INPG Grenoble*;
 - ▷ Inbar Fijalkow, *ENSEA Cergy-Pontoise*;
 - ▷ Pascal Chevalier, *THALÈS Communications*;
 - ▷ Philippe Loubaton, *Université de Marne la Vallée*.
2. **Japão** - um grupo formado por pesquisadores do Instituto RIKEN, com várias de suas abordagens voltadas para o emprego de redes neurais artificiais, tem grande destaque na comunidade devido ao imenso número de proposições de algoritmos para desconvolução e separação de sinais através do emprego de estruturas não-lineares e do uso de técnicas

³Na verdade, é necessário ter um número mínimo de sensores, conforme será discutido nas seções seguintes.

de abordagem geométrica com álgebra em espaços de Hilbert [Haykin, 2000a]. Alguns dos nomes de maior projeção do grupo são:

- ▷ Sun-ichi Amari;
- ▷ Andrzej Cichocki.

3. Finlândia - dentro da Universidade Tecnológica de Helsinki, o grupo liderado pelo Prof. Erkki Oja, originalmente criado pelo Prof. Teuvo Kohonen (criador dos mapas auto-organizáveis), tem se dedicado de maneira bastante peculiar ao uso de técnicas de teoria da informação e separação de fontes para a aplicação em problemas práticos (tomografia, comunicações, desconvolução multicanal, recuperação de imagens). Alguns dos principais nomes são:

- ▷ Erkki Oja;
- ▷ Aapo Hyvärinen;
- ▷ Juha Karhunen.

4. Estados Unidos - neste país as pesquisas iniciais foram bem mais relacionadas a problemas de inversão de sistemas multivariável. Atualmente, os maiores nomes atuando em separação de fontes concentram-se nas áreas de redes neurais, sistemas de misturas convolutivas e aplicações em telecomunicação (detecção multiusuário, decomposição trilinear e separação de misturas convolutivas atrasadas). Vários nomes podem ser citados de todos os lugares do país, embora valha a pena ressaltar que alguns deles são originários de grupos na Europa:

- ▷ Terrence J. Sejnowski, *University of California*;
- ▷ Anthony J. Bell, *Reedwood Neuroscience Institut*;
- ▷ José Carlos Príncipe, *University of Florida*;
- ▷ Nicholas D. Sidiropoulos e Georgios B. Giannakis, *University of Minnesota*;
- ▷ Kari Torkkola, *Motorola Corporation*;
- ▷ Tulay Adali, *University of Maryland*;
- ▷ Constantinos B. Papadias, *Bell Labs, Lucent Technologies*.

5. Bélgica - apresenta uma escola mais recente, cujo principal expoente é o Dr. Lieven de Lathauwer da *Katholieke Universiteit Leuven*, cujos trabalhos em álgebra multilinear permitiram uma extensão dos problemas de estatísticas de ordem superior através da utilização de tensores, facilitando a generalização de conceitos e critérios.

No Brasil, o campo ainda é por demais incipiente. O interesse dos pesquisadores no problema de separação de fontes surge, gradativamente, suportado por aplicações específicas, tais como telecomunicações, modelagem estatístico multivariável, aprendizado não-supervisionado, dentre outros. Não é possível, pelo menos no momento, listar nomes de pesquisadores que realmente se dediquem à área, salvo algumas exceções como o Prof. A. K. Barros da Universidade Federal do Maranhão (UFMA). Deseja-se fortemente que dentro em breve esta situação seja revertida.

Atualmente, o interesse em pesquisa sobre BSS permanece significativo. Cada vez mais pesquisadores encontram no tema um problema de interesse prático e caráter generalista,

sendo, desta forma, um tema em que a prospecção teórica deve encontrar por vários anos um campo fértil de idéias e soluções extensíveis a várias áreas.

A seção seguinte vem ao encontro de formalizar matematicamente as idéias e conceitos inseridos na discussão anterior.

2.2 Modelagem matemática: um pouco de formalismo

Seja um mapeamento de um número K de fontes (sinais) e V sinais de ruído dado por:

$$\mathbf{x}(n) = \mathcal{F}(\mathbf{a}(n), \mathbf{v}(n), n), \quad (2.1)$$

em que o índice n indica dependência temporal do mapeamento e os diversos sinais são representados pelos vetores dados por:

$$\begin{aligned} \mathbf{a}(n) &= [a_1(n) \ a_2(n) \ \cdots \ a_K(n)]^T \\ \mathbf{v}(n) &= [v_1(n) \ v_2(n) \ \cdots \ v_V(n)]^T, \end{aligned} \quad (2.2)$$

em que $\mathbf{a}(n)$ e $\mathbf{v}(n)$ são os vetores das fontes e das amostras de ruído, respectivamente.

Então, o sistema $\mathcal{F}(\cdot, \cdot, \cdot)$ fornece sinais de saída, chamados de *misturas* dados por sua representação vetorial como:

$$\mathbf{x}(n) = [x_1(n) \ x_2(n) \ \cdots \ x_M(n)]^T. \quad (2.3)$$

A formulação permite representar o problema da maneira mais genérica possível, isto é:

- ▶ o mapeamento pode ser não-linear;
- ▶ pode haver dependência temporal;
- ▶ as fontes e sinais de ruído podem ser inseridos tanto de forma aditiva quanto multiplicativa.

Entretanto, algumas suposições são geralmente feitas para tornar o problema mais “tratável” e também mais próximo da maioria dos problemas práticos.

É bastante usual a consideração de que o sistema \mathcal{F} é linear e invariante com o tempo. Além disto, as fontes são geralmente consideradas mutuamente independentes e independentes do ruído. Outras considerações que, eventualmente, simplificam a formulação são feitas de acordo com o problema tratado.

Com isto, devido em grande parte à sua simplicidade mas também à sua aplicabilidade, o modelo de um sistema linear invariante no tempo com fontes estacionárias, tem sido o grande foco de estudo da grande maioria dos trabalhos dedicados à BSS.

Enfim, considerando-se linearidade do sistema de mistura e estacionariedade das fontes, pode-se escrever o modelo resultante da seguinte forma:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{v}(n), \quad (2.4)$$

em que \mathbf{H} é uma *matriz de mistura convolutiva*⁴ de ordem $M \times K$ e $V = M$, o que significa que o ruído é inserido somente nos sensores e de forma aditiva. As outras hipóteses consideradas de forma bastante usual são:

- H_{BSS1}**. Os sinais (fontes) $a_i(n)$ e $v_i(n)$ são estacionários e de média nula.
- H_{BSS2}**. As fontes $a_i(n)$ são estatisticamente independentes.
- H_{BSS3}**. Os sinais de ruído $v_i(n)$ são estatisticamente independentes e independentes das fontes $a_i(n)$.
- H_{BSS4}**. O número de sensores é maior ou igual ao número de fontes: $M \geq K$.

Pode-se então esboçar o esquema de separação cega de fontes conforme a Figura 2.1.

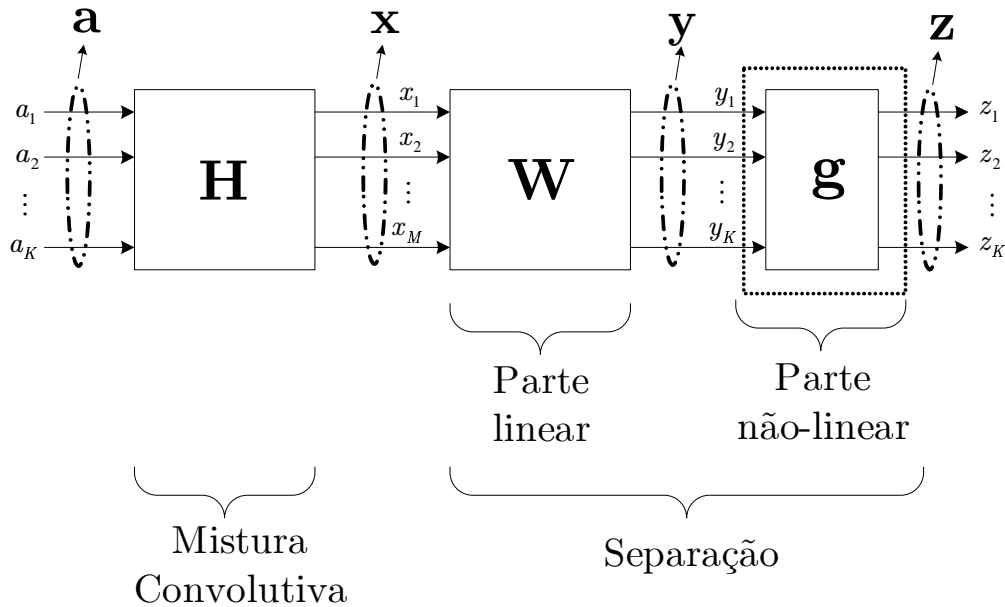


Figura 2.1: Esquema do processamento de separação cega de fontes.

⁴O termo foi cunhado pelos pesquisadores franceses Pierre Comon e Jean-François Cardoso. Na literatura americana é mais comum encontrarmos o termo matriz de mistura.

Assim, a meta da separação linear é obter uma *matriz de separação* \mathbf{W} tal que

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \hat{\mathbf{a}}(n) \quad (2.5)$$

seja uma boa estimativa dos sinais das fontes.

De acordo com a hipótese H_{BSS4} pode-se ver que, uma vez que o sistema não é sub-determinado, e se a matriz \mathbf{H} é identificada, pode-se ter uma inversão perfeita $\mathbf{W} = \mathbf{H}^{-1}$ na ausência de ruído. Quando o ruído não tem potência baixa o suficiente para ser considerado desprezível, não é possível inverter o sistema perfeitamente, e recuperar exatamente as fontes não é factível. Uma maneira de considerar o ruído de tal forma que o sistema tenha a mesma representação do caso sem ruído é escrever a Equação (2.4) da seguinte forma:

$$\mathbf{u}(n) = \mathbf{H}^\dagger \mathbf{a}^\dagger(n), \quad (2.6)$$

em que

$$\begin{aligned} \mathbf{H}^\dagger &= [\mathbf{H} \quad | \quad \mathbf{I}_M] \\ \mathbf{a}^\dagger(n) &= [\mathbf{a}^T(n) \quad | \quad \mathbf{v}^T(n)]^T, \end{aligned} \quad (2.7)$$

em que \mathbf{I}_M é a matriz identidade de ordem M .

Esta maneira de reescrever o modelo facilita visualizar que, no caso com ruído, não é possível inverter perfeitamente de forma linear o sistema (matriz de mistura convolutiva), já que a matriz \mathbf{H}^\dagger tem ordem $M \times (M + K)$. Diante da impossibilidade natural de disponibilizar um aumento na dimensionalidade (ordem) da matriz de mistura convolutiva, o que permitiria uma separação de fontes por um hiperplano, se faz necessária a inserção de um dispositivo não-linear (\mathbf{g}) na parte responsável pela separação. Este requisito deve-se ao fato que técnicas não-lineares são capazes de realizar a separação de sistemas sub-determinados, caso da Equação (2.6) [Hyvärinen et al., 2001; Cichocki & Amari, 2002]. Vale a pena ressaltar que a real localização do dispositivo não-linear pode não ser exatamente aquela representada na Figura 2.1 e estar localizado antes da parte linear da separação [Kofidis, 2001; Hyvärinen & Oja, 2000].

Um ponto de extrema importância na resolução do problema de BSS diz respeito à indeterminação quanto ao escalonamento e à ordem das fontes na saída da matriz de separação \mathbf{W} [Haykin, 2000a]. Esta indeterminação ocorre sempre que não se possui alguma informação adicional sobre o tipo de sistema (caso comum em BSS). Isto se deve ao fato de que a solução da Equação (2.5) não será alterada se uma fonte sofrer uma multiplicação por um escalar e a coluna correspondente da matriz \mathbf{H} for dividida pelo mesmo escalar. O mesmo acontece se houver uma permutação das fontes e uma correspondente permutação das colunas da matriz \mathbf{H} . Assim, o melhor que se pode conseguir é uma solução escalonada e permutada das fontes a_i de tal forma que [Hyvärinen & Oja, 2000]:

$$\mathbf{y}(n) = \mathbf{P}\mathbf{D}\mathbf{a}(n) \quad (2.8)$$

em que \mathbf{P} é uma matriz de permutação de ordem $K \times K$ e \mathbf{D} é uma matriz diagonal e inversível de ordem $K \times K$.

Em problemas de natureza prática, a questão do escalonamento não impõe um severo empecilho já que pode-se sempre realizar, de maneira bem simples, um ajuste no ganho dos sinais até atingir-se o nível original através de um controle automático de ganho (CAG). Já no caso da indeterminação da ordem, na maioria das aplicações isto não é relevante, e nos casos em que a identificação da fonte necessita ser realizada, como por exemplo em problemas de detecção multiusuário, codificações especiais são utilizadas para permitir uma classificação direta entre as estimativas e as fontes.

2.2.1 Quais estatísticas são necessárias para resolver o problema?

Seja o caso sem ruído e com $K = M$, a recuperação das fontes é feita uma vez que a matriz de mistura convolutiva é identificada. Assim sendo, resta saber quais informações estatísticas são apropriadas para que a identificação mencionada seja alcançada.

Seja a matriz de autocorrelação do vetor de observações $\mathbf{x}(n)$ dada por

$$\mathbf{R}_x = \mathbb{E} \{ \mathbf{x}(n) \mathbf{x}^T(n) \} = \mathbf{H} \mathbf{R}_a \mathbf{H}^T = \mathbf{H} \mathbf{H}^T, \quad (2.9)$$

em que o índice n é suprimido devido à estacionariedade da matriz de autocorrelação; $\mathbf{R}_a = \mathbf{I}$, de acordo com as hipóteses (H_{BSS1}-H_{BSS4}) consideradas; e supondo as fontes com potência unitária.

A Equação (2.9) sugere que a matriz de mistura convolutiva pode ser identificada através da matriz de correlação dos dados. Também é fácil verificar que uma multiplicação matricial $\mathbf{H} \mathbf{Q}^T$, em que \mathbf{Q} é uma matriz ortogonal, também soluciona a Equação (2.9). Isto significa que o uso de \mathbf{R}_x permite \mathbf{H} ser identificado a menos de um fator ortogonal. Uma importante observação é que esta limitação pode ser vista como um caso particular da indeterminação de fase contida na fatoração espectral, baseada no método da predição linear em que as fontes correspondem às amostras de inovação [Kofidis, 2001].

Observando-se novamente a Equação (2.9), pode-se intuir que a matriz de mistura convolutiva pode ser identificada como uma raiz quadrada de \mathbf{R}_x , ou seja $\mathbf{H} = \mathbf{R}_x^{\frac{1}{2}}$. A extração da raiz quadrada de uma matriz é realizada através da decomposição da matriz em função de seus autovalores (SVD, *Singular Value Decomposition*). De fato, escrevendo-se a matriz \mathbf{H} como

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{V}^T, \quad (2.10)$$

em que \mathbf{U} e \mathbf{V} são matrizes retangulares de ordem $K \times M$, tais que $\mathbf{U} \mathbf{U}^T = \mathbf{V} \mathbf{V}^T = \mathbf{I}_M$ e $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_K$, compostas dos autovetores esquerdos e direitos, respectivamente. $\mathbf{\Lambda}$ é

uma matriz diagonal contendo os K autovalores não-nulos da matriz $\mathbf{H}\mathbf{H}^T$ [Golub & Loan, 1996; Lacoume et al., 1997].

Então, extrair a raiz quadrada de uma matriz é extrair a raiz quadrada de cada um dos seus autovalores da sua representação em SVD. Reescrevendo a matriz de autocorrelação utilizando-se a Equação (2.10) tem-se:

$$\begin{aligned}\mathbf{R}_x &= \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}\mathbf{V}^T\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U} \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T.\end{aligned}\tag{2.11}$$

Da equação acima, verifica-se então que a matriz de autocorrelação permite estimar as matrizes \mathbf{U} e $\mathbf{\Lambda}$. Como a matriz \mathbf{V} não aparece na Equação (2.11), esta matriz e, por consequência, a matriz \mathbf{H} , não são identificáveis através de SOS [Lacoume et al., 1997].

Desta forma, o máximo que se consegue realizar através do uso da matriz de correlação é encontrar uma matriz de separação que é a inversa da matriz de mistura convolutiva identificada, ou seja, $\mathbf{T} = \mathbf{R}_x^{-\frac{1}{2}}$ as observações na saída do sistema de mistura são representadas por:

$$\bar{\mathbf{x}}(n) = \mathbf{T}\mathbf{x}(n),\tag{2.12}$$

de tal forma que

$$\begin{aligned}\mathbf{R}_{\bar{x}} &= \mathbf{T}\mathbf{R}_x\mathbf{T}^T \\ &= \mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}^T\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U} \\ &= \mathbf{I}_K,\end{aligned}\tag{2.13}$$

o que corresponde a *branquear* os dados, já que sinais de média zero são considerados *brancos* se a sua matriz de autocorrelação é a matriz identidade [Papoulis, 1991; Hyvärinen et al., 2001; Kofidis, 2001; Haykin, 2000a; Lacoume et al., 1997].

O processamento realizado pela Equação (2.12) corresponde à projeção dos dados $\mathbf{x}(n)$ sobre as principais direções orientadas por \mathbf{R}_x , que são determinadas pelos autovetores associados aos autovalores mais importantes da SVD da matriz de autocorrelação dos dados (não-esferatizados), o que corresponde à conhecida técnica estatística de análise em componentes principais (PCA, *Principal Component Analysis*) [Coppi & Bolasco, 1989]. De uma forma mais direta, os dados são projetados nas direções determinadas pelos principais autovetores de \mathbf{R}_x . O vetor $\bar{\mathbf{x}}$, cujos elementos são os componentes principais de \mathbf{x} , é freqüentemente referenciado como padronizado ou esferatizado⁵ [Lacoume et al., 1997; Haykin, 2000a; Kofidis, 2001]. A Figura 2.2 ilustra a composição da estrutura da matriz de dados esferatizados, a qual é ortogonal, representando elementos não-nulos como áreas destacadas.

⁵O termo foi dado por Pierre Comon e em inglês chama-se *sphered*, nesta tese, em português, será utilizado o termo esferatizado.

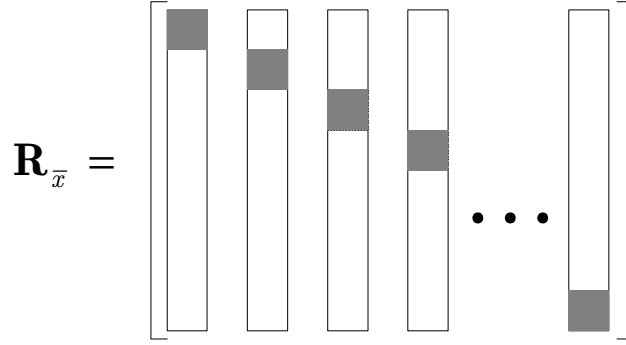


Figura 2.2: Estrutura da matriz de autocorrelação de dados esferatizados.

Como a matriz de autocorrelação é definida positiva, a matriz de branqueamento, também conhecida por *transformação de Mahalanobis* pode ser calculada por [Cichocki & Amari, 2002]:

$$\mathbf{T} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^T. \quad (2.14)$$

O processamento de esferatização reduz o problema para um no qual a matriz de mistura é ortogonal:

$$\bar{\mathbf{x}}(n) = \mathbf{THa}(n) = \mathbf{Qa}(n) \quad (2.15)$$

em que $\mathbf{QQ}^T = \mathbf{I}$.

Então, como obter as informações restantes (matriz \mathbf{V}) sobre a matriz \mathbf{H} ?

Uma alternativa é considerar um atraso arbitrário ℓ para o qual não haja duas fontes, i e j , com a mesma autocorrelação, ou seja:

$$\mathbb{E} \{a_i(n)a_i(n-\ell)\} \neq \mathbb{E} \{a_j(n)a_j(n-\ell)\} \quad \forall i \neq j, \quad (2.16)$$

o que significa que duas fontes não podem possuir a mesma densidade espectral de potência.

Refletindo a influência deste atraso na autocorrelação dos dados branqueados tem-se

$$\begin{aligned} \mathbf{R}_{\bar{\mathbf{x}}}(\ell) &= \mathbb{E} \{ \bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n-\ell) \} \\ &= \mathbf{THR}_a(\ell) \mathbf{HT}^T \\ &= \underbrace{\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^T}_{\mathbf{T}} \underbrace{\mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{V}^T}_{\mathbf{H}} \mathbf{R}_a(\ell) \underbrace{\mathbf{V} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{U}^T}_{\mathbf{H}^T} \underbrace{\mathbf{U} \mathbf{\Lambda}^{-\frac{1}{2}}}_{\mathbf{T}^T} \\ &= \mathbf{V}^T \mathbf{R}_a(\ell) \mathbf{V}. \end{aligned} \quad (2.17)$$

Através da hipótese H_{BSS2} , sabe-se que a matriz $\mathbf{R}_a(\ell)$ é diagonal (fontes independentes). Então, a Equação (2.17) representa uma decomposição em valores singulares de $\mathbf{R}_{\bar{\mathbf{x}}}(\ell)$. Através

da expressão (2.16) sabe-se que os autovalores são distintos e isto implica que as colunas da matriz \mathbf{Q} são determinadas unicamente através da decomposição de $\mathbf{R}_{\bar{x}}(\ell)$ sujeito, talvez, a uma mudança de ordem e/ou escala [Hyvärinen & Oja, 2000].

Desta forma, a identificação do restante da informação da matriz \mathbf{H} depende da matriz $\mathbf{R}_a(\ell)$ não ser identidade, ou seja, a correlação das fontes não pode ser a mesma para um atraso arbitrário ℓ , o que implica que algoritmos que utilizam somente SOS estão limitados a separar somente fontes que possuem espectro diferente. Este tipo de algoritmo, conhecido na literatura como *Algorithm for Multiple Unknown Signals Extraction* (AMUSE) utiliza somente estatísticas de segunda ordem para encontrar a solução do sistema de mistura [Tong et al., 1991].

Na prática, mesmo para espectros de potência diferentes mas que são similares, a separação não ocorre de maneira satisfatória, mesmo sendo teoricamente possível⁶ [Kofidis, 2001]. Isto é, o uso apenas das estatísticas de segunda ordem não será aplicável para os casos de fontes consideradas brancas, ou ainda independentes e identicamente distribuídas (i.i.d.), uma consideração muito usual e realista em sistemas de comunicação digital.

Este ponto ilustra o limite do método através da utilização de PCA, que emprega somente estatísticas de segunda ordem. Neste caso, o uso de SOS somente permite a separação cega quando as fontes são gaussianas, pois neste caso os sinais podem ser completamente descritos através de suas médias e variâncias. De fato, nas considerações anteriores, apenas a condição de decorrelação foi usada e isto confirma o uso para sinais gaussianos, nos quais a condição de independência e decorrelação coincidem [Picinbono, 1993; Papoulis, 1991]. Além do mais, a maioria dos sinais encontrados em aplicações práticas, tais como voz, música, dados e imagens, são não-gaussianos.

Bem, isto responde à pergunta inicial de quais estatísticas são necessárias para a separação cega de fontes, mas não qual método pode então ser utilizado para realizá-la. A próxima seção é dedicada a uma ferramenta que se propõe a resolver o problema sob a consideração de uma maior liberdade na consideração da natureza das fontes em contraposição ao método de PCA. Este método chama-se *Independent Component Analysis* e é atualmente bastante conhecido através de sua sigla, ICA.

⁶Aspectos numéricos não possibilitam a diferenciação de fontes com espectros similares através do cálculo dos autovalores da matriz de autocorrelação do sinal recebido [Lacoume et al., 1997].

2.3 A técnica de análise por componentes independentes - *Independent Component Analysis*

O trabalho de Pierre Comon [Comon, 1994] iniciou uma nova era no campo do estudo de técnicas de separação de fontes. A partir de suas idéias, a restrição sobre a fdp dos sinais das fontes pode ser relaxada de maneira a tornar as hipóteses mais próximas daquelas que são observadas em situações práticas. Sua contribuição de formalização e esclarecimento matemático neste aspecto foi tão importante que a técnica em si é, com alguma frequência, confundida com o problema de BSS em sua forma geral.

Mas em que consiste a diferença de ICA para PCA?

A técnica ICA tem como premissa a hipótese de que as fontes são mutuamente independentes. Procura-se então obter um conjunto de fontes também independentes na saída dos filtros de separação mesmo que seus espectros não sejam distinguíveis, premissa básica da técnica PCA [Comon, 1994; Kofidis, 2001; Cao & Liu, 1996; Hyvärinen & Oja, 2000]. Para isso, a única restrição imposta sobre a função de densidade de probabilidade das fontes é de que as mesmas não sejam gaussianas, ou então que, *no máximo uma das fontes seja gaussiana* [Kofidis, 2001; Hyvärinen & Oja, 2000].

A imposição de não-gaussianidade se deve ao fato de que uma soma de variáveis aleatórias gaussianas fornece uma distribuição de variáveis aleatórias conjuntamente gaussianas, o que impossibilita inferir qualquer informação sobre as fontes somente a partir das observações [Kofidis, 2001; Hyvärinen & Oja, 2000].

De forma bastante intuitiva, pode-se afirmar que o ponto chave da ICA é a não-gaussianidade das fontes. Então, de maneira muito natural, é esperado que os critérios utilizados para obter a separação cega levem em consideração esta característica dos sinais e busquem explorá-la.

O problema que surge a partir de agora reflete-se na seguinte pergunta:



Como medir a não-gaussianidade das fontes e utilizar o fato de que as mesmas são estatisticamente independentes para separá-las?

Para responder a tal questão, são colocados, a seguir, os fundamentos matemáticos e estatísticos da técnica de análise por componentes independentes que se baseiam na não-gaussianidade e na independência estatística das fontes.

2.3.1 Independência estatística

Supor que os sinais das fontes são independentes significa, estatisticamente, que a fdp conjunta do vetor $\mathbf{a}(n)$ é igual ao produto das fdps marginais dos componentes do vetor K -dimensional. Ou seja, deve-se ter:

$$p_{\mathbf{a}}(\mathbf{a}) \triangleq \prod_{i=1}^K p_{a_i}(a_i). \quad (2.18)$$

No caso das estimativas das fontes, os sinais de saída oriundos da aplicação de uma transformação sobre sinais contaminados visando a separação das fontes também devem apresentar a mesma relação entre as fdps conjunta e marginais para que a separação seja assegurada, ou seja:

$$p_{\mathbf{y}}(\mathbf{y}) \triangleq \prod_{i=1}^K p_{y_i}(y_i), \quad (2.19)$$

Em termos mais precisos, a meta de ICA é, sob a hipótese de independência das fontes, encontrar uma matriz de separação cujas saídas sejam também fontes mutuamente independentes.

Em termos de estatísticas dos sinais, se as Equações (2.18) e (2.19) são válidas, tem-se, entre outros resultados:

$$E\{y_1 \cdot y_2 \cdots y_K\} = E\{y_1\} \cdot E\{y_2\} \cdots E\{y_K\}. \quad (2.20)$$

De um ponto de vista geométrico, pode-se caracterizar a diferença entre ICA e PCA da seguinte maneira: PCA procura reduzir a dimensão (para efetuar a separação) dos dados projetando-os em uma dimensão de maneira que a *variância dos dados seja máxima* (direções determinadas pelos maiores autovalores) e ICA procura reduzir a dimensão dos dados de tal forma que a *estrutura dos dados seja preservada ao máximo*. A Figura 2.3 ilustra este conceito para dados bidimensionais. Neste caso, as projeções estão representadas por eixos ortogonais, o que não é regra geral.

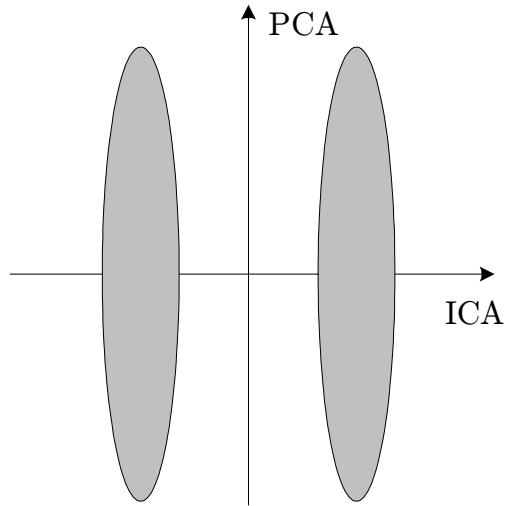


Figura 2.3: Visão geométrica da diferença entre ICA e PCA.

Deve-se, ainda, levantar alguns pontos importantes:

Descorrelação estatística

Um conceito menos restritivo que o de independência é o de *descorrelação*. Por definição, um conjunto K -dimensional de variáveis aleatórias (v.a.) é formado de variáveis descorrelacionadas se a sua covariância é zero, ou seja,

$$\mathbb{E} \{y_1 \cdot y_2 \cdots y_K\} - \mathbb{E} \{y_1\} \cdot \mathbb{E} \{y_2\} \cdots \mathbb{E} \{y_K\} = 0. \quad (2.21)$$

Se as variáveis são independentes, elas são descorrelacionadas, todavia, descorrelação não implica em independência. A única classe de v.a. na qual a descorrelação implica em independência é a de fdp gaussiana que é completamente descrita por seus momentos de primeira e de segunda ordens [Papoulis, 1991]. Desta forma, para v.a. não gaussianas, vale o seguinte mapeamento

$$\text{Independência} \not\Rightarrow \text{Descorrelação}$$

Como será visto na seqüência, uma vez que independência implica em descorrelação, vários métodos de ICA restringem o problema de estimação das fontes de tal forma que as estimativas dos termos independentes sejam sempre descorrelacionadas⁷.

⁷Posteriormente, a necessidade da estimação dos sinais ou de suas fdps ficará evidente.

Mas e se as fontes forem gaussianas?

Como já mencionado, se as fontes forem gaussianas não há como separá-las através de métodos que utilizem somente a independência estatística das fontes.

A título de ilustração, considera-se duas variáveis a_1 e a_2 gaussianas normalizadas⁸ e uma matriz de ordem (2×2) de mistura convolutiva ortogonal, então, a fdp conjunta das duas variáveis x_1 e x_2 no receptor é dada por:

$$p_{x_1, x_2}(x_1, x_2) = \frac{1}{2\pi} \cdot \exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

Como a fdp conjunta na saída da matriz de mistura é simétrica, a partir das amostras recebidas não se pode inferir nenhuma informação sobre as colunas da matriz e a mesma não pode ser estimada. Com mais rigor matemático, qualquer transformação ortogonal sobre v.a. gaussianas têm a mesma fdp conjunta dos dados gaussianos e, além disso, as variáveis são independentes [Hyvärinen & Oja, 2000]. Este é o motivo da impossibilidade do uso de v.a. gaussianas em problemas de BSS.

A Figura 2.4 ilustra o conceito acima através das densidades conjuntas de misturas de duas fontes em um sistema de mistura instantânea aleatória com fontes gaussianas e fontes de distribuição uniforme. Observa-se claramente a simetria das misturas quando as fontes são gaussianas (Figura 2.4(a)) e a existência de direções preferenciais quando as fontes são de distribuição uniforme (Figura 2.4(b)).

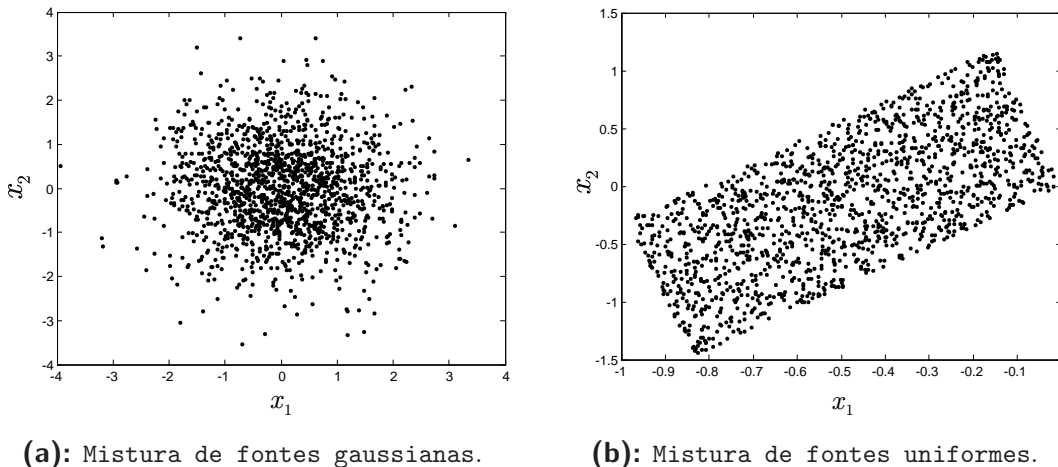


Figura 2.4: Funções de distribuição de probabilidade conjuntas de variáveis gaussianas e uniformes em um sistema de mistura instantânea.

⁸Média nula e variância unitária.

No caso de variáveis não-gaussianas entretanto, a independência garante que as fontes são realmente distintas e é neste critério que são baseados a maioria dos métodos de separação cega de fontes.

Mas quais as ferramentas que possibilitam medir a independência das fontes? Esta questão é abordada a seguir nos itens 2.3.2 a 2.3.6.

2.3.2 Entropia e informação mútua

Sabendo que a *entropia* de uma variável aleatória, segundo a teoria de Shannon [Shannon, 1948], fornece uma medida da quantidade média de incerteza contida na v.a., pode-se definir, para um vetor de v.a. contínuas⁹ \mathbf{x} , sua entropia $\mathcal{H}(\mathbf{x})$ da seguinte forma:

$$\mathcal{H}(\mathbf{x}) \triangleq -\mathbb{E} \{ \ln [p_x(\mathbf{x})] \} = - \int_{-\infty}^{\infty} p_x(\mathbf{x}) \cdot \ln [p_x(\mathbf{x})] d\mathbf{x}. \quad (2.22)$$

Na verdade, a Equação (2.22) designa a *entropia diferencial* de \mathbf{x} , uma vez que a incerteza contida numa v.a. contínua é infinita.

Seja agora outro vetor de v.a. \mathbf{y} , e $p_{x|y}(\mathbf{x}|\mathbf{y})$ a fdp de \mathbf{x} condicionada a \mathbf{y} . Uma medida da incerteza remanescente em \mathbf{x} após a observação de \mathbf{y} é dada pela *entropia condicional*, representada por:

$$\mathcal{H}(\mathbf{x}|\mathbf{y}) = -\mathbb{E} \{ \ln [p_{x|y}(\mathbf{x}|\mathbf{y})] \} = - \int p_{x,y}(\mathbf{x}, \mathbf{y}) \cdot \ln [p_{x|y}(\mathbf{x}|\mathbf{y})] d\mathbf{x} d\mathbf{y}, \quad (2.23)$$

em que

$$p_{x,y}(\mathbf{x}, \mathbf{y}) = p_{x|y}(\mathbf{x}|\mathbf{y}) p_y(\mathbf{y}) \quad (2.24)$$

é a fdp conjunta de \mathbf{x} e \mathbf{y} .

Então, a conhecida relação [Gibson, 1993; Haykin, 1998]

$$I(\mathbf{x}, \mathbf{y}) = \mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{x}|\mathbf{y}) \quad (2.25)$$

representa a informação contida em \mathbf{x} após a observação de \mathbf{y} . Esta grandeza é chamada de *informação mútua* entre \mathbf{x} e \mathbf{y} .

⁹Para o caso de valores discretos, tem-se a seguinte forma:

$$\mathcal{H}(\mathbf{x}) \triangleq -\mathbb{E} \{ \ln [p_x(\mathbf{x})] \} = - \sum_i p_i \cdot \ln [p_i],$$

em que p_i é a probabilidade do evento i .

A informação mútua será de grande importância porque ela é sempre não-negativa e assume o valor zero se e somente se \mathbf{x} e \mathbf{y} são independentes. De fato, isso é esperado já que, para variáveis independentes, a observação de uma delas não fornece nenhuma informação sobre a outra. Isto é fácil de verificar através da Equação (2.25) que, para $I(\mathbf{x}, \mathbf{y}) = 0$, tem-se $\mathcal{H}(\mathbf{x}) = \mathcal{H}(\mathbf{x}|\mathbf{y})$. Logo, a informação mútua apresenta-se como uma referência confiável para independência estatística.

2.3.3 Divergência de Kullback-Leibler

Um problema freqüentemente encontrado em inferência estatística é a definição da medida de similaridade entre funções. A Divergência de Kullback-Leibler (KLD, *Kullback-Leibler Divergence*) é uma medida de similaridade entre duas funções estritamente positivas [Kullback & Leibler, 1951]. É também chamada de *entropia diferencial* ou ainda *entropia cruzada* [Deco & Obradovic, 1996].

Matematicamente, a definição da KLD pode ser explicitada como:

$$D(p_x(\mathbf{x})||g_x(\mathbf{x})) \triangleq \int_{-\infty}^{\infty} p_x(\mathbf{x}) \cdot \ln \left[\frac{p_x(\mathbf{x})}{g_x(\mathbf{x})} \right] dx, \quad (2.26)$$

em que $p_x(\cdot)$ e $g_x(\cdot)$ são duas funções estritamente positivas. Para o caso discreto tem-se:

$$D(p_x(\mathbf{x})||g_x(\mathbf{x})) \triangleq \sum_{x \in \aleph} p_x \cdot \ln \left[\frac{p_x(\mathbf{x})}{g_x(\mathbf{x})} \right], \quad (2.27)$$

em que \aleph é o conjunto domínio da variável x .

É bastante usual a utilização da KLD para a comparação entre duas *funções de densidade de probabilidade*. Neste caso, a fdp $g_x(\mathbf{x})$ realiza o papel de uma *função de referência*.

Deve-se observar que $D(p_x(\mathbf{x})||g_x(\mathbf{x})) \neq D(g_x(\mathbf{x})||p_x(\mathbf{x}))$, significando que $D(\bullet||\circ)$ não é uma medida de distância¹⁰, embora seja comum na literatura ser referenciada como distância de Kullback-Leibler. Encontra-se ainda a denominação de divergência assimétrica por ser, matematicamente, uma semi-distância [Kofidis, 2001].

A KLD tem algumas propriedades importantes [Haykin, 1998]:

¹⁰Não é uma medida de distância no espaço euclidiano, mas pode ser apresentada no contexto de geometria diferencial como uma métrica de Riemman no espaço das distribuições de probabilidade [Deco & Obradovic, 1996].

$\mathbf{P}_{\text{KLD}}1.$ é sempre de valor positivo ou zero; KLD é zero para o caso específico de $p_x(\mathbf{x}) = g_x(\mathbf{x})$.

$\mathbf{P}_{\text{KLD}}2.$ é invariante com relação às seguintes mudanças nos componentes do vetor \mathbf{x} ;

- permutação de ordem
- escalonamento de amplitude
- transformação monotônica não-linear

Utilizando-se duas funções estritamente positivas, preserva-se ainda a propriedade da KLD possuir apenas valores positivos ou zero.

A propriedade $\mathbf{P}_{\text{KLD}}1$ é de extremo interesse para a construção de funções de otimização, pois proporciona a elaboração de uma classe de funcionais que têm apelo em teoria da informação, por utilizar o princípio de mínima entropia através da comparação entre funções [Cavalcante, 2001; Montalvão, 2000].

Usando o espaço das fdps como o espaço da métrica das mesmas, podemos escrever, através do uso da KLD, a seguinte relação entre três fdps distintas, $p_a(\mathbf{a})$, $\tilde{p}_a(\mathbf{a})$ e $p_y(\mathbf{y})$ [Haykin, 1998]:

$$D(p_y(\mathbf{y})||p_a(\mathbf{a})) = D(p_y(\mathbf{y})||\tilde{p}_a(\mathbf{a})) + D(\tilde{p}_a(\mathbf{a})||p_a(\mathbf{a})). \quad (2.28)$$

Para a equação acima ser válida, deve-se ter respeitadas as seguintes condições, que são facilmente atendidas em problemas típicos de BSS:

- 1) $\mathbf{y} = \mathbf{B}\mathbf{a}$, em que \mathbf{y} e \mathbf{a} são vetores e \mathbf{B} é uma matriz não-diagonal.
- 2) $\tilde{p}_a(\mathbf{a})$ é a fdp marginal de cada elemento a_i do vetor \mathbf{a} , obtida a partir de $p_a(\mathbf{a})$.

A relação descrita pela Equação (2.28) é também conhecida por *decomposição de Pitágoras* por poder ser vista como uma extensão, para o espaço das fdps, do teorema de Pitágoras para triângulos retângulos no espaço Euclidiano [Kofidis, 2001; Haykin, 1998].

Usando a KLD, pode-se ainda escrever a informação mútua como:

$$\begin{aligned} I(\mathbf{x}, \mathbf{y}) &= \int p_{x,y}(\mathbf{x}, \mathbf{y}) \cdot \ln \left[\frac{p_{x,y}(\mathbf{x}, \mathbf{y})}{p_x(\mathbf{x}) p_y(\mathbf{y})} \right] d\mathbf{x} d\mathbf{y} \\ &= D(p_{x,y}(\mathbf{x}, \mathbf{y}) || p_x(\mathbf{x}) p_y(\mathbf{y})), \end{aligned} \quad (2.29)$$

o que indica que a informação mútua é uma medida da dependência estatística. Além disso, pode-se ainda definir de forma mais elegante a informação mútua entre os componentes de um vetor K -dimensional (qualquer dimensão) \mathbf{y} como sendo a KLD entre $p_y(\mathbf{y})$ e $\tilde{p}_y(\mathbf{y})$ dada

por [Kofidis, 2001]:

$$\begin{aligned}
 I(\mathbf{y}) &= D(p_{\mathbf{y}}(\mathbf{y}) \parallel \tilde{p}_{\mathbf{y}}(\mathbf{y})) \\
 &= \int \int \cdots \int p_{\mathbf{y}}(y_1, y_2, \dots, y_K) \cdot \ln \left[\frac{p_{\mathbf{y}}(y_1, y_2, \dots, y_K)}{\prod_{i=1}^K p_{y_i}(y_i)} \right] dy_1 dy_2 \cdots dy_K, \quad (2.30)
 \end{aligned}$$

que assume valor igual à zero se e somente se todos os componentes de \mathbf{y} são mutuamente independentes.

Através da equação anterior, pode-se ainda escrever que [Haykin, 1998]:

$$I(\mathbf{y}) = -\mathcal{H}(\mathbf{y}) + \sum_{i=1}^K \mathcal{H}(y_i), \quad (2.31)$$

o que significa que minimizar a informação mútua entre os componentes do vetor \mathbf{y} representa tornar a entropia de \mathbf{y} o mais próxima possível da soma de suas entropias marginais.

2.3.4 Negentropia

A *negentropia* é uma medida de “não-gaussianidade” baseada na medida de entropia diferencial [Touzni, 1998; Hyvärinen & Oja, 2000; Comon, 1994].

Matematicamente, a negentropia é a diferença entre a entropia da v.a. multidimensional¹¹ \mathbf{y} e a entropia de uma v.a. \mathbf{y}^G de distribuição gaussiana e com os mesmos momentos de ordem um e dois (média e variância) de \mathbf{y} . Assim, temos que $p_{\mathbf{y}^G}(\mathbf{y}) = N(\mu_{\mathbf{y}}, \sigma_{\mathbf{y}})$, em que $\mu_{\mathbf{y}}$ e $\sigma_{\mathbf{y}}^2$ são respectivamente a média e a variância da variável \mathbf{y} , sendo sua negentropia dada por:

$$N_G(\mathbf{y}) \triangleq \mathcal{H}(\mathbf{y}^G) - \mathcal{H}(\mathbf{y}). \quad (2.32)$$

Desta forma, a negentropia é uma grandeza não-negativa pois uma variável gaussiana possui sempre a maior entropia [Picinbono & Barret, 1990]. A prova de que a variável gaussiana possui maior entropia, para uma determinada variância, é dada no Apêndice A.

Através da divergência de Kullback-Leibler, a definição de negentropia também pode ser dada por:

$$N_G(\mathbf{y}) \triangleq D(p_{\mathbf{y}}(\mathbf{y}) \parallel p_{\mathbf{y}^G}(\mathbf{y})). \quad (2.33)$$

¹¹A negentropia pode, logicamente, ser expressa para variáveis unidimensionais.

O uso desta medida torna o processamento de separação das fontes mais robusto, uma vez que a medida de negentropia é o estimador ótimo da não-gaussianidade [Hyvärinen & Oja, 2000]. Além disso, a construção de um funcional que maximize a negentropia e, por conseqüência, torne as fontes estimadas na saída do sistema de separação o menos gaussianas possível, é de fácil dedução [Hyvärinen et al., 2001].

Entretanto, a desvantagem de tal método reside no fato de que a maximização de um funcional que leva em conta a negentropia pressupõe o conhecimento ou a estimativa da função de densidade de probabilidade das fontes na saída do sistema de separação. Como geralmente não se dispõe do conhecimento *a priori* das fontes, esta estimativa possivelmente deve ser realizada de forma não-paramétrica. Este fato limita bastante a utilização da negentropia como critério de otimização para separação autodidata de fontes, motivando o uso de critérios sub-ótimos de medida de não-gaussianidade ou de aproximações no cálculo da negentropia, como será discutido a seguir.

2.3.5 Kurtosis

A *kurtosis*, ou cumulante de quarta ordem, é uma medida clássica de gaussianidade e, para uma variável real y , é dada por [Papoulis, 1991; Picinbono, 1993]:

$$\mathcal{K}\{y\} \triangleq \mathbb{E}\{y^4\} - 3 \cdot (\mathbb{E}\{y^2\})^2. \quad (2.34)$$

De acordo com a definição¹² da Equação (2.34), a *kurtosis* tem a seguinte faixa de valores dependendo do tipo de fdp:

- ▶ Distribuição gaussiana: $\mathcal{K}\{y\} = 0$
- ▶ Distribuição sub-gaussiana: $\mathcal{K}\{y\} < 0$
- ▶ Distribuição super-gaussiana¹³: $\mathcal{K}\{y\} > 0$

Uma questão importante que surge diz respeito a como se obtêm distribuições destes três tipos.

¹²Na literatura encontra-se também a *kurtosis* dada por:

$$\mathcal{K}\{y\} = \frac{\mathbb{E}\{y^4\}}{(\mathbb{E}\{y^2\})^2},$$

que corresponde ao momento de quarta ordem normalizado [Papadias, 2000a; Papoulis, 1991].

¹³A notação aqui descrita é mais usada na literatura de engenharia. Na literatura de estatística é comum encontrar-se as expressões *mesokurtica*, *platykurtica* e *leptokurtica* para descrever distribuições gaussianas, sub-gaussiana e super-gaussiana, respectivamente.

As distribuições gaussiana, sub-gaussiana e super-gaussiana podem ser obtidas através da *distribuição Gaussiana generalizada* para variáveis de média nula [Haykin, 1994; Cichocki & Amari, 2002] dada por:

$$p_x(x) = \frac{r}{2\sigma_x^r \Gamma\left(\frac{1}{r}\right)} \exp\left(-\frac{1}{r} \cdot \left|\frac{x}{\sigma_x}\right|^r\right), \quad (2.35)$$

em que r é um parâmetro variável que determina o tipo da distribuição (normal, super-gaussiana e sub-gaussiana), $\Gamma(r) = \int_0^\infty x^{r-1} \exp(-x) dx$ é a função Gamma e $\sigma_x^r = \mathbb{E}\{|x|^r\}$ é uma medida generalizada da variância conhecida como *dispersão da distribuição* [Cichocki & Amari, 2002].

A Figura 2.5 ilustra as distribuições obtidas através da utilização da Equação (2.35).

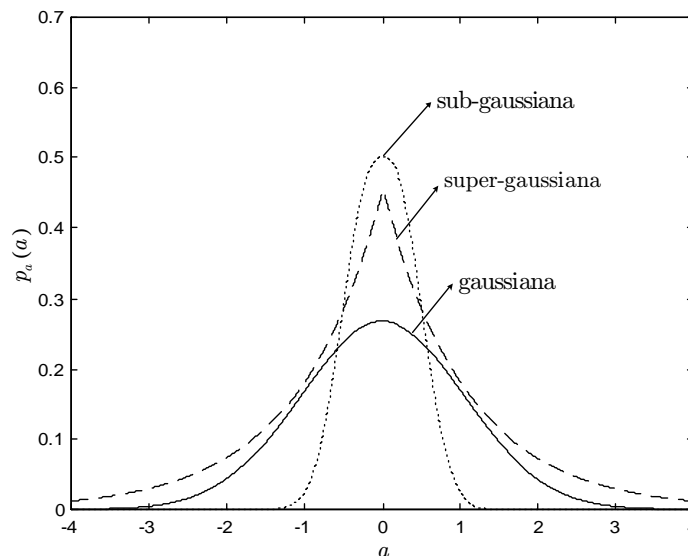


Figura 2.5: Exemplos de fdps gaussiana, sub-gaussiana e super-gaussiana.

A *kurtosis* ainda possui algumas propriedades importantes, utilizadas na concepção de um critério para separação de fontes. Dentre elas, para variáveis aleatórias y_1 e y_2 , tem-se [Hyvärinen & Oja, 2000; Hyvärinen et al., 2001]:

$$\mathcal{K}\{y_1 + y_2\} = \mathcal{K}\{y_1\} + \mathcal{K}\{y_2\} \quad (2.36a)$$

$$\mathcal{K}\{\alpha \cdot y\} = \alpha^4 \cdot \mathcal{K}\{y\}, \quad (2.36b)$$

em que α é um escalar.

As propriedades descritas pela Equação (2.36) são, basicamente, aquelas que devem ser calculadas na saída do filtro (matriz) de separação, uma vez que os sinais processados são, na maioria dos casos estudados, composições lineares de outras fontes.

Na linha de buscar respostas à questão colocada em destaque no início da Seção 2.3, o interesse aqui é derivar um algoritmo que, por exemplo, maximize a *kurtosis* de modo que a gaussianidade dos sinais à saída seja minimizada. Esta foi a abordagem usada para separação em [Papadias, 2000a, 2000b], a qual será vista mais adiante. Porém, embora a *kurtosis* seja uma medida relativamente simples de ser realizada, o uso deste cumulante apresenta algumas limitações.

2.3.6 Funções de contraste

As *funções de contraste* foram formalmente definidas por [Comon, 1994] e se referem a um conjunto de funções utilizadas como critérios de otimização, as quais atingem seu mínimo somente quando a total separação das fontes é alcançada. De uma certa maneira, as funções de contraste já eram utilizadas no processamento autodidata de sinais, mas não havia um critério específico para determinar quais funções poderiam ser utilizadas como contraste.

Uma função $\Psi(\cdot)$, no espaço de K fdps (distintas ou não) é dita ser um *contraste* se respeita as seguintes condições [Comon, 1994; Kofidis, 2001]:

C_{cont}1. $\Psi(p_y)$ é invariante a permutações:

$$\Psi(p_{Py}) = \Psi(p_y) \text{ para qualquer matriz de permutação } \mathbf{P}$$

em que $\Psi(p_y)$ representa uma função definida a partir das variáveis da fdp p_y e Py representa uma variável permutada de y .

C_{cont}2. $\Psi(p_y)$ é invariante a mudanças de escala:

$$\Psi(p_{Dy}) = \Psi(p_y) \text{ para qualquer matriz diagonal } \mathbf{D}$$

em que Dy representa uma variável escalonada de y .

C_{cont}3. Se \mathbf{y} possui componentes independentes, então:

$$\Psi(p_{Wy}) \leq \Psi(p_y) \text{ para qualquer matriz inversível } \mathbf{W}$$

C_{cont}4. A igualdade em C_{cont}3 deve ser respeitada se e somente se a matriz \mathbf{W} é uma matriz de permutação generalizada, ou seja, $\mathbf{W} = \mathbf{PD}$, em que \mathbf{P} é uma matriz de permutação e \mathbf{D} é uma matriz diagonal inversível.

Na verdade, basta que a função respeite as condições C_{cont}1-C_{cont}3 para ser considerada contraste. Entretanto, a condição C_{cont}4 é necessária para se fornecer uma correta solução

para o problema de BSS. Um contraste que satisfaça a última condição é dito *discriminante*. Em geral, utiliza-se o termo contraste referindo-se ao contraste discriminante [Comon, 1994].

Uma importante função de contraste pode ser definida através da informação mútua. A Figura 2.6 ilustra a informação mútua e a negentropia num espaço de funções de densidade de probabilidade que se encontram em subespaços ortogonais, tais como as fdps gaussianas e as fdp fatoráveis em fdps marginais [Cardoso, 2000].

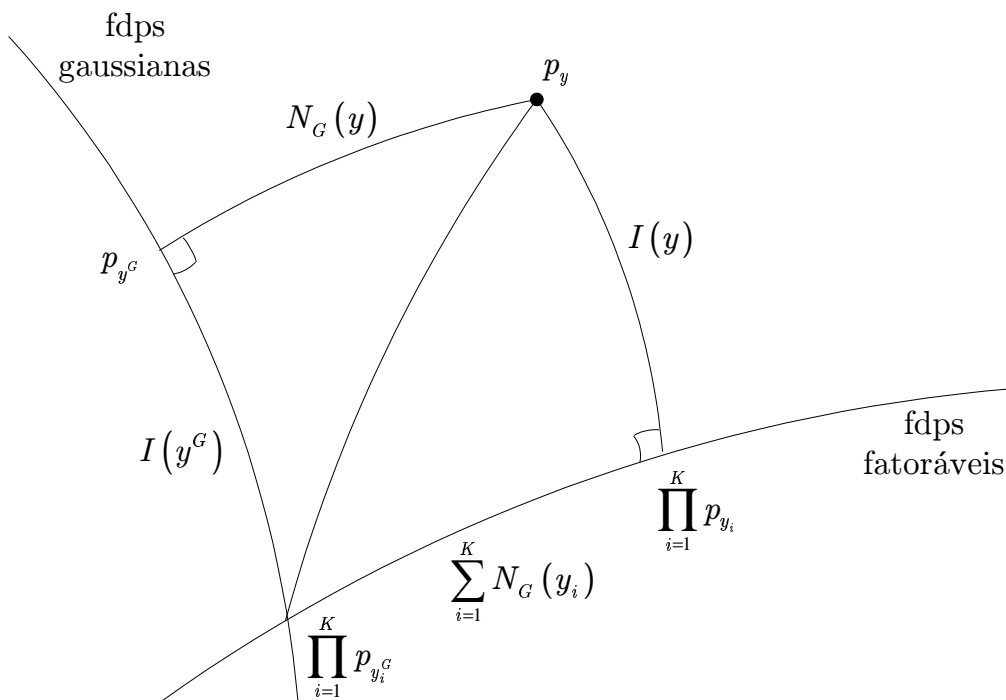


Figura 2.6: Interpretação geométrica no espaço das distribuições entre a informação e a negentropia.

Reescrevendo a Equação (2.28) para as distribuições representadas na Figura 2.6, tem-se [Kofidis, 2001]:

$$\begin{aligned} D\left(p_y \parallel \prod_{i=1}^K p_{y_i^G}\right) &= D(p_y \parallel p_{y^G}) + D\left(p_{y^G} \parallel \prod_{i=1}^K p_{y_i^G}\right) \\ &= N_G(\mathbf{y}) + I(\mathbf{y}^G), \end{aligned} \quad (2.37)$$

e também

$$\begin{aligned} D \left(p_y \parallel \prod_{i=1}^K p_{y_i^G} \right) &= D \left(p_y \parallel \prod_{i=1}^K p_{y_i} \right) + D \left(\prod_{i=1}^K p_{y_i} \parallel \prod_{i=1}^K p_{y_i^G} \right) \\ &= I(\mathbf{y}) + \sum_{i=1}^K N_G(y_i), \end{aligned}$$

em que $I(\mathbf{y}^G)$, na Equação (2.37), representa a informação mútua da “versão gaussiana” de \mathbf{y} , ou seja, a informação mútua de \mathbf{y}^G . Das equações anteriores pode-se, então, escrever:

$$I(\mathbf{y}) = I(\mathbf{y}^G) + \left(N_G(\mathbf{y}) - \sum_{i=1}^K N_G(y_i) \right). \quad (2.38)$$

A partir da relação descrita na Equação (2.38), pode-se verificar que a informação mútua possui dois termos responsáveis pela redundância em p_y . O termo $I(\mathbf{y}^G)$ mede a informação fornecida pelas SOS e o termo $\left[N_G(\mathbf{y}) - \sum_{i=1}^K N_G(y_i) \right]$ mede a informação provida pelas HOS, conseqüentemente medindo a não-gaussianidade do sinal.

Relembrando que separar as fontes corresponde a obter-se $I(\mathbf{y}) = 0$, neste ponto pode-se observar que as transformações de pré-branqueamento, discutidas na Seção 2.2.1, visam anular o primeiro termo da Equação (2.38), enquanto os métodos baseados em HOS são empregados para minimizar o segundo termo, sujeito à condição de que a matriz de separação seja ortogonal (condição criada pelo pré-branqueamento).

Este tipo de processamento em duas etapas, não é necessariamente ótimo. Idealmente, os termos da Equação (2.38) deveriam ser minimizados conjuntamente. Entretanto, o uso do pré-branqueamento antes do uso de técnicas de HOS, para minimizar o termo relativo à negentropia, pode evitar o surgimento de problemas de precisão numérica de operações não-lineares para a inversão da matriz de mistura convolutiva [Lacoume et al., 1997]. Além disto, o fato da matriz de separação ser ortogonal pode, ocasionalmente, facilitar a derivação de algoritmos de separação.

Finalmente, utilizando-se a definição de informação mútua através da KLD da Equação (2.29), e a propriedade $P_{\text{KLD}2}$ da KLD descrita na seção 2.3.3, mostra-se de forma direta que o negativo¹⁴ da informação mútua é uma função contraste para o caso de análise por componentes independentes [Comon, 1994]. Ou seja, pode-se escrever

$$\Psi_{\text{ICA}}(p_y) = -I(\mathbf{y}). \quad (2.39)$$

¹⁴O valor negativo advém do fato de uma função contraste ser *maximizada* e no caso da informação mútua buscar *minimizá-la*.

Esta equação é conhecida como contraste da informação mútua negativa (NMI, *Negative Mutual Information*).

Pode-se, ainda, simplificar a expressão do contraste para a NMI de forma a explicitar a dependência com a matriz de separação. Através da hipótese de inversibilidade da matriz \mathbf{W} , a fdp do sinal na saída da matriz de separação pode ser dada através da fórmula de inversão de fdps [Therrien, 1992; Papoulis, 1991; Picinbono, 1993]:

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{p_{\mathbf{x}}(\mathbf{W}^{-1}\mathbf{y})}{|\det(\mathbf{W})|}. \quad (2.40)$$

Isto implica que se pode tentar escrever a entropia de \mathbf{y} em termos de \mathbf{x} como:

$$\mathcal{H}(\mathbf{y}) = \mathcal{H}(\mathbf{x}) + \ln [|\det(\mathbf{W})|]. \quad (2.41)$$

Daí, a Equação (2.31) pode ser escrita como

$$I(\mathbf{y}) = -\mathcal{H}(\mathbf{x}) - \ln [|\det(\mathbf{W})|] + \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K p_{y_i}(y_i) \right] \right\}, \quad (2.42)$$

e, como $\mathcal{H}(\mathbf{x})$ não depende de \mathbf{W} , o contraste relacionado à NMI escreve-se na forma:

$$\Psi_{\text{ICA}}(\mathbf{W}) = \ln [|\det(\mathbf{W})|] - \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K p_{y_i}(y_i) \right] \right\}, \quad (2.43)$$

que é a função que se busca maximizar.

Deve-se notar que o primeiro termo de $\Psi_{\text{ICA}}(\mathbf{W})$, quando os sinais são branqueados e \mathbf{W} é ortogonal, torna-se zero. Então, a maximização da função contraste $\Psi_{\text{ICA}}(\mathbf{W})$ corresponde à minimização da soma de entropias dos componentes de \mathbf{y} . E ainda, como já mencionado anteriormente, máxima entropia corresponde à gaussianidade. Deve-se então procurar tornar as fontes o menos gaussianas possíveis, o que necessita, logicamente, de estatísticas de ordem maior que dois.

Este conceito de mínima entropia ou mínima gaussianidade foi o ponto central de trabalhos pioneiros em desconvolução cega [Donoho, 1981; Benveniste et al., 1980; Bell & Sejnowski, 1995], os quais serão comentados posteriormente.

Aproximações da negentropia

Conforme discutido anteriormente, embora a negentropia seja teoricamente uma função de contraste factível para problemas de BSS, a sua utilização baseia-se na estimativa da função densidade de probabilidade dos sinais na saída da matriz de separação. Claramente, esta estimativa nem sempre é fácil ou mesmo factível. Sendo assim, opta-se por utilizar aproximações da negentropia baseada na utilização direta dos dados nos sensores, para que seja construída uma função contraste adequada.

Uma aproximação clássica é através do uso de momentos de ordem superior, dada por [Hyvärinen, 1999]:

$$N_G(\mathbf{y}) \approx \frac{1}{12} [E \{\mathbf{y}^3\}]^2 + \frac{1}{48} [\mathcal{K} \{\mathbf{y}\}]^2, \quad (2.44)$$

na qual é considerado que a variável \mathbf{y} tem média nula e variância unitária. Devido a esta mesma hipótese, a validade desta aproximação é bastante limitada.

Em [Hyvärinen, 1999] é mostrada uma classe de aproximações que fornecem bom desempenho, da forma:

$$N_G(\mathbf{y}) \approx \sum_{i=1}^p \varrho_i \cdot [E \{g_i(\mathbf{y})\} - E \{g_i(\nu)\}]^2, \quad (2.45)$$

em que $g_i(\cdot)$ são funções não-quadráticas adequadas, ϱ_i são constantes arbitrárias e ν é uma variável gaussiana normalizada, ou seja, de média zero e variância unitária. A questão agora recai sobre a escolha das funções $g_i(\cdot)$ de tal forma que a aproximação seja acurada.

As propriedades estatísticas de estimadores baseados na otimização da negentropia foram analisadas em [Hyvärinen, 1997] e, para algumas funções, foi mostrado que as características de estimadores que usam funções do tipo da Equação (2.45) apresentaram melhores propriedades estatísticas (robustez e variância assintótica) que aqueles que usam aproximações baseadas em momentos de ordem superior. Desta forma, as seguintes funções não-quadráticas foram propostas:

$$g_1(u) = \frac{1}{\varrho_1} \log [\cosh (a_1 u)] \quad \text{e} \quad (2.46a)$$

$$g_2(u) = -\exp \left(-\frac{u^2}{2} \right), \quad (2.46b)$$

em que $1 \leq \varrho_1 \leq 2$ é uma constante adequada.

Estas funções fornecem um bom compromisso entre as características de duas medidas de não-gaussianidade, *kurtosis* e negentropia. Além disso, são de cálculo rápido e simples. A aproximação da negentropia fornecida com o auxílio da Equação (2.46) resultou no conhecido

algoritmo *FastICA*, alvo de estudo em sistemas práticos [Hyvärinen & Oja, 2000], e que será discutido posteriormente.

2.4 Conclusões e síntese

Este capítulo introduz os conceitos e modelos fundamentais para a área de separação cega de fontes. São explicitadas as principais limitações decorrentes do desconhecimento dos sinais em um sistema de múltiplas fontes, visando, a seguir, prover uma análise posterior dos critérios que servem ao problema em questão.

Uma atenção especial é dada ao conceito de análise por componentes independentes, visto que o mesmo permeia várias aplicações em sistemas reais e também apresenta uma rica quantidade de critérios associados aos conceitos de não-gaussianidade e de mínima entropia. Estes, por sua vez, estão intimamente ligados aos critérios que buscam tornar as estimativas das fontes de um sistema independentes entre si.

Finalmente, o conteúdo descrito no capítulo é apenas uma pequena amostra, embora representativa, da literatura produzida na área recentemente. Algumas referências possuem caráter tutorial e se apresentam de forma bastante didática de tal forma que são fortemente indicadas para aqueles que desejam se inserir na área através de um formalismo mais refinado e detalhado. Como sugestão, algumas referências são listadas como pontos de partida:

- ◆ [Kofidis, 2001];
- ◆ [Hyvärinen & Oja, 2000];
- ◆ [Smaragdis, 1997];
- ◆ [Haykin, 2000a];
- ◆ [Hyvärinen, 1999].

Além destas publicações, um grande ponto de congregação de excelentes trabalhos é a edição do Proceedings of the IEEE, vol. 86, no. 10, Outubro de 1998.

De forma a completar a parte tutorial sobre separação de fontes da tese, os principais critérios e alguns algoritmos que são encontrados na vasta literatura da área, bem como aplicações práticas de BSS, são apresentados no capítulo seguinte.

– “Saber onde se quer ir é importante, mas ainda é importante mostrar que se pode chegar.”

Emile Zola, 1877

3

Separação Cega de Fontes: Estratégias e Aplicações

O CONJUNTO de estratégias de separação cega de fontes é extremamente amplo e provavelmente seriam necessários vários capítulos para esgotar tal assunto.

Este capítulo propõe-se a discutir *algumas* das principais estratégias existentes na literatura, bem como suas principais características e limitações. Algumas aplicações bastante clássicas, mas não menos realistas ou importantes, são também ilustradas para exemplificar a potencialidade da técnica e abrangência dos modelos teóricos até então descritos.

O capítulo é organizado da seguinte forma. As Seções 3.1 e 3.2 são dedicadas à descrição dos principais métodos de separação de fontes e alguns dos principais algoritmos utilizados. Na Seção 3.3, problemas práticos de BSS e estratégias de solução são comparadas. As conclusões são então apresentadas na Seção 3.4.

3.1 Principais estratégias de separação de fontes

Depois de definidos os fundamentos matemáticos que determinam a factibilidade de encontrar uma solução para o problema de separação cega de fontes, faz-se necessário listar as principais estratégias utilizadas para separação de fontes.

Conforme poderá ser visto, várias delas podem ser consideradas equivalentes, segundo algumas condições.

3.1.1 *MaxEnt e InfoMax*

Na Seção 2.3 do Capítulo 2, foi mostrado que as fontes são completamente separadas quando as saídas do sistema de separação apresentam sinais independentes, o que equivale a tornar a informação mútua nula.

Relembrando a Equação (2.25) e a Figura 2.1, tem-se que

$$I(\mathbf{y}, \mathbf{x}) = \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{z}|\mathbf{x}). \quad (3.1)$$

Uma vez que o mapeamento entre \mathbf{x} e \mathbf{y} é determinístico¹, o conhecimento de uma variável determina completamente a outra e anula a entropia condicional. Logo, neste contexto, tem-se:

$$I(\mathbf{y}, \mathbf{x}) = \mathcal{H}(\mathbf{y}). \quad (3.2)$$

Neste ponto, faz-se necessário o uso de uma transformação não-linear sobre as saídas da matriz de separação, pois uma vez que não há restrição para $\mathbf{y} = \mathbf{W}\mathbf{x}$, a entropia de \mathbf{y} não é limitada. Desta forma, considera-se uma transformação não-linear:

$$\mathbf{z} = \mathbf{g}(\mathbf{W}\mathbf{x}), \quad (3.3)$$

em que

$$\mathbf{g}(\mathbf{y}) \triangleq [g_1(y_1) \quad g_2(y_2) \quad \cdots \quad g_K(y_K)]^T,$$

e as funções $g_i(\cdot)$ são monotonicamente crescentes, limitadas de tal forma que $g_i(-\infty) = 0$ e $g_i(\infty) = 1$, tais como as funções típicas de ativação dos neurônios de RNA [Haykin, 1998; Cavalcante, 2001].

¹Considerando o caso sem ruído esta implicação é direta. No caso da presença de ruído, a mesma afirmativa é válida nos casos em que o ruído é passível de identificação e conseqüente separação dos sinais [Bell & Sejnowski, 1995].

Obviamente, se a escolha de $g_i(\cdot)$ for igual à função de distribuição cumulativa (fdc) da i -ésima fonte, que também tem a característica desejada de $g_i(-\infty) = 0$ e $g_i(\infty) = 1$, tem-se que

$$p_z(\mathbf{z}) = U[0, 1], \quad \text{para}$$

- ◇ $z_i = a_i$, o que corresponde à identificação correta da fonte a_i , ou
- ◇ $z_i = a_j, i \neq j$, o que corresponde à condição de indeterminação da recuperação com relação a uma permutação.

O processo de igualar as densidades da entrada e saída através da utilização de funções não-lineares fornece uma adaptação de termos que forçam as fdps o mais próximo possível da distribuição uniforme entre zero e um², ou seja, $p_z(\mathbf{z}) = U[0, 1]$ [Bell & Sejnowski, 1995].

Esta abordagem baseada na maximização da entropia é chamada de *MaxEnt*³ [Bell & Sejnowski, 1995; Haykin, 2000a]. Como, de acordo com a Equação (3.2), maximizar a entropia corresponde também a maximizar a informação mútua, este critério é também chamado de *InfoMax*⁴ [Haykin, 2000a; Bell & Sejnowski, 1995]. Curiosamente, este tipo de abordagem equivale ao de igualdade entre fdps, realizada em trabalhos de desconvolução não-supervisionada, que proporcionaram a fundamentação matemática desta área [Benveniste et al., 1980].

Em [Bell & Sejnowski, 1995], a derivação de um algoritmo adaptativo também é realizada, resultando na seguinte equação de adaptação:

$$\Delta \mathbf{W} \propto (\mathbf{W}^{-T})^{-1} - 2 \cdot \tanh(\mathbf{W}\mathbf{x}) \mathbf{x}^T, \quad (3.4)$$

na qual a função \tanh é aplicada separadamente em cada componente do vetor $\mathbf{y} = \mathbf{W}\mathbf{x}$ e foi escolhida por ser a derivada da log-densidade da distribuição “logística” [Bell & Sejnowski, 1995; Hyvärinen, 1999]. Esta função mostrou-se adequada ao uso na estimação da maioria das distribuições super-gaussianas.

Um ponto importante, também discutido em [Bell & Sejnowski, 1995], é que os critérios InfoMax/MaxEnt são, sob certas condições, equivalentes ao critério NMI. Será mostrado aqui apenas uma dedução simplificada de como a equivalência entre os critérios pode ser visualizada. Uma prova mais formal, bem como uma proposição de algoritmo adaptativo, é feita em [Yang & Amari, 1997].

Seja \mathbf{y} a parte linear do sistema de separação, dado por

$$\mathbf{y} = \mathbf{W}\mathbf{x}.$$

²Devido à função não-linear que limita os valores entre zero e um.

³Do inglês, *Maximization of Entropy*.

⁴Do inglês, *Information Maximization*.

Então, sabendo que $\mathbf{g}(\cdot)$ é inversível, tem-se, através do teorema de funções de v.a. [Papoulis, 1991]:

$$p_{\mathbf{z}}(\mathbf{z}) = \frac{p_{\mathbf{y}}(\mathbf{y})}{\left| \prod_{i=1}^K g'_i(y_i) \right|},$$

em que $g'_i(\cdot)$ representa a derivada de $g_i(\cdot)$. A partir daí e lembrando-se de que $g_i(\cdot)$ é não-negativa, pode-se escrever a entropia de \mathbf{z} como:

$$\mathcal{H}(\mathbf{z}) = \mathcal{H}(\mathbf{x}) + \ln [|\det(\mathbf{W})|] - \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K g'_i(y_i) \right] \right\}. \quad (3.5)$$

Escrevendo-se a solução do critério InfoMax, ou MaxEnt, como uma função contraste, obtém-se:

$$\max_{\mathbf{W}} \left(\Psi_{\text{InfoMax}}(\mathbf{W}) \triangleq \ln [|\det(\mathbf{W})|] - \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K g'_i(y_i) \right] \right\} \right). \quad (3.6)$$

Comparando-se as Equações (2.43) e (3.6), pode-se observar que os critérios são equivalentes se as funções $g_i(\cdot)$ coincidem com as fdc das fontes⁵. De maneira indireta, pode-se também deduzir que os critérios InfoMax/MaxEnt possuem uma equivalência também com a maximização da negentropia, já que esta corresponde à minimização da NMI.

O problema que ainda persiste diz respeito à escolha correta das funções $g_i(\cdot)$, o qual será abordado na seção sobre algoritmos.

3.1.2 Máxima verossimilhança e abordagem bayesiana

Diante da necessidade de fazer uso de teoria de estimação, é interessante estabelecer relações entre a abordagem de teoria da informação e de máxima verossimilhança (MV). Neste caso, será ainda utilizada uma função não-linear após a parte linear do sistema de separação, de tal forma que as funções $g_i(\cdot)$ forneçam sempre valores limitados no intervalo $[0, 1]$. Desta forma, a entropia de \mathbf{z} pode ser escrita como o negativo da distância⁶ entre sua fdp e a fdp uniforme, ou seja,

$$\mathcal{H}(\mathbf{z}) = - \int p_{\mathbf{z}}(\mathbf{z}) \ln \left[\frac{p_{\mathbf{z}}(\mathbf{z})}{\prod_{i=1}^K U(z_i)} \right] d\mathbf{z} = -D(p_{\mathbf{z}}(\mathbf{z}) \| U_N(\mathbf{z})), \quad (3.7)$$

⁵Isto é devido ao fato de que a derivada da fdc é igual à fdp e também que a esperança matemática é calculada para os dois termos em relação ao termo linear do sistema de separação.

⁶No espaço das distribuições.

em que $U(z_i) = U[0, 1]$ e $U_N(\mathbf{z}) = \prod_{i=1}^K U(z_i)$.

Para um critério MV, faz-se necessário a elaboração de um modelo de geração dos dados observáveis que inclua alguma hipótese das estatísticas da fonte. Desconsiderando o ruído, por questões de simplificação, o modelo estabelecido é dado por:

$$\mathbf{x} = \Theta \tilde{\mathbf{a}}, \quad (3.8)$$

em que idealmente, de acordo com a Equação (2.4), deve-se ter $\Theta = \mathbf{H}$ e $\tilde{\mathbf{a}} = \mathbf{a}$. Assim, representa-se a fdp de \mathbf{x} condicionada aos parâmetros do modelo como $p_x(\mathbf{x}|\Theta)^7$.

O princípio do critério MV é encontrar, dentre os conjuntos possíveis de parâmetros Θ , aquele que maximize $p_x(\mathbf{x}|\Theta)$. Considerando-se um número Q de realizações de \mathbf{x} , ou seja, $\mathbf{x}(1), \dots, \mathbf{x}(Q)$, a verossimilhança de que estas amostras obedeçam a uma fdp $p_x(\mathbf{x}|\Theta)$ é dada por $\prod_{q=1}^Q p_x(\mathbf{x}(q)|\Theta)$, e a log-verossimilhança normalizada por:

$$\mathcal{L}_Q(\Theta) = \frac{1}{Q} \ln \left[\prod_{q=1}^Q p_x(\mathbf{x}(q)|\Theta) \right] = \frac{1}{Q} \sum_{q=1}^Q \ln [p_x(\mathbf{x}(q)|\Theta)].$$

Escrevendo-se a fdp do sinal recebido condicionada ao modelo em função do sinal na saída da parte linear do sistema de separação, tem-se

$$\begin{aligned} p_x(\mathbf{x}(q)|\Theta) &= \frac{p_{\tilde{\mathbf{a}}}(\Theta^{-1}\mathbf{x}(q)|\Theta)}{|\det(\Theta)|}, \\ &= p_{\tilde{\mathbf{a}}}(\mathbf{y}(q)) \cdot \det(\mathbf{W}) \end{aligned}$$

em que, vale lembrar, $\mathbf{y} = \mathbf{W}\mathbf{x}$ e, idealmente, $\mathbf{W} = \Theta^{-1}$. Assim:

$$\mathcal{L}_Q(\Theta) = \frac{1}{Q} \sum_{q=1}^Q \ln [p_{\tilde{\mathbf{a}}}(\mathbf{y}(q))] + \ln [|\det(\mathbf{W})|],$$

e, através da lei dos grandes números [Papoulis, 1991; Picinbono, 1993; Therrien, 1992], tem-se a seguinte relação:

$$\mathcal{L}_Q(\Theta) \xrightarrow{Q \rightarrow \infty} \mathcal{L}(\mathbf{W}) \triangleq \int p_y(\mathbf{y}|\mathbf{W}) \ln [p_{\tilde{\mathbf{a}}}(\mathbf{y})] d\mathbf{y} + \ln [|\det(\mathbf{W})|].$$

Utilizando a KLD para escrever a relação acima, obtém-se:

$$\begin{aligned} \mathcal{L}(\mathbf{W}) &= -D(p_x(\mathbf{y}|\mathbf{W}) \| p_{\tilde{\mathbf{a}}}(\mathbf{y})) - \mathcal{H}(\mathbf{y}|\mathbf{W}) + \ln [|\det(\mathbf{W})|] \\ &= -D(p_x(\mathbf{y}|\mathbf{W}) \| p_{\tilde{\mathbf{a}}}(\mathbf{y})) - \mathcal{H}(\mathbf{x}). \end{aligned}$$

⁷A não explicitação do termo $\tilde{\mathbf{a}}$ deve-se ao fato de que o critério será otimizado em relação aos parâmetros Θ .

Daí, pode-se escrever o critério MV como sendo:

$$\max_{\mathbf{W}} \left(\Psi_{\text{MV}}(\mathbf{W}) \triangleq -D(p_y(\mathbf{y}|\mathbf{W})\|p_{\tilde{\mathbf{a}}}(\mathbf{y})) \right). \quad (3.9)$$

Isto significa que a meta é igualar as fdps das estimativas das fontes com as fdps das fontes hipotéticas $\tilde{\mathbf{a}}$. A título de maior clareza, pode-se fazer um abuso de notação e representar por $D(\mathbf{y}|\mathbf{z})$ a KLD entre as fdps de \mathbf{y} e \mathbf{z} . Com esta notação, a função custo MV pode ser escrita como:

$$\Psi_{\text{MV}}(\mathbf{W}) = -D(\mathbf{W}\mathbf{H}\mathbf{a}|\tilde{\mathbf{a}}). \quad (3.10)$$

Ao assumir-se que as funções $g_i(\cdot)$ são as fdes correspondentes às funções de probabilidade das fontes hipotéticas $\tilde{\mathbf{a}}$, o vetor $\tilde{\mathbf{z}} = \mathbf{g}(\tilde{\mathbf{a}})$ é uniformemente distribuído em $[0, 1]$ [Kofidis, 2001]. Desta forma, o critério InfoMax pode ser visto através da seguinte função de contraste:

$$\begin{aligned} \Psi_{\text{InfoMax}}(\mathbf{W}) &= \mathcal{H}(\mathbf{z}) \\ &= -D(\mathbf{z}|\tilde{\mathbf{z}}) \\ &= -D(\mathbf{g}(\mathbf{W}\mathbf{H}\mathbf{a})\|\mathbf{g}(\tilde{\mathbf{a}})) \\ &= -D(\mathbf{W}\mathbf{H}\mathbf{a}|\tilde{\mathbf{a}}), \end{aligned} \quad (3.11)$$

Comparando-se as Equações (3.10) e (3.11) pode-se deduzir uma equivalência entre os critérios MV e InfoMax (MaxEnt) [Cardoso, 1997]. Logicamente, se as fontes são perfeitamente conhecidas, ou seja, $\tilde{\mathbf{a}} = \mathbf{a}$, ambas as funções custo são maximizadas quando $\mathbf{W} = \mathbf{H}^{-1}$ ou, de forma equivalente, $\Theta = \mathbf{H}$, ao que corresponde à perfeita solução para separação.

Uma interpretação bastante interessante é apresentada em [Cardoso, 1998], que consiste em observar o critério MV da Equação (3.10) escrevendo-o através da decomposição de Pitágoras. Na verdade, se os componentes de $\tilde{\mathbf{a}}$ são independentes e $\tilde{\mathbf{z}}$ representa o vetor com componentes distribuídas de acordo com as distribuições marginais de \mathbf{z} , tem-se

$$D(\mathbf{z}|\tilde{\mathbf{a}}) = D(\mathbf{z}|\tilde{\mathbf{z}}) + D(\tilde{\mathbf{z}}|\tilde{\mathbf{a}}), \quad (3.12)$$

e, daí, a Equação (3.10) torna-se:

$$\begin{aligned} -\Psi_{\text{MV}}(\mathbf{W}) &= D(\mathbf{z}|\tilde{\mathbf{z}}) + D(\tilde{\mathbf{z}}|\tilde{\mathbf{a}}) \\ &= -\Psi_{\text{ICA}}(\mathbf{W}) + \sum_{i=1}^K D(\tilde{z}_i|\tilde{a}_i), \end{aligned} \quad (3.13)$$

o que leva à seguinte interpretação da MV [Cardoso, 1998; Kofidis, 2001]:

$$\left(\begin{array}{c} \text{Desvio} \\ \text{total} \end{array} \right) = \left(\begin{array}{c} \text{Desvio da} \\ \text{independência} \end{array} \right) + \left(\begin{array}{c} \text{Desvio} \\ \text{marginal} \end{array} \right). \quad (3.14)$$

Isto ocorre pois, uma vez que a matriz de separação é encontrada baseando-se em uma hipótese sobre as fontes, pode ocorrer que as distribuições das fontes hipotéticas sejam muito diferentes daquelas das fontes verdadeiras. Sendo assim, a minimização da divergência entre o modelo assumido e os dados disponíveis deve ser feita com relação à matriz de parâmetros e com respeito ao modelo de distribuição das fontes.

Então, pode-se ver a maximização da medida de verossimilhança, dada na Equação (3.13), como a otimização de dois objetivos [Cardoso, 1998]:

1. tornar as fontes na saída do sistema de separação independentes, através da minimização da informação mútua, o que corresponde a maximizar Ψ_{ICA} ;
2. fazer com que as fontes sejam mais próximas possíveis de um produto de densidades marginais previamente assumidas, minimizando a divergência entre as fontes obtidas e o modelo.

Outro ponto ilustrado a partir da relação (3.14) é que o critério ICA é resultante da otimização da KLD da Equação (3.10) em relação a ambos os termos, \mathbf{W} e o modelo de probabilidade das fontes. Matematicamente tem-se:

$$\Psi_{ICA}(\mathbf{W}) = \max_{\mathbf{a}} (\Psi_{MV}(\mathbf{W})).$$

Este resultado, discutido em [Cardoso, 1998], é a prova de que a informação mútua é a medida de dependência decorrente do princípio de máxima verossimilhança [Kofidis, 2001].

Um outro critério equivalente ao MV pode ser obtido através da abordagem bayesiana [Mohammad-Djafari, 1999]. Pelo teorema de Bayes [Picinbono, 1993], pode-se escrever a fdp *a posteriori* do modelo (Θ, \mathbf{a}) em termos da verossimilhança [Kofidis, 2001]:

$$p(\Theta, \mathbf{a} | \mathbf{x}, \mathcal{I}) = \frac{\overbrace{p(\mathbf{x} | \Theta)}^{p_x(\mathbf{x} | \Theta)} \cdot p(\Theta, \mathbf{a} | \mathcal{I})}{p(\mathbf{x} | \mathcal{I})}, \quad (3.15)$$

em que \mathcal{I} é utilizado para denotar qualquer informação *a priori* possível sobre o modelo de BSS. Simplificando a expressão através de

$$p(\Theta, \mathbf{a} | \mathbf{x}, \mathcal{I}) \propto p(\mathbf{x} | \Theta, \mathbf{a}, \mathcal{I}) \cdot p(\Theta, \mathbf{a} | \mathcal{I}), \quad (3.16)$$

pois, uma vez que a probabilidade *a priori* dos dados não depende do modelo, esta foi incorporada através de uma constante de proporcionalidade. Outro fato é que, geralmente, a matriz de mistura convolutiva é independente das fontes, o que permite escrever:

$$p(\Theta, \mathbf{a} | \mathcal{I}) = p(\Theta | \mathcal{I}) \cdot p(\mathbf{a} | \mathcal{I}). \quad (3.17)$$

Como a meta é determinar o sistema de separação Θ^{-1} , o seguinte modelo *a posteriori* é considerado:

$$p(\Theta|\mathbf{x}, \mathcal{I}) \propto p(\Theta, \mathcal{I}) \cdot \int p(\mathbf{x}|\Theta, \mathbf{a}, \mathcal{I}) \cdot p(\mathbf{a}|\mathcal{I}) d\mathbf{a}, \quad (3.18)$$

e, tomando-se a log-verossimilhança, tem-se

$$\ln [p(\Theta|\mathbf{x}, \mathcal{I})] = \ln [p(\Theta, \mathcal{I})] + \ln \left[\int p(\mathbf{x}|\Theta, \mathbf{a}, \mathcal{I}) \cdot p(\mathbf{a}|\mathcal{I}) d\mathbf{a} \right], \quad (3.19)$$

em que a constante de proporcionalidade foi omitida porque não tem efeito na otimização.

Sendo o critério de *máxima probabilidade a posteriori* (MAP) dado por

$$\max_{\Theta} \left(\Psi_{\text{MAP}}(\Theta) \triangleq \mathbb{E} \{ \ln [p(\Theta|\mathbf{x}, \mathcal{I})] \} \right), \quad (3.20)$$

é possível construir funções de densidade de probabilidade que expressam o conhecimento prévio de um determinado problema de separação de fontes. Sabendo-se que a essência da abordagem bayesiana é fazer alguma inferência sobre os parâmetros desconhecidos com base na informação disponível [Rowe, 2002], a hipótese de que o sistema é linear, estacionário e instantâneo, como descrito pela equação $\mathbf{x} = \mathbf{H}\mathbf{a}$, e possui fontes de sinais independentes, pode ser expressa através da verossimilhança dada por [Knuth, 1999; Rowe, 2002]:

$$p(\mathbf{x}|\Theta, \mathbf{a}, \mathcal{I}) = \prod_{i=1}^K \delta \left(x_i - \sum_{j=1}^K \Theta_{ij} a_j \right), \quad (3.21)$$

em que $\delta(\cdot)$ é a função de Dirac, e o conhecimento relativo à independência da fonte é utilizado através de

$$p(\mathbf{a}|\mathcal{I}) = \prod_{i=1}^K p_{a_i}(a_i|\mathcal{I}).$$

Já que o conhecimento da matriz de mistura convolutiva é, em geral, restrito às suas dimensões e não-singularidade, $p(\Theta|\mathcal{I})$ é escolhido ser uma constante e é omitido do critério MAP. Ou seja, pode-se escrever:

$$\begin{aligned} \int p(\mathbf{x}|\Theta, \mathbf{a}, \mathcal{I}) p(\mathbf{a}|\mathcal{I}) d\mathbf{a} &= \int \int \cdots \int \prod_{i=1}^K \delta \left(x_i - \sum_{j=1}^K \Theta_{ij} a_j \right) \prod_{i=1}^K p_{a_i}(a_i|\mathcal{I}) da_1 da_2 \cdots da_K \\ &= |\det(\mathbf{W})| \prod_{i=1}^K p_{a_i} \left(\sum_{j=1}^K W_{ij} x_j \right) \end{aligned}$$

em que $\mathbf{W} = \Theta^{-1}$. Desta maneira, a função custo MAP é escrita como:

$$\Psi_{\text{MAP}}(\mathbf{W}) = \ln [|\det(\mathbf{W})|] - \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K p_{a_i}(y_i) \right] \right\} \quad (3.22)$$

Comparando-se a Equação (3.22) com a Equação (3.6), verifica-se a equivalência da abordagem bayesiana com o critério InfoMax, desde que as funções $g_i(\cdot)$ sejam iguais às fdc das fontes [Kofidis, 2001].

3.1.3 Critério “universal”

Tendo em vista a discussão das seções anteriores, é possível escrever as técnicas de BSS sob a luz de um critério que congrega as propriedades das técnicas até então discutidos, formando, desta maneira, um critério “universal” [Kofidis, 2001].

Assim, pode-se escrever de uma forma geral:

$$J_{\text{BSS}}(\mathbf{W}) = \ln [|\det(\mathbf{W})|] - \mathbb{E} \left\{ \ln \left[\prod_{i=1}^K \phi_i(y_i) \right] \right\}, \quad (3.23)$$

em que $\mathbf{y} = \mathbf{W}\mathbf{x}$ e, idealmente, as funções $\phi_i(\cdot)$ são as fdc das fontes. Entretanto, nem tudo é tão simples. Uma vez que as fontes e suas fdps são, geralmente, desconhecidas, algumas aproximações são propostas, conforme será visto a seguir.

Aproximações de funções de densidade de probabilidade

Como é necessário o conhecimento das fdps dos sinais na saída da parte linear do sistema de separação para a função custo da Equação (3.23), e a função custo da Equação (2.43) necessita das fdps marginais de \mathbf{y} , sendo que este conhecimento não está disponível, uma abordagem alternativa é empregar expansões truncadas das fdps desconhecidas e reduzir o desconhecimento a um número de cumulantes possíveis de serem calculados a partir das saídas da matriz de separação.

Uma decomposição conhecida é a *expansão de Edgeworth*⁸ de uma fdp $p_y(\mathbf{y})$ de média zero e variância unitária ao redor de uma fdp gaussiana normalizada $p_G(\mathbf{y})$ [Lacoume et al., 1997; McCullagh, 1987; Lindsey, 1996]:

$$p_y(\mathbf{y}) = p_G(\mathbf{y}) \left[1 + \frac{1}{3!} c_3(\mathbf{y}) \mathfrak{h}_3(\mathbf{y}) + \frac{1}{4!} c_4(\mathbf{y}) \mathfrak{h}_4(\mathbf{y}) + \frac{10}{6!} c_3^2(\mathbf{y}) \mathfrak{h}_6(\mathbf{y}) + \frac{1}{5!} c_5(\mathbf{y}) \mathfrak{h}_5(\mathbf{y}) + \frac{35}{7!} c_3(\mathbf{y}) c_4(\mathbf{y}) \mathfrak{h}_7(\mathbf{y}) + \dots \right], \quad (3.24)$$

em que $c_k(\mathbf{y})$ é o cumulante de k -ésima ordem de $p_y(\mathbf{y})$ e $\mathfrak{h}_k(\mathbf{y})$ é o polinômio de Hermite de grau k definido através da k -ésima derivada da função de densidade gaussiana dada por

⁸Esta expansão polinomial será discutida em maiores detalhes no Capítulo 6.

[Haykin, 1998; Lacoume et al., 1997]:

$$\frac{\partial^k p_G(\mathbf{y})}{\partial \mathbf{y}^k} = (-1)^k p_G(\mathbf{y}) \mathfrak{h}_k(\mathbf{y}). \quad (3.25)$$

Se forem preservados apenas os primeiros termos do somatório da Equação (3.24) tem-se uma aproximação polinomial de $p_{y_i}(y_i)$ na Equação (2.43). Através de uma aproximação de Taylor para a função logarítmica, têm-se uma expressão aproximada da função custo em termos de alguns cumulantes dos y_i . Quando um pré-branqueamento é realizado nos dados, supostos simetricamente distribuídos (cumulantes de ordem ímpar são nulos), mostra-se que uma boa aproximação da fdp é dada por [Comon, 1994; Kofidis, 2001; Lee et al., 2000]:

$$\Psi_{\text{ICA}}(\mathbf{W}) \approx \Psi_4(\mathbf{W}) \triangleq \sum_{i=1}^K c_4^2(y_i), \quad (3.26)$$

em que se considera ainda que \mathbf{W} é ortogonal.

Conforme mostrado em [Comon, 1994], Ψ_4 é um contraste se no máximo uma fonte tiver o momento de quarta ordem igual à zero. Esta condição é menos restritiva que a de se supor no máximo uma fonte gaussiana. Isto implica que *todos* os momentos de ordem superior sejam iguais a zero. No caso particular dos sistemas de comunicação digital, esta fonte pode ser entendida como de origem ruidosa⁹ no *front-end* dos receptores.

A mesma função custo é encontrada se for aplicada uma expansão truncada de Gram-Charlier para o método MV [Lacoume et al., 1997]. Um ponto interessante é que, devido à hipótese da ortogonalidade, $\max_{\mathbf{W}} \Psi_4(\mathbf{W})$ equivale a forçar a zero os cumulantes cruzados de \mathbf{y} , o que equivale a tornar seus componentes mutuamente independentes.

Embora as aproximações polinomiais forneçam uma abordagem interessante para a aproximação das fdps dos sinais, elas podem não ser suficientemente acuradas na prática [Kofidis, 2001]. Isto se deve ao fato de que a estimação de cumulantes sofre bastante com o desvio de amostras (poucas amostras com valores altos promovem problemas de estimação). Métodos baseados na estimação da entropia tentam escapar deste problema [Hyvärinen et al., 2001; Kofidis, 2001].

Uma vez que os critérios de BSS foram descritos, os algoritmos que deles derivam necessitam ser apresentados. Este tópico é abordado na seção seguinte.

⁹Na verdade, o ruído nos sistemas de comunicação digital é uma composição de várias fontes de ruído.

3.2 Algoritmos

Embora a construção de critérios seja de interesse teórico, a derivação e avaliação de algoritmos possui grande impacto em aplicações que necessitam de procedimentos de rastreo para as técnicas de BSS.

Inicialmente, tendo em vista a generalidade do critério da Equação (3.23), é interessante derivar para ele um algoritmo do gradiente estocástico. Tomando a derivada em relação a matriz \mathbf{W} , obtém-se [Kofidis, 2001]:

$$\nabla J(\mathbf{W}) = (\mathbf{W}^T)^{-1} - \mathbf{f}(\mathbf{z}) \mathbf{x}^T, \quad (3.27)$$

em que o vetor $\mathbf{f}(\cdot)$ é definido como

$$\mathbf{f}(\mathbf{z}) = \left[-\frac{\phi'_1(z_1)}{\phi_1(z_1)} \quad -\frac{\phi'_2(z_2)}{\phi_2(z_2)} \quad \dots \quad -\frac{\phi'_K(z_K)}{\phi_K(z_K)} \right]^T. \quad (3.28)$$

As derivadas do logaritmo de ϕ_i , $\frac{\phi'_i(z_i)}{\phi_i(z_i)}$, são conhecidas como *funções score* [Cardoso, 1998] e o algoritmo do gradiente ascendente¹⁰ é dado por:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \Delta \mathbf{W}(n), \quad (3.29)$$

em que

$$\Delta \mathbf{W}(n) \propto (\mathbf{W}^T(n))^{-1} - \mathbf{f}[\mathbf{z}(n)] \mathbf{x}^T(n). \quad (3.30)$$

Ao se utilizar a função logística sigmóide para as funções ϕ_i ,

$$\phi_i(z_i) = \frac{1}{1 + \exp(-z_i)},$$

o algoritmo de recursão torna-se o algoritmo InfoMax [Smaragdis, 1997].

Outros algoritmos que fazem, ocasionalmente, algumas suposições diferentes sobre o problema são, a seguir, descritos.

3.2.1 Algoritmo Jutten-Hérault e variantes

O trabalho pioneiro de separação de fontes, realizado por Jutten e Héroult, tinha como inspiração os modelos de RNA. Ele se baseava no cancelamento das correlações cruzadas dos sinais na saída do dispositivo não-linear da etapa de separação, de forma a obter sinais

¹⁰Vale a pena lembrar que o objetivo é *maximizar* a função contraste.

independentes [Hyvärinen, 1999]. Tais correlações cruzadas são do tipo $\mathbb{E} \{g_1(y_i)g_2(y_j)\}$, em que $g_1(\cdot)$ e $g_2(\cdot)$ são funções não-lineares adequadamente escolhidas.

O algoritmo se baseia no fato de que, se y_i e y_j forem independentes, suas correlações cruzadas serão iguais à zero, supondo que as fontes possuam densidades simétricas. Desta forma, uma representação da solução analítica do problema tratado¹¹ pode ser expressa através da Figura 3.1 [Hérault & Jutten, 1994].

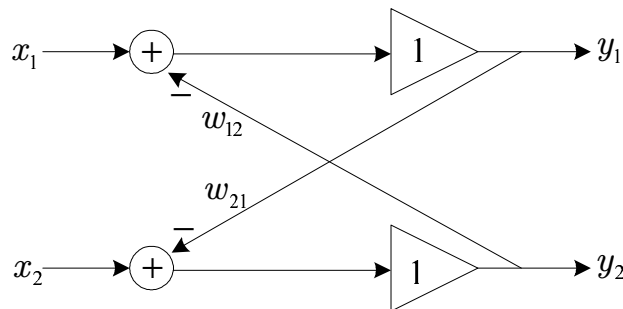


Figura 3.1: Rede neural artificial de dois neurônios para solução do problema de separação de mistura de duas variáveis.

Desta forma, a variação dos termos não-diagonais ($i \neq j$) da matriz de separação \mathbf{W} é calculada através da seguinte regra [Hérault & Jutten, 1994; Hyvärinen, 1999]:

$$\Delta[w_{ij}] \propto g_1(y_i) \cdot g_2(y_j) \quad \forall i \neq j, \quad (3.31)$$

em que cada um dos sinais y são calculados a cada instante como $\mathbf{y} = (\mathbf{I} + \mathbf{W})^{-1} \mathbf{x}$, e os termos diagonais $[w_{ii}]$ são colocados inicialmente em zero. Então o vetor \mathbf{y} , após convergência, fornecerá uma estimativa dos componentes independentes. Infelizmente, conforme mostrado em [Delfosse & Loubaton, 1995], este algoritmo possui condições de convergência bastante restritivas, impossibilitando uma ampla aplicação nos problemas de BSS. Isto decorre da derivação de uma solução *ad hoc* para o problema inicial, que necessita apenas forçar uma decorrelação entre as saídas dos dispositivos não-lineares.

Outros algoritmos que também são baseados no cancelamento das correlações cruzadas foram propostos e levam em consideração a abordagem apresentada por [Hérault & Jutten, 1994]. A principal diferença é que eles tentam reduzir a complexidade computacional evitando inversões de matrizes, o que aumenta a estabilidade [Hyvärinen, 1999].

De acordo com [Cichocki & Amari, 2002], o seguinte algoritmo foi proposto:

$$\Delta \mathbf{W} \propto (\mathbf{I} - g_1(\mathbf{y}) \cdot g_2(\mathbf{y}^T)) \cdot \mathbf{W}, \quad (3.32)$$

¹¹Convém lembrar que o problema inicial apresentado em [Hérault et al., 1985] era composto de duas entradas (posição angular e velocidade) e que por isso a representação é feita somente para duas entradas e duas saídas.

em que $\mathbf{y} = \mathbf{W}\mathbf{x}$ e as não-linearidades $g_1(\cdot)$ e $g_2(\cdot)$ são aplicadas separadamente em cada componente do vetor \mathbf{y} . A matriz identidade pode ainda ser substituída por qualquer matriz definida positiva [Hyvärinen, 1999].

Visando uma maior adequação dos algoritmos adaptativos a problemas mais genéricos, algumas propostas incorporaram hipóteses e considerações mais abrangentes, como os descritos na seqüência.

3.2.2 Equivariância e gradiente relativo/natural

Uma desvantagem do algoritmo mostrado na Equação (3.30) é a necessidade de inversão de uma matriz de dimensão $K \times K$ a cada iteração, sendo que a inversibilidade desta matriz não é sequer assegurada. Além do mais, o desempenho é altamente condicionado ao tipo de matriz de mistura convolutiva envolvida no processo.

É interessante, então, obter um critério que seja independente do condicionamento da matriz \mathbf{H} .

Isto pode ser obtido através de um estimador $\hat{\mathbf{E}}$, chamado *equivariante*, que produza uma transformação nos dados de tal forma que [Cardoso & Laheld, 1996; Cardoso, 2000]:

$$\hat{\mathbf{E}} \langle \mathbf{W}\mathbf{x} \rangle = \mathbf{W} \hat{\mathbf{E}} \langle \mathbf{x} \rangle \quad (3.33)$$

para qualquer matriz \mathbf{W} não-singular, em que $\langle \cdot \rangle$ é uma operação sobre vetores e matrizes. Assim, a estimativa das fontes será dada por:

$$\begin{aligned} \hat{\mathbf{a}} &= \left(\hat{\mathbf{E}} \langle \mathbf{x} \rangle \right)^{-1} \mathbf{x} \\ &= \left(\hat{\mathbf{E}} \langle \mathbf{H}\mathbf{a} \rangle \right)^{-1} \mathbf{H}\mathbf{a} \\ &= \left(\mathbf{H} \hat{\mathbf{E}} \langle \mathbf{a} \rangle \right)^{-1} \mathbf{H}\mathbf{a} \\ &= \left(\hat{\mathbf{E}} \langle \mathbf{a} \rangle \right)^{-1} \mathbf{a}, \end{aligned} \quad (3.34)$$

o que não depende, explicitamente, da matriz de mistura convolutiva.

Então, o algoritmo da Equação (3.30) pode ser transformado para ser equivariante através de uma multiplicação à direita por $\mathbf{W}^T(n)\mathbf{W}(n)$. Como a matriz $\mathbf{W}(n)$ é definida positiva, esta operação não afeta a estabilidade [Kofidis, 2001], e a nova regra de recursão é dada por:

$$\Delta \mathbf{W}(n) \propto [\mathbf{I} - \mathbf{f}[\mathbf{z}(n)] \mathbf{z}^T(n)] \mathbf{W}(n). \quad (3.35)$$

Este algoritmo é chamado de EASI (*Equivariant Adaptive Separation via Independence*).

Uma importante implicação da propriedade da equivariância pode ser vista através da *resposta combinada* (ou global) do sistema

$$\mathbf{G} = \mathbf{H}\mathbf{W}. \quad (3.36)$$

Realizando uma multiplicação à esquerda na Equação (3.35), vem o seguinte resultado em termos da resposta combinada:

$$\Delta\mathbf{G}(n) \propto (\mathbf{I} - \mathbf{f}(\mathbf{z}(n))\mathbf{z}^T(n))\mathbf{G}(n). \quad (3.37)$$

A equivariância é então observada pois este processo depende de \mathbf{H} apenas na inicialização $\mathbf{G}(0) = \mathbf{W}(0) \cdot \mathbf{H}$, o que é aplicável para modificações em \mathbf{H} através de uma modificação na inicialização, excetuando-se, logicamente, os casos em que \mathbf{H} não é inversível.

A equação acima pode ser interpretada através do chamado *gradiente relativo* [Cardoso & Laheld, 1996; Cardoso, 2000], definido para uma função genérica Λ como

$$\tilde{\nabla}\Lambda(\mathbf{W}) \triangleq \left. \frac{\partial\Lambda([\mathbf{I} + \boldsymbol{\Xi}] \cdot \mathbf{W})}{\partial\boldsymbol{\Xi}} \right|_{\boldsymbol{\Xi}=\mathbf{0}}. \quad (3.38)$$

O termo gradiente relativo vem do fato de $\tilde{\nabla}\Lambda(\mathbf{W})$ medir variações de Λ em relação a mudanças relativas de \mathbf{W} ¹².

Desta forma, o gradiente relativo se aplica bem à adaptação necessária na Equação (3.35) e seu papel na transformação equivariante é visto na relação com o gradiente de Λ , que é derivado diretamente da sua definição:

$$\tilde{\nabla}\Lambda(\mathbf{W}) = \nabla\Lambda(\mathbf{W}) \cdot \mathbf{W}^T. \quad (3.39)$$

A utilização do gradiente relativo propocionou uma significativa melhoria no desempenho do algoritmo InfoMax, Equação (3.4), em termos de velocidade de convergência, tornando-o da forma [Hyvärinen, 1999]:

$$\Delta\mathbf{W} \propto [\mathbf{I} - 2 \tanh(\mathbf{y}\mathbf{y}^T)] \mathbf{W}. \quad (3.40)$$

Com esta modificação o algoritmo não necessita de esferatização, que é obtida através de uma operação de pré-branqueamento ou ainda através de PCA . Além disso, o algoritmo da Equação (3.40) pode ser visto como um caso especial do algoritmo de decorrelação não-linear da Equação (3.32) [Hyvärinen, 1999].

¹²Este gradiente também recebe o nome de *gradiente natural*, proposto de forma independente por Amari [Haykin, 2000a]. Os dois autores publicaram, posteriormente, trabalhos em conjunto analisando as questões de desempenho do gradiente relativo/natural.

3.2.3 Deflation

A técnica de *deflation*¹³ proposta em [Delfosse & Loubaton, 1995] e consiste, basicamente, da construção de um critério que realiza a separação das fontes de maneira serial, mas não utiliza diretamente o conhecimento de uma fonte previamente identificada para encontrar outra.

De fato, este trabalho utilizou o critério de Shalvi-Weinstein [Shalvi & Weinstein, 1990] para mostrar que é possível separar *uma* fonte da mistura apenas através de um conjunto de restrições sobre a geração de funções para que seja obtida a separação. A partir da identificação de uma fonte, o processamento é feito sempre na mesma base de observação, ou seja, não há uma redução no conjunto de observações de tal maneira que a mistura observada é sempre a mesma.

Utiliza-se, então, o fato de haver conhecimento de uma fonte e de que as demais sejam ortogonais àquela previamente separada para garantir a independência. Assim, a próxima etapa deve considerar as novas fontes identificadas e fornecer um receptor ortogonal a todas as fontes identificadas até um dado instante, até que todas as fontes tenham sido identificadas. A Figura 3.2 ilustra o esquema de separação através da abordagem de *deflation*.

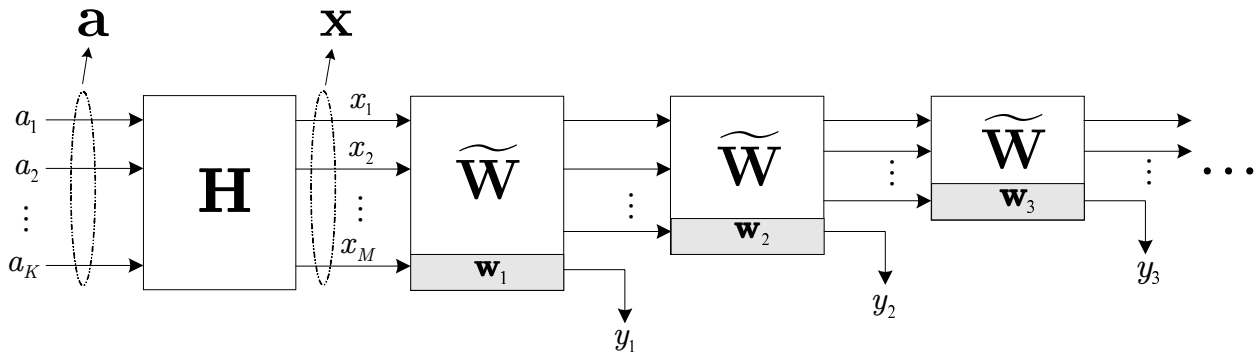


Figura 3.2: Esquema de separação de fontes através do procedimento *deflation*.

De uma forma bem simplificada, o algoritmo utiliza o seguinte critério para cada linha da matriz de separação \mathbf{W}

$$J(\mathbf{w}_i) = \frac{\mathbb{E}^4 \{ \mathbf{w}_i^T \mathbf{x}(n) \}}{4}, \quad (3.41)$$

e mostra que se o vetor \mathbf{w}_i tiver norma unitária e for gerado pela minimização de $J(\mathbf{w}_i)$ na hiper-esfera unitária do \mathbb{R}^K , o sinal $y(n) = \mathbf{w}_i^T(n)\mathbf{x}(n)$ convergirá para uma das fontes. O

¹³O termo poderia ser, talvez, traduzido para o português como *deflação*, entretanto, nesta tese será utilizado o termo em inglês que é derivado do método de Schur para extração de autovalores e autovetores de uma matriz através de uma transformação ortogonal de similaridade.

restante do algoritmo é baseado numa parametrização dos vetores de norma unitária, que permite gerar a cada instante n uma seguinte matriz $\widetilde{\mathbf{W}}$ de ordem $(K - 1) \times K$ de tal forma que a matriz

$$\begin{bmatrix} \widetilde{\mathbf{W}} \\ \mathbf{w} \end{bmatrix}$$

seja ortogonal. Além disso, mostra-se que o sinal residual $y_1(n) = \widetilde{\mathbf{W}}(n)\mathbf{x}(n)$ é um sinal decorrelacionado de \mathbf{w}_i ¹⁴.

Este tipo de processamento é interessante quando não se deseja obter todas as fontes e/ou quando a matrix \mathbf{H} não é inversível, o que obviamente impossibilita qualquer método que busque encontrar a sua inversa [Kofidis, 2001].

O problema do algoritmo de *deflation* é que ele não é derivado em função dos parâmetros do filtro e sim da resposta global, ou combinada, do sistema, o que dificulta sua análise de convergência global para um número qualquer de fontes [Papadias, 2000a, 2000b].

3.2.4 *FastICA*

Conforme mencionado em seções anteriores, devido ao uso da negentropia necessitar do conhecimento ou estimação da fdp dos sinais, métodos de aproximação da negentropia foram propostos e aqueles baseados em expansões polinomiais não-lineares apresentaram-se com melhor desempenho em relação a desvios de estimação [Hyvärinen & Oja, 2000].

Para utilizar este fato, o seguinte critério, baseado na aproximação da Equação (2.45), foi proposto [Hyvärinen, 1999]:

$$\max_{\mathbf{G}^T \mathbf{G} = \mathbf{I}} \left(\Psi_{\text{FastICA}}(\mathbf{W}) \triangleq |\mathbb{E} \{g(\mathbf{y}) - g(\mathbf{y}^{\mathbf{G}})\}|^2 \right), \quad (3.42)$$

em que $g(\cdot)$ é uma função não-linear, não-quadrática e derivável, aplicada a cada um dos componentes do vetor \mathbf{y} e $\mathbf{y}^{\mathbf{G}}$ é uma versão gaussiana de \mathbf{y} . Além disso, é suposto que os dados são esferatizados, o que implica na matriz \mathbf{W} ser ortogonal.

Com isso, o critério da Equação (3.42) tem como princípio maximizar a divergência da variável de saída \mathbf{y} de sua versão gaussiana através do uso de uma função não-linear que tenta extrair as informações das HOS. De uma forma equivalente, para cada um dos componentes de \mathbf{y} e cada linha \mathbf{w}_i da matriz \mathbf{W} , pode-se escrever o seguinte critério com restrição:

$$\left\{ \begin{array}{l} \max_{\mathbf{w}_1, \dots, \mathbf{w}_K} \sum_{i=1}^K \Psi_{\text{FastICA}}(\mathbf{w}_i) \\ \text{sujeito a } \mathbf{w}_i \mathbf{w}_j = \delta_{ij}, \end{array} \right. \quad (3.43)$$

¹⁴Por não ser o principal escopo desta tese, a demonstração e provas que decorrem do algoritmos são apenas citadas e sugere-se a leitura da referência [Delfosse & Loubaton, 1995] para provas e demonstrações formais.

em que δ_{ij} é o delta de Kronecker e

$$\Psi_{\text{FastICA}}(\mathbf{w}_i) \triangleq (\mathbb{E} \{g(\mathbf{w}_i^T \mathbf{x}) - g(y_i^G)\}). \quad (3.44)$$

O processamento da Equação (3.43) ocorre através da abordagem de *deflation* comentada na subseção anterior, de forma que após uma das fontes ter sido encontrada/identificada o restante pode ser calculado com a restrição adicional de ser ortogonal àquela anteriormente encontrada [Kofidis, 2001].

Considerando agora a maximização da Equação (3.43) sujeita à restrição de que o vetor de separação (linha da matriz de separação) apresente norma unitária, e ainda considerando o termo $g(y_i^G)$ constante, o problema de otimização é reduzido para

$$\max_{\|\mathbf{g}_i\|=1} \mathbb{E} \{g(\mathbf{w}_i^T \mathbf{x})\}, \quad (3.45)$$

e, desta forma, os pontos que solucionam a equação são dados por [Hyvärinen et al., 2001; Kofidis, 2001]

$$\mathbb{E} \{\mathbf{x}g'(\mathbf{w}_i^T \mathbf{x})\} - \beta \mathbf{w}_i = \mathbf{0}, \quad (3.46)$$

em que $g'(\cdot)$ é a derivada da função $g(\cdot)$ e $\beta = \mathbb{E} \{yg'(y)\}$ é o multiplicador de Lagrange.

Um algoritmo do gradiente ascendente é dado por [Kofidis, 2001]:

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu(n) [\mathbb{E} \{\mathbf{x}g'(\mathbf{w}_i^T(n)\mathbf{x})\} - \beta(n)\mathbf{w}_i(n)], \quad (3.47)$$

em que $\mu(n)$ é um fator de passo variante no tempo. E através da aproximação de Newton no algoritmo mostrado na Equação (3.47), obtém-se o seguinte resultado [Hyvärinen, 1999]:

$$\mathbf{w}_i^\dagger(n+1) = \mathbb{E} \{\mathbf{x}g'(\mathbf{w}_i^T(n)\mathbf{x})\} - \mathbb{E} \{g''(\mathbf{w}_i^T(n)\mathbf{x})\} \mathbf{w}_i(n) \quad (3.48a)$$

$$\mathbf{w}_i(n+1) = \frac{\mathbf{w}_i^\dagger(n+1)}{\|\mathbf{w}_i^\dagger(n+1)\|}, \quad (3.48b)$$

que é o conhecido algoritmo *FastICA* e $g''(\cdot)$ é a derivada segunda de $g(\cdot)$ [Hyvärinen, 1997]. É interessante notar que o algoritmo dado pela expressão (3.48) pode ser visto como obtido a partir da Equação (3.47) sob a condição da escolha do fator de passo igual a [Kofidis, 2001]:

$$\mu(n) = \frac{1}{\beta(n) - \mathbb{E} \{g''(\mathbf{w}_i^T(n)\mathbf{x})\}}, \quad (3.49)$$

mas alguns trabalhos mostraram que o fator de passo deveria ser da forma [Kofidis, 2001]:

$$\mu(n) = \frac{1}{\beta(n)}, \quad (3.50)$$

o que resulta na variante rápida do *FastICA* na Equação (3.48):

$$\mathbf{w}_i^\dagger(n+1) = \mathbb{E} \{ \mathbf{x} g'(\mathbf{w}_i^T(n)\mathbf{x}) \} \quad (3.51a)$$

$$\mathbf{w}_i(n+1) = \frac{\mathbf{w}_i^\dagger(n+1)}{\|\mathbf{w}_i^\dagger(n+1)\|} . \quad (3.51b)$$

Pode-se também observar que, para a escolha de $g(y) = y^{2p}$, e daí $g'(y) = 2py^{2p-1}$, é fácil verificar que a Equação (3.51) reduz-se ao *algoritmo super-exponencial* (SEA, *Super-Exponential Algorithm*) [Shalvi & Weinstein, 1993; Haykin, 1994; Mboup & Regalia, 2000]. E como o SEA possui uma equivalência com o critério CM (*Constant Modulus*) [Regalia, 1999] é possível mostrar que o algoritmo *FastICA*, aplicando-se o gradiente estocástico para \mathbf{g} de norma unitária, é equivalente ao conhecido algoritmo do módulo constante (CMA, *Constant Modulus Algorithm*) [Kofidis, 2001].

Uma relação interessante é dada pelo fato de que SEA é o método mais rápido para maximizar a razão

$$\left| \frac{c_{2p}(y)}{(c_{2p}(y))^p} \right|,$$

conhecida como razão de *mínima entropia* ou *critério de Donoho* [Donoho, 1981]. Geralmente, $p = 2$ e daí $c_{2p}(y) = c_4(y) = \mathbb{E} \{y^4\} - 3 \cdot (\mathbb{E} \{y^2\})^2$, e para os dados esferatizados tem-se $c_4(y) = \mathbb{E} \{y^4\} - 3$ e com isso a relação do critério de Donoho com a Equação (3.45) é direta. Outro fato é que, para $p = 2$, $g(y) = y^4$, o critério da Equação (3.43) torna-se o critério dado na Equação (3.26).

3.2.5 Aprendizado anti-Hebbiano não-linear

Uma poderosa ferramenta de construção de estratégias de aprendizado é a da regra de Hebb [Haykin, 1998], que tem inspiração nos sistemas biológicos. Várias estratégias e algoritmos de adaptação constituem casos de aprendizado baseados em regras anti-Hebbianas. Alguns trabalhos utilizaram critérios de aprendizado auto-organizável para soluções de problemas de identificação e equalização (desconvolução) [Montalvão, 2000; Cavalcante, 2001].

Uma forma de avaliar os algoritmos até então descritos é de mostrá-los como sendo casos do princípio não-linear do aprendizado anti-Hebbiano. A regra linear de ajuste anti-Hebbiano dos pesos sinápticos entre dois neurônios com atividades z_i e z_j é dada por [Haykin, 1998]:

$$\Delta \mathbf{W}_{ij} \propto -z_i z_j. \quad (3.52)$$

Neste caso, se z_i e z_j são correlacionados positivamente (negativamente), esta regra fornece um peso negativo (positivo) entre eles, o que torna a ativação simultânea mais difícil e,

conseqüentemente, elimina a correlação cruzada. Um modelo de RNA com dois neurônios com esta finalidade é mostrado na Figura 3.3. Os pesos de auto-conexões são incluídos para destacar que a matriz \mathbf{W} é uma matriz de separação.

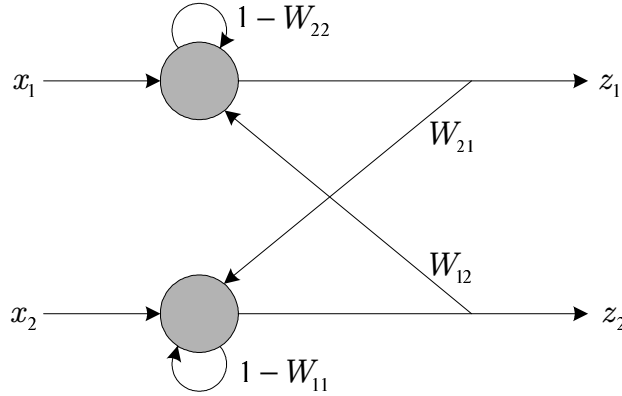


Figura 3.3: Rede neural artificial com dois neurônios e inibição lateral para implementação da regra anti-Hebbiana.

Logo, se o vetor $\mathbf{z} = [z_1 \ z_2]^T$ é espacialmente branco, ou seja, $\mathbb{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, a versão normalizada da Equação (3.52) é dada [Kofidis, 2001]:

$$\Delta \mathbf{W}_{ij} \propto \mathbf{I} - \mathbf{z}\mathbf{z}^T. \quad (3.53)$$

Mas a Equação (3.53) corresponde somente a obter descorrelação entre os componentes de \mathbf{z} . Para detalhar mais esta tarefa, ou seja, tornar as variáveis na saída independentes, deve-se recorrer às HOS. De uma forma implícita, isto é feito através da inserção de funções não-lineares na regra anti-Hebbiana. Pode-se obter:

$$\Delta \mathbf{W} \propto \mathbf{I} - \mathbf{k}(\mathbf{z}) \cdot \mathbf{g}(\mathbf{z}^T), \quad (3.54)$$

em que $\mathbf{k}(\cdot)$ e $\mathbf{g}(\cdot)$ são funções não-lineares.

Alguns casos interessantes:

- algoritmos em sua forma equivariante foram deduzidos como [Cichocki & Amari, 2002]:

$$\Delta \mathbf{W} \propto [\mathbf{I} - \mathbf{k}(\mathbf{z})\mathbf{g}(\mathbf{z}^T)] \cdot \mathbf{W}$$

- $\mathbf{g}(\mathbf{z}) = \mathbf{z}$ resulta no algoritmo anti-Hebbiano não-linear a partir da equação do item anterior;
- a Figura 3.3 com as auto-conexões iguais a zero e a regra de ajuste $\Delta \mathbf{W} \propto -\mathbf{k}(\mathbf{z}) \cdot \mathbf{g}(\mathbf{z}^T)$ corresponde à arquitetura do algoritmo Jutten-Hérault [Hérault & Jutten, 1994].

3.2.6 Análise não-linear por componentes principais - *Nonlinear PCA*

Conforme foi mostrado anteriormente, as estatísticas de segunda ordem não são suficientes para resolver o problema de BSS. A técnica de PCA é baseada na aproximação, no sentido de mínimos quadrados, de uma projeção linear. Esta aproximação é necessária pois da maneira usual tal técnica não é aplicável a problemas genéricos de BSS.

O grupo da Universidade Tecnológica de Helsinki propôs uma generalização para o critério de PCA envolvendo uma projeção não-linear [Hyvärinen et al., 2001]. O critério mostrou-se eficaz nos problemas de BSS e foi chamado de *Nonlinear PCA* (nPCA), o qual pode ser descrito em termos da minimização da seguinte função custo:

$$\Psi_{\text{nPCA}}(\mathbf{W}) = \mathbb{E} \{ \|\mathbf{x} - \mathbf{W}^T \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 \} \quad (3.55)$$

Ao considerar-se o pré-branqueamento dos dados, deve-se ter \mathbf{W} ortogonal. Logo, pode-se reescrever a equação acima como:

$$\begin{aligned} \Psi_{\text{nPCA}}(\mathbf{W}) &= \mathbb{E} \{ \|\mathbf{W}\mathbf{x} - \mathbf{g}(\mathbf{W}\mathbf{x})\|^2 \} \\ &= \mathbb{E} \{ \|\mathbf{y} - \mathbf{g}(\mathbf{y})\|^2 \} \\ &= \sum_{i=1}^K \mathbb{E} \{ |y_i - g_i(y_i)|^2 \} \end{aligned} \quad (3.56)$$

A Equação (3.56) revela que o critério nPCA tem uma forte ligação com os critérios de Bussgang para equalização cega [Haykin, 1994, 1996]. Um aspecto importante a ser mencionado é que, para a correta escolha de \mathbf{g} , o critério nPCA pode ser aproximado pela maximização da soma de valores absolutos dos cumulantes de \mathbf{y} , o que torna o critério equivalente àquele descrito pela Equação (3.26) [Kofidis, 2001].

Em termos de equações dinâmicas, uma versão hierárquica da regra de adaptação é dada por [Hyvärinen, 1999]:

$$\Delta \mathbf{w}_i \propto g(y_i) \mathbf{x} - g(y_i) \sum_{j=1}^i g(y_j) \mathbf{w}_j \quad , \quad (3.57)$$

em que g é uma função escalar não-linear.

Uma simplificação do algoritmo nPCA leva ao chamado *algoritmo bigradiente* [Hyvärinen, 1999]:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(n)g[\mathbf{W}(n)\mathbf{x}(n)] \mathbf{x}^T(n) + \varrho [\mathbf{I} - \mathbf{W}(n)\mathbf{W}^T(n)] \mathbf{W}(n), \quad (3.58)$$

em que $\mu(n)$ é o fator de passo (possivelmente variável no tempo), ρ é uma constante no intervalo $[0.5, 1]$, a função g é aplicada separadamente a cada elemento do vetor $\mathbf{y} = \mathbf{W}\mathbf{x}$ e os dados são assumidos “esferatizados” (branqueados).

De uma forma equivalente, a variação da matriz de separação pode ser dada na seguinte forma [Kofidis, 2001]:

$$\Delta \mathbf{W}(n) \propto [\mathbf{x}^T(n) - \mathbf{W}(n)\mathbf{f}(\mathbf{z}^T(n))] \mathbf{f}(\mathbf{z}(n)), \quad (3.59)$$

em que \mathbf{W} é suposta ser ortogonal e, para relembrar, $\mathbf{f}(\cdot)$ é o vetor com as funções escore, definido na página 53. Ao utilizar-se a aproximação

$$E \{ \mathbf{f}(\mathbf{z}(n))\mathbf{f}(\mathbf{z}^T(n)) \} \approx \mathbf{I},$$

válida quando as estimativas z_i são amostras aproximadamente independentes, a Equação (3.59) pode ser escrita como

$$\Delta \mathbf{W}(n) \propto - [\mathbf{W}(n) - \mathbf{f}(\mathbf{z}(n))\mathbf{x}^T(n)]. \quad (3.60)$$

Assim, supondo a ortogonalidade de \mathbf{W} , a Equação (3.60) corresponde à Equação (3.30) com sinal algébrico contrário.

O número de algoritmos em separação cega de fontes é bastante significativo. Uma vez que várias abordagens e considerações podem ser combinadas para dar origem a regras de aprendizado de estruturas de separação. Uma lista de referências sugerida como ponto de partida para vários outros algoritmos e estratégias pode ser encontrada em [Cardoso, 1998; Haykin, 2000a; Cichocki & Amari, 2002].

A Tabela 3.1 resume os algoritmos discutidos até o momento destacando suas principais características.

Algoritmo	Características
Jutten-Hérault	Algoritmo simples que utiliza uma ortogonalização da matriz de separação através de funções não-lineares. Possui a desvantagem de ter a convergência assegurada apenas em condições bastante restritas.
Equivariante	Permite utilizar um critério independente da matriz de mistura através da utilização de funções não-lineares como estimador equivariante. Ainda permite a utilização do gradiente natural para melhoria de desempenho.
<i>Deflation</i>	Algoritmo de extração serial de fontes, bastante utilizado quando somente uma parte das fontes é desejada. Utiliza um critério de ortogonalização das colunas da matriz de mistura, uma vez que as projeções das fontes são ortogonais, para assegurar separação.
<i>FastICA</i>	Mede a negentropia através de aproximações polinomiais. Apresenta, sob certas condições, equivalência com critérios de desconvolução cega.
Aprendizado anti-Hebbiano não-linear	Eliminação das correlações cruzadas entre as fontes pela aplicação da regra de Hebb.
<i>Nonlinear PCA</i>	Aproximação da técnica de projeção linear utilizando um operador não-linear na saída da matriz de separação, forçando com isso a ortogonalização da matriz de separação, de maneira a encontrar a melhor superfície, não necessariamente linear, que fornece separação.

Tabela 3.1: Resumo das características de alguns algoritmos de separação cega de fontes.

Na seção seguinte, são ilustradas algumas aplicações mais clássicas de forma a permitir uma verificação das técnicas existentes e de alguns pontos fracos de algumas delas.

3.3 Aplicações

Prover uma avaliação completa e objetiva das técnicas de separação cega de fontes é tarefa excessivamente árdua e talvez equivocada. O número de soluções, e de combinações de diferentes soluções, é muito vasto; assim como a gama de possíveis aplicações, em vista das quais uma ou outra técnica pode se mostrar mais adequada.

Nesta seção, a meta é ilustrar aplicações práticas representativas do problema e onde estratégias de BSS já se mostraram capazes de recuperação em ambientes onde se dá interferência mútua entre as fontes. A escolha das estratégias foi feita de tal maneira a mostrar a diversidade de algumas abordagens e também que várias aplicações podem ser contempladas a partir do mesmo modelo.

Outro ponto que merece destaque é o fato de que alguns dos algoritmos utilizados nas simulações não foram explicitamente descritos nesta tese, mas sua metodologia encontra-se permeada nas estratégias de separação cega de fontes, sejam elas baseadas em SOS ou em HOS. Uma breve descrição dos algoritmos é realizada em cada uma das aplicações consideradas.

Finalmente, deve-se mencionar que algumas das simulações foram executadas através do uso de um conjunto de programas reunidos em um pacote chamado *ICALAB* (*Independent Component Analysis Laboratory*) desenvolvido no *Laboratory for Advanced Brain Signal Processing* do Instituto RIKEN do Japão [Cichocki & Amari, 2002].

3.3.1 Processamento de sinais de voz

Uma das aplicações clássicas em separação de fontes é derivada diretamente do problema *cocktail party*. A detecção e separação de sinais de voz em um ambiente qualquer apresenta-se como um cenário típico para a aplicação de técnicas de BSS. A Tabela 3.2 mostra os parâmetros do sistema de mistura/separação.

A potência do ruído em cada sensor é dada através da relação sinal-ruído (SNR, *Signal-to-Noise Ratio*), definida como:

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{\sigma_a^2}{\sigma_v^2} \right), \quad (3.61)$$

em que σ_v^2 é a potência do ruído e σ_a^2 é a potência do sinal de interesse.

Na Figura 3.4 observa-se o conjunto de fontes de sinais de voz. Embora a análise espectral de tais sinais possa fornecer alguma informação que permita a separação através de uma

Número de fontes	4
Número de sensores	4
Ruído	gaussiano com SNR = 20 dB em cada sensor
Matriz de mistura	mistura instantânea com coeficientes reais aleatórios
Algoritmo de separação	TICA

Tabela 3.2: Parâmetros do sistema de mistura/separação de processamento de sinais de voz.

filtragem em frequência, em um sistema que não dispõe das fontes nem mesmo de suas características espectrais é imperativa a utilização de técnicas de BSS.

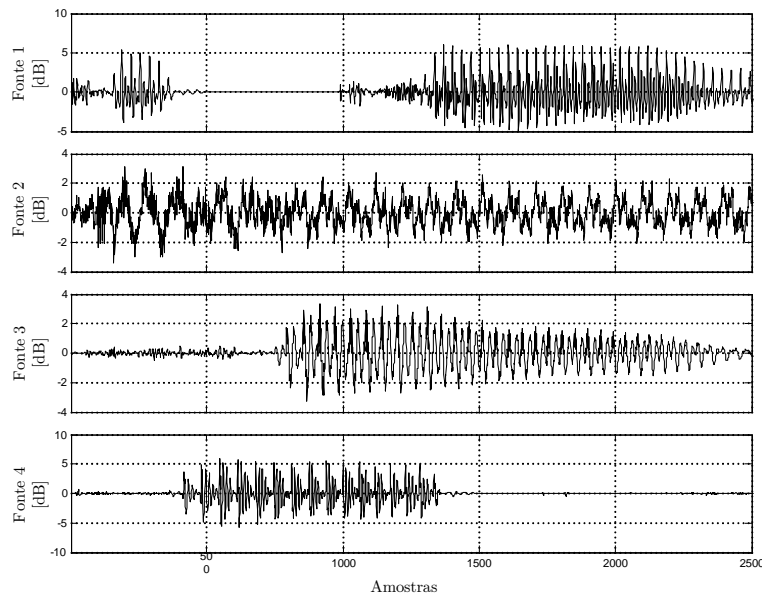


Figura 3.4: Fontes de sinais de voz.

O conjunto de sinais disponíveis no conjunto de receptores é então uma combinação linear das amostras instantâneas das diversas fontes, conforme ilustrado na Figura 3.5. Como não se dispõe do sistema de mistura, não é possível inferir nenhuma informação *a priori* sobre as fontes desconhecidas.

Neste exemplo, aplicou-se um algoritmo que utiliza funções contraste definidas como função dos cumulantes e das matrizes de correlação dos dados para diferentes atrasos. Desta forma, este algoritmo combina as características de HOS e SOS para extração dos diversos sinais. A sigla TICA significa *Thin algorithm for Independent Component Analysis* [Cichocki & Amari, 2002]. Pode-se observar, na Figura 3.6, como o algoritmo é capaz de fornecer boas estimativas

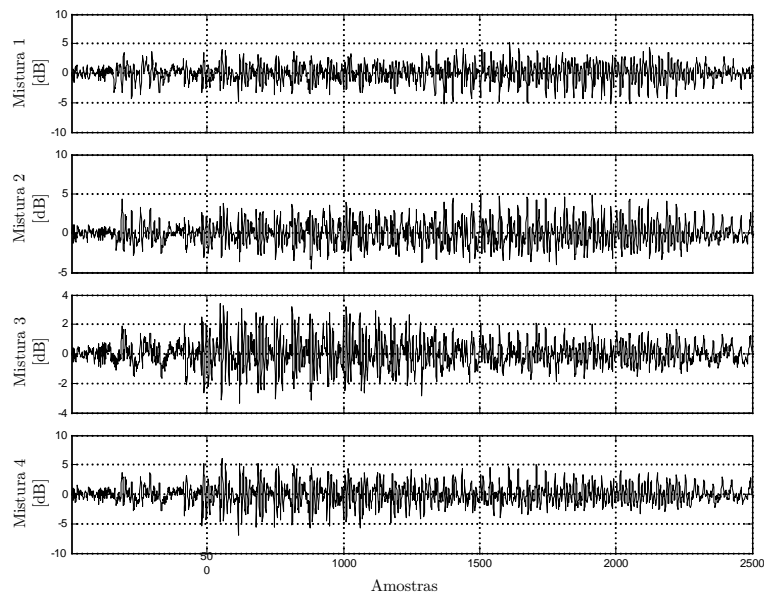


Figura 3.5: Misturas disponíveis nos sensores provenientes de sinais de voz.

das fontes mesmo com a presença de um ruído de potência igual a 20 dB. Claramente, há uma permutação da ordem e escalonamento das estimativa das fontes, conforme previsto pelo modelo teórico.

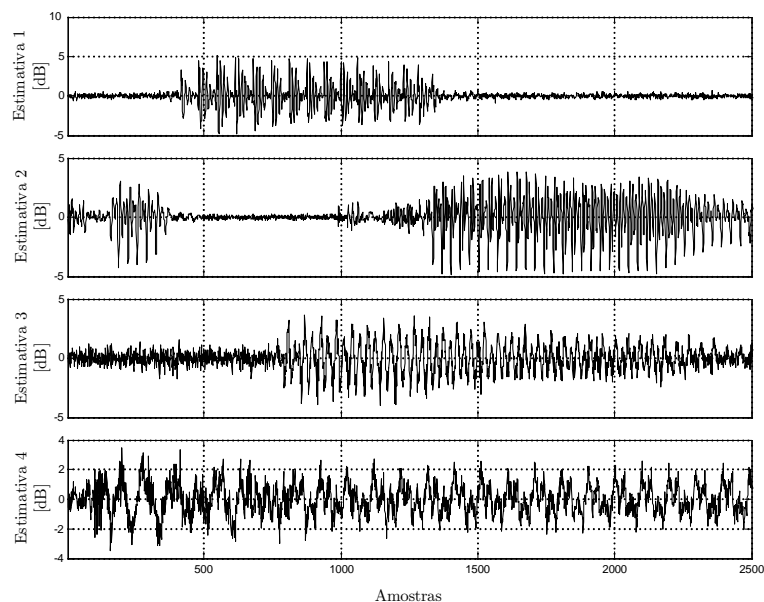


Figura 3.6: Estimativas das fontes de voz.

Uma forma de medir objetivamente a eficiência da separação é através da matriz global

de separação \mathbf{G} , que corresponde à resposta conjunta do sistema, dada por $\mathbf{G} = \mathbf{HW}$. Uma vez que esta deveria ser, idealmente, a matriz identidade ou uma permutação de suas colunas, a matriz obtida após a convergência ilustra quão próximo da separação perfeita o sistema se encontra. No exemplo em questão, o módulo dos coeficientes da matriz de separação global é dado por

$$\mathbf{G} = \begin{bmatrix} 0.0094 & 0.0336 & 0.1571 & 1 \\ 1 & 0.0195 & 0.0472 & 0.0079 \\ 0.0048 & 0.0315 & 1 & 0.0124 \\ 0.0095 & 1 & 0.0266 & 0.0132 \end{bmatrix}.$$

Vale ressaltar que esta determinação da matriz de separação global só é possível em termos de simulação para se verificar que a separação foi atingida, uma vez que na prática não se dispõe do sistema de mistura para o cálculo da matriz $\mathbf{G} = \mathbf{HW}$.

3.3.2 Sinais de comunicação digital

Viabilizar e maximizar a capacidade de comunicação de múltiplas fontes é uma das principais metas dos modernos sistemas de comunicação digital. Um dos maiores limitantes para se atingir esta meta é a interferência produzida por várias fontes atuando no sistema de comunicação o que torna a aplicação de técnicas de cancelamento de interferência ou recuperação de informação necessária para o funcionamento satisfatório do sistema.

Neste exemplo, são utilizadas fontes com símbolos de uma modulação QPSK (*Quadrature Phase Shift Keying*) de módulo unitário, transmitidas simultaneamente e recuperadas de forma síncrona, ou seja, sem atrasos relativos entre as diversas fontes. A Tabela 3.3 ilustra os parâmetros do sistema de comunicação digital.

Número de fontes	3
Número de sensores	4
Ruído	sem ruído
Matriz de mistura	mistura instantânea com coeficientes complexos aleatórios
Algoritmo de separação	EASI

Tabela 3.3: Parâmetros do sistema de mistura/separação de sinais de comunicação digital com modulação QPSK.

A escolha da matriz de mistura, que neste caso representa o meio de transmissão das diversas fontes, com coeficientes complexos deve-se ao fato de que desta forma é possível modelar prováveis desvios de fase e não somente de amplitude, uma vez que tal fenômeno é

bastante comum em sistemas de comunicação digital. Neste exemplo, foi suprimido o ruído para facilitar a visualização das constelações dos sinais recebidos e processados.

Na Figura 3.7, pode-se observar a constelação obtidas nos diversos sensores. Os símbolos \diamond , $+$, \times e \bullet representam os sinais em cada um dos diferentes sensores. As notações $\text{Re}(\cdot)$ e $\text{Im}(\cdot)$ representam, respectivamente, a parte real e imaginária dos sinais .

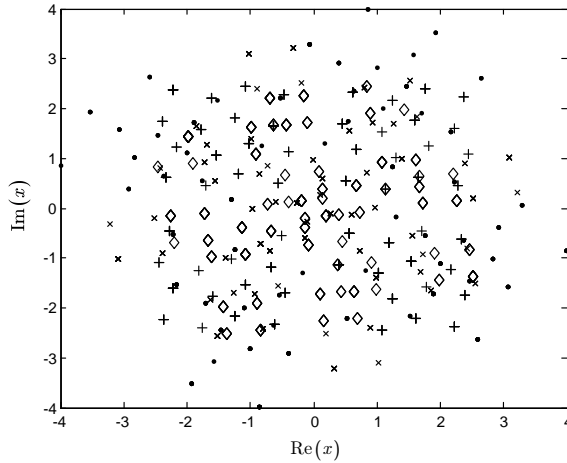


Figura 3.7: Sinais de mistura de um sistema de comunicação digital com modulação QPSK.

Uma vez que as várias fontes possuem a mesma fdp, a tarefa de separação de fontes consiste em retirar de cada uma delas a interferência promovida pelas demais. Assim, após o processamento com o algoritmo equivariante EASI [Cardoso & Laheld, 1996], que utiliza funções não-lineares para promover a separação, obtém-se a distribuição na saída do sistema de separação representada na Figura 3.8. Novamente, os símbolos \times , \diamond e \bullet representam os sinais provenientes de cada uma das diferentes saídas do sistema.

Vale relembrar que os sinais QPSK de potência unitária são do tipo $\frac{1}{\sqrt{2}}(\pm 1 \pm j)$. Desta forma, a Figura 3.8 mostra constelações QPSK (para cada uma das estimativas das fontes) com rotações de fase decorrentes da ambiguidade de permutação do processo de BSS.

A matriz de separação global do sistema obtida após a convergência é, em módulo,

$$\mathbf{G} = \begin{bmatrix} 0.9817 & 0.0752 & 0.0499 \\ 0.0619 & 0.9758 & 0.1287 \\ 0.0436 & 0.0762 & 0.9686 \end{bmatrix}.$$

Uma questão recorrente em sistemas de comunicação digital é referente à taxa de convergência do algoritmo, uma vez que vários sistemas operam com cancelamento de interferência e separação de fontes em tempo real. Desta forma, uma figura de mérito que se

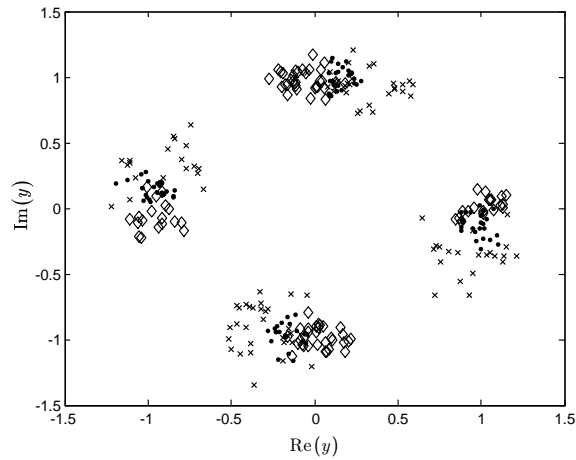


Figura 3.8: Fontes estimadas de sinais QPSK.

pode utilizar é avaliar a evolução temporal do módulo dos coeficientes da matriz de separação global. De forma ideal, os coeficientes deveriam convergir para os valores 1 e 0. Na Figura 3.9 observa-se a evolução dos coeficientes de separação (da matriz de separação global) em função do número de iterações. Neste caso, um número bastante pequeno de símbolos foi necessário para que o sistema convergisse para uma solução. Para aproximadamente 100 símbolos a convergência é atingida. Em alguns casos mais complexos, ou seja, com um número maior de fontes, presença de ruído, número de sensores, tipo de receptor, etc, o número de iterações necessárias pode ser bem maior.

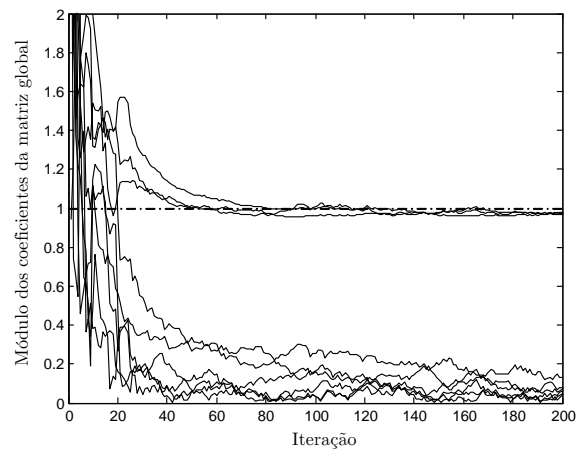


Figura 3.9: Evolução da matriz de separação global para um sistema de fontes de comunicação com modulação QPSK.

3.3.3 Processamento de sinais biomédicos

Uma outra área na qual a aplicação de técnicas de processamento de sinais é de grande importância para a melhoria da resolução dos tipos de sinais envolvidos é a área biomédica.

A complexidade do corpo humano faz dele um ambiente cuja modelagem apresenta severas dificuldades. A tentativa de isolar sinais provenientes de alguma parte do corpo é permeada de dificuldades, uma vez que cada organismo apresenta sinais com características diferentes. Desta maneira, a utilização de estratégias que não necessitam conhecimento *a priori* dos sinais envolvidos, mais que interessante torna-se necessária.

Este exemplo busca exemplificar a utilização de estratégias de BSS em problemas de recuperação de dados de sinais biomédicos através de um conjunto de sensores dispostos na superfície corporal de um paciente¹⁵. A Tabela 3.4 mostra os parâmetros do sistema de separação utilizados.

Número de fontes	4
Número de sensores	4
Ruído	gaussiano com SNR = 20 dB em cada sensor
Matriz de mistura	mistura instantânea com coeficientes reais aleatórios
Algoritmo de separação	NG-FICA

Tabela 3.4: Parâmetros do sistema de mistura/separação de sinais biomédicos.

A Figura 3.10 ilustra 4 sinais biomédicos típicos captados por sensores dispostos na superfície corporal, são eles, respectivamente:

- (1) sinal gaussiano filtrado - emulação de ruído de medida correlacionado;
- (2) eletrocardiograma;
- (3) artefato de movimento - medida de sinal em nervo eferente;
- (4) respiração.

Pode-se observar que os sinais possuem características bastante distintas e que alguns deles apresentam comportamento similar ao de ruído branco, o que torna ainda mais difícil uma separação/identificação de tais sinais por técnicas clássicas de filtragem.

Na Figura 3.11, observam-se os diversos sinais disponíveis nos sensores, e pode-se verificar que os sinais recebidos não apresentam grande correlação com os sinais das fontes. Na

¹⁵Um caso particular clássico diz respeito ao eletrocardiograma fetal discutido no Capítulo 1. Ver também, por exemplo, [Cichocki & Amari, 2002].

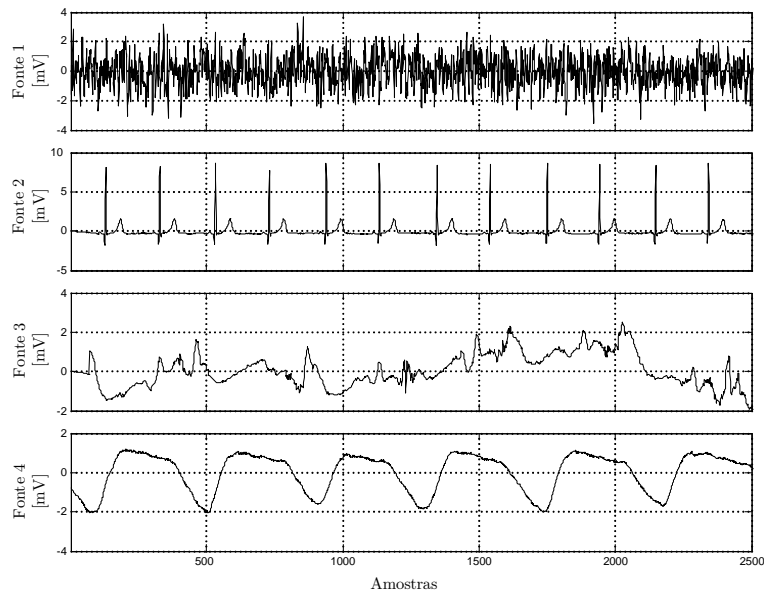


Figura 3.10: Alguns sinais biomédicos típicos.

verdade, seria possível supor erroneamente que há apenas um tipo de fonte no sistema, dada a similaridade dos sinais, a menos de fatores de escala.

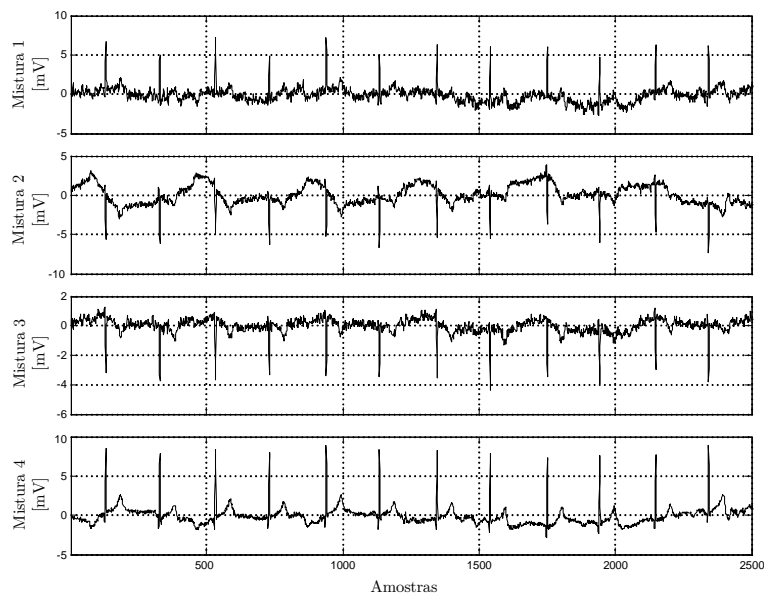


Figura 3.11: Sinais disponíveis nos sensores de um sistema de processamento de sinais biomédicos.

Após o processamento com o algoritmo NG-FICA, sigla de *Natural Gradient Flexible ICA*

[Cichocki & Amari, 2002], pode-se verificar a separação dos sinais, através das estimativas das fontes, mostradas na Figura 3.12. Novamente, é possível observar que a recuperação (estimativas) dos sinais das fontes apresenta, além de versões ruidosas dos sinais originais, permutação da ordem e alterações de escala.

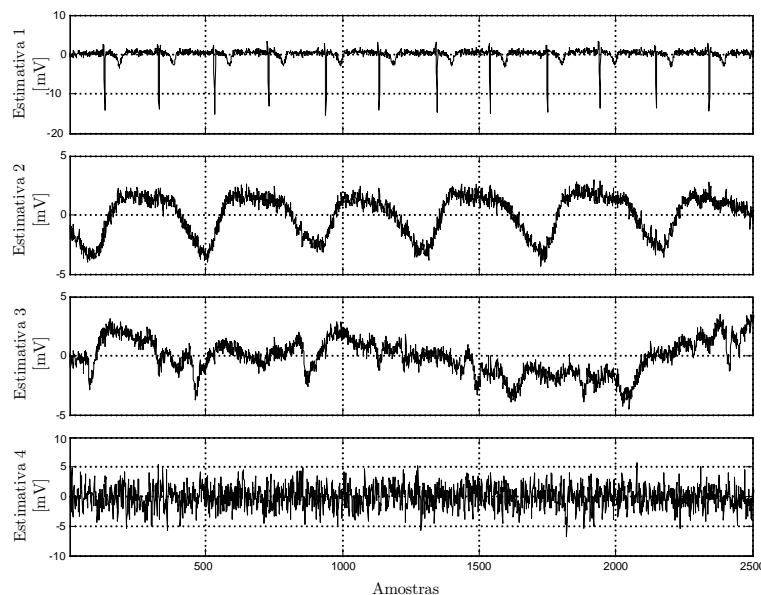


Figura 3.12: Estimativas das fontes de sinais biomédicos.

O algoritmo é baseado, como a maioria dos outros, na estimativa de funções não-lineares. A diferença deste caso específico é que ele é baseado na utilização da distribuição gaussiana generalizada, mostrada na Equação (2.35), para a estimativa da *kurtosis* e inferência sobre o modelo da distribuição, se gaussiana, sub-gaussiana ou super-gaussiana. Além disto, é utilizado o gradiente natural, no lugar do gradiente estocástico, para realizar a busca adaptativa da solução [Cichocki & Amari, 2002].

A matriz (módulo) de separação global obtida após o processo de separação é

$$\mathbf{G} = \begin{bmatrix} 0.0825 & 1 & 0.1337 & 0.1433 \\ 0.0055 & 0.0489 & 1 & 0.0950 \\ 1 & 0.0613 & 0.0432 & 0.0024 \\ 0.1271 & 0.0938 & 0.0975 & 1 \end{bmatrix},$$

na qual se observa pouca interferência de outros sinais naquele recuperado em cada filtro (colunas da matriz de separação), como indicam os baixos coeficientes que compõem as colunas da matriz.

3.3.4 Aquisição de sinais eletroencefalográficos

Ainda no campo das aplicações em biomédicas, um dos principais objetos de estudo de estratégias de separação de fontes é no processamento de sinais eletroencefalográficos.

Devido à extrema complexidade do sistema nervoso humano, a coleta de dados específicos é uma tarefa de difícil realização. Uma das principais técnicas para extração dos sinais desejados é o isolamento dos sinais através da tentativa de eliminar a influência de outros sinais existentes na coleta de dados. Este isolamento permite então estimar a região de ativação do cérebro correspondente àquela excitação, conforme descrito no Capítulo 1.

Por exemplo, uma situação comum na realização de um exame para coleta de sinais eletroencefalográficos é solicitar ao paciente que ele não se mova, em nenhuma hipótese, durante toda a duração do exame que dura em média 1 hora. Isto visa solucionar o problema de separação de fontes *eliminando* as fontes¹⁶.

Além do incômodo de se ficar imóvel durante tanto tempo, tal estratégia é também de desempenho questionável, já que podem ocorrer combinações seja durante alguns instantes específicos em que o paciente não respeite a instrução de não se mover, seja por alguns “espamos” musculares capazes de corromper os sinais, tais como piscar de olhos, movimento do maxilar, espirros, etc. Há até mesmo fontes artificiais externas como relógios e outros equipamentos. Desta forma, estratégias de BSS são de extrema importância no campo de aquisição de sinais eletroencefalográficos.

A Tabela 3.5 ilustra os parâmetros de uma aplicação simples de um sistema de BSS para aquisição de sinais eletroencefalográficos.

Número de fontes	4
Número de sensores	4
Ruído	sem ruído
Matriz de mistura	mistura instantânea com coeficientes reais aleatórios
Algoritmo de separação	Pearson

Tabela 3.5: Parâmetros do sistema de mistura/separação de sinais eletroencefalográficos.

Na Figura 3.13, pode-se observar alguns sinais tipicamente envolvidos no processamento de sinais em um exame de eletroencefalograma. Duas das fontes são sinais provenientes do cérebro e outras duas são sinais interferentes provenientes do piscar de olhos e ativação da musculatura do maxilar.

¹⁶Na verdade, o que se tenta através disto é minimizar o máximo possível a influência de outros sinais naqueles de interesse.

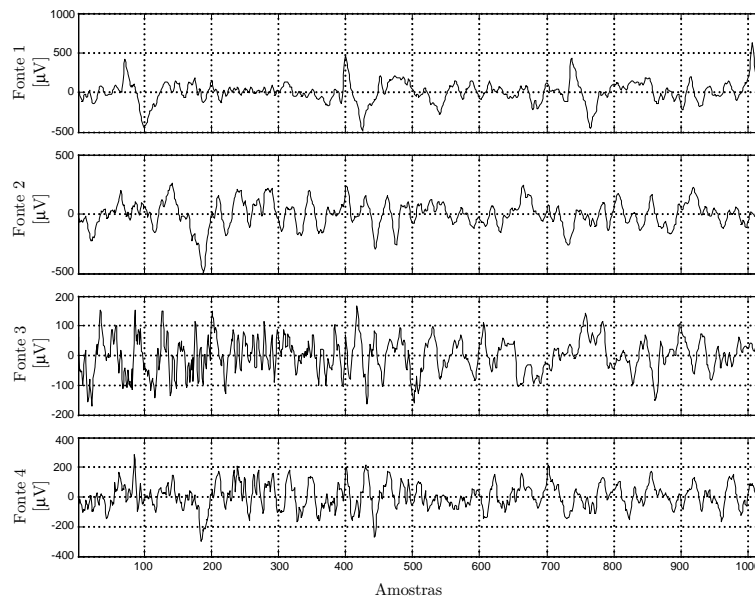


Figura 3.13: Fontes de um sistema de aquisição de sinais eletroencefalográficos.

Os sinais coletados nos sensores são ilustrados na Figura 3.14.

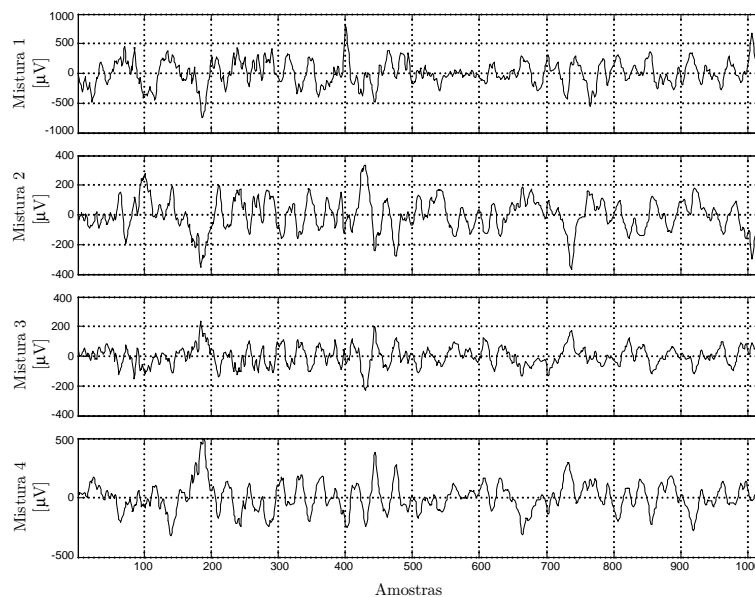


Figura 3.14: Misturas de sinais eletroencefalográficos.

O algoritmo de Pearson utiliza uma função dos cumulantes de terceira e quarta ordem para calcular a negentropia das fontes e o gradiente natural para adaptação dos parâmetros de separação [Haykin, 2000a; Cichocki & Amari, 2002]. Desta forma, o algoritmo é bastante

similar ao algoritmo *FastICA* para cálculo da máxima verossimilhança [Cichocki & Amari, 2002; Mutihac & Hulle, 2003].

As estimativas das fontes podem ser visualizadas na Figura 3.15. Conforme pode ser também observado a partir da matriz de separação global dada por

$$\mathbf{G} = \begin{bmatrix} 0.0986 & 0.4941 & 0.3514 & 1 \\ 1 & 0.0393 & 0.1586 & 0.2587 \\ 0.0188 & 0.1554 & 1 & 0.2120 \\ 0.0169 & 1 & 0.3310 & 0.0904 \end{bmatrix},$$

as estimativas apresentam ainda forte interferência, dificultando a identificação das fontes originais envolvidas no processo. Além disso da interferência, a indeterminação em relação à permutação e escalonamento da solução, conforme pode ser observado na Figura 3.15, torna a identificação dos sinais das fontes ainda mais difícil neste caso.

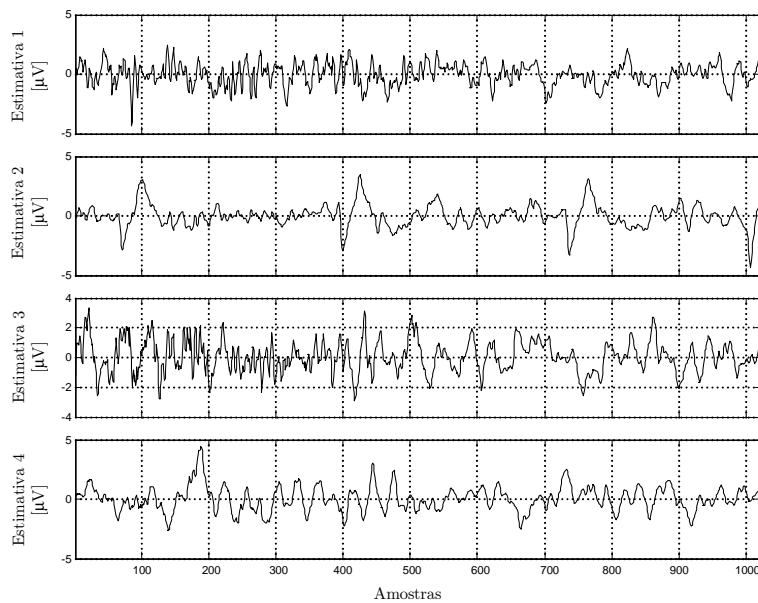


Figura 3.15: Estimativas das fontes de sinais eletroencefalográficos.

Obviamente, as estimativas das fontes poderiam ser melhoradas por modificações na estratégia de separação, tal como aumento do número de sensores, utilização de outros algoritmos ou modificação de parâmetros do algoritmo utilizado [Cichocki & Amari, 2002].

3.3.5 Extração de imagens

A extração de imagens, não conhecidas *a priori*, em ambientes contaminados por interferência é mais um problema de destaque na área de BSS. Tal problema diz respeito a várias aplicações onde há incerteza mesmo sobre a existência ou não de fontes. Pode-se citar, entre outras aplicações, os casos de ultra-sonografia, recuperação de imagens deterioradas, tratamento de imagens de astronomia, etc.

A Tabela 3.6 ilustra as características de um exemplo de sistema de separação para a extração de imagens. Deve-se mencionar que as imagens são combinadas *pixel a pixel* através da matriz de mistura obtida de forma aleatória. Na Figura 3.16, pode-se visualizar as duas fontes de imagem que são empregadas para a simulação. As duas imagens são extremamente conhecidas no campo do reconhecimento de imagens.

Número de fontes	2
Número de sensores	2
Ruído	gaussiano com SNR = 20 dB em 30% dos <i>pixels</i>
Matriz de mistura	mistura instantânea com coeficientes reais aleatórios
Algoritmo de separação	<i>FastICA</i>

Tabela 3.6: Parâmetros do sistema de mistura/separação de extração de imagens.

As misturas obtidas nos sensores são ilustradas na Figura 3.17. Pode-se observar claramente que as imagens em cada um dos sensores apresentam características das duas fontes. É importante lembrar que esta informação não pode ser utilizada, uma vez que não se conhece, no receptor, quais os sinais originais.



Figura 3.16: Imagens fonte.

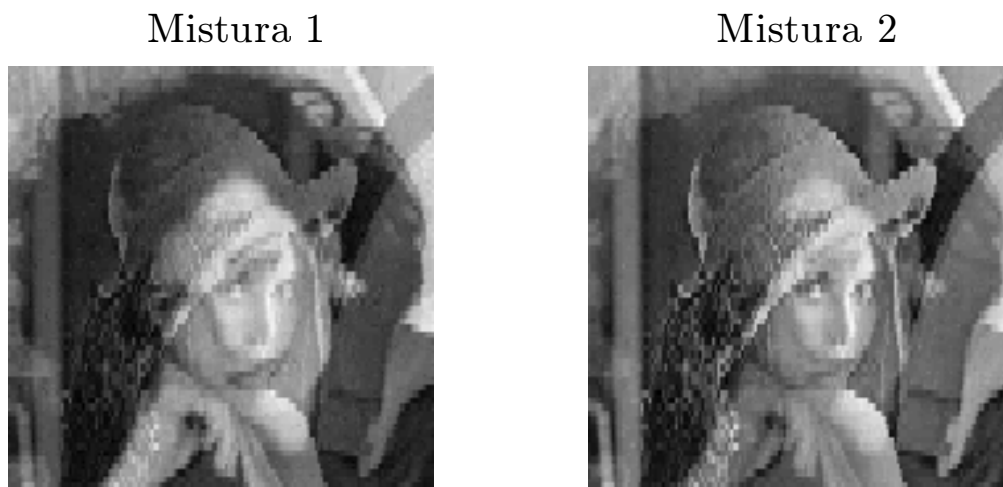


Figura 3.17: Misturas bidimensionais resultantes da composição de imagens em um sistema de mistura instantânea.

A matriz de separação global, obtida após a convergência, é dada por

$$\mathbf{G} = \begin{bmatrix} 1 & 0.0171 \\ 0.0608 & 1 \end{bmatrix},$$

sendo que na Figura 3.18 podem ser observadas as imagens estimadas a partir dos sinais bidimensionais recebidos.

Estimativas 1



Estimativas 2

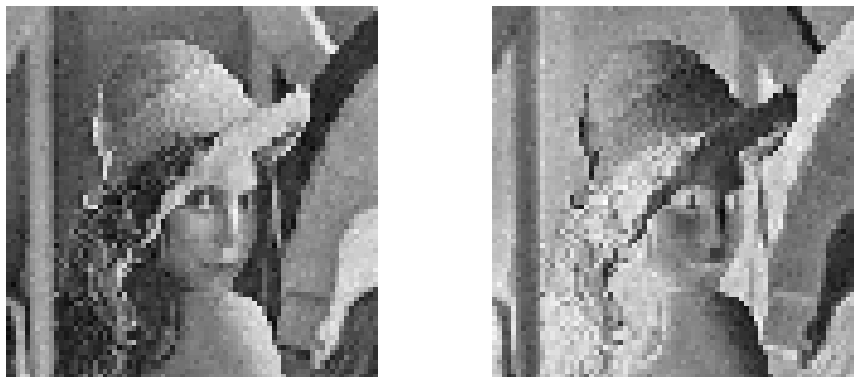


Figura 3.18: Estimativas dos sinais imagem fontes.

Uma importante informação a ser destacada é que a existência de duas estimativas para cada sinal original é um fato já esperado. Isto ocorre devido ao fato de que a estrutura de separação apresentam pontos de equilíbrio para os sinais originais e para os sinais duais, ou seja, aqueles com os *pixels* com tons de cinza invertidos. Este fato é observado também nas redes neurais artificiais auto-organizáveis [Haykin, 1998].

3.4 Conclusões e síntese

Discutem-se, no presente capítulo, algumas das principais estratégias, compreendendo tanto critérios e abordagens quanto algoritmos, relacionadas ao problema de separação cega de fontes. Não é meta do capítulo fazer uma listagem exaustiva dos diversos métodos existentes

na literatura. Os poucos aqui contemplados representam uma parte bastante importante que permite a assimilação de um conhecimento necessário para o entendimento das diversas variantes já referenciadas.

As aplicações aqui apresentadas servem apenas para mostrar o potencial das estratégias frente a problemas práticos nos quais a disponibilidade dos dados, ou de parte deles, é impossível ou limitada por aspectos técnicos e/ou financeiros. Várias outras aplicações seriam passíveis de ser discutidas, entretanto julga-se que aquelas contidas neste texto sejam bastante representativas.

A apresentação de alguns critérios e algoritmos é, ocasionalmente, realizada sem as devidas provas ou demonstrações matemáticas que compõem a derivação dos mesmos. Isto é feito propositadamente para que um maior número de informações, embora com certeza, mais superficiais, possa ser inserida no âmbito do texto. Entretanto, tais pontos que encontram-se com descrição limitada apresentam-se destacados no texto com as devidas referências, cuja leitura preenche as lacunas deixadas neste capítulo.

Novamente, uma lista de publicações com caráter tutorial é sugerida para os interessados:

- ◆ [Haykin, 2000a];
- ◆ [Hyvärinen, Oja, & Karhunen, 2001];
- ◆ [Cichocki & Amari, 2002];
- ◆ [Mutihac & Hulle, 2003].

Um ponto de extrema importância é que os modelos até então considerados permitem aplicação em uma vasta gama de ambientes, considerando fontes com diferentes distribuições, para os quais o uso de estruturas não-lineares é imperativo, conforme foi mostrado nas seções anteriores.

Entretanto, quando algumas das condições sobre as fontes e/ou do sistema de mistura são relaxadas, é possível a utilização de estratégias lineares para solucionar o problema de BSS. Isto é o que se propõe discutir e explorar no capítulo seguinte.

– “A cada época, a ciência deseja vencer uma verdade que a importuna.”

Joseph Arthur Gobineau, 1870

4

Processamento Multiusuário

TÉCNICAS de múltiplo acesso têm sido bastante estudadas nas últimas décadas como uma maneira de prover aumento na capacidade em sistemas de comunicação digital. Entretanto, a complexidade computacional no receptor cresce na proporção direta do aumento do número de usuários, tornando necessária a detecção (identificação e separação) desses usuários.

De um ponto de vista global, o problema de recuperação de sinal em um sistema no qual vários usuários compartilham recursos é o mesmo de separação de fontes. No caso do sistema de múltiplos usuários, as fontes correspondem aos sinais emitidos por cada um dos usuários e a matriz de mistura convolutiva o meio de comunicação, no que diz respeito a distorção, ruído e suas características sistêmicas. Uma diferença crucial entre os dois problemas é que, classicamente, as técnicas de múltiplo acesso utilizam-se da ortogonalidade entre as fontes a partir de estratégias de alocação de recursos.

Neste capítulo, serão descritas algumas técnicas de remoção de interferência e identificação multiusuário clássicas, bem como técnicas de separação de fontes que se utilizam de características estatísticas dos sinais, permitindo o emprego, somente de estratégias lineares e provendo ganhos em alguns aspectos. Características particulares de um sistema de comunicação multiusuário e

algumas ferramentas usuais também são discutidas de forma a prover uma compreensão adequada do problema de detecção multiusuário como um caso particular do de separação de fontes.

O conteúdo é então organizado da seguinte maneira. Na Seção 4.1, são apresentadas sucintamente as estratégias de múltiplo acesso, destacando-se as diferenças entre elas e os problemas que levam à necessidade de processamento do sinal recebido para permitir a correta identificação e recuperação dos usuários. A Seção 4.2 é dedicada ao assunto de antenas adaptativas, que apresentam uma grande flexibilidade de utilização em sistemas de múltiplos usuários. A Seção 4.3 é dedicada a descrever as principais características do canal de comunicação quando utiliza-se um arranjo de antenas no receptor. As características das fontes de um sistema de comunicação multiusuário, que particularizam um sistema de separação de fontes, são descritas na Seção 4.4. Na seqüência, a Seção 4.5 é dedicada à discussão de algumas estratégias que utilizam conceitos de BSS e que são aplicáveis a problemas de detecção multiusuário. Finalmente, as principais conclusões e uma síntese do capítulo são apresentadas na Seção 4.6.

4.1 Esquemas de múltiplo acesso

Como vários usuários podem enviar informações para o mesmo receptor, de tal forma que este seja capaz de identificar e recuperar as informações dos diferentes usuários e, desta forma, maximizar a capacidade de transmissão do sistema?

Existem várias respostas possíveis para a pergunta acima. E a resposta fornecida para um dado sistema caracteriza a estratégia de múltiplo acesso necessária para que a capacidade de transferência de informação em um dado sistema seja a maior possível. Desta forma, é interessante caracterizar alguns dos sistemas de múltiplo acesso para justificar a escolha do sistema de detecção.

Sistemas de múltiplo acesso, de uma forma geral, são classificados quanto à estratégia de alocação de usuários aos recursos disponíveis para a transmissão. Assim, uma vez que o meio de comunicação é compartilhado pelos usuários, alguns procedimentos de diferenciação entre estes usuários são necessários para tornar possível uma recepção correta dos dados transmitidos.

Uma divisão bastante comum é a seguinte [Proakis, 1995; Godara, 1997b]:

- ***Acesso múltiplo por divisão em frequência***

A estratégia de acesso múltiplo por divisão em frequência (FDMA, *Frequency Division Multiple Access*) é utilizada quando uma faixa de frequência é alocada para cada usuário,

de tal forma que não haja superposição das faixas de cada um dos usuários. Desta maneira, a identificação do usuário no receptor é realizada através de uma seleção na frequência dos sinais recebidos.

- ***Acesso múltiplo por divisão no tempo***

Na técnica de TDMA (*Time Division Multiple Access*) cada usuário recebe um intervalo de tempo para transmitir. Após o período para ele alocado, outro usuário recebe a permissão de utilização do canal durante outro intervalo de tempo, utilizando-se a mesma faixa de frequência para todos os usuários.

- ***Acesso múltiplo por divisão em código***

A técnica de espalhamento espectral tornou-se uma alternativa bastante atrativa para fins civis com o uso da técnica de CDMA (*Code Division Multiple Access*). Nela associa-se uma seqüência de código específica, chamada *assinatura*, para cada usuário. Com isto, espalha-se o sinal em toda a faixa de frequência enquanto que no receptor é realizada uma medida de correlação entre cada sinal e as respectivas seqüências de código, para determinar o usuário.

- ***Acesso múltiplo por divisão no espaço***

O caso do SDMA (*Space Division Multiple Access*), possui grande similaridade com o dos sistemas CDMA. A diferença básica reside no fato de que, no SDMA, a assinatura é dita *espacial* e diferencia os usuários por sua posição no espaço, geralmente através da informação angular [Cavalcanti, 1999]. Neste sistema, há também compartilhamento do canal, em seus recursos temporal e frequencial, tal como no CDMA.

A Figura 4.1 ilustra quais os recursos que são compartilhados pelos usuários dependendo da estratégia de múltiplo acesso empregada.

Os sistemas FDMA e TDMA, devido a suas estrutura, são implementados com os mesmos métodos que para sistemas mono-usuário. Assim, o problema de múltiplo acesso nestes casos consiste em alocar um novo usuário para um canal (temporal ou frequencial) disponível. Todavia, esta simplicidade tem um preço a ser pago. A capacidade destes sistemas é baixa uma vez que o número de usuários permitidos é relativamente pequeno devido à ortogonalidade provida através da inserção das faixas e tempos de guarda, necessários para que não ocorra sobreposição dos usuários.

No caso dos sistemas CDMA e SDMA, esta capacidade pode ser consideravelmente aumentada. É importante notar porém que, devido ao compartilhamento do canal, as assinaturas de código ou espacial devem ser projetadas de tal maneira que a correlação entre elas seja nula ou muito pequena, para não permitir interferência de múltiplo acesso (MAI, *Multiple Access Interference*). No CDMA, isto se reflete no uso de seqüências de códigos com correlações cruzadas baixas; no caso SDMA, significa uma restrição sobre a proximidade (angular) entre os usuários.

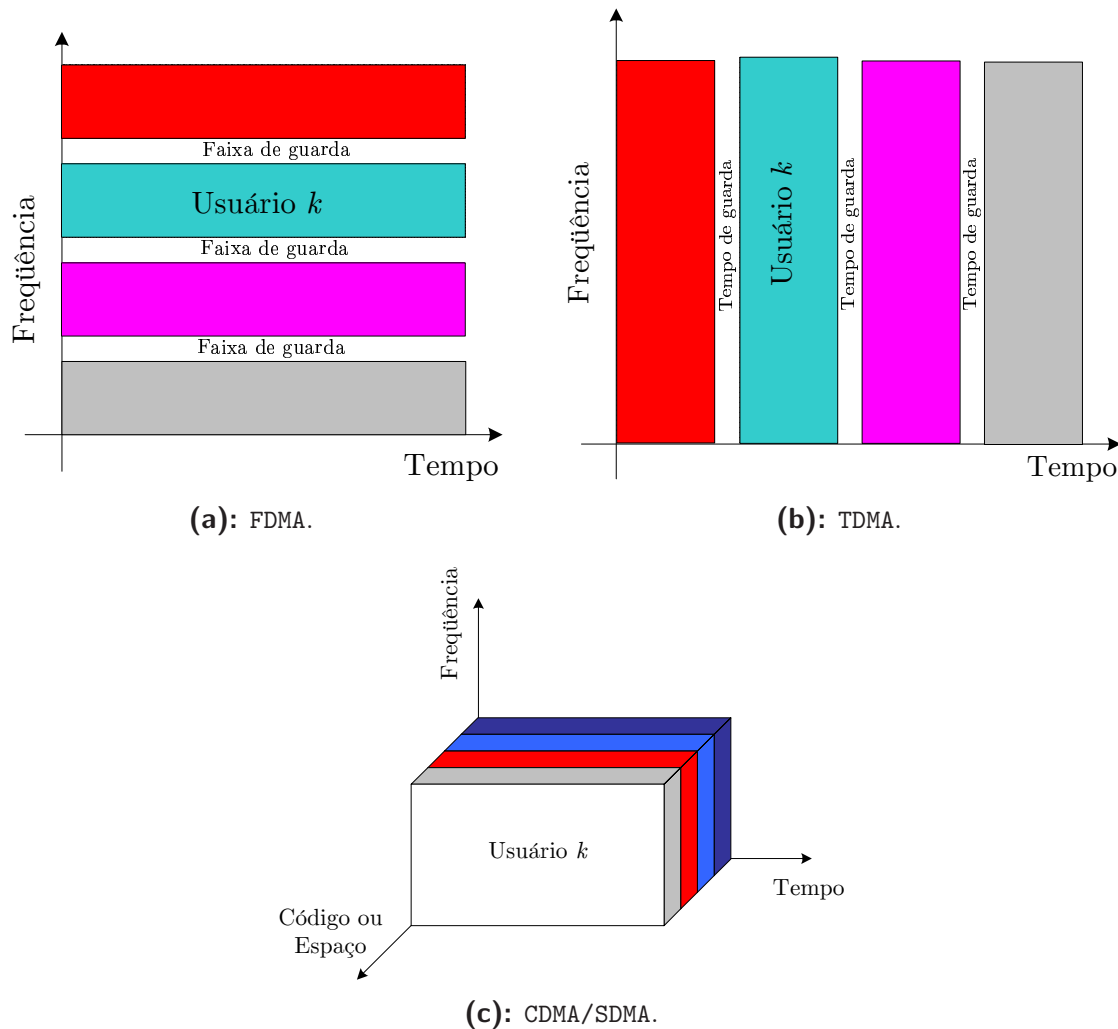


Figura 4.1: Recursos compartilhados em estratégias de múltiplo acesso.

Entretanto, nem sempre é possível prover seqüências de código ortogonais ou garantir a separação angular necessária. Isto implica na necessidade da utilização de técnicas de processamento de sinais que possibilitem recuperar o sinal de interesse, a partir de um sinal corrompido por interferência inter-simbólica (ISI, *InterSymbol Interference*) e/ou MAI.

De uma forma geral, as técnicas de detecção multiusuário se aplicam a qualquer sistema de múltiplo acesso. Entretanto, o sistema CDMA tornou-se bastante popular devido à sua capacidade de alocação de usuários [Proakis, 1995; Verdú, 1998] e o termo *detecção multiusuário* (MUD, *Multiuser Detection*) é, praticamente, vinculado à aplicação em sistemas de espectro espalhado. Deste modo, nesta tese será utilizada a expressão **processamento multiusuário** para designar a aplicação de técnicas de recuperação da informação em um sistema multiusuário qualquer.

Analisando-se um caso genérico, ou seja, sem a preocupação imediata do tipo de processamento multiusuário a ser realizado, pode-se verificar que o receptor de um sistema de múltiplos usuários deve ser capaz de retirar dos sinais recebidos os sinais das diversas fontes para que sejam extraídas as características desejadas e ainda eliminadas as interferências para cada um dos sinais fonte. Em sistemas de comunicação sem fio, uma configuração do receptor capaz de aproveitar as diversas propriedades dos sinais para permitir uma seleção das particularidades de cada uma das fontes é a combinação de elementos em uma estrutura chamada de *antena inteligente*, também conhecida como antena adaptativa. A descrição de tal ferramenta é o assunto da seção seguinte.

4.2 Antenas adaptativas

Um *arranjo de antenas* é um conjunto de sensores espacialmente separados, cujas saídas são combinadas para se extrair alguma informação desejada sobre determinado sinal. Este tipo de estrutura pode ser implementada tanto como transmissor quanto como receptor de um sistema de comunicação. Algumas hipóteses são geralmente consideradas quando se analisa um arranjo de antenas [Petrus, 1997]:

- H_{AA1}**. Todos os sinais incidentes no arranjo receptor são compostos de um número finito de ondas planas que resultam dos diversos multipercursos, incluindo a direção principal;
- H_{AA2}**. O transmissor (receptor) e objetos que provocam multipercursos estão localizados distantes o suficiente do receptor (transmissor);
- H_{AA3}**. Os sensores estão suficientemente próximos, de tal maneira que as amplitudes e as direções de chegada dos sinais em dois (ou mais) elementos do arranjo não diferem significativamente;
- H_{AA4}**. Cada sensor possui o mesmo diagrama de radiação;
- H_{AA5}**. O acoplamento mútuo entre os elementos do arranjo é desprezível.

A geometria do arranjo pode ser arbitrária (linear, circular, planar), sendo que a resposta do arranjo para cada uma das configurações é diferente. Entretanto, por sua simplicidade, o modelo linear é preferido e será o modelo tratado nesta tese. Assim, a Figura 4.2 ilustra um arranjo de antenas e os planos incidentes das ondas [Petrus, 1997].

Através da hipótese H_{AA2}, na qual se supõe que a distância do arranjo para o transmissor (receptor) é muito maior que a separação entre os elementos, a direção de chegada (DOA, *Direction Of Arrival*) do sinal incidente no arranjo pode ser descrita apenas pela sua coordenada azimutal φ , uma vez que seu componente de elevação é considerado no horizonte, ou seja, proveniente da direção 90°.

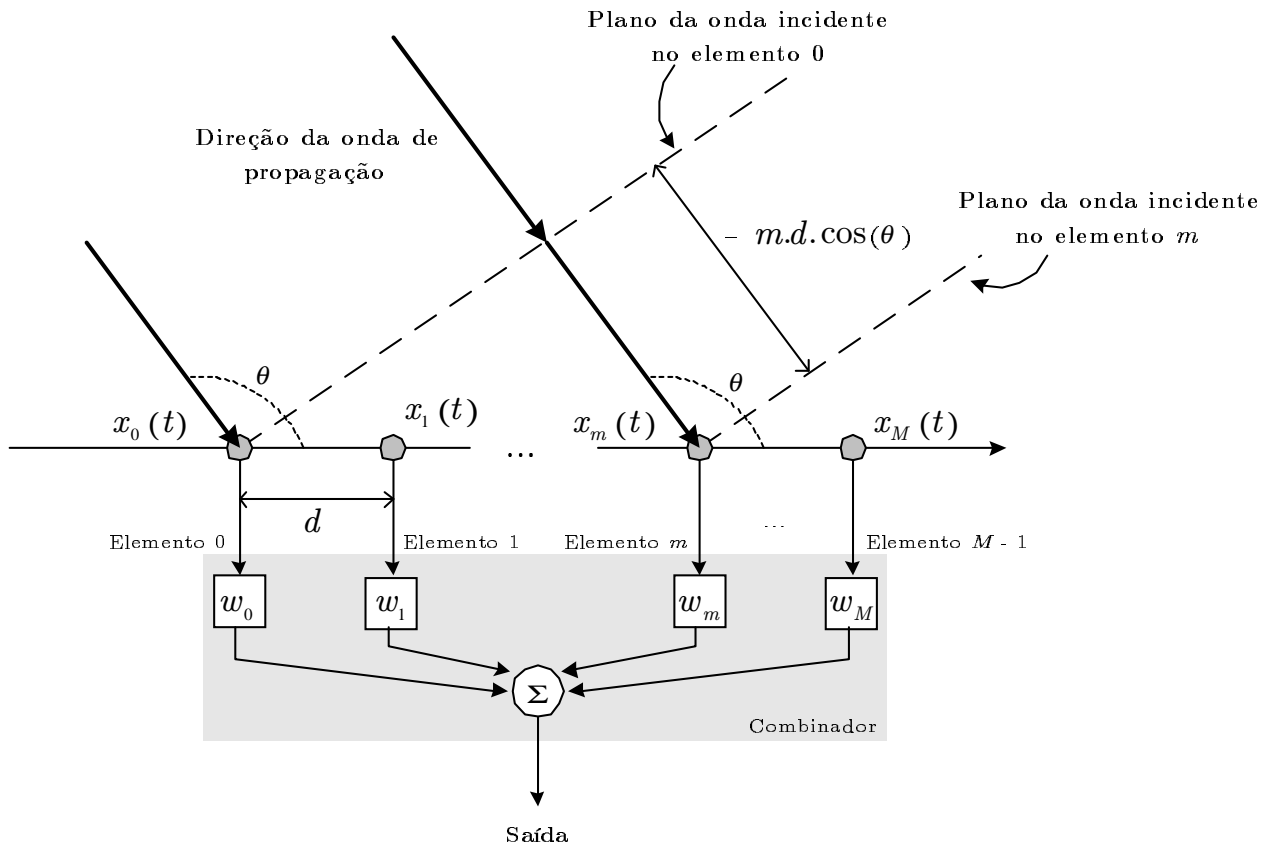


Figura 4.2: Ilustração do plano de onda incidente em um arranjo linear de antenas.

De maneira a simplificar a manipulação matemática, pode-se representar o sinal incidente no m -ésimo elemento do arranjo como sendo dado por

$$x_m(t) = \epsilon_m(t) \cdot \exp [j\varphi(t)], \quad (4.1)$$

em que $\epsilon_m(t)$ é a magnitude e $\varphi(t)$ é a fase do sinal e j representa o número complexo $\sqrt{-1}$.

É ainda definida uma variável complexa $f_m(\theta)$, que representa a razão entre o sinal recebido no elemento m e o sinal recebido no elemento de referência (considerado o elemento 0) quando uma onda plana incide no arranjo. Desta forma, para a configuração linear, obtém-se [Petrus, 1997; Cavalcanti, 1999; Liberti & Rappaport, 1999]:

$$f_m(\theta) = \exp \left(j \frac{2\pi \cdot (m-1) \cdot d \cdot \sin(\theta)}{\lambda} \right), \quad (4.2)$$

em que d é a distância entre os elementos do arranjo e λ é o comprimento de onda da portadora. Uma vez que, por questões de simplicidade, é adotada uma fase de referência nula no elemento

0 do arranjo, chega-se ao *vetor de resposta do arranjo*¹ dado por [Godara, 1997a]:

$$\mathbf{f}(\theta) = \begin{bmatrix} 1 \\ f_2(\theta) \\ \vdots \\ f_M(\theta) \end{bmatrix}. \quad (4.3)$$

Um caso de particular interesse para sistemas que empregam arranjos de antenas é aquele no qual a distância entre os elementos é igual a meio comprimento de onda da portadora. Logo, para este caso, tem-se $d = \frac{\lambda}{2}$ e a resposta relativa do elemento m é então dada por

$$f_m(\theta) = \exp(j \cdot \pi \cdot (m - 1) \cdot \sin(\theta)). \quad (4.4)$$

Em situações práticas, a escolha do valor do espaçamento dos elementos do arranjo determina o compromisso entre menor acoplamento mútuo e menor amplitude de lóbulos laterais, os quais prejudicam a remoção da interferência [Petrus, 1997; Cavalcanti, 1999; Liberti & Rappaport, 1999].

Uma tarefa comumente associada às antenas adaptativas é a capacidade de *formatação de feixes*. Uma vez que o arranjo recebe amostras espaciais, um filtro chamado de conformador de feixes é responsável pela separação dos sinais que se encontram utilizando a mesma faixa frequencial e temporal mas que são originados de diferentes posições espaciais.

Desta forma, a capacidade de um arranjo de antenas remover a interferência e identificar os usuários está intimamente relacionada ao tipo de processamento que irá fazer o ajuste dos filtros conformadores. Este tópico é abordado nas seções seguintes.

4.3 Caracterização da resposta do canal e do sistema de separação dos usuários

Conforme mencionado na seção anterior, a redundância do sinal recebido no receptor permite que, aliado a estratégias de processamento, um arranjo de antenas seja capaz de extrair informações dos sinais envolvidos em um sistema de múltiplos usuários para separá-los e identificá-los.

Em um modelo genérico em banda básica, pode-se escrever a resposta ao impulso do canal

¹Na literatura o termo é amplamente conhecido por *steering vector*.

entre um usuário k e o elemento de referência do arranjo como sendo

$$h_{0,k}(t) = \sum_{i=0}^{L_k-1} \alpha_{k,i} \cdot \delta(t - \tau_{k,i}), \quad (4.5)$$

em que L_k é o número de multipercursos para o usuário k , $\alpha_{k,i}$ é a variável complexa que modela a atenuação e o deslocamento de fase inseridos pelo canal, e $\tau_{k,i}$ são os atrasos relativos para cada multipercurso do k -ésimo usuário.

Ao tomar-se a resposta ao impulso do canal do usuário k para o m -ésimo elemento do arranjo tem-se:

$$h_{m,k}(t) = f_m(\theta_k) \cdot \sum_{i=0}^{L_k-1} \alpha_{k,i} \cdot \delta(t - \tau_{k,i}), \quad (4.6)$$

em que $f_m(\theta_k)$ é definido pela Equação (4.2).

Deve-se observar que $\alpha_{k,i}$ e $\tau_{k,i}$ são independentes de m . Então, o vetor de resposta impulsiva é dado por:

$$\mathbf{h}_{m,k}(t) = \sum_{i=0}^{L_k-1} \mathbf{f}(\theta_{k,i}) \cdot \alpha_{k,i} \cdot \delta(t - \tau_{k,i}), \quad (4.7)$$

em que $\mathbf{f}(\theta_{k,i})$ é o vetor de resposta do arranjo definido na Equação (4.3) para o DOA do i -ésimo multipercurso do usuário k .

A Equação (4.7) representa um canal invariante no tempo, ou seja, usuários estáticos ou ainda uma observação de um curto intervalo de tempo, no qual as variações dos parâmetros do canal são desprezíveis. No caso dos usuários estarem em movimento ou do tempo de observação não ser desprezível, $\alpha_{k,i}$ varia com o tempo segundo alguma função de autocorrelação que obedece uma certa distribuição de probabilidade [Winters, 1998].

Ao utilizar-se um arranjo de antenas pode-se classificar o tipo de processamento necessário na recepção de sinais de múltiplos usuários em função dos parâmetros do canal de cada usuário. Os itens seguintes descrevem os dois tipos de processamento possíveis.

4.3.1 Processamento espacial

Também chamado de processamento de faixa estreita, este tipo de processamento se caracteriza pelo fato de que todos os componentes de frequência do sinal na saída do receptor apresentam apenas um deslocamento de fase sem nenhuma modificação na amplitude [Petrus, 1997]. Isto significa que a largura de faixa do canal é muito pequena se comparada com a faixa do sinal transmitido.

Isto implica em dizer que o sinal apresenta uma largura de faixa muito pequena se comparada à frequência da portadora.

Desta forma, quando o sinal é considerado de faixa estreita, todos os atrasos de multipercusos do k -ésimo usuário são aproximadamente iguais, ou seja, $\tau_{k,i} = \tau_k$ [Proakis, 1995; Petrus, 1997; Cavalcanti, 1999]. Com isso, a Equação (4.7) pode ser reescrita como:

$$\mathbf{h}_{m,k}(t) = \delta(t - \tau_k) \cdot \mathbf{f}(\theta_k) \cdot \alpha_k = \delta(t - \tau_k) \cdot \mathbf{s}_k, \quad (4.8)$$

em que \mathbf{s}_k é a chamada *assinatura espacial* do k -ésimo usuário, dada por

$$\mathbf{s}_k = \mathbf{f}(\theta_k) \cdot \alpha_k. \quad (4.9)$$

Ao escrever-se o sinal recebido no m -ésimo elemento do arranjo tem-se

$$x_m(t) = \sum_{k=1}^K \sqrt{P_k} a_k(t - \tau_k) \cdot \alpha_{k,m} \cdot f_m(\theta_k) + v(t), \quad (4.10)$$

em que K é o número total de usuários, P_k é a potência do sinal transmitido pelo k -ésimo usuário, a_k é o símbolo transmitido pelo k -ésimo usuário e $v(t)$ é uma v.a. de um processo gaussiano.

Se for considerado que os atrasos τ_k são múltiplos do período de símbolo T e que há perfeito sincronismo de relógio, o seguinte modelo discreto do sinal recebido pode ser escrito [Cavalcanti & Romano, 1999; Cavalcanti, 1999]:

$$\mathbf{x}(n) = \sum_{k=1}^K \sqrt{P_k} a_k(n) \cdot \alpha_{k,m} \cdot \mathbf{f}(\theta_k) + \mathbf{v}(n). \quad (4.11)$$

Ao utilizar-se equações matriciais, pode-se escrever o sinal recebido como sendo:

$$\mathbf{x}(n) = \mathbf{F}\mathbf{a}(n) + \mathbf{v}(n), \quad (4.12)$$

em que

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ \vdots \\ x_M(n) \end{bmatrix}, \quad \mathbf{F} = [\mathbf{f}(\theta_1) \mid \cdots \mid \mathbf{f}(\theta_K)]_{M \times K} \quad \text{e} \quad (4.13)$$

$$\mathbf{a}(n) = \begin{bmatrix} \sqrt{P_1} a_1(n) \\ \vdots \\ \sqrt{P_K} a_K(n) \end{bmatrix}.$$

Desta forma, pode-se ver que o modelo de um sistema de múltiplos usuários é o mesmo que o de um sistema de separação de fontes, quando compara-se as Equações (2.6) e (4.12). Assim, a matriz de mistura convolutiva para o processamento espacial será a matriz \mathbf{F} .

De fato, é possível escrever-se as Equações (4.11) e (4.12) graças à hipótese de que não há ISI, uma vez que todos os atrasos dos diferentes usuários são iguais entre si (sincronismo) e todos os multipercursos têm o mesmo atraso relativo [Verdú, 1998]. Neste caso, a única interferência existente é a MAI. Com isso, uma estrutura de processamento espacial capaz de separar e identificar vários usuários é representada na Figura 4.3. E, conforme mencionado anteriormente, o conjunto de filtros é chamado de filtro conformador de feixes, e uma questão a ser respondida é: como encontrar o conjunto de valores dos coeficientes de cada filtro?

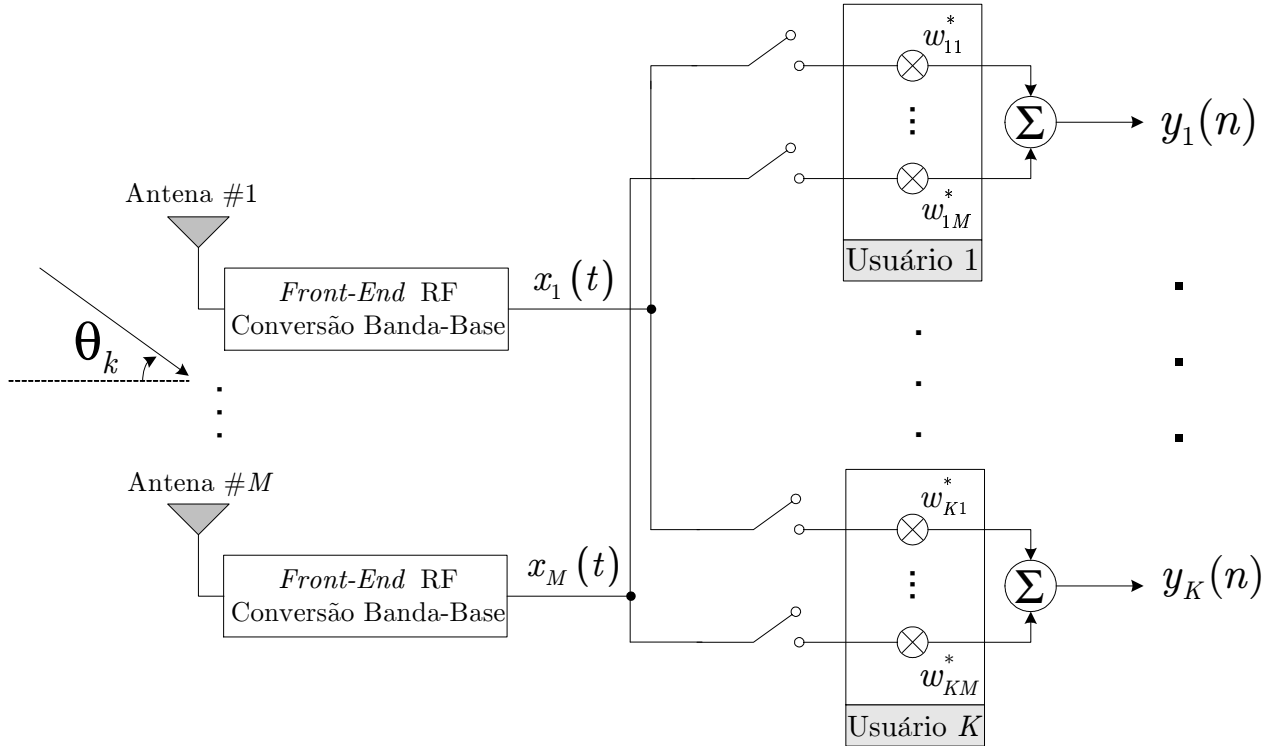


Figura 4.3: Arranjo de antenas e processamento espacial.

Novamente, podemos escrever o sinal recuperado no k -ésimo filtro espacial como sendo dado, na sua forma discreta, por:

$$y_k(n) = \mathbf{w}_k^H(n)\mathbf{x}(n), \quad (4.14)$$

em que o sobrescrito $(\cdot)^H$ indica transposição Hermitiana (transposto conjugado). Ao utilizar-se a forma matricial tem-se então:

$$\mathbf{y}(n) = \mathbf{W}^H(n)\mathbf{x}(n), \quad (4.15)$$

em que

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_K(n) \end{bmatrix} \quad \text{e} \quad \mathbf{W} = [\mathbf{w}_1(n) \mid \cdots \mid \mathbf{w}_K(n)]_{M \times K}. \quad (4.16)$$

Conforme mostrado nas equações anteriores, para separar K usuários são necessários, pelo menos, K filtros espaciais os quais extraem as informações contidas no sinal recebido.

No caso de haver ISI, o modelo e a estrutura de separação são modificadas, como será visto a seguir.

4.3.2 Processamento espaço-temporal

Também chamado de processamento de faixa larga, caracteriza-se por possuir todos os componentes em frequência do sinal com variações tanto em fase como em amplitude [Petrus, 1997]. Isto significa que a largura de faixa do canal é da mesma ordem que a do sinal transmitido, desta forma, há presença de interferência inter-simbólica além da MAI.

Sendo assim, a hipótese dos atrasos de múltiplos percursos para cada usuário serem aproximadamente os mesmos não é mais válida. Então, o vetor da resposta impulsiva do canal de um usuário k para o m -ésimo elemento do arranjo é dado por [Cavalcanti, 1999]:

$$\mathbf{h}_{m,k}(t) = \sum_{i=0}^{L_k-1} \mathbf{f}(\theta_{k,i}) \cdot \alpha_{k,i} \cdot \delta(t - \tau_{k,i}). \quad (4.17)$$

Com isto, o sinal recebido é representado como [Petrus, 1997]:

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=1}^K a_k(t) \star \mathbf{h}_k(t) + \mathbf{v}(t) \\ &= \sum_{k=1}^K \sum_{i=0}^{L_k-1} \mathbf{f}(\theta_{k,i}) \cdot \alpha_{k,i} \cdot a_k(t - \tau_{k,i}) + \mathbf{v}(t), \end{aligned} \quad (4.18)$$

em que a operação \star representa a convolução. Tem-se ainda que o vetor representando o modelo discreto dos sinais incidentes no arranjo de antenas é dado por:

$$\mathbf{x}(n) = \sum_{k=1}^K \sum_{i=0}^{L_k-1} \alpha_{k,i} \cdot a_k(n-i) \cdot \mathbf{f}(\theta_{k,i}) + \mathbf{v}(n), \quad (4.19)$$

em que $h_k(i)$ representa o i -ésimo coeficiente da resposta impulsiva do k -ésimo usuário, representando a variável complexa $\alpha_{k,i}$. Vale ressaltar que na Equação (4.19) está sendo considerado sincronismo entre os sinais das fontes, já que é considerado $\tau_k \propto T$, em que T é o intervalo de símbolo.

Assim, a Equação (4.19) representa a convolução do sinal do usuário k com seu canal, incluindo-se no modelo a ISI inserida pelos diversos multipercursos existentes para cada usuário.

Para se escrever o modelo linear de processamento espaço-temporal em sua forma matricial, faz-se necessária a definição de algumas matrizes e vetores que possuem estruturas especiais, de forma a permitir modelar o sistema através de equações lineares.

A matriz de canal do k -ésimo usuário é dada por [Cavalcanti, 1999]

$$\mathbf{H}_k = \mathbf{F}(\theta_k) \cdot \boldsymbol{\alpha}_k, \quad (4.20)$$

em que:

$$\mathbf{F}(\theta_k) = \left[\mathbf{f}(\theta_{k,0}) \mid \mathbf{f}(\theta_{k,1}) \mid \cdots \mid \mathbf{f}(\theta_{k,L_k-1}) \right]_{M \times L} \quad (4.21)$$

é uma matriz composta dos vetores de resposta do arranjo para o k -ésimo usuário, e

$$\boldsymbol{\alpha}_k = \text{diag} \left(\left[\alpha_k(0) \quad \alpha_k(1) \quad \cdots \quad \alpha_k(L_k - 1) \right] \right) \quad (4.22)$$

é a matriz diagonal de ganhos complexos (atenuações e rotações de fase) do canal relativo ao usuário k , de dimensão $L_k \times L_k$.

Uma estrutura diagramática capaz de realizar a separação de usuários em um sistema que insere MAI e ISI é representada na Figura 4.4. Pode-se observar que para cada antena há um filtro temporal FIR (*Finite Impulse Response*) associado. Desta maneira, o arranjo inclui também a diversidade espacial necessária para mitigar os efeitos da ISI, além da diversidade temporal para cancelar a MAI.

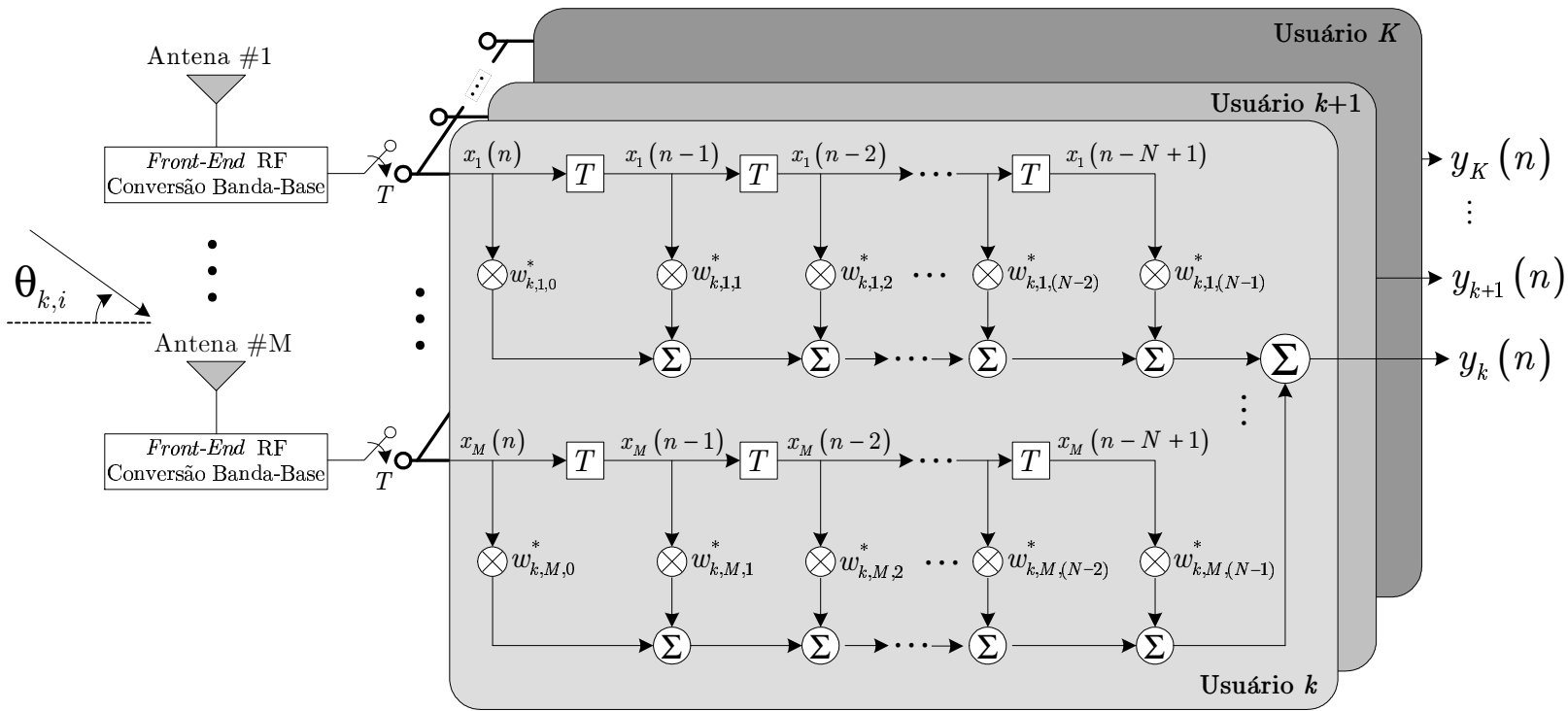


Figura 4.4: Arranjo de antenas e processamento espaço-temporal a partir do k -ésimo usuário.

Na Figura 4.4, observa-se uma notação tensorial, uma vez que são necessários três índices para descrever os coeficientes do receptor espaço-temporal, ou seja, $w_{k,m,\hat{n}}$ em que k representa o índice do usuário, m o índice do elemento do arranjo e \hat{n} o índice do coeficiente do filtro temporal associado a cada antena do arranjo.

De forma a tentar escrever o sinal recebido no arranjo e o sinal na saída do receptor para o k -ésimo usuário de uma forma mais compacta, tem-se então o vetor de sinais recebidos dado por [Paulraj & Papadias, 1997; Cavalcanti, 1999; Papadias, 2000a]

$$\mathcal{X}(n) = \sum_{k=1}^K \mathcal{H}_k \mathcal{A}_k(n) + \mathcal{V}(n) \quad (4.23)$$

em que:

$$\begin{aligned} \mathcal{X}(n) &= [\mathbf{x}^T(n) \mid \mathbf{x}^T(n-1) \mid \cdots \mid \mathbf{x}^T(n-N+1)]_{MN \times 1} \\ \mathbf{x}(n) &= [x_1(n) \ x_2(n) \ \cdots \ x_M(n)]^T, \end{aligned} \quad (4.24)$$

$$\mathcal{A}_k(n) = \begin{bmatrix} a_k(n) \\ a_k(n-1) \\ \vdots \\ a_k(n-L_k-N+2) \end{bmatrix}_{(L_k+N-1) \times 1}, \quad (4.25)$$

$$\begin{aligned} \mathcal{V}(n) &= [\mathbf{v}^T(n) \mid \mathbf{v}^T(n-1) \mid \cdots \mid \mathbf{v}^T(n-N+1)]_{MN \times 1} \\ \mathbf{v}(n) &= [v_1(n) \ v_2(n) \ \cdots \ v_M(n)]^T \end{aligned} \quad (4.26)$$

e a matriz \mathcal{H}_k , chamada de matriz de convolução do canal espaço-temporal, é uma matriz Toeplitz por blocos, escrita da seguinte maneira [Haykin, 2000b; Zanatta, 2002]:

$$\mathcal{H}_k = \begin{bmatrix} \mathbf{H}_k(n) & \mathbf{0}_{M \times (N-1)} \\ \mathbf{0}_{M \times 1} & \mathbf{H}_k(n-1) & \mathbf{0}_{M \times (N-2)} \\ \mathbf{0}_{M \times 1} & \mathbf{0}_{M \times 1} & \mathbf{H}_k(n-2) & \mathbf{0}_{M \times (N-3)} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0}_{M \times (N-1)} & & & \mathbf{H}_k(n-N+1) \end{bmatrix}_{MN \times (N+L-1)}, \quad (4.27)$$

em que $\mathbf{0}_{k \times l}$ representa uma matriz de zeros com k linhas e l colunas.

Para tornar mais clara a notação e a estrutura da matriz de convolução do canal espaço-temporal, pode-se utilizar um exemplo. Seja um caso em que $M = 4$, $L = 3$ e $N = 3$,

desta forma tem-se

$$\mathcal{H}_k = \begin{bmatrix} \alpha_{k,1,0}(n) & \alpha_{k,1,1}(n) & \alpha_{k,1,2}(n) & 0 & 0 \\ \alpha_{k,2,0}(n) & \alpha_{k,2,1}(n) & \alpha_{k,2,2}(n) & 0 & 0 \\ \alpha_{k,3,0}(n) & \alpha_{k,3,1}(n) & \alpha_{k,3,2}(n) & 0 & 0 \\ \alpha_{k,4,0}(n) & \alpha_{k,4,1}(n) & \alpha_{k,4,2}(n) & 0 & 0 \\ 0 & \alpha_{k,1,0}(n-1) & \alpha_{k,1,1}(n-1) & \alpha_{k,1,2}(n-1) & 0 \\ 0 & \alpha_{k,2,0}(n-1) & \alpha_{k,2,1}(n-1) & \alpha_{k,2,2}(n-1) & 0 \\ 0 & \alpha_{k,3,0}(n-1) & \alpha_{k,3,1}(n-1) & \alpha_{k,3,2}(n-1) & 0 \\ 0 & \alpha_{k,4,0}(n-1) & \alpha_{k,4,1}(n-1) & \alpha_{k,4,2}(n-1) & 0 \\ 0 & 0 & \alpha_{k,1,0}(n-2) & \alpha_{k,1,1}(n-2) & \alpha_{k,1,2}(n-2) \\ 0 & 0 & \alpha_{k,2,0}(n-2) & \alpha_{k,2,1}(n-2) & \alpha_{k,2,2}(n-2) \\ 0 & 0 & \alpha_{k,3,0}(n-2) & \alpha_{k,3,1}(n-2) & \alpha_{k,3,2}(n-2) \\ 0 & 0 & \alpha_{k,4,0}(n-2) & \alpha_{k,4,1}(n-2) & \alpha_{k,4,2}(n-2) \end{bmatrix}_{12 \times 5}. \quad (4.28)$$

A saída do conjunto de filtros espaço-temporais, referente ao k -ésimo usuário é dada por:

$$y_k(n) = \mathcal{W}_k^H(n) \mathcal{X}(n), \quad (4.29)$$

em que

$$\mathcal{W}_k = \begin{bmatrix} \mathbf{w}_{k0} \\ \mathbf{w}_{k1} \\ \vdots \\ \mathbf{w}_{k(N-1)} \end{bmatrix}_{MN \times 1}, \quad \mathbf{w}_{k\hat{n}} = \begin{bmatrix} w_{k,1,\hat{n}} \\ w_{k,2,\hat{n}} \\ \vdots \\ w_{k,M,\hat{n}} \end{bmatrix}_{M \times 1}. \quad (4.30)$$

Obviamente, a equivalência entre o problema de separação de fontes descrito no Capítulo 2 e o problema de processamento espaço-temporal não é visualizada de forma direta pelo simples fato de que no modelo de BSS não foi inserida a dependência de fontes com atrasos nas misturas observadas no receptor (sensores). Entretanto, conforme pode ser visto em [Haykin, 2000a, Cap. 9] e outras referências lá citadas, o modelo de misturas com dependência temporal das fontes recai em um tipo de sistema que também pode ser representado por tensores, no qual uma das dimensões do tensor é associada ao instante de amostragem.

Assim, o problema de processamento multiusuário é o mesmo de separação de fontes quando as hipóteses consideradas para as fontes e o sistema de mistura são as mesmas para ambos os problemas. A questão chave é: como processar um sistema multiusuário com estratégias computacionalmente simples, uma vez que é desejado que tais estratégias operem em tempo real? Tal questão pode ser respondida quando certas condições acerca das fontes são tomadas.

A seção seguinte é dedicada a discutir quais são as características das fontes de um sistema de múltiplos usuários que permitem adotar estratégias simples de separação de fontes.

4.4 Caracterização das fontes de um sistema multiusuário

Uma vez que o interesse recai sobre sistemas de transmissão de dados digitais artificiais², as características das fontes, que são projetadas para que a transmissão seja realizada da maneira mais eficaz possível, tornam-se importante.

Estas características influenciam no tipo de processamento presente no receptor para assegurar que a transmissão seja confiável, ou seja, que a interferência seja removida, ou pelo menos mitigada, de tal forma que os dados sejam confiáveis.

A seguir, são listadas algumas das características das fontes de um sistema de processamento multiusuário que tornam possível a aplicação de estratégias simples (lineares) de separação de fontes.

4.4.1 Fontes discretas

O fato de considerar-se um sistema de comunicação digital implica em dizer que modulações digitais são empregadas. Isto significa que para cada esquema de modulação escolhido, tem-se uma *constelação* associada. Cada constelação apresenta características próprias e carrega informações estatísticas da fonte. A Figura 4.5 ilustra as constelações para algumas modulações típicas, a saber: BPSK (*Binary Phase Shift Keying*), QPSK e 8-PSK (*8-Phase Shift Keying*).

Conforme pode ser intuído na Figura 4.5, os diferentes tipos de constelação possuem diferentes tipos de características estatísticas. Embora a Figura 4.5 ilustre apenas constelações nas quais os sinais apresentam a propriedade de módulo constante, isto não é uma regra e vários tipos de modulação apresentam sinais com módulo não constante o que permite um melhor aproveitamento espectral. Pode-se citar, por exemplo, modulações 16-QAM, 256-QAM e 8-PAM, dentre outras [Lee & Messerschmitt, 1993; Proakis, 1995].

Uma consequência do uso de modulações digitais é decorrente do número de sinais diferentes em cada constelação. Este tópico é abordado a seguir.

²A explicitação de que os dados são artificiais é somente para destacar que as fontes envolvidas podem ser projetadas para atuar de acordo com certas características do sistema, diferenciando-se daqueles casos em que não se tem controle sobre as fontes, tais como, por exemplo, sistemas biomédicos e de aquisição de imagens.

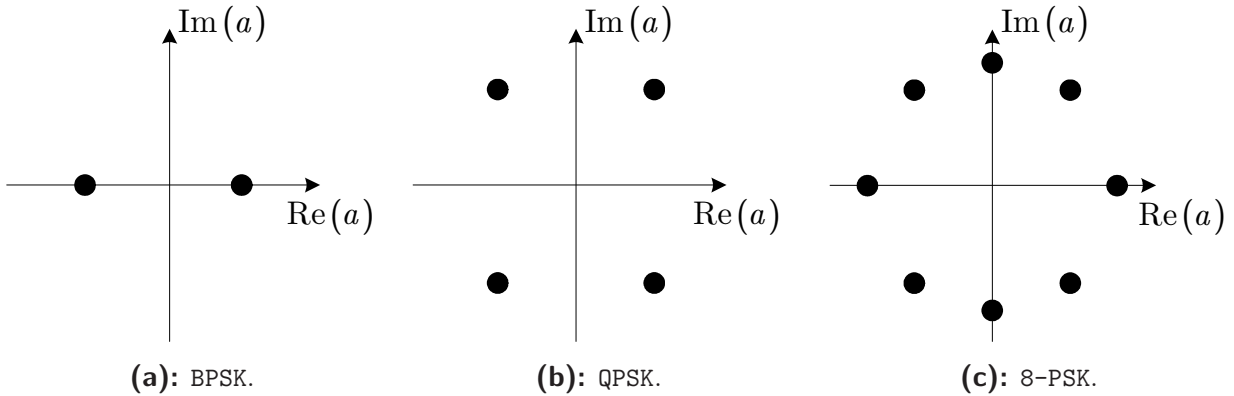


Figura 4.5: Alguns tipos de constelações de modulações digitais.

4.4.2 Alfabeto finito

Devido ao uso de um esquema de modulação digital, existe um número finito de possibilidades para os sinais provenientes das fontes. O conjunto que compreende todas as possibilidades, ou seja, todos os sinais diferentes provenientes da fonte, é chamado de *alfabeto*³ da modulação ou somente alfabeto da fonte.

Tratando-se de estratégias de recuperação de sinais, a existência de um alfabeto finito simplifica bastante a tarefa do receptor, uma vez que há um número limitado de seqüências possíveis de serem transmitidas e, logicamente, recuperadas.

Com isso, o alfabeto da modulação é representado por \mathfrak{A} , e a cardinalidade representada por $\mathfrak{C} = \text{card}(\mathfrak{A})$.

4.4.3 Mesma distribuição estatística

Uma das principais diferenças entre um sistema de processamento multiusuário e aqueles que modelam o problema genérico de separação de fontes diz respeito às distribuições de probabilidade das fontes. No primeiro caso, como já descrito anteriormente, usuários compartilham os mesmos recursos sujeitos às características sistêmicas.

Desta forma, as características das fontes obedecem às mesmas restrições para que o máximo de capacidade seja alcançado. Assim, uma configuração bastante aceita é a de que todos os usuários possuam fontes com as mesmas características, o que torna a distribuição

³O conjunto que compreende todos os sinais de um sistema de modulação é também denominado na literatura de *suporte*.

estatística uma só para todas as fontes envolvidas.

Esta hipótese simplifica bastante o problema e permite que as estratégias utilizadas no receptor sejam realizadas com estruturas mais simples, uma vez que somente uma distribuição deve ser identificada e o restante da tarefa é alocar qual seqüência ou símbolo foi transmitida por cada fonte. Outra hipótese comum consiste em supor que o sistema *conhece* a distribuição de probabilidade da fonte.

Também há o fato de que, em alguns sistemas modernos, nos quais há adaptação de enlace, ou seja, adequação da modulação e codificação da fonte para aproveitar melhor as características do canal, fontes com diferentes distribuições estatísticas coexistam. Entretanto, a inclusão de tal fato nos modelos não modifica consideravelmente o sistema de separação, que deve então ser capaz de encontrar um modelo abrangente para adequar todos os tipos de modulação. Além disso, o número de distribuições diferentes é geralmente pequeno, a exemplo do sistema EDGE de comunicação móvel, que utiliza modulações GMSK e 8-PSK [Freitas, 2002; Freitas et al., 2002].

Quando fontes discretas são consideradas, várias são as estratégias existentes na literatura propondo-se a tratar do problema de cancelamento de interferência, dentre elas pode-se citar [Gamboa & Gassiat, 1997; Papadias & Paulraj, 1997; Grellier & Comon, 1998; Macchi & Moreau, 1999]. Vale salientar que quando o problema de separação de fontes reduz-se ao de desconvolução (para $K = M = 1$), vários trabalhos na literatura utilizam-se de fontes serem discretas para construir critérios adequados. Dentre eles, pode-se citar [Li, 1992, 1995; Li & Mbarek, 1997; Yellin & Porat, 1993].

Neste âmbito, são discutidas nas seções seguintes algumas estratégias possíveis de tratar o problema de remoção de interferência em sinais de sistemas multiusuário.

4.5 Estratégias de separação de fontes em sistemas multiusuário

Uma vez que o problema de processamento multiusuário está definido, é interessante analisar algumas das principais técnicas de solução propostas na literatura. Embora o foco principal desta tese seja a abordagem de técnicas de separação cega de fontes, é necessário também estabelecer alguns limites de desempenho através de estratégias supervisionadas.

As seções seguintes discorrem sobre alguns dos algoritmos de separação cega de fontes que trabalham com o problema de remoção de interferência em sistemas de múltiplos usuários.

4.5.1 Solução ótima e algoritmos supervisionados

A solução ótima ou de Wiener para um sistema linear de recepção de sinais em um sistema de processamento multiusuário é obtida através da utilização do critério no sentido MMSE (*Minimum Mean Square Error*) [Haykin, 1996; Diniz, 1997].

A solução de Wiener é calculada para o k -ésimo usuário através de um filtro linear. Como já dito, esta solução é baseada na minimização do erro quadrático médio (EQM) definido, para o processamento espacial, como

$$e_k^2(n) = |a_k(n) - \mathbf{w}_k^H(n)\mathbf{x}(n)|^2, \quad (4.31)$$

em que $a_k(n)$ é o sinal transmitido pelo k -ésimo usuário no instante n .

Assim, o critério de Wiener é definido, para o usuário k , através da função custo dada por:

$$J_{\text{Wiener}}(\mathbf{w}_k) = \mathbb{E} \{e_k^2(n)\}. \quad (4.32)$$

Considerando-se as seguintes hipóteses:

- H_{Op}1.** as fontes possuem média nula;
- H_{Op}2.** $\mathbf{x}(n)$ é estacionário e
- H_{Op}3.** os sinais $\mathbf{x}(n)$ e $a_k(n)$ são conjuntamente estacionários,

a minimização da Equação (4.32) fornece a seguinte resposta para a solução ótima ou de Wiener:

$$\mathbf{w}_k^{\text{op}} = \mathbf{R}_x^{-1} \mathbf{p}_k, \quad (4.33)$$

em que \mathbf{R}_x é a matriz de autocorrelação dos dados de entrada no receptor, dada por

$$\begin{aligned} \mathbf{R}_x &= \mathbb{E} \{ \mathbf{x}(n)\mathbf{x}^H(n) \} \\ &= \sigma_a^2 \sum_{k=1}^K \mathbf{h}_k \mathbf{h}_k^H + \sigma_v^2 \mathbf{I}_M, \end{aligned} \quad (4.34)$$

em que $\sigma_a^2 = \mathbb{E} \{a_k^2(n)\}$ é a potência média dos sinais das fontes, σ_v^2 é a potência do ruído gaussiano em cada sensor e \mathbf{p}_k é o vetor de correlação cruzada, dado por

$$\begin{aligned} \mathbf{p}_k &= \mathbb{E} \{ \mathbf{x}(n)a_k^*(n) \} \\ &= \sigma_a^2 \mathbf{h}_k. \end{aligned} \quad (4.35)$$

O processamento descrito nas equações anteriores é válido para o caso no qual não ocorre ISI e o sinal desejado do usuário no instante n é aquele transmitido no mesmo instante, a

menos de um atraso de propagação. Ao considerar-se dispersão temporal dentro de um bloco de dados, ou seja a existência de ISI, tem-se que:

$$\mathcal{W}_k^{\text{op}} = \mathbf{R}_{\mathcal{X}}^{-1} \mathbf{p}_k, \quad (4.36)$$

em que

$$\begin{aligned} \mathbf{R}_{\mathcal{X}} &= \mathbb{E} \{ \mathcal{X}(n) \mathcal{X}^H(n) \} \\ &= \sigma_a^2 \sum_{k=1}^K \mathcal{H}_k \mathcal{H}_k^H + \sigma_v^2 \mathbf{I}_{MN}, \end{aligned} \quad (4.37)$$

e

$$\begin{aligned} \mathbf{p}_k &= \mathbb{E} \{ \mathcal{X}(n) a_k^*(n - \ell_k) \} \\ &= \sigma_a^2 \mathcal{H}_k^{(\ell_k)}, \end{aligned} \quad (4.38)$$

com ℓ_k sendo o atraso de decisão do k -ésimo usuário e $\mathcal{H}_k^{(\ell_k)}$ indicando a $(\ell_k + 1)$ -ésima coluna de \mathcal{H}_k [Cavalcanti, 1999].

Desta forma, algoritmos adaptativos podem ser utilizados para a atualização dos coeficientes do filtro de separação do k -ésimo usuário. O algoritmo LMS (*Least Mean Squares*) para o processamento espacial é dado por [Haykin, 1996; Diniz, 1997]:

$$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) - \mu [y_k(n) - a_k(n)]^* \mathbf{x}(n), \quad (4.39)$$

em que μ é um fator de passo.

Outro algoritmo adaptativo derivado do critério de Wiener é o algoritmo DMI (*Direct Matrix Inversion*) dado por [Cavalcanti & Romano, 1999; Cavalcanti, 1999]:

$$\widehat{\mathbf{R}}_x(n+1) = \varsigma \widehat{\mathbf{R}}_x(n) + (1 - \varsigma) \mathbf{x}(n) \mathbf{x}^H(n) \quad (4.40a)$$

$$\widehat{\mathbf{p}}_k(n+1) = \varsigma \widehat{\mathbf{p}}_k(n) + (1 - \varsigma) a_k^*(n) \mathbf{x}(n) \quad (4.40b)$$

$$\mathbf{w}_k(n) = \widehat{\mathbf{R}}_x^{-1} \widehat{\mathbf{p}}_k(n), \quad (4.40c)$$

em que ς é um *fator de esquecimento* responsável por eliminar os erros inseridos pelos componentes instantâneos.

Para o caso de processamento espaço-temporal, a equação de adaptação do LMS é dada por [Cavalcanti, 1999]:

$$\mathcal{W}_k(n+1) = \mathcal{W}_k(n) - \mu [y_k(n) - a_k(n - \ell_k)]^* \mathcal{X}(n). \quad (4.41)$$

e o algoritmo DMI resultante é dado por

$$\widehat{\mathbf{R}}_{\mathcal{X}}(n+1) = \varsigma \widehat{\mathbf{R}}_{\mathcal{X}}(n) + (1 - \varsigma) \mathcal{X}(n) \mathcal{X}^H(n) \quad (4.42a)$$

$$\widehat{\mathbf{p}}_k(n+1) = \varsigma \widehat{\mathbf{p}}_k(n) + (1 - \varsigma) a_k^*(n - \ell_k) \mathcal{X}(n) \quad (4.42b)$$

$$\mathcal{W}_k(n) = \widehat{\mathbf{R}}_{\mathcal{X}}^{-1} \widehat{\mathbf{p}}_k(n). \quad (4.42c)$$

Entretanto, os algoritmos descritos até então necessitam do conhecimento do sinal transmitido, pelo menos por algum tempo durante a adaptação. As seções seguintes mostram algoritmos cegos aplicados ao contexto de processamento multiusuário.

4.5.2 *Multiuser Constant Modulus Algorithm*

O algoritmo *Multiuser Constant Modulus Algorithm* (MU-CMA) foi proposto como uma generalização do CMA para o caso de múltiplos usuários. A proposta de [Papadias & Paulraj, 1997] foi a de utilizar o critério CM como função de custo para a remoção da MAI e/ou ISI e utilizar um critério auxiliar para garantir a separação de todas as fontes.

A necessidade de um critério auxiliar ocorre devido ao fato de que o CMA irá recuperar o sinal do usuário que tiver maior potência de modo que este seja privilegiado durante o processo de adaptação. Desta forma, pode-se ter uma situação que somente um usuário seja recuperado e replicado em todos os filtros de separação, uma vez que em sistemas de comunicação digital, e principalmente em sistemas de processamento espacial (sistema SDMA), a situação de um usuário apresentar potência maior que os outros, o chamado efeito *near-far*, é bastante comum.

Para solucionar este problema, a proposição de um critério adicional que elimina a possibilidade de replicação foi realizada em [Papadias & Paulraj, 1997]. O critério visa decorrelacionar as saídas dos diversos filtros de separação conjuntamente com a otimização do critério CM.

Assim, a função custo do MU-CMA para o k -ésimo usuário, é dada por [Papadias & Paulraj, 1997]:

$$J_{\text{MU-CMA}}(\mathbf{w}_k) = J_{\text{CMA}}(\mathbf{w}_k) + \gamma \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K |r_{ij}|^2, \quad (4.43)$$

em que $r_{ij} = \mathbb{E} \{y_i(n)y_j^*(n)\}$ é a correlação cruzada entre as saídas do i -ésimo e do j -ésimo filtro de separação, γ é o fator de regularização do termo de decorrelação e $J_{\text{CMA}}(\mathbf{w}_k) = \mathbb{E} \left\{ (|y(n)|^2 - \rho_2)^2 \right\}$, em que $\rho_2 = \frac{\mathbb{E} \{a_k^4(n)\}}{\mathbb{E} \{a_k^2(n)\}}^4$, é a função custo do CMA [Godard, 1980; Treichler & Agee, 1983].

Ao tomar-se o gradiente estocástico da Equação (4.43) tem-se [Papadias & Paulraj, 1997;

⁴A constante ρ_2 é conhecida como *dispersão* ou raio médio da constelação. Quando a constelação possui módulo constante, ρ_2 fornece a energia média do sinal de entrada.

Cavalcanti & Romano, 1999]

$$\nabla J_{\text{MU-CMA}}(\mathbf{w}_k) = \mathbb{E} \{ y^*(n) \cdot [|y_k(n)|^2 - 1] \mathbf{x}(n) \} + \gamma \sum_{\substack{i=1 \\ i \neq k}}^K r_{ik} \mathbb{E} \{ y_i^*(n) \mathbf{x}(n) \}, \quad (4.44)$$

na qual assume-se que a constelação das fontes têm potência unitária.

Deve-se então notar que os valores de $r_{ij} = \mathbb{E} \{ y_i(n) y_j^*(n) \}$ e $\mathbb{E} \{ y_i^*(n) \mathbf{x}(n) \}$ necessitam ser estimados através de médias temporais. Logo, tais estimativas podem ser realizadas através das seguintes equações:

$$\widehat{\mathbf{R}}_y(n+1) = \varsigma \widehat{\mathbf{R}}_y(n) + (1 - \varsigma) \mathbf{y}(n) \mathbf{y}^H(n) \quad (4.45a)$$

$$\widehat{\mathbf{P}}(n+1) = \varsigma \widehat{\mathbf{P}}(n) + (1 - \varsigma) \mathbf{x}(n) \mathbf{y}^H(n), \quad (4.45b)$$

em que $\mathbf{y}(n) = [y_1(n) \ \cdots \ y_K(n)]^T$.

Com isso, em uma notação mais resumida, a adaptação do algoritmo será dada através da seguinte equação:

$$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) + \mu \cdot (1 - |y(n)|^2) \cdot y_k^* \cdot \mathbf{x}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \widehat{r}_{ik}(n) \widehat{\mathbf{p}}_i(n), \quad (4.46)$$

em que $\widehat{r}_{ik}(n)$ é o (i, k) -ésimo elemento da matriz $\widehat{\mathbf{R}}_y(n)$ e $\widehat{\mathbf{p}}_i(n)$ é a i -ésima coluna de matriz $\widehat{\mathbf{P}}(n)$, dadas na Equação (4.45).

As equações anteriores referem-se ao processamento espacial, uma vez que, neste caso, foi suposto não haver a presença de interferência inter-simbólica no vetor de dados do receptor nem das saídas do sistema de separação. A inclusão de ISI modifica as equações, para que o termo de decorrelação compreenda os diferentes instantes temporais e cancele a replicação da seqüência de um mesmo usuário em um diferente instante temporal. Assim, as equações tornam-se [Cavalcanti, 1999]:

$$J_{\text{MU-CMA}}(\mathcal{W}_k) = J_{\text{CMA}}(\mathcal{W}_k) + \gamma \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} |r_{ij}(\ell)|^2, \quad (4.47)$$

em que $r_{ij}(\ell) = \mathbb{E} \{ y_i(n) y_j^*(n - \ell) \}$ é a correlação cruzada entre os sinais das saídas dos filtros espaço-temporais i e j , com diferença de tempo ℓ , e $\frac{\Delta}{2}$ é o atraso máximo estimado para o qual os sinais dos diversos usuários devem estar decorrelacionados.

Desta forma, e apelando para a similaridade entre o processamento espacial e o espaço-temporal, pode-se escrever:

$$\mathcal{W}_k(n+1) = \mathcal{W}_k(n) + \mu \cdot (1 - |y_k(n)|^2) \cdot y(n)_k^* \cdot \mathcal{X}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \hat{r}_{ik,\ell}(n) \hat{\mathbf{P}}_{i,\ell}(n), \quad (4.48)$$

e

$$\mathbf{R}_{y,\ell}(n+1) = \varsigma \mathbf{R}_{y,\ell}(n) + (1 - \varsigma) \mathbf{y}(n) \mathbf{y}^H(n - \ell) \quad (4.49a)$$

$$\mathbf{P}_\ell(n+1) = \varsigma \mathbf{P}_\ell(n) + (1 - \varsigma) \mathcal{X}(n) \mathbf{y}^H(n - \ell) \quad (4.49b)$$

$$\mathbf{y}(n - \ell) = [y_1(n - \ell) \quad \cdots \quad y_K(n - \ell)]^T \quad (4.49c)$$

$$\ell = -\frac{\Delta}{2}, \dots, \frac{\Delta}{2}, \quad (4.49d)$$

em que $\hat{r}_{ik,\ell}(n)$ é o termo de correlação cruzada entre os conformadores de pulso dos i -ésimo e j -ésimo usuários com atraso ℓ e corresponde ao (i, j) -ésimo elemento da matriz $\mathbf{R}_{y,\ell}(n)$, e $\hat{\mathbf{p}}_{i,\ell}(n)$ é a i -ésima coluna da matriz $\mathbf{P}_\ell(n)$.

Apesar do MU-CMA funcionar bem para vários tipos de processamento multiusuário, há algumas limitações devido à própria estrutura do CMA, conforme discutido em [Haykin, 1996; Papadias, 1995]. São elas:

- baixa velocidade de convergência;
- o fator de regularização do termo de decorrelação tem de ser escolhido de forma a encontrar um compromisso entre o erro em estado permanente e o número de capturas errôneas, ou seja, de que usuários não identificados.

Algoritmos que buscam melhorar estes aspectos são discutidos nas seções a seguir.

4.5.3 *Fast Multiuser Constant Modulus Algorithm*

O algoritmo proposto em [Cavalcanti & Romano, 1999; Cavalcanti, 1999] promove uma melhoria de desempenho ao custo de uma maior complexidade computacional. Seu desenvolvimento consiste em uma versão recursiva do MU-CMA e é chamado de *Fast Multiuser Constant Modulus Algorithm* (FMU-CMA) ou ainda de *Least-Squares with Adaptive Decorrelation - CMA* (LSAD-CMA).

O FMU-CMA também utiliza o critério de decorrelação explícita⁵ e utiliza recursão para derivar o algoritmo a partir do critério dado na Equação (4.43).

⁵O termo *decorrelação explícita* para denominar a iniciativa de forçar a decorrelação das diversas saídas para os K usuários foi cunhado por [Cavalcanti, 1999].

Desta forma, o algoritmo pode ser descrito pelo seguinte conjunto de equações:

$$\mathbf{w}_k(n) = \mathbf{R}_{xy,k}^{-1}(n)\mathbf{d}_{xy,k}(n) \quad (4.50a)$$

$$\mathbf{R}_{xy,k}(n+1) = \zeta\mathbf{R}_{xy,k}(n) + (1-\zeta)|y_k(n)|^2\mathbf{x}(n)\mathbf{x}^H(n) \quad (4.50b)$$

$$\mathbf{d}_{xy,k}(k+1) = \zeta\mathbf{d}_{xy,k}(n) + (1-\zeta) \cdot \rho_2 \cdot y_k^*(n)\mathbf{x}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \hat{r}_{ik}(n)\hat{\mathbf{p}}_i(n), \quad (4.50c)$$

em que ζ é um termo de suavização e $\hat{r}_{ik}(n)$ e $\hat{\mathbf{p}}_i(n)$ são obtidos da Equação (4.45).

O procedimento recursivo na Equação (4.50) melhora a velocidade de convergência aumentando-se a complexidade de implementação do algoritmo. A questão do erro em estado permanente necessita de uma modificação na função custo de forma a permitir que a solução encontrada minimize ambos, o erro em estado permanente e o número de capturas errôneas.

Assim, a inclusão de um fator de regularização adaptativo do termo de decorrelação promove um auto-ajuste da interferência de múltiplo acesso. Este procedimento melhora o erro em estado permanente pois, de forma bastante intuitiva, pode-se perceber que o fator de decorrelação necessita ser maior nos instantes iniciais. Quando os usuários já atingiram a separação, ou seja, quando as saídas dos filtros encontram-se decorrelacionadas, o fator de regularização pode ser diminuído de maneira a melhorar o desempenho em estado permanente.

Com isso, foi proposto em [Cavalcanti & Romano, 1999; Cavalcanti, 1999] um critério no qual o valor do γ varia com o tempo e é dependente do nível de correlação cruzada entre os usuários, ou melhor, entre os sinais estimados dos usuários.

Inicialmente, mede-se o nível médio de correlação cruzada por usuário através da seguinte equação:

$$\bar{r}_k(n) = \frac{1}{K-1} \sum_{\substack{i=1 \\ i \neq k}}^K |\hat{r}_{ik}(n)|^2. \quad (4.51)$$

A média calculada sobre o número de usuários tem como meta tornar a medida independente do número de usuários ativos no sistema [Cavalcanti, 1999]. O fator de regularização adaptativo é obtido, finalmente, pela utilização de uma transformação que permite a saturação dos valores de $\bar{r}_k(n)$. Esta saturação é obtida através de [Cavalcanti & Romano, 1999]:

$$\gamma_k(n) = \tanh[\bar{r}_k(n)], \quad (4.52)$$

em que $\tanh(\cdot)$ representa a função tangente hiperbólica. Outra medida tomada foi a de considerar uma normalização na matriz $\mathbf{R}_y(n)$ da Equação (4.45a) e obter $\bar{r}_k(n)$ a partir dos valores normalizados de $\mathbf{R}_y(n)$, dispensando a utilização da transformação através da tangente

hiperbólica. Com isso tem-se [Cavalcanti, 1999][Cavalcanti et al., 1999, em SPAWC'99]:

$$\gamma_k(n) = \bar{r}_k(n). \quad (4.53)$$

Desta forma, a adaptação do FMU-CMA é dado pela escolha do termo de regularização adaptativo para cada usuário nas Equações (4.52) ou (4.53), substituindo o fator de regularização fixo da Equação (4.50). A Tabela 4.1 resume o algoritmo FMU-CMA.

<p>(1). Inicialize $\mathbf{W}(0) = \mathbf{I}$, $\mathbf{P}(0) = \mathbf{0}$, $\mathbf{d}_{xy,k} = \mathbf{0}$, $\mathbf{R}_y(0) = \mathbf{I}$ e $\mathbf{R}_{xy,k}(0) = \mathbf{I}$</p> <p>(2). Para $n > 0$</p> <p>(3). Calcule matrizes de autocorrelação e correlação cruzada das saídas através de</p> $\begin{aligned} \mathbf{R}_y(n+1) &= \zeta \mathbf{R}_y(n) + (1-\zeta) \mathbf{y}(n) \mathbf{y}^H(n) \\ \mathbf{P}(n+1) &= \zeta \mathbf{P}(n) + (1-\zeta) \mathbf{y}(n) \mathbf{x}(n) \end{aligned}$ <p>(4). Para $k = 1 : K$, calcule</p> <ul style="list-style-type: none"> - Saída do k-ésimo filtro $y_k(n) = \mathbf{w}_k^H(n) \mathbf{x}(n)$ <ul style="list-style-type: none"> - Fator de regularização para k-ésimo usuário $\begin{aligned} \bar{r}_k(n) &= \frac{1}{K-1} \sum_{\substack{i=1 \\ i \neq k}}^K \hat{r}_{ik}(n) ^2 \\ \gamma_k(n) &= \tanh[\bar{r}_k(n)] \end{aligned}$ <ul style="list-style-type: none"> - Atualização filtro separação $\begin{aligned} \mathbf{w}_k(n) &= \mathbf{R}_{xy,k}^{-1}(n) \mathbf{d}_{xy,k}(n) \\ \mathbf{R}_{xy,k}(n+1) &= \zeta \mathbf{R}_{xy,k}(n) + (1-\zeta) y_k(n) ^2 \mathbf{x}(n) \mathbf{x}^H(n) \\ \mathbf{d}_{xy,k}(k+1) &= \zeta \mathbf{d}_{xy,k}(n) + (1-\zeta) \cdot \rho_2 \cdot y_k^*(n) \mathbf{x}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \hat{r}_{ik}(n) \hat{\mathbf{p}}_i(n), \end{aligned}$ <p>(5). Retorne para Passo 4</p> <p>(6). Retorne para Passo 2</p>
--

Tabela 4.1: Algoritmo FMU-CMA.

Embora o algoritmo forneça bons resultados comparativos com o MU-CMA, não é possível provar a convergência para um número qualquer de usuários e sua complexidade é mais elevada que a de algoritmos do tipo LMS. A seção seguinte discute um algoritmo cuja convergência é provada para um número qualquer de usuários.

4.5.4 *Multiuser Kurtosis Maximization*

O critério de maximização da *kurtosis* para múltiplos usuários é baseado no critério de Shalvi-Weinstein (SW), derivado diretamente do teorema de equalização cega proposto em [Shalvi & Weinstein, 1990]. O teorema mostra que, para que a desconvolução cega seja possível, deve-se igualar *um* momento de ordem superior mantendo o de segunda ordem seja constante.

O critério *Multiuser Kurtosis Maximization* (MUK) foi proposto em [Papadias, 2000a, 2000b] e utiliza um conjunto de condições necessárias para prover a recuperação dos dados de vários usuários. São elas:

- C_{RS1}**. $a_l(n)$ é i.i.d. e de média zero ($l = 1, \dots, K$);
- C_{RS2}**. $a_l(n)$ e $a_q(n)$ são estatisticamente independentes para $l \neq q$, com a mesma fdp;
- C_{RS3}**. $|\mathcal{K}[y_l(n)]| = |\mathcal{K}_a|$ ($l = 1, \dots, K$);
- C_{RS4}**. $\mathbb{E}\{|y_l(n)|^2\} = \sigma_a^2$ ($l = 1, \dots, K$);
- C_{RS5}**. $\mathbb{E}\{y_l(n)y_q^*(n)\} = 0$, $l \neq q$,

em que \mathcal{K}_a e σ_a^2 são, respectivamente, a *kurtosis* e a variância da seqüência transmitida e $\mathcal{K}[\cdot]$ é o operador *kurtosis*, definido na Equação (2.34). Deve-se notar que a Condição C_{RS5} tem como meta garantir o mesmo comportamento que o procedimento de decorrelação explícita da Equação (4.43). A prova de que as Condições C_{RS1}-C_{RS5} são suficientes e necessárias para garantir a recuperação (separação e identificação) dos sinais é apresentada no Apêndice B.

Um aspecto fundamental do algoritmo derivado do critério MUK diz respeito a prover a identificação correta dos diferentes usuários envolvidos no sistema. Enquanto a maioria dos algoritmos que têm por base critérios de equalização cega, tais como o MU-CMA, utilizam um critério de decorrelação explícita [Papadias & Paulraj, 1997], o algoritmo MUK utiliza, como critério para garantir que todas as fontes serão corretamente identificadas, uma ortogonalização da matriz de separação global, de tal forma que $\mathbf{G}^H \mathbf{G} = \mathbf{I}$. Desta maneira, o critério resultante da combinação das Condições C_{RS1}-C_{RS5} pode ser escrito como [Papadias, 2000a, 2000b]:

$$\begin{cases} \max_{\mathbf{G}} J_{\text{MUK}}(\mathbf{G}) = \sum_{k=1}^K |\mathcal{K}[y_k]| \\ \text{sujeito a : } \mathbf{G}^H \mathbf{G} = \mathbf{I} \end{cases} \quad (4.54)$$

Na verdade, também baseado na maioria das abordagens de BSS inspiradas em critérios de equalização cega, o critério MUK pode ser dividido em duas etapas:

1. etapa de equalização – que maximiza a *kurtosis*, respeitando com isso o teorema de SW. A esta etapa estará associada uma matriz \mathbf{W}^e ;
2. etapa de separação – responsável por prover a decorrelação dos conformadores de pulso para os diversos usuários. A esta etapa corresponde uma matriz \mathbf{W} ⁶.

Para forçar que a matriz de separação global seja ortogonal, é utilizado uma *ortogonalização de Gram-Schmidt* [Golub & Loan, 1996] sobre a matriz \mathbf{W}^e [Papadias, 2000b]. Este procedimento, realizado de forma interativa, força que as saídas para os diferentes usuários, estejam decorrelacionadas. A definição e derivação da ortogonalização iterativa de Gram-Schmidt é realizada no Apêndice B.

Assim, tem-se que o gradiente estocástico do critério MUK é dado por [Papadias, 2000a, 2000b]:

$$\nabla J_{\text{MUK}}(\mathbf{G}) = 4 \text{sign}(\mathcal{K}_a) \cdot \sum_{k=1}^K \mathbb{E} \{ |y_k(n)|^2 \cdot y_k(n) \cdot \mathbf{y}^*(n) \}, \quad (4.55)$$

em que $\text{sign}(\cdot)$ é o operador que extrai o sinal algébrico de uma variável.

Então, na primeira etapa (equalização) é realizada uma adaptação de $\mathbf{W}(n)$ na direção do gradiente instantâneo, obtendo-se:

$$\mathbf{W}^e(n+1) = \mathbf{W}(n) + \mu \text{sign}(\mathcal{K}_a) \mathbf{x}^*(n) \mathbf{y}(n), \quad (4.56)$$

em que $\mathbf{y}(n) = [|y_1(n)|^2 y_1(n) \quad \cdots \quad |y_K(n)|^2 y_K(n)]$.

Uma vez executada a etapa de equalização, a restrição sobre a ortogonalidade de \mathbf{G} deve ser respeitada. Entretanto, inicialmente uma outra hipótese deve também ser respeitada, a de que os dados no receptor sejam *branqueados* (especialmente ou espaço-temporalmente), ou seja, a matriz \mathbf{H} deve ser unitária.

Esta hipótese é a mesma que necessita ser respeitada para que uma série de critérios de BSS possam ser aplicados: a de que os dados sejam esferatizados. Isto pode ser obtido, como já anteriormente comentado no Capítulo 2, através de PCA [Cichocki & Amari, 2002] ou de outros métodos de segunda ordem que garantem a convergência da matriz de mistura instantânea para uma mistura unitária das entradas [Deneire, 1998]. Uma decorrência do método de branqueamento é que a matriz de separação torna-se de ordem $M \times M$, uma vez que um número menor de componentes são selecionados, o que reduz a complexidade computacional.

⁶Devido vários algoritmos utilizarem este procedimento em duas etapas, esta notação será a mesma para quaisquer outros que apresentem o mesmo comportamento.

Esta hipótese também tem por meta garantir que a variância, ou o momento de segunda ordem, seja mantida constante durante o processo para que as condições de recuperação dos sinais sejam respeitadas.

Logo, a etapa de ortogonalização para o k -ésimo usuário é dada por:

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1)}{\left\| \mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1) \right\|}, \quad (4.57)$$

em que $\|\cdot\|$ representa a norma de vetor.

Com isso, o algoritmo MUK pode ser resumido conforme o pseudo-código apresentado na Tabela 4.2.

(1). Inicialize $\mathbf{W}(0)$ e $\mathbf{W}^e(0)$

(2). Para $n > 0$

(3). Calcule $\mathbf{W}^e(n+1)$ através de

$$\begin{aligned} \mathbf{W}^e(n+1) &= \mathbf{W}(n) + \mu \text{sign}(\mathcal{K}_a) \mathbf{x}^*(n) \mathbf{y}(n) \\ \mathbf{y}(n) &= [|y_1(n)|^2 y_1(n) \quad \cdots \quad |y_K(n)|^2 y_K(n)] \end{aligned}$$

(4). Calcule

$$\mathbf{w}_1(n+1) = \frac{\mathbf{w}_1^e(n+1)}{\|\mathbf{w}_1^e(n+1)\|}$$

(5). Para $k = 2 : K$

(6). Calcule $\mathbf{w}_j(n+1)$ através de

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1)}{\left\| \mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1) \right\|}$$

(7). Retorne para Passo 5

(8). Retorne para Passo 2

Tabela 4.2: Algoritmo MUK.

É importante notar que o algoritmo MUK pode ser visto como um algoritmo de BSS que utiliza uma função contraste baseada no momento de quarta ordem (*kurtosis* não-normalizada). Além disto, o processamento executado de uma maneira desacoplada, diferenciando as etapas de equalização e separação, permite a generalização e o desenvolvimento de vários critérios de BSS baseados na mesma abordagem, conforme será visto no capítulo seguinte.

Um aspecto importante que merece uma atenção especial, refere-se ao pré-branqueamento do sinal no receptor. Uma vez que o processamento cego não provê conhecimento sobre o sistema de mistura, ou do canal, é bastante intuitivo que o pré-branqueamento é sempre necessário para garantir o funcionamento adequado do algoritmo. Além disso, quando se utilizam arranjos de antenas no receptor, o canal equivalente gerado é geralmente não unitário, o que reforça a questão da necessidade de pré-branqueamento.

Um algoritmo utilizado para este fim é o *algoritmo de Schur* [Deneire, 1998; Deneire & Slock, 1999b, 1999a] baseado na decomposição de Schur [Haykin, 1989], aplicado à matriz de covariância do sinal sem ruído $\widehat{\mathbf{R}}_{\tilde{x}}$, a qual, devido à sua simetria, fornece a seguinte decomposição:

$$\widehat{\mathbf{R}}_{\tilde{x}} = \mathbf{L}\mathbf{D}\mathbf{L}^H, \quad (4.58)$$

em que \tilde{x} representa os sinais recebidos sem ruído, \mathbf{L} é uma matriz unitária e \mathbf{D} é uma matriz diagonal com entradas reais. Uma estimativa não-ruidosa de $\widehat{\mathbf{R}}_x = E\{\mathbf{x}\mathbf{x}^H\}$ pode ser realizada através da estimativa da variância do ruído, σ_v^2 , pela média dos menores autovalores de $\widehat{\mathbf{R}}_x$ e então obtendo $\widehat{\mathbf{R}}_{\tilde{x}} = \widehat{\mathbf{R}}_x - \sigma_v^2\mathbf{I}_M$.

Idealmente, deve-se ter $\widehat{\mathbf{R}}_{\tilde{x}} = \mathbf{H}\mathbf{H}^H$; então, como no caso descrito na Seção 2.2.1, tem-se que $\mathbf{L}\mathbf{D}^{\frac{1}{2}}$ é igual a \mathbf{H} , a menos de uma matriz de permutação \mathbf{P} , ou seja,

$$\tilde{\mathbf{L}} = \mathbf{H}\mathbf{P}, \quad (4.59)$$

em que $\tilde{\mathbf{L}}$ é construída como a matriz formada pelos K autovetores associados aos maiores autovalores da matriz $\mathbf{L}\mathbf{D}^{\frac{1}{2}}$ [Haykin, 1989; Deneire, 1998; Papadias, 2004].

Finalmente, o pré-branqueamento é dado por $\tilde{\mathbf{L}}^{\#}$, em que $\#$ denota pseudo-inversão da matriz, dada por $\tilde{\mathbf{L}}^{\#} = \left(\tilde{\mathbf{L}}^H\tilde{\mathbf{L}}\right)^{-1}\tilde{\mathbf{L}}^H$ [Haykin, 1996].

Deve-se notar que este procedimento é análogo ao fornecido por PCA, no qual os autovetores fornecidos pelos principais autovalores determinam as direções de projeção dos dados.

Para o caso de processamento espaço-temporal, além da definição da matriz de convolução de canal espaço-temporal, que é uma matriz de Toeplitz por blocos ou uma matriz de Sylvester generalizada [Haykin, 1996], deve-se escrever a equação do sinal recebido conforme a

Equação (4.23). Com isso, a modificação que necessita ser feita é que um pré-branqueamento espaço-temporal é imperativo para permitir ao MUK atuar de forma correta na recuperação do sinal. Uma vez que seja executado o algoritmo de pré-branqueamento, o modelo do sinal resultante é similar àquele do processamento espacial. A diferença se dá nas dimensões dos vetores envolvidos.

O desenvolvimento da transformação de branqueamento é feito como na Equação (4.58), respeitando-se a ordem dos vetores envolvidos. Conforme mostrado em [Papadias, 2000c] e também recentemente em [Papadias, 2004], pode-se utilizar um método de predição linear temporal para substituir a etapa de equalização temporal no processamento. Com isso, o algoritmo MUK pode ser utilizado normalmente na tarefa de BSS.

Assim, considerando-se um preditor de um passo para $\mathbf{y}(n)$, a partir das amostras atrasadas $\mathbf{y}(n-1), \dots, \mathbf{y}(n-L+1)$, o preditor linear ótimo é dado pela matriz [Papadias, 2000c, 2004]

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathbf{I}_M & -\mathbf{\Upsilon}_1 & \cdots & -\mathbf{\Upsilon}_{L-1} \end{bmatrix}_{M \times ML}, \quad (4.60)$$

que opera em $\mathbf{y}(n)$ e minimiza a variância do seguinte erro de predição

$$\min_{\mathbf{\Upsilon}} \sigma_{\mathbf{e}, L-1}^2 = \mathbb{E} \{ \mathbf{e}_{L-1} \mathbf{e}_{L-1}^H \}, \quad (4.61)$$

em que

$$\mathbf{e}_{L-1}(n) = \mathbf{y}(n) - \mathbf{\Upsilon} \mathcal{X}(n). \quad (4.62)$$

A solução da Equação (4.61) fornece um conjunto de equações normais que fornecem o preditor ótimo como [Gorokhov & Loubaton, 1997, 1999; Papadias, 2004]

$$\mathbf{\Upsilon} = \mathbf{v} \mathbf{R}_{x, L-1}^{-1}, \quad (4.63)$$

em que a matriz de covariância dos dados é particionada da seguinte forma [Papadias, 2000c, 2004]:

$$\mathbf{R}_x = \left[\begin{array}{c|c} \mathbf{v}_0 & \mathbf{v} \\ \hline \mathbf{v}^H & \mathbf{R}_{x, L-1} \end{array} \right]. \quad (4.64)$$

Então, pode-se construir o sinal pré-branqueado de dimensão $M \times 1$ por

$$\bar{\mathbf{x}}(n) = \mathbf{\Upsilon} \mathcal{X}(n). \quad (4.65)$$

A partir disso, pode-se novamente executar o algoritmo de Schur da Equação (4.58) e utilizar o algoritmo MUK para encontrar a separação dos usuários.

Levando em conta todas as etapas do algoritmo, a saber:

- pré-branqueamento,
- equalização,
- ortogonalização;

é difícil afirmar que a complexidade do algoritmo é menor que a de outros, como por exemplo, o FMU-CMA. No caso de levar-se em conta ainda aspectos de robustez, a análise pode ser bastante tendenciosa se os parâmetros corretos não forem utilizados. Desta forma a complexidade, quando referida nesta tese, é relacionada à existência ou não de inversão de matrizes no algoritmo. Outros aspectos que envolvem etapas prévias (pré-branqueamento, por exemplo) serão tratadas à parte da complexidade computacional, mas como requisitos teóricos envolvidos.

Dizer que os algoritmos de processamento multiusuário aqui citados são necessariamente os mais importantes seria tendencioso. Entretanto, eles representam a classe de algoritmos e critérios que servirão de base para a proposta da tese. A seção seguinte é dedicada a discutir os algoritmos descritos até o momento e fazer uma análise das características de outros existentes na literatura.

4.5.5 Discussão e comentários

Algoritmos de processamento multiusuário baseados em critérios de equalização cega têm sido bastante investigados na última década. O principal interesse vem do fato de várias técnicas e algoritmos de equalização cega estarem bastante sedimentados de tal forma que sua aplicação em vários cenários torna-se factível e atraente.

Devido a sua simplicidade de implementação e utilização em um vasto número de situações práticas [Treichler et al., 1998], o CMA tem sido preferido para composição de critérios multiusuário. Desde o trabalho [Gooch & Lundell, 1986], que foi o pioneiro da utilização do CMA para remoção de interferências em sistemas de múltiplos usuários, vários outros enveredaram pelo mesmo caminho ao tentarem, através do CMA, encontrar uma solução para o problema de tratar simultaneamente vários usuários em sistemas cooperativos.

Dentre várias referências que podem ser citadas, [Petrus, 1997] apresenta um breve tutorial sobre os trabalhos que utilizaram critérios de equalização cega para processamento espaço-temporal em sistemas celulares, particularmente utilizando antenas adaptativas como receptor. São também propostas versões adaptativas (gradiente descendente e mínimos quadrados) do algoritmo de múltiplos usuários, sob uma notação de *multi-objetivo* para acesso em sistemas de processamento espaço-temporal e CDMA. Complementando este trabalho, [Liberti & Rappaport, 1999] descrevem vários algoritmos adaptativos e autodidatas que utilizam arranjos de antenas no receptor para remoção/mitigação de interferência em sistemas celulares sem fio.

Ainda nesta linha de utilização do CMA, outros trabalhos buscaram aplicar o critério CM no contexto de remoção de ISI e MAI em sistemas sem fio. Podem ser listados por exemplo [Castedo et al., 1997; Tugnait, 1997a; Míguez & Castedo, 1998; Touzni et al., 2001; Lambbotharan et al., 1999]. A principal diferença entre eles é basicamente a maneira de abordar o critério que permite uma maior ou menor capacidade de generalização da análise feita para o caso mono-usuário, e garantir assim convergência para um número maior de fontes no sistema. Além disto, vários deles aplicam o algoritmo decorrente para cancelamento de interferência em sistemas CDMA. Uma análise de convergência de um algoritmo para processamento de sistemas multiusuário baseado no CMA é também apresentada em [Lambbotharan & Chambers, 1999], no qual a análise é realizada através das características da função custo do critério resultante.

Outro artigo que aborda com muita precisão a convergência e condição de recuperação dos sinais do MU-CMA⁷, bem como a proposição de um novo algoritmo de BSS para sistemas MIMO (*Multiple-Input Multiple-Output*), é [Li & Liu, 1998]. Além destes, [Papadias & Huang, 2001] discute algoritmos de processamento espaço-temporal, treinados e supervisionados, e seu desempenho em sistemas CDMA de seqüência direta (DS-CDMA, *Direct Sequence Code Division Multiplex Access*), bem como aspectos de convergência e desempenho em termos de taxa de erro de bit (BER, *Bit Error Rate*).

Outros trabalhos enveredaram na direção de derivar versões recursivas de algoritmos para os critérios multiusuário baseados no CMA. Dentre eles, tem-se [Leary, 1997] como precursor de um estudo de algoritmos recursivos do CMA e [Cavalcanti et al., 1999, em ICASSP'99].

Outro tipo de abordagem, que não é a desta tese, é a de *identificação serial* dos usuários. Neste tipo de processamento, um usuário é identificado por vez e retirado do sinal recebido, o sinal é processado novamente agora com $K - 1$ usuários e o processo se repete até todos serem identificados. Este tipo de abordagem é geralmente denominada *multi-estágio, cancelamento sucessivo* ou ainda *identificação serial*. Este processamento difere do *deflation* devido ao fato de que, no cancelamento serial, o sinal recebido vai tendo cada vez menos fontes devido à retirada de sua contribuição no sinal recebido, enquanto que no *deflation* o sinal recebido possui sempre a contribuição de todas as fontes enquanto que o dispositivo de separação vai identificando quais são as fontes envolvidas, assumindo-se a ortogonalidade entre elas. Dentre os trabalhos que utilizaram a abordagem de cancelamento sucessivo tem-se [Shynk & Gooch, 1993], que lançou a utilização do CMA neste tipo de processamento, e [Li & Sidiropoulos, 2000], que utiliza técnicas de projeção em sub-espacos para recuperar um usuário por vez.

Ainda na linha de critérios baseados no CMA, em [Cavalcanti et al., 1999, em SPAWC'99] e [Cavalcanti, 1999] é proposto um critério autodidata de equalização que modifica o critério CMA através da substituição da constante de dispersão da constelação por um valor estimado a partir das amostras temporais envolvidas no processo de desconvolução. Isto leva a um

⁷Neste trabalho, o MU-CMA é denominado MIMO-CMA.

conjunto de soluções possíveis, uma vez que a dispersão da constelação é determinada a partir dos dados disponíveis no receptor. O algoritmo resultante é denominado *Generalized Constant Modulus Algorithm* (GCMA). Também em [Cavalcanti et al., 1999, em SPAWC'99] é proposto um critério de processamento espaço-temporal para sistemas multiusuário que apresenta ganhos de desempenho quando comparado ao MU-CMA. Posteriormente, em [Neves, 2001] foi mostrado que o GCMA é na verdade derivado do critério SW em termos de maximização do momento de quarta ordem, mas que a função custo também apresenta problemas de mínimos locais, o que condiciona a convergência para um mínimo global a uma inicialização adequada.

Outros critérios que derivam de estratégias da abordagem direta de igualdade entre cumulantes também têm sido aplicados no contexto de processamento multiusuário com bastante sucesso. Estes critérios se caracterizam por utilizar explicitamente um determinado número de momentos de ordem superior para garantir que a recuperação e identificação dos sinais seja atingida. Dentre os vários, tem-se [Yang, 1998; Tang et al., 1999; Tugnait, 1999; Reynolds et al., 2002; Chi et al., 2002].

Finalmente, visando características específicas de sistemas multiusuário várias propostas de critérios também têm sido registradas na literatura. Dentre elas pode-se citar [Bugallo et al., 2000, 2001], que apresentam propostas baseadas em estratégias semi-cegas, ou seja, com curtas seqüências de treinamento, para cancelamento de interferência em sistemas multiusuário CDMA. O trabalho [Sala-Alvarez & Vázquez-Grau, 1994] também apresenta um critério de separação que utiliza a decorrelação dos dados na saída dos filtros de separação dos usuários, através de estratégias preditivas e utilização de técnicas de super-amostragem (aproveitando a cicloestacionaridade do sinal), para prover a separação de fontes. Medidas e avaliações de algoritmos cegos em sistemas reais são apresentadas em [Samardzija et al., 2002], na qual o critério MUK é utilizado como critério para redução de interferência de múltiplo acesso em interface aérea, comparando-o com estratégias amplamente utilizadas em sistemas MIMO, inclusive soluções supervisionadas. Na linha da utilização de técnicas de análise por componentes independentes, alguns trabalhos têm buscado um receptor que utiliza ICA para a retirada de interferência em sistemas com espalhamento espectral. Entre os poucos, tem-se [Ristaniemi & Joutsensalo, 1999; Ristaniemi, 2000; Ristaniemi & Joutsensalo, 2002] que ilustram a capacidade e os ganhos a se combinar o algoritmo *FastICA* para mitigação da interferência e um correlacionador (receptor *rake*) para identificação dos usuários por suas seqüências de código.

A lista de contribuições nesta área é realmente extensa e poderiam ser citados vários outros trabalhos. Entretanto, como ponto de partida para uma busca inicial, a lista fornecida nesta seção pretende ser representativa e bastante abrangente.

Vale a pena lembrar que as estratégias cegas de processamento multiusuário também possuem o problema de ambiguidade de permutação presente nas estratégias de separação cega de fontes. Esta questão é considerada solucionada na camada de acesso, onde protocolos

de acesso definem quais usuários possuem os sinais recuperados.

4.6 Conclusões e síntese

Este capítulo tem como um dos principais objetivos a caracterização do problema de processamento multiusuário como uma particularidade do problema de separação de fontes, dadas as características das fontes e dos canais que fornecem as misturas disponíveis no receptor. Também se propõe a mostrar a abrangência do problema através de soluções existentes na literatura.

A descrição de alguns dos tipos de múltiplo acesso existente nos sistemas de comunicação atuais é feita para ilustrar em quais situações ocorre a necessidade da utilização de técnicas de remoção de interferência de múltiplo acesso e/ou interferência inter-simbólica. Outro aspecto apresentado com detalhes é a estrutura de um receptor com múltiplos sensores implementada através de um arranjo de antenas, que é uma das principais tecnologias em destaque dos sistemas de comunicação sem fio de última geração.

Na seqüência do presente capítulo, a caracterização do tipo do canal de um sistema de processamento multiusuário é feita de maneira a mostrar a necessidade de utilizar diversidade espacial (processamento de faixa estreita) ou diversidade espaço-temporal (processamento de faixa larga) na tarefa de eliminar as interferências em um sistema de vários usuários. Posteriormente, a caracterização das fontes que são utilizadas em sistemas cooperativos com múltiplo acesso é também realizada visando permitir a utilização de estratégias que tratam as condições específicas do processamento multiusuário.

Finalmente, alguns critérios e algoritmos cegos utilizados no processamento multiusuário são descritos e analisados e suas diferenças são enfatizadas, buscando apresentar a evolução das técnicas envolvidas no processamento não-supervisionado. Outros algoritmos também reportados na literatura são descritos e comentados de forma bastante sucinta, mas permitindo ao leitor ter uma visão geral dos aspectos considerados na pesquisa em termos de algoritmos cegos para processamento multiusuário.

Quanto à concepção de critérios e algoritmos, ainda há espaço para técnicas que apresentem uma melhor relação de compromisso complexidade \times desempenho. Na tentativa de se encontrar e analisar soluções mais adequadas no atendimento deste compromisso, desenvolveu-se a segunda parte desta tese.

PARTE **II**

*Proposições e
Análises*

– “A imaginação é mais importante que o conhecimento.”

Albert Einstein, 1930

5

Métodos Multiusuário Baseados na Estimação de Densidade de Probabilidade

VÁRIOS critérios para processamento multiusuário são derivados do problema de desconvolução autodidata, conforme discutido no Capítulo 4. Neste capítulo, é apresentada e analisada uma proposta de critério de otimização para problemas de detecção multiusuário, derivada a partir de uma idéia inicialmente sugerida em equalização cega. O critério e o algoritmo dele derivado são analisados no que diz respeito à convergência, tecendo-se também comparações quanto à generalidade da proposta.

Outro aspecto avaliado é a garantia de obtenção de uma solução que atinja a separação de todos os usuários envolvidos no processamento. Além disso, análises teóricas e algumas simulações computacionais são feitas para explicitar as diferenças e similaridades com outros critérios e algoritmos de processamento multiusuário, propostos a partir de diferentes abordagens. Visando uma maior abrangência, a proposta é discutida à luz de consideração sobre métodos de estimação da densidade de probabilidade, em particular o método de Parzen, que tem uma inserção direta no processamento de sinais discretos característicos de

sistemas de comunicação digital. Já na discussão sobre equivalências de critérios, surgem resultados originais de análise, inclusive sobre métodos já clássicos de equalização cega e treinada.

O presente capítulo é dividido da seguinte forma. Na Seção 5.1 é realizada uma breve revisão sobre o critério para equalização cega que norteia o critério de processamento multiusuário aqui proposto. Na Seção 5.2, a análise teórica de convergência e aspectos relativos à equivalência com outras técnicas são discutidos, permitindo uma clara visão da generalidade da proposta e de seus impactos em processamento adaptativo. A discussão de métodos e aspectos de estimação da função de densidade de probabilidade e de alguns critérios que também utilizam esta abordagem é realizada na Seção 5.3. A proposta de uma família de critérios multiusuário baseados na estimação da densidade de probabilidade é realizada na Seção 5.4. Uma análise do critério de minimização da probabilidade de erro como uma medida de informação é apresentada na Seção 5.5 e as conclusões e síntese do capítulo são listadas na Seção 5.6.

5.1 Revisão sobre método para equalização cega

Inicialmente, é importante apresentar o critério de equalização cega baseado na estimação da função de densidade de probabilidade, preliminar, proposto em [Cavalcante, 2001]. Naquele trabalho, a proposta foi bastante incipiente, sem maiores análises teóricas sobre o critério nem o comportamento do algoritmo dele decorrente, baseando a avaliação de desempenho apenas em simulações computacionais comparativas. É importante também destacar que a motivação maior do trabalho em [Cavalcante, 2001] é a proposta de um algoritmo auto-organizável para RNA, que tem como meta encontrar algumas particularidades da fdp do sinal recebido.

No sentido de propor e analisar um critério para processamento multiusuário, pretende-se aqui entender e avaliar as hipóteses e desenvolvimentos do critério original de equalização cega.

Considera-se um sistema de comunicação digital em seu modelo em banda básica equivalente representado na Figura 5.1.

Considera-se que o equalizador é otimizado de forma a cancelar a ISI, ou seja, satisfaz o critério *Zero Forcing* (ZF) que fornece $H(z) \cdot W(z) = z^{-\ell}$, em que $H(z)$ e $W(z)$ são as funções de transferência (FT) do canal e do equalizador, respectivamente, e ℓ é um atraso de decisão.

De forma equivalente, utilizando-se notação matricial, a condição ZF pode ser escrita como:

$$\mathcal{H}^H \mathbf{w} = \delta_{\ell}, \quad (5.1)$$

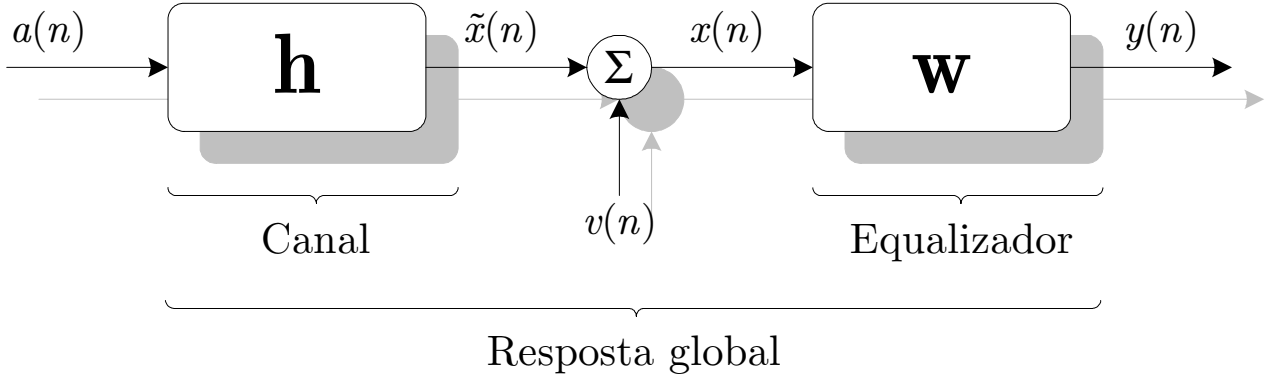


Figura 5.1: Modelo em banda básica equivalente de um sistema de comunicação digital.

em que δ_ℓ é o vetor, do tipo da função de Kronecker, com o ℓ -ésimo componente não-nulo e \mathcal{H} é a matriz de convolução do canal dada por

$$\mathcal{H} = \begin{bmatrix} h_0 & & \mathbf{0} \\ \vdots & \ddots & \\ h_{L-1} & \ddots & h_0 \\ & \ddots & \vdots \\ \mathbf{0} & & h_{L-1} \end{bmatrix}_{(L+N-1) \times (N)}, \quad (5.2)$$

para um canal de L coeficientes e um equalizador com N coeficientes [Cavalcante, 2001].

O vetor com as amostras temporais do sinal recebido no equalizador é dado por:

$$\mathbf{x}(n) = \mathcal{H}^H \mathbf{a}(n) + \mathbf{v}(n), \quad (5.3)$$

sendo $\mathbf{x}(n) = [x(n) \ \cdots \ x(n-N+1)]^T$ e $\mathbf{a}(n) = [a(n) \ \cdots \ a(n-L+1)]^T$, enquanto que à saída, considerando o equalizador ideal, tem-se [Cavalcante, 2001]:

$$\begin{aligned} y(n) &= (\mathcal{H}^H \mathbf{a}(n) + \mathbf{v}(n))^H \mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^H(n) \mathcal{H} \mathbf{w}_{\text{ideal}} + \mathbf{v}^H(n) \mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^H(n) \underbrace{\mathcal{H} \mathbf{w}_{\text{ideal}}}_{\mathbf{g}_{\text{ideal}}} + \mathbf{v}^H(n) \mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^H(n) \mathbf{g}_{\text{ideal}} + \vartheta(n) \\ &= a(n-\ell) + \vartheta(n), \end{aligned} \quad (5.4)$$

o vetor \mathbf{g} é a resposta global do sistema SISO e $\vartheta(n)$ é uma v.a. gaussiana de média nula, uma vez que é uma combinação linear de v.a. gaussianas. Este tipo de hipótese é considerada também nos trabalhos sobre métodos de Bussgang para equalização cega [Haykin, 1994].

A Equação (5.4) continua sendo válida mesmo que o equalizador não seja ideal resultando na ocorrência de alguma ISI residual. Para tanto, é suficiente assumir que a interferência existente (ruído e ISI) no sinal na saída do equalizador apresente uma distribuição gaussiana [Haykin, 1994, 1996]. Tal hipótese é sustentada pelo Teorema Central do Limite [Papoulis, 1991]. Em sistemas multiusuário esta hipótese pode ser também suportada como sendo decorrente de todos os sinais interferentes a um usuário específico [Poor & Verdú, 1997].

Assim, é possível calcular a fdp do sinal idealmente recuperado, considerando lineares o canal e o equalizador temporal. Considerando-se que $a(n - \ell)$ pertence ao alfabeto da modulação de símbolos transmitidos \mathfrak{A} , a fdp de y é dada por:

$$p_{Y,\text{ideal}}(y) = \frac{1}{\sqrt{2\pi\sigma_{\vartheta}^2}} \sum_{i=1}^c \exp\left(-\frac{|y(n) - \mathbf{a}_i|^2}{2\sigma_{\vartheta}^2}\right) \Pr(\mathbf{a}_i), \quad (5.5)$$

em que \mathbf{a}_i é o i -ésimo elemento (escalar) de \mathfrak{A} e a variância de cada uma das gaussianas é dada por

$$\sigma_{\vartheta}^2 = \sigma_v^2 \cdot \sum_i |w_i|^2. \quad (5.6)$$

É interessante observar da Equação (5.6) que σ_v^2 pode ser estimada a partir da estimação de σ_{ϑ}^2 , impondo a normalização dos coeficientes do equalizador; ou seja, $\sigma_{\vartheta}^2 = \sigma_v^2$ se $\sum_i |w_i|^2 = 1$.

A partir do conhecimento da fdp do sinal que se deseja obter na saída do equalizador, a meta é forçar os sinais na saída do equalizador a obedecerem à distribuição da Equação (5.5).

A proposta em [Cavalcante, 2001] é de abordar o problema como um problema de *classificação*, ou seja, a questão da equalização consiste agora em separações entre classes que representem os sinais do alfabeto da modulação [Montalvão, 2000; Cavalcante, 2001].

De acordo com esta abordagem, encontrar a separação entre as classes corresponde a *minimizar a dispersão intraclassa*. Esta corresponde à variação que símbolos de uma mesma classe possuem em torno do baricentro da mesma. Por exemplo, num sistema digital em que duas classes associadas aos símbolos +1 ou -1 existem, a dispersão intraclassa é dada pela variância dos sinais que pertencem à classe +1 (-1), conforme pode ser visto na Figura 5.2 na qual os símbolos + e o pertencem às classes +1 e -1, respectivamente. Esta dispersão é também chamada de autocorrelação condicional [Montalvão, 2000] e é responsável por uma má classificação dos padrões uma vez que sinais pertencentes a uma dada classe podem estar contidos ou muito próximos à outra, o que torna a tarefa de separação mais complexa. Uma minimização da dispersão intraclassa pode ser obtida através da estimação das características da fdp do sinal de saída do equalizador, tornando possível estimar os sinais de cada classe.

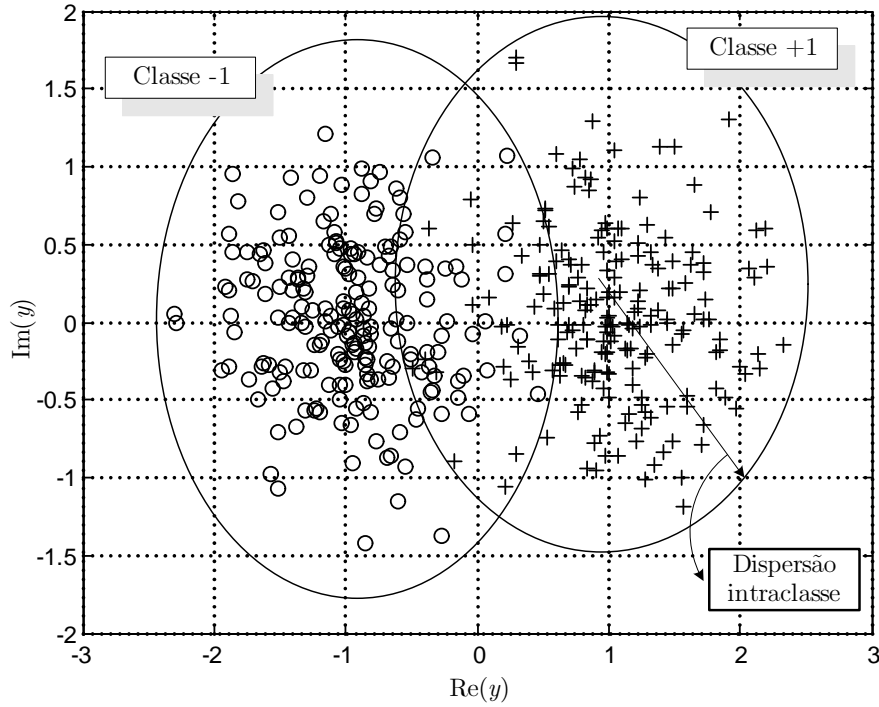


Figura 5.2: Exemplo de dispersão intraclasse para um sistema digital contendo duas classes.

Um possível critério de medida de similaridade entre funções é dado por [Small & McLeish, 1994]

$$J = - \int_{-\infty}^{\infty} f(r)g(r)dr, \quad (5.7)$$

em que $f(\cdot)$ e $g(\cdot)$ são funções quaisquer. Quando as funções são iguais (máximo grau de similaridade) $J = 0$. Neste caso, para encontrar a mínima dispersão intraclasse, a função dada por

$$J_{\text{ideal}}(\mathbf{w}_{\text{ideal}}, \sigma_{\vartheta}^2, \mathbf{w}, \sigma_r^2) = - \int p_{Y,\text{ideal}}(y, \mathbf{w}_{\text{ideal}}, \sigma_{\vartheta}^2) \cdot f(y, \mathbf{w}, \sigma_r^2) dy, \quad (5.8)$$

deve ser mínima, ou seja, $p_{Y,\text{ideal}}(y, \mathbf{w}_{\text{ideal}}, \sigma_{\vartheta}^2) = f(y, \mathbf{w}, \sigma_r^2)$, em que $f(y, \mathbf{w}, \sigma_r^2)$ é uma função paramétrica adequada.

Desta forma, como é necessário mensurar a similaridade entre as funções, lança-se mão da clássica medida entre fdps, a divergência de Kullback-Leibler como alternativa à métrica dada na Equação (5.7).

Além da divergência, utiliza-se o fato do conhecimento da fdp do sinal idealmente equalizado para construir um modelo paramétrico da função que se deseja obter. Desta forma,

o modelo que retém as características do sinal desejado (estatisticamente) é escrito como:

$$\Phi(y) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_r^2}}}_A \cdot \sum_{i=1}^c \exp\left(-\frac{|y(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right) \cdot \Pr(\mathbf{a}_i), \quad (5.9)$$

em que $y(n) = \mathbf{w}^H(n)\mathbf{x}(n)$ e σ_r^2 é uma estimativa de σ_θ^2 , que foi também incluída na Equação (5.8) para explicitar a necessidade de uma estimativa do ruído na saída do equalizador para o modelo do sinal idealmente recebido.

Na realidade, ao buscar-se um sinal idealmente equalizado, seria interessante obter um modelo *sem ruído* e tornar o modelo paramétrico da Equação (5.9) um somatório de funções de Dirac para os valores dos símbolos de \mathfrak{A} , o que corresponde a fazer $\sigma_r^2 \rightarrow 0$. Entretanto, como observado em simulações e também em [Amara, 2001], este modelo causa muitos problemas numéricos, o que impossibilita seu uso em aplicações práticas.

Assim, é possível construir uma função custo que minimize a dispersão intraclasse, utilizando-se o fato de se conhecer a estatística do sinal recebido, através da seguinte medida:

$$\begin{aligned} D(p_Y(y) || \Phi(y)) &= \int_{-\infty}^{\infty} p_Y(y) \cdot \ln \left[\frac{p_Y(y)}{\Phi(y)} \right] dy \\ &= \int_{-\infty}^{\infty} p_Y(y) \cdot \ln [p_Y(y)] dy - \int_{-\infty}^{\infty} p_Y(y) \cdot \ln [\Phi(y)] dy. \end{aligned} \quad (5.10)$$

Então, o funcional da Equação (5.10) pode ser simplificado para ser dado somente pelo termo dependente do modelo paramétrico $\Phi(y)$, o que fornece:

$$\begin{aligned} J(\mathbf{w}) &= \int_{-\infty}^{\infty} p_Y(y) \cdot \ln \left(\frac{1}{\Phi(y)} \right) dy \\ &= -\mathbb{E} \{ \ln [\Phi(y)] \} \\ &= -\mathbb{E} \left\{ \ln \left[A \cdot \sum_{i=1}^c \exp \left(-\frac{|y - \mathbf{a}_i|^2}{2\sigma_r^2} \right) \right] \right\}. \end{aligned} \quad (5.11)$$

O critério acima é denominado **Fitting pdf Criterion (FPC)** e será representado por J_{FPC} . Para proceder a uma otimização estocástica, necessita-se da estimativa do vetor

gradiente dada por:

$$\nabla J_{\text{FPC}}(\mathbf{w}) = \frac{\sum_{i=1}^{\mathfrak{c}} \exp\left(-\frac{|y(n)-\mathbf{a}_i|^2}{2\sigma_r^2}\right) [y(n) - \mathbf{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{c}} \exp\left(-\frac{|y(n)-\mathbf{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n), \quad (5.12)$$

de modo que a equação de adaptação é então escrita como:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J_{\text{FPC}}(\mathbf{w}). \quad (5.13)$$

Este algoritmo de adaptação será denominado *Fitting pdf Algorithm* (FPA).

Na seção seguinte, é realizada uma análise mais detalhada do método *Fitting pdf* (FP)¹, no sentido de levar à concepção de um critério para processamento multiusuário tomando por base o FPC.

5.2 Análise do método FP

Vários aspectos importantes quando da concepção de um método de equalização cega devem ser abordados para que sua aplicabilidade seja possível para o maior número de cenários possível. Serão visto aqui diversos aspectos relativos ao critério FP e às propriedades de convergência do algoritmo estocástico a ele associado.

5.2.1 Propriedades de convergência

Um aspecto importante é mostrar a convergência do algoritmo para um ponto o mais próximo possível da solução ótima.

Tal análise é inspirada na análise de estabilidade local apresentada em [Barbarossa & Scaglione, 1997]. Naquele trabalho, uma estratégia de equalização cega baseada no chaveamento do CMA para um critério baseado em um somatório de gaussianas é apresentada.

Inicialmente, sendo $\mathbf{x}(n) = \mathcal{H}\mathbf{a}(n) + \mathbf{v}(n)$ o sinal recebido no instante n , conforme discutido na Equação (5.4), deve-se ter para um equalizador ideal

$$\mathbf{w}_{\text{ideal}}^H \mathcal{H}\mathbf{a}(n) = a(n - \ell),$$

¹Por questões de simplificação da redação, serão utilizados os termos critério FP e FPC para se referenciar ao critério descrito nesta seção.

em que $a(n - \ell) = \mathbf{a}_i$ para $1 \leq i \leq \mathfrak{C}$.

Para avaliar a convergência do algoritmo, considera-se uma perturbação na estimação do filtro ideal, de forma a incluir os efeitos da adaptação. Assim, o filtro equalizador obtido passa a ser modelado na forma [Cavalcante et al., 2002a]:

$$\mathbf{w} = \mathbf{w}_{\text{ideal}} + \Delta \mathbf{w}. \quad (5.14)$$

Escrevendo-se então o valor do critério para os parâmetros descritos na Equação (5.14) tem-se, a partir da Equação (5.11):

$$\begin{aligned} J_{\text{FPC}}(\mathbf{w}) &= -\mathbb{E} \left\{ \ln \left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp \left(-\frac{|\mathbf{w}^H \mathcal{H}^H \mathbf{a}(n) + \mathbf{w}^H \mathbf{v}(n) - \mathbf{a}_i|^2}{2\sigma_r^2} \right) \right] \right\} \\ &= -\mathbb{E} \left\{ \ln \left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp \left(-\frac{|(\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathcal{H}^H \mathbf{a}(n) + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) - \mathbf{a}_i|^2}{2\sigma_r^2} \right) \right] \right\} \\ &= -\mathbb{E} \left\{ \ln \left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp \left(-\frac{|(\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathcal{H}^H \mathbf{a}(n) + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) - \mathbf{a}_i|^2}{2\sigma_r^2} \right) \right] \right\} \\ &= -\mathbb{E} \left\{ \ln \left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp \left(-\frac{|\mathbf{w}_{\text{ideal}} \mathcal{H}^H \mathbf{a}(n) + \Delta \mathbf{w} \mathcal{H}^H \mathbf{a}(n) + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) - \mathbf{a}_i|^2}{2\sigma_r^2} \right) \right] \right\}, \end{aligned} \quad (5.15)$$

Então, considerando que o algoritmo está próximo o suficiente de uma região próxima da solução ideal, pode-se considerar apenas um dos termos do somatório já que a contribuição para os demais termos do somatório pode ser considerada desprezível. Esta região corresponde ao símbolo estimado ser próximo de um dos valores de \mathbf{a}_i . Por exemplo, para $i = 1$, chega-se a

$$\begin{aligned}
& J_{\text{FPC}}(\mathbf{w}) \\
& \approx -\mathbb{E} \left\{ \ln \left[A \cdot \exp \left(- \frac{\left| \underbrace{\mathbf{w}_{\text{ideal}} \mathcal{H}^H \mathbf{a}(n)}_{\mathbf{a}_1} + \Delta \mathbf{w} \mathcal{H}^H \mathbf{a}(n) + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) - \mathbf{a}_1 \right|^2}{2\sigma_r^2} \right) \right] \right\} \\
& \approx -\mathbb{E} \left\{ - \frac{\left| \Delta \mathbf{w} \mathcal{H}^H \mathbf{a}(n) + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) \right|^2}{2\sigma_r^2} + \ln[A] \right\}.
\end{aligned} \tag{5.16}$$

Expandindo-se o termo do módulo ao quadrado, obtém-se

$$\Delta \mathbf{w}^H \mathcal{H} \mathbf{a}(n) \mathbf{a}^H(n) \mathcal{H}^H \Delta \mathbf{w} + (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H \mathbf{v}(n) \mathbf{v}^H(n) (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w}), \tag{5.17}$$

com o qual, tomando-se a esperança matemática da Equação (5.16), tem-se:

$$J_{\text{FPC}}(\mathbf{w}) \approx \frac{\sigma_a^2 \Delta \mathbf{w}^H \mathcal{H} \mathcal{H}^H \Delta \mathbf{w} + \sigma_v^2 (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})^H (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w})}{2\sigma_r^2} - \ln(A). \tag{5.18}$$

A partir da Equação (5.18), o gradiente do FPC pode ser dado, de forma aproximada, por [Cavalcante et al., 2002a]:

$$\nabla J_{\text{FPC}}(\mathbf{w}) \approx \frac{\sigma_a^2}{2\sigma_r^2} \left(\mathcal{H} \mathcal{H}^H \Delta \mathbf{w} + \frac{\sigma_v^2}{\sigma_a^2} (\mathbf{w}_{\text{ideal}} + \Delta \mathbf{w}) \right) \tag{5.19}$$

em que $\frac{\sigma_v^2}{\sigma_a^2}$ é o inverso da relação sinal-ruído.

Para avaliar a convergência do algoritmo, considera-se o caso sem ruído, ou seja, $\frac{\sigma_v^2}{\sigma_a^2} \rightarrow 0$. Assim, a Equação (5.19) torna-se

$$\begin{aligned}
\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu \frac{\sigma_a^2}{\sigma_r^2} \mathcal{H} \mathcal{H}^H \Delta \mathbf{w}(n) \\
\mathbf{w}_{\text{ideal}}(n+1) + \Delta \mathbf{w}(n+1) &= \mathbf{w}_{\text{ideal}}(n) + \Delta \mathbf{w}(n) - \mu \frac{\sigma_a^2}{\sigma_r^2} \mathcal{H} \mathcal{H}^H \Delta \mathbf{w}(n),
\end{aligned} \tag{5.20}$$

e, uma vez que $\mathbf{w}_{\text{ideal}}$ não depende do instante temporal, segue que:

$$\Delta \mathbf{w}(n+1) = \Delta \mathbf{w}(n) \left(\mathbf{I} - \mu \frac{\sigma_a^2}{\sigma_r^2} \mathcal{H} \mathcal{H}^H \right), \tag{5.21}$$

em que \mathbf{I} é a matriz identidade.

A recursão da Equação (5.21) converge se o fator de passo respeita a seguinte condição [Haykin, 1996; Attux, 2001]:

$$0 < \mu < \frac{4\sigma_r^2}{\sigma_a^2 \lambda_{\max}}, \quad (5.22)$$

em que λ_{\max} é o maior autovalor de $\mathcal{H}\mathcal{H}^H$.

O resultado mostrado na Equação (5.22) mostra que o parâmetro σ_r^2 apresenta uma forte influência no desempenho do algoritmo, uma vez que ele atua como um *fator de controle* da taxa de convergência.

Além de atuar como fator de controle da convergência, por ser a variância de cada uma das gaussianas que compõem o modelo do sinal que se deseja obter, o parâmetro σ_r^2 também desempenha um papel fundamental na função custo. Isto porque se σ_r^2 assumir um valor muito elevado, o conjunto das gaussianas tende a se tornar só uma devido à proximidade dos seus centros. Quando o valor de σ_r^2 é muito pequeno, o vetor gradiente na Equação (5.12) assume valores elevados, causando divergência.

Esta dependência da função custo com σ_r^2 pode ser facilmente visualizada. Para a obtenção do funcional $J_{\text{FPC}}(\mathbf{w})$ em função dos parâmetros do equalizador, pode-se considerar um caso simples com os seguintes parâmetros: modulação BPSK, equalizador de dois coeficientes e canal com dois coeficientes. Visando avaliar um maior número de casos, $J_{\text{FPC}}(\mathbf{w})$ é calculado para duas categorias de canais.

Na Figura 5.3, ilustra-se $J_{\text{FPC}}(\mathbf{w})$ para um canal AR (*Auto Regressive*), uma SNR = 10 dB e diferentes valores de σ_r^2 . O modelo discreto da FT do canal é dado por

$$H_{\text{AR}}(z) = \frac{1}{1 + 0.6z^{-1}}. \quad (5.23)$$

Como pode ser observado na Figura 5.3, o aumento do valor do parâmetro σ_r^2 faz o funcional modificar sua forma e assemelhar-se a uma função unimodal. Entretanto, o mínimo de tal solução não corresponde a uma solução para o problema de equalização. Nestes casos, o valor de σ_r^2 faz com que o somatório das gaussianas seja aproximado, através do Teorema Central do Limite, por somente uma gaussiana de centro igual à média dos centros das gaussianas individuais do modelo da fdp do sinal desejado. Neste ponto, o ponto de mínimo corresponde a tal valor do centro da gaussiana resultante e a otimização do critério força os dados a estarem concentrados em torno da média nula, uma vez que os \mathbf{a}_i possuem valor médio zero.

Este tipo de resultado reforça a dedução da importância do valor de σ_r^2 para a composição do critério.

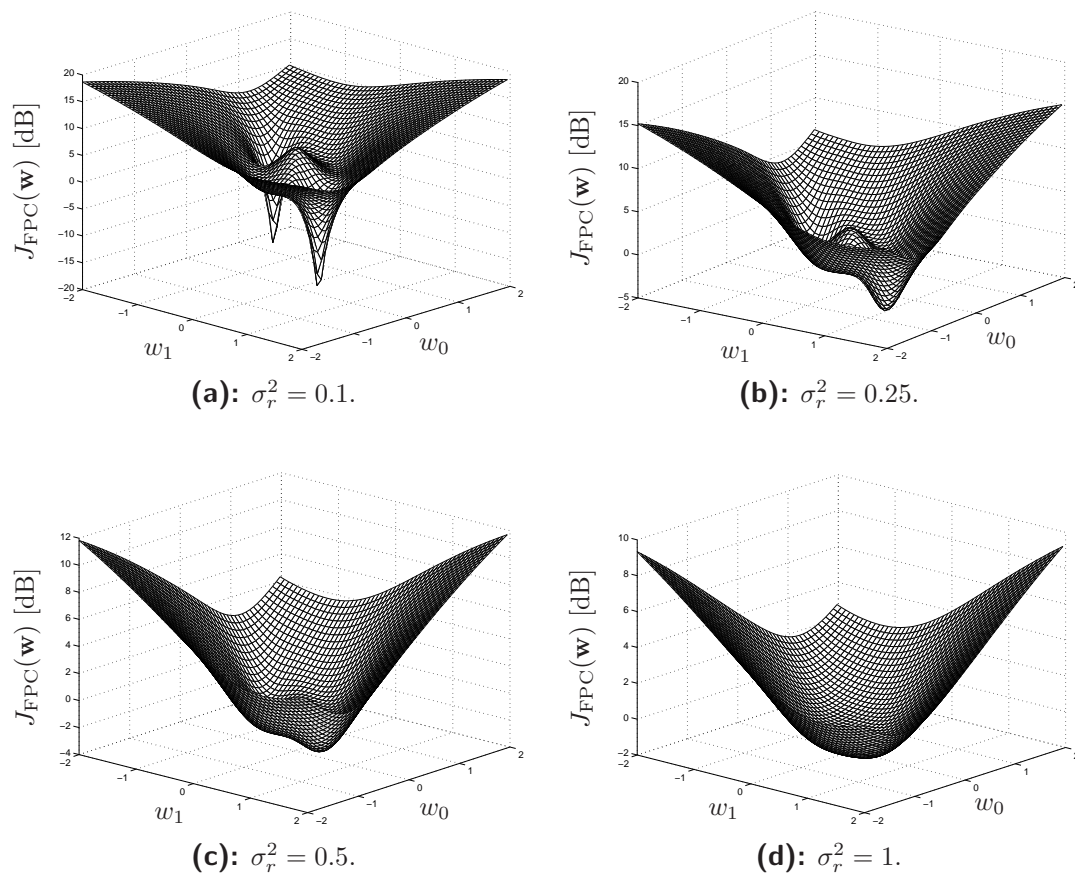


Figura 5.3: Comportamento da função custo de FP para vários valores de σ_r^2 utilizando canal $H_{\text{AR}}(z)$.

Ao utilizar-se um canal MA (*Moving Average*) dado por

$$H_{\text{MA}}(z) = 1 + 0.6z^{-1}, \quad (5.24)$$

a geometria de $J_{\text{FPC}}(\mathbf{w})$ apresenta diferenças, conforme pode ser visualizado na Figura 5.4.

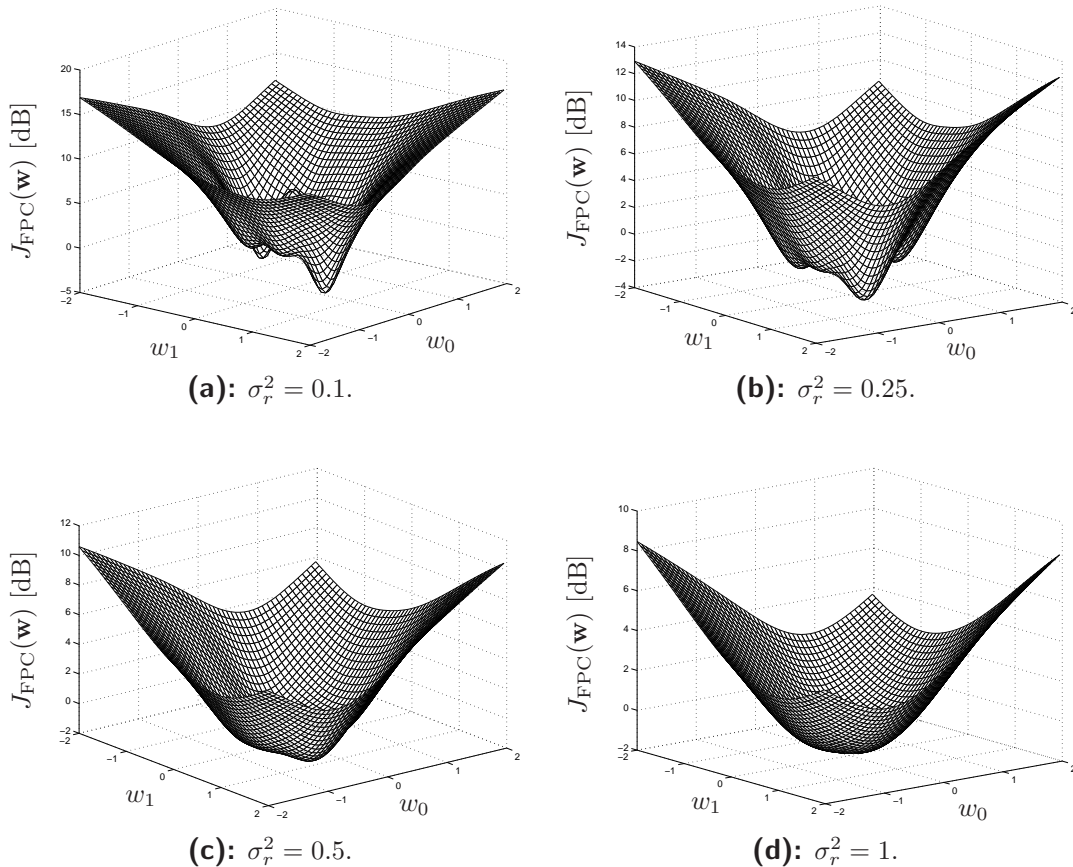


Figura 5.4: Comportamento da função custo de FP para vários valores de σ_r^2 utilizando canal $H_{\text{MA}}(z)$.

Pode-se observar o mesmo comportamento que para o caso utilizando H_{AR} . A diferença se dá pelo fato de que σ_r^2 pode assumir menores valores para que a função seja aproximada por uma única gaussiana. Isto ocorre devido à proximidade dos mínimos (locais e globais) da função custo.

Uma vez observada a existência de mínimos globais e locais que perturbam a tarefa de recuperação do sinal, pode-se ainda visualizar a questão da dependência da convergência para mínimos globais ou locais dependendo da inicialização do algoritmo. Para este tipo de simulação, escolhe-se o valor de $\sigma_r^2 = 0.1$, modulação BPSK, SNR = 10 dB e 20 trajetórias com

diferentes inicializações em um círculo de raio igual a 2. Estas inicializações são dividindo-se o círculo em um número de setores iguais ao número de trajetórias que se deseja avaliar.

A Figura 5.5 mostra as convergências, nas curvas de nível da função custo do critério FP, para os diferentes valores de μ para o canal $H_{AR}(z)$, como função da evolução temporal dos parâmetros w_0 e w_1 do equalizador.

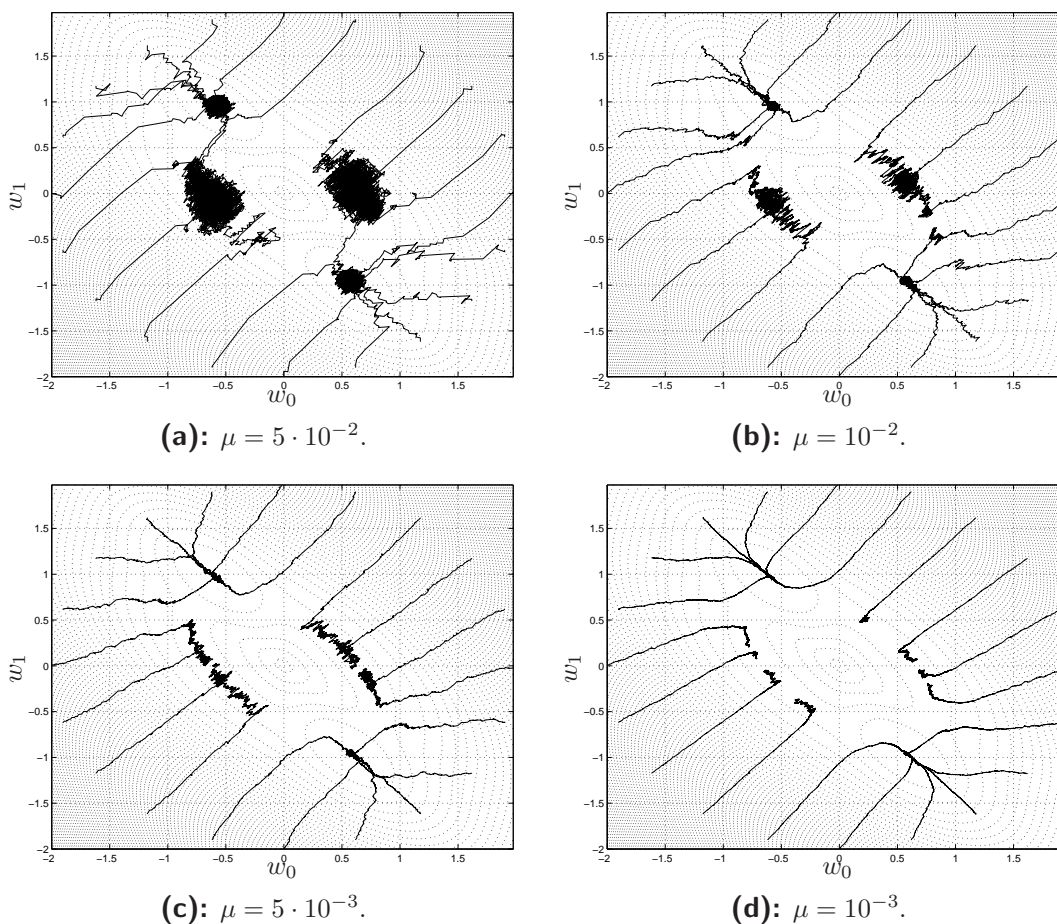


Figura 5.5: Trajetórias de convergência para diferentes inicializações e valores de fator de passo utilizando canal $H_{AR}(z)$ e modulação BPSK.

O caso do canal $H_{MA}(z)$ é ilustrado na Figura 5.5, também para diferentes valores de μ , modulação BPSK e 20 inicializações distribuídas uniformemente sob um círculo de raio 2; a relação sinal-ruído é de 10 dB.

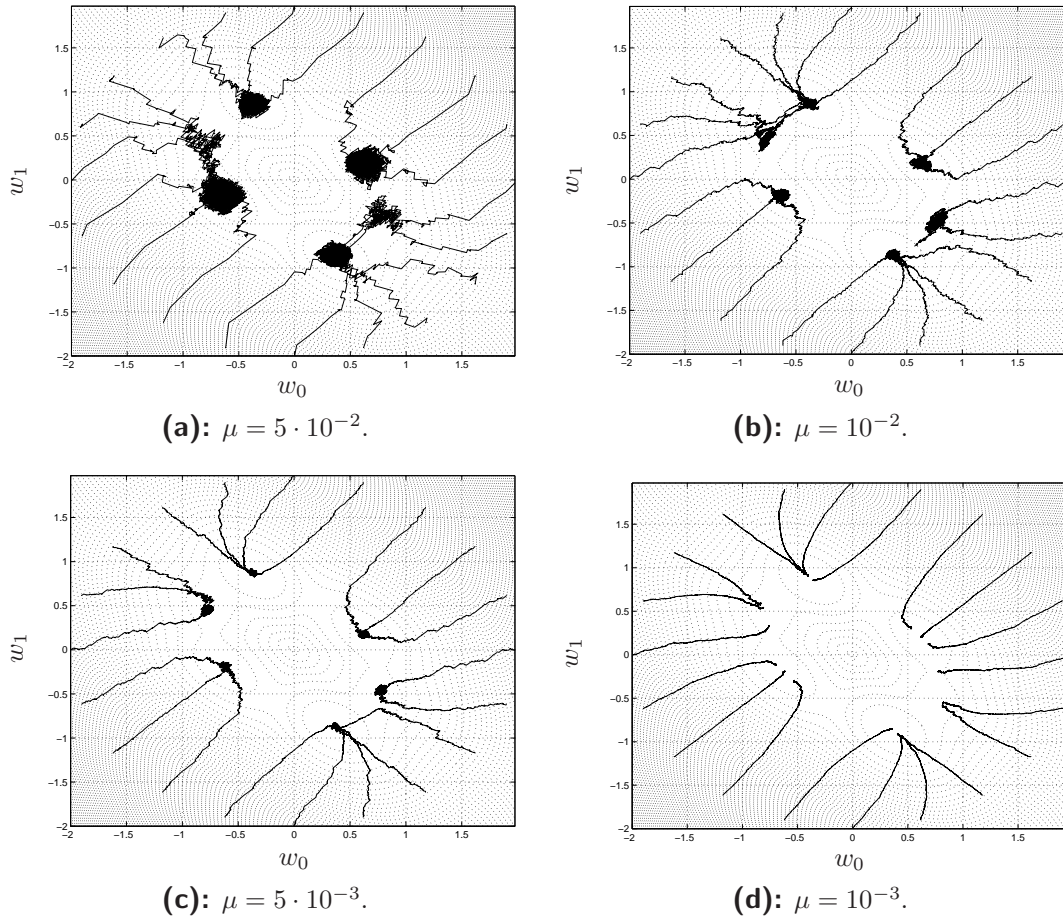


Figura 5.6: Trajetórias de convergência para diferentes inicializações e valores de fator de passo utilizando canal $H_{MA}(z)$ e modulação BPSK.

Como pode ser visualizado, com mais clareza, a partir da Figura 5.6, a função custo apresenta no espaço dos parâmetros do equalizador pontos que são denominados de *mínimos espúrios*, ou seja, mínimos que não são solução para nenhuma escolha de atraso [Suyama, 2003].

Este comportamento é também apresentado por outras classes de critérios cegos, o que torna interessante uma comparação do critério FP com outros critérios. Além disso, é importante verificar se os pontos de mínimo locais ou globais apresentam-se como soluções aceitáveis para o problema de desconvolução. Estes são os assuntos tratados na seção seguinte.

5.2.2 Equivalência entre o critério FP e outros critérios

Para discorrer sobre similaridades entre o critério FP e outros critérios existentes na literatura, é importante fazer algumas observações sobre sua função custo. A partir daí, alguns resultados originais de análise serão apresentados nesta seção.

Em se tratando de um procedimento cego, não se conhece o símbolo transmitido no instante n . Desta forma, o modelo paramétrico $\Phi(y)$ é na verdade o somatório das probabilidades do sinal recebido condicionado ao envio de um símbolo \mathbf{a}_i . Desta forma, pode-se escrever o modelo paramétrico da seguinte forma:

$$\Phi(y) = \sum_{i=1}^c p(y(n)|\mathbf{a}_i). \quad (5.25)$$

De fato, para se re-obter a Equação (5.9) a partir da Equação (5.25), basta considerar que o sinal ϑ na saída do equalizador possui uma fdp gaussiana e utilizar a regra de Bayes [Therrien, 1992].

O somatório da Equação (5.25) é referente à incerteza decorrente do não conhecimento do símbolo transmitido, os quais são ponderados para fornecer a métrica de adaptação do critério.

Um critério supervisionado, por sua vez, não sofre da incerteza dos dados, uma vez que o dado transmitido é disponível. Neste caso, os termos do somatório da Equação (5.25) relativos aos símbolos não enviados. Assim sendo, a função custo do critério FP, para o caso supervisionado, é escrita como:

$$J_{\text{FPC}}(\mathbf{w}) = p(y(n)|\mathfrak{d}(n)), \quad (5.26)$$

em que $\mathfrak{d}(n)$ é o símbolo desejado no instante n . Logo, considerando novamente a hipótese de ruído gaussiano na saída do equalizador e expandindo-se a fdp do sinal, tem-se, utilizando a regra de Bayes

$$\begin{aligned} J_{\text{FPC}}(\mathbf{w}) &= -\mathbb{E} \left\{ \ln \left[A \cdot \exp \left(-\frac{|y(n) - \mathfrak{d}(n)|^2}{2\sigma_r^2} \right) \right] \right\} \\ &= \frac{1}{2\sigma_r^2} \mathbb{E} \{ |y(n) - \mathfrak{d}(n)|^2 \} - \ln[A], \end{aligned} \quad (5.27)$$

que é o critério MMSE, a menos de fatores de escala. Em termos de otimização, o valor a ser obtido será o mesmo, uma vez que o gradiente possuirá o mesmo valor.

Assim, partindo-se do fato do modelo paramétrico ser escrito como um modelo de mistura de fdps condicionais, estabelece-se uma ligação entre o critério FP (cego) e um procedimento supervisionado. Ou seja, o FPC se reduz ao critério MMSE quando há disponibilidade do símbolo transmitido.

Outra maneira de explicitar esta equivalência pode ser realizada, utilizando-se o gradiente do FPC para mostrar que se chega à solução de Wiener.

Inicialmente, para facilitar o entendimento, será utilizado um abuso de notação para representar a esperança matemática. Assim, a notação $E_Y\{\cdot\}$ representa a esperança matemática em relação à variável Y , permitindo a representação da esperança em relação a outras v.a. para funções de várias variáveis [Papoulis, 1991].

Inicialmente, tomando-se o gradiente da função custo do FPC, sem realizar a aproximação estocástica da média pelo valor instantâneo das amostras, tem-se

$$\begin{aligned}\nabla J_{\text{FPC}}(\mathbf{w}) &= -\frac{\partial}{\partial \mathbf{w}} (E_Y \{\ln [\Phi(y)]\}) \\ &= -E_Y \left\{ \frac{\partial}{\partial \mathbf{w}} (\ln [\Phi(y)]) \right\} \\ &= -E_Y \left\{ \frac{\Phi(y) \cdot [y - \mathbf{a}]}{\sigma_r^2 \cdot \Phi(y)} \cdot \mathbf{x}^* \right\}.\end{aligned}\tag{5.28}$$

O ponto fundamental para demonstrar a equivalência entre a solução de Wiener e a do FPC está na maneira de escrever o modelo paramétrico. Como mencionado anteriormente, o modelo paramétrico pode ser escrito como uma soma das probabilidades condicionais do sinal recebido em relação aos símbolos transmitidos. Isto implica que o modelo pode ser escrito como uma esperança matemática em relação à variável de suporte.

Utilizando-se deste fato, pode-se escrever o modelo paramétrico como:

$$\begin{aligned}\Phi(y) &= \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \frac{1}{\mathfrak{C}} \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right) \\ &= E_{\mathfrak{A}} \left\{ A \cdot \exp\left(-\frac{|y - \mathbf{a}|^2}{2\sigma_r^2}\right) \right\},\end{aligned}\tag{5.29}$$

em que \mathfrak{A} é o espaço da variável \mathbf{a} . Como \mathfrak{A} é um conjunto discreto, chega-se também da Equação (5.29) à expressão de $\Phi(y)$ dada na Equação (5.9).

Então, utilizando a Equação (5.29) para reescrever o gradiente de J_{FPC} , obtém-se

$$\nabla J_{\text{FPC}}(\mathbf{w}) = -E_Y \left\{ \frac{E_{\mathfrak{A}} \left\{ A \cdot \exp\left(-\frac{|y - \mathbf{a}|^2}{\sigma_r^2}\right) [y - \mathbf{a}^*] \right\}}{\sigma_r^2 \cdot E_{\mathfrak{A}} \left\{ A \cdot \exp\left(-\frac{|y - \mathbf{a}|^2}{\sigma_r^2}\right) \right\}} \cdot \mathbf{x} \right\}.\tag{5.30}$$

Com esta expressão, pode-se ainda escrever, conforme sugerido em [Sala-Alvarez &

Vázquez-Grau, 1997], uma função auxiliar dada por

$$\psi(y, \mathbf{a}) = \frac{A \cdot \exp\left(-\frac{|y-\mathbf{a}|^2}{\sigma_r^2}\right)}{\mathbb{E}_{\mathfrak{A}}\left\{A \cdot \exp\left(-\frac{|y-\mathbf{a}|^2}{\sigma_r^2}\right)\right\}}, \quad (5.31)$$

de forma a se chegar à seguinte expressão para $\nabla J_{\text{FPC}}(\mathbf{w})$:

$$\nabla J_{\text{FPC}}(\mathbf{w}) = -\frac{1}{\sigma_r^2} \mathbb{E}_Y \left\{ \mathbb{E}_{\mathfrak{A}} \left\{ \psi(y, \mathbf{a}) \right\} [y - \mathbf{a}^*] \mathbf{x} \right\}. \quad (5.32)$$

Realizando então uma mudança de variável, $y = \mathbf{w}^H \mathbf{x}$, na Equação (5.32), muda-se também a variável no operador esperança. Considerando então a situação ótima, ou seja, gradiente nulo, tem-se:

$$\begin{aligned} \mathbb{E}_X \left\{ \mathbb{E}_{\mathfrak{A}} \left\{ \psi(y, \mathbf{a}) \right\} [\mathbf{x}^H \mathbf{w} - \mathbf{a}^*] \mathbf{x} \right\} &= 0 \\ \mathbb{E}_X \left\{ \mathbb{E}_{\mathfrak{A}} \left\{ \psi(y, \mathbf{a}) \right\} \mathbf{x}^H \mathbf{w} \mathbf{x} \right\} &= \mathbb{E}_X \left\{ \mathbb{E}_{\mathfrak{A}} \left\{ \psi(y, \mathbf{a}) \right\} \mathbf{a}^* \mathbf{x} \right\}. \end{aligned} \quad (5.33)$$

Verificando ainda, através da Equação (5.31), que [Sala-Alvarez & Vázquez-Grau, 1997]

$$\mathbb{E}_{\mathfrak{A}} \left\{ \psi(y, \mathbf{a}) \right\} = 1, \quad (5.34)$$

a Equação (5.33) reduz-se a

$$\begin{aligned} \mathbb{E}_X \left\{ \mathbf{x} \mathbf{x}^H \mathbf{w} \right\} &= \mathbb{E}_X \left\{ \mathbf{a}^* \mathbf{x} \right\} \\ \mathbb{E}_X \left\{ \mathbf{x} \mathbf{x}^H \right\} \mathbf{w} &= \mathbb{E}_X \left\{ \mathbf{a}^* \mathbf{x} \right\}. \end{aligned} \quad (5.35)$$

No contexto de Wiener o sinal \mathbf{a} possui as características do sinal desejado, através do qual se faz a supervisão no processo de filtragem. Assim, pode-se escrever [Haykin, 1996]

$$\begin{aligned} \mathbf{R}_x \mathbf{w} &= \mathbf{p} \\ \mathbf{w} &= \mathbf{R}_x^{-1} \mathbf{p}, \end{aligned} \quad (5.36)$$

em que \mathbf{p} é o vetor de correlação cruzada entre o sinal desejado e o sinal recebido e \mathbf{R}_x é a matriz de autocorrelação do sinal recebido.

Desta forma, mostra-se que a anulação do gradiente do FPC leva ao mesmo vetor de coeficientes do filtro que a solução de Wiener no caso supervisionado.

Uma outra equivalência diz ainda respeito a aspectos estruturais da função custo.

Como ilustrado na Seção 5.2.1, o critério FP apresenta alguns pontos de *mínimos espúrios*. Este mesmo comportamento é observado no critério de *Decisão Dirigida* (DD), o qual apresenta a seguinte função custo [Haykin, 1996; Ding & Li, 2001]:

$$J_{\text{DD}}(\mathbf{w}) = \mathbb{E} \left\{ (y(n) - \text{dec}[y(n)])^2 \right\}, \quad (5.37)$$

em que $\text{dec}[\cdot]$ é a função de decisão do sinal², implementada após o equalizador para recuperar

²Na literatura, é comum encontrar-se a função de decisão também denominada de *quantizador* ou *slicer*.

um sinal com as mesmas características do sinal transmitido.

Para ilustrar a comparação entre os critérios DD e o FP, são mostradas, na Figura 5.7, as funções custo para o canal $H_{MA}(z)$, modulação BPSK e $\text{SNR} = 10$ dB. Para o FPC é também utilizado $\sigma_r^2 = 0.1$.

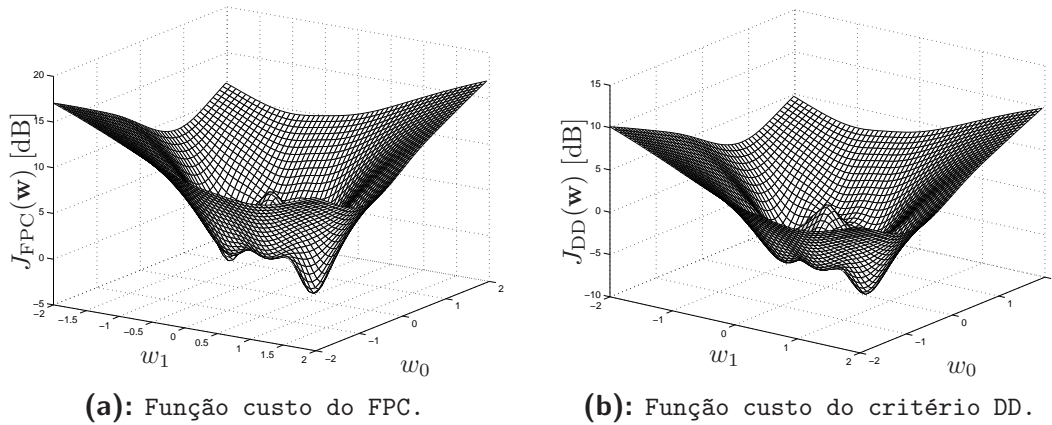


Figura 5.7: Comparação das funções de custo dos critérios DD e FP.

A Figura 5.8 ilustra as curvas de nível, em escala logarítmica, para as funções custo anteriores.

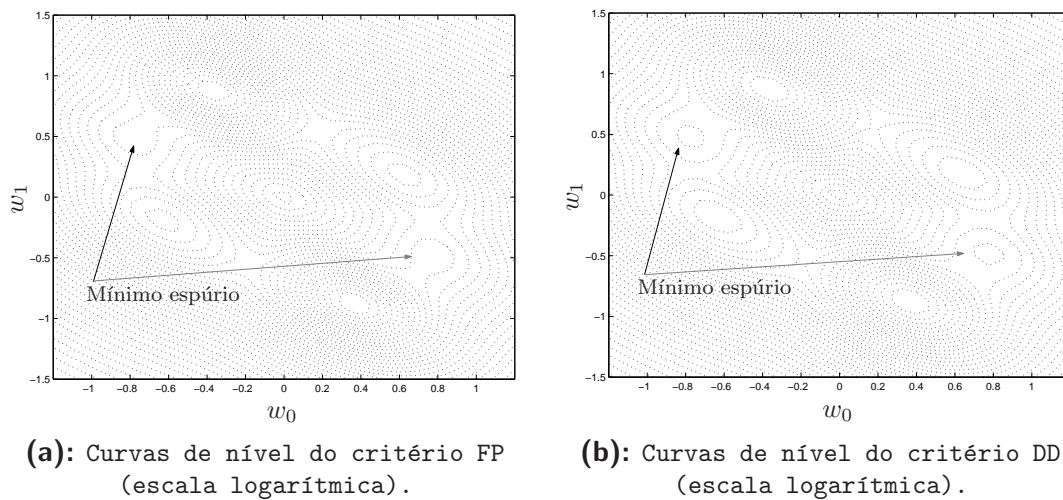


Figura 5.8: Comparação das curvas de nível dos critérios DD e FP.

Como pode ser observado nas curvas de nível, todos os mínimos de ambas as funções custo

são os mesmos. Além disso, também a função custo de ambos os critérios apresenta a mesma geometria, a menos de um ganho. Isto levanta alguns aspectos interessantes de discussão.

É sabido que o algoritmo DD converge para a solução MMSE, ou de Wiener, se as seguintes condições são satisfeitas [Macchi & Eweda, 1984; Haykin, 1996; Brossier, 1997]:

1. o diagrama de olho encontra-se aberto;
2. o fator de passo do algoritmo é fixo;
3. a seqüência das observações na saída do canal é ergódica.

Embora as condições 2 e 3 são de fácil consideração, a condição 1 coloca-se como um grande problema para técnicas de equalização cega, já que não há garantia da condição de “olho aberto”.

Entretanto, conforme verificado em [Cavalcante, 2001; Cavalcante et al., 2002a], o FPA não apresenta problemas de convergência para os casos em que a situação de olho aberto ainda não está configurada.

Surgem com isso duas perguntas importantes:

- (1) como dois critérios tão diferentes apresentam a mesma geometria para seu funcional?
- (2) já que os dois critérios apresentam a mesma geometria da função de custo, por que esta diferença na capacidade de rastreamento de canais com a condição de “olho fechado”?

A resposta para ambas está num ponto fundamental da concepção de ambos os critérios: *as estatísticas de ordem superior*.

Como vimos, o FPC é concebido para maximizar a similaridade entre a fdp do sinal na saída do equalizador e um modelo paramétrico que incorpora todas as características estatísticas do sinal desejado. Ou seja, ele visa estimar através do modelo paramétrico a fdp do sinal desejado e forçar que o sinal na saída do equalizador tenha a mesma fdp do sinal desejado.

Assim, como está sendo realizada uma estimativa de fdp, **todas** as estatísticas do sinal estão sendo utilizadas para a sua representação.

No caso do critério DD, a análise segue o mesmo caminho. O critério DD tenta minimizar uma “distância” entre o sinal na saída do equalizador e uma transformação não-linear que recupera a característica do sinal originalmente transmitido. O ponto chave é entender o comportamento de tal transformação.

A função não-linear de decisão é capaz de fazer uma transformação de um sinal que não tem uma fdp conhecida para outro sinal que tem **todas** as características e, por conseqüência, todas as estatísticas do sinal desejado. Isto se deve ao fato de que a quantização do sinal permite

a recuperação de um conjunto de dados com as mesmas características do sinal transmitido, entre elas, entropia limitada (visto que o número de possibilidades é finito) e valores discretos. Com isso, a densidade de probabilidade dos sinais transmitidos e após o dispositivo de decisão são as mesmas.

Desta maneira, ambos os critérios conservam a mesma idéia:

► *Minimizar a diferença entre todas as estatísticas do sinal na saída do equalizador e as do sinal desejado.*

Logo, analisando-se pelo ponto de vista de equalização de estatísticas, os critérios DD e FP são equivalentes. Isto explica porque os critérios apresentam a mesma geometria da função custo em função dos parâmetros do equalizador.

A outra questão, referente à capacidade de tratar canais na condição de “olho fechado”, deve-se a outro aspecto intrínseco da estrutura das funções não-lineares empregadas.

Uma maneira fácil de visualizar esta diferença é reescrever a função custo do critério DD como um critério que minimiza a probabilidade do $y(n)$ ser o símbolo transmitido, dado que a decisão é dada por $\text{dec}[y(n)]$, ou seja,

$$J_{\text{DD}}(\mathbf{w}) = \arg \min_{\mathbf{w}} [p(\mathbf{a}) - p(y(n) | \text{dec}[y(n)])]. \quad (5.38)$$

Desta forma, o critério DD pode ser expresso como uma diferença de fdps para facilitar a comparação com o FPC.

O grande diferencial dos dois critérios está justamente na *confiabilidade* da transformação linear. O critério DD utiliza como estimativa apenas a decisão da amostra temporal. Se esta decisão estiver errada, o filtro é adaptado sem nenhuma medida da confiabilidade de tal transformação.

O FPC por sua vez considera uma medida de confiabilidade das decisões serem feitas para cada um dos símbolos do alfabeto da modulação. Isto pode ser visto através da Equação (5.32) em que a função

$$\psi(y, \mathbf{a}) = \frac{A \cdot \exp\left(-\frac{|y-\mathbf{a}|^2}{\sigma_r^2}\right)}{E_{\mathfrak{A}} \left\{ A \cdot \exp\left(-\frac{|y-\mathbf{a}|^2}{\sigma_r^2}\right) \right\}},$$

é uma medida de probabilidade de proximidade de y dos possíveis valores de \mathbf{a} . Convém observar que tal função é bastante similar à medida de informação extrínseca empregada como medida de confiabilidade de decisão nos algoritmos de decisão suave [Hagenauer & Hoehner, 1989; Kühn, 1999].

Desta forma, dois aspectos muito importantes decorrem do descrito acima; não só a equivalência entre os critérios FP e o DD como também uma análise do critério DD, sob a ótica da teoria da informação, ainda não apresentada na literatura.

Na seqüência, são ainda abordados outros pontos importantes sobre o FPC.

5.2.3 Alguns outros aspectos relevantes

Um ponto de extrema relevância é a existência de um critério semelhante na literatura. A proposta apresentada em [Sala-Alvarez & Vázquez-Grau, 1997] também tem como meta forçar a fdp do sinal na saída do equalizador ao mais próximo possível da solução desejada. Vale ressaltar que, embora a proposição do critério/ algoritmo FP e o critério descrito em [Sala-Alvarez & Vázquez-Grau, 1997] tenham sido realizadas de maneira independente, há várias semelhanças entre os dois. Ambos utilizam como meta a obtenção de sinais que obedeçam a uma densidade de probabilidade adequada. A maior diferença entre eles é na obtenção do algoritmo decorrente do critério. Enquanto o FPC considera o valor instantâneo para a aproximação estocástica, o algoritmo em [Sala-Alvarez & Vázquez-Grau, 1997] necessita de um bloco de amostras para estimação dos valores médios das grandezas envolvidas e uma função de regeneração para fornecer uma estimativa recursiva da fdp dos dados, considerando que o ruído é gaussiano. Além disto, [Sala-Alvarez & Vázquez-Grau, 1997] considera que a escolha do parâmetro σ_r^2 não tem impacto considerável no desempenho do algoritmo, o que difere totalmente das conclusões sobre o FPC. Embora as estruturas de ambos os critérios sejam bastante similares, os algoritmos derivados de cada um deles é bastante diferente.

De fato, critérios que têm por base a otimização da entropia, através de estruturas auto-organizáveis, apresentam um termo relativo à variância da variável cuja entropia deseja-se otimizar. Dentre outras questões, a escolha do parâmetro σ_r^2 implica num melhor controle da otimização da entropia, realizada através de uma regra de aprendizado anti-Hebbiana [Schraudolph, 1995].

Esta observação vem ao encontro do fato de que o parâmetro σ_r^2 atua como um “regulador” do fator de passo, já que este é normalizado por $\frac{1}{\sigma_r^2}$, conforme pode ser observado nas Equações (5.12) e (5.13). Assim, o fator de passo efetivo é dado por

$$\mu_{\text{efet}} = \frac{\mu}{\sigma_r^2}.$$

Outra análise, previamente realizada em [Cavalcante, 2001] e posteriormente em [Cavalcante et al., 2002a], ilustra a importância de σ_r^2 quando compara-se o FPC com critérios clássicos de equalização.

Novamente, para possibilitar uma análise comparativa entre diferentes critérios, faz-se apelo à teoria da classificação [Duda & Hart, 1973]. Ao tentar-se unificar as abordagens, uma das maneiras passíveis de utilização é encontrar uma função de referência que preserve as características do sinal. No caso do FPC esta função de referência é o modelo paramétrico $\Psi(y)$. Desta forma, pode-se utilizar funcionais que são decorrentes da aplicação da medida de similaridade entre funções, dada pela Equação (5.7), e verificar as diferenças entre as funções de referência.

Então, para critérios como o CM e o critério de Sato, pode-se escrever as seguintes funções paramétricas que têm o papel da função de referência:

- $f_{\text{CM}}(y, \mathbf{w})$;
- $f_{\text{Sato}}(y, \mathbf{w})$.

Logo, utilizando o mesmo critério dado pela Equação (5.7) tem-se, para o critério CM, que:

$$J_{\text{CM}}(\mathbf{w}) = -\mathbb{E} \{ \ln [f_{\text{CM}}(y, \mathbf{w})] \}, \quad (5.39)$$

que sabe-se ser igual a

$$J_{\text{CM}}(\mathbf{w}) = \mathbb{E} \left\{ (|y|^2 - \rho_2)^2 \right\}. \quad (5.40)$$

Igualando-se então as Equações (5.39) e (5.40), tem-se a seguinte função de referência:

$$f_{\text{CM}}(y, \mathbf{w}) = -\exp \left[(|y|^2 - \rho_1)^2 \right]. \quad (5.41)$$

Para o critério de Sato, tem-se que [Haykin, 1996]:

$$J_{\text{Sato}}(\mathbf{w}) = \mathbb{E} \left\{ (\rho_1 \cdot \text{sign}[y] - y)^2 \right\}, \quad (5.42)$$

em que $\rho_1 = \frac{\mathbb{E} \{ a^2(n) \}}{\mathbb{E} \{ a(n) \}}$. Logo, através do mesmo desenvolvimento realizado para o critério CM, a função de referência para o critério de Sato é dada por:

$$f_{\text{CM}}(y, \mathbf{w}) = -\exp \left[(\rho_1 \cdot \text{sign}[y] - y)^2 \right]. \quad (5.43)$$

A partir das Equações (5.41) e (5.43), pode-se ver que as funções de referência desses algoritmos clássicos são casos particulares do FPC, uma vez que as funções apresentam o mesmo tipo de estrutura (exponencial). A diferença está no número de *kernels*³ que são considerados e como o sinal de saída é ponderado na função. No caso do FPC, o número de

³*Kernel* é uma notação estatística para a função de base radial gaussiana, que possui as mesmas características da função exponencial.

$kernel$ s é igual ao número de símbolos do alfabeto da modulação e as características do sinal são extraídas a partir da utilização da função exponencial e também do valor quadrático. No caso dos critérios CM e de Sato, o número de $kernel$ s é menor que o número de símbolos e as informações do sinal são extraídas, respectivamente, pelo valor quadrático e módulo do sinal de saída. Esta diferença entre o tipo de característica do sinal na saída do equalizador que cada critério considera será discutida em maiores detalhes no capítulo seguinte.

Para efeito de diferenciação entre os critérios, comparando-se as funções de referência dos critérios CM e de Sato com a função de referência $\Phi(y, \mathbf{w}, \sigma_r^2)$ do FPC, observa-se que este último possui um grau de liberdade a mais que os outros critérios. Algumas simulações comparativas dos valores logarítmicos da função de referência $\Phi(y, \mathbf{w}, \sigma_r^2)$ para vários valores de σ_r^2 , e das funções f_{CM} e f_{Sato} , para uma modulação BPSK, podem ser visualizadas na Figura 5.9.

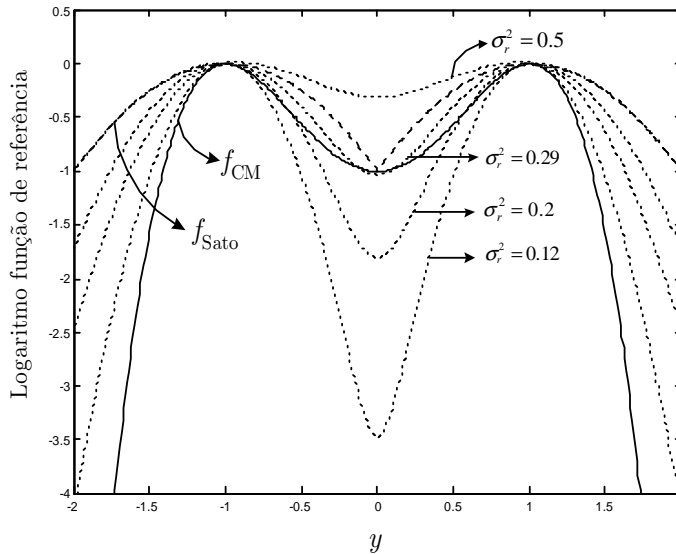


Figura 5.9: Comparação entre funções de referência para os critérios de Sato, CM e FP.

Evidencia-se assim mais uma vez a importância do parâmetro σ_r^2 para o critério FPC. Com a variação do parâmetro é possível atingir o desempenho dos critérios CM e de Sato, bem como melhorar a própria recuperação dos sinais através de uma seleção adequada do valor de σ_r^2 .

Uma questão interessante de ser ao menos mencionada é a possibilidade da verificação de uma relação do FPC com o critério CM por uma abordagem indireta. Isto é possível devido à existência de um limite superior do funcional de Wiener com o critério CM, como mostrado em [Suyama, 2003]. Entretanto, este tópico não está no escopo da tese e é deixado apenas como indicação para verificação futura.

Dentre as várias características do FPC, uma das mais destacadas se relaciona a sua capacidade de forçar a saída do equalizador a ter a mesma fdp de um sinal idealmente equalizado. Deste modo, pode-se enxergar o critério como sendo uma estimativa da densidade de probabilidade desejada, o que permite utilizar algumas técnicas estatísticas para tal finalidade. Este tópico é abordado na seção seguinte.

5.3 Estimação de densidade de probabilidade

O princípio de recuperação de sinais através da estimativa da densidade de probabilidade do sinal transmitido remonta ao Teorema de Benveniste-Goursat-Rouget (BRG) [Benveniste et al., 1980], o qual pode ser enunciado da seguinte forma:

Teorema BGR: *Seja um sinal cuja distribuição de probabilidade é não-gaussiana, transmitido através de um canal linear. O sinal na saída de um equalizador linear será aquele da entrada do sistema, se as distribuições da entrada e da saída do sistema forem iguais.*

Prova: Ver Apêndice B.

Este teorema baseou o estudo de equalização cega durante vários anos e levantou questões práticas sobre a factibilidade da equalização cega, uma vez que a estimação da fdp de um sinal era realizada através da estimativa dos infinitos cumulantes do sinal, necessários para compor a distribuição de probabilidade. Estes cumulantes, por sua vez, são de difícil e alta complexidade de estimação, o que dificulta bastante um método de equalização baseado na estimativa de todos os momentos de ordem superior.

Nos últimos anos, técnicas baseadas na estimativa explícita da fdp do sinal transmitido têm sido propostas com a finalidade de evitar a alta complexidade computacional necessária para estimativa dos cumulantes envolvidos.

Sendo uma área de bastante interesse na estatística aplicada, vários métodos podem ser listados como ferramentas de identificação da distribuição de probabilidade de sinais envolvidos no processamento da informação.

Os tipos de estimativa de fdps são classificados em dois grandes grupos:

- paramétricas;
- não-paramétricas.

Métodos paramétricos são utilizados quando se conhece a estrutura da fdp que se deseja estimar. Neste caso, apenas os parâmetros do modelo previamente adotado são estimados, uma vez que todo o restante da estrutura já é informado *a priori*, restando apenas fornecer corretamente a posição e o escalonamento da fdp.

Já na estimação não-paramétrica, não apenas os parâmetros são indeterminados mas também a estrutura dos dados cuja fdp deseja-se estimar. Além disso, como há mais graus de liberdade na estimação, a representação da densidade de probabilidade dos dados é assegurada com uma maior probabilidade. Isto porque o número de classes de distribuições que podem ser representadas por uma estimativa não-paramétrica é maior que com estimativas paramétricas [Wegman, 1972].

Vários métodos de estimação não-paramétrica podem ser encontrados na literatura, como por exemplo em [Silverman, 1986; Devroye, 1987; Lindsey, 1996] e também nos tutoriais [Zabin & Wright, 1994b, 1994a] e nas suas referências internas. Um método que ganha bastante destaque é a estimação através de *kernels* [Phillips, 1999].

A idéia por trás da estimação de *kernels* é a de utilizar uma generalização por meio de funções para a estimativa de fdp através de um histograma. O histograma é uma das ferramentas de estimação de fdp mais simples e, embora tenha sido desenvolvido em 1895 [Devroye, 1987], permanece em uso bastante freqüente até a atualidade. O método de estimativa consiste em dividir os dados em janelas pequenas, chamadas *bin*, representando para cada *bin* o número de dados naquele intervalo. Matematicamente, tem-se

$$\hat{p}(x) = \frac{1}{N} \cdot \frac{\text{número de dados no bin}}{\text{largura do bin}} ,$$

em que N é o número total de dados.

O método de estimativa por *kernels* tenta generalizar a estimativa utilizando uma função que permite tomar as amostras em diferentes intervalos, o que não é possível através do uso de um histograma.

Os trabalhos iniciais utilizando-se desta estratégia de estimação originaram-se na década de 1950 [Phillips, 1999], mas ainda de uma forma pouco pretensiosa, apenas buscando uma melhoria da estimativa através do histograma. A estimativa era realizada segundo a seguinte equação [Phillips, 1999]:

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{l} \cdot f\left(\frac{x - X_i}{l}\right), \quad (5.44)$$

em que a função de ponderação é dada por

$$f(x) = \begin{cases} \frac{1}{2} & \text{para } |x| < 1 \\ 0 & \text{caso contrário} \end{cases} . \quad (5.45)$$

Desta forma, é possível ter estimativas mais suaves que aquelas fornecidas pelo histograma.

Embora este tipo de estratégia tenha apresentado ganhos relativos ao uso de histogramas, foi somente com o trabalho de Parzen que muitas outras funções foram discutidas e utilizadas para estimação não-paramétrica. Este tópico é abordado na seqüência.

5.3.1 Estimação pelo método de Parzen

Em 1962, Parzen desenvolveu o formalismo da possibilidade de representação de uma fdp através da combinação de funções capazes de representar as principais características da fdp envolvida [Parzen, 1962]. O trabalho de Parzen, baseou-se na sua própria experiência de métodos de estimação de densidade espectral de séries temporais estacionárias.

O estimador de fdp de Parzen, para um conjunto de N pontos, é definido como:

$$\hat{p}(x) = \hat{p}(x|x_1, \dots, x_N) \triangleq \frac{1}{N \cdot l(N)} \sum_{i=1}^N K\left(\frac{x - x_i}{l(N)}\right), \quad (5.46)$$

em que $K(\cdot)$ é um *kernel*⁴ e l é um fator de suavização. Para que $\hat{p}(x)$ seja uma fdp válida, o *kernel* e o fator de suavização devem respeitar as seguintes condições [Parzen, 1962]:

$$\begin{aligned} l(N) &> 0 \\ K(u) &\geq 0 \\ \int_{-\infty}^{\infty} K(u) du &= 1 \\ |u \cdot K(u)| &\rightarrow 0, \quad \text{quando } |u| \rightarrow \infty \\ \int_{-\infty}^{\infty} u^2 K^2(u) du &< \infty \end{aligned} \quad (5.47)$$

Desta forma, a variância e a polarização (bias) do estimador são dados por [Laster, 1997;

⁴A título de curiosidade, vale mencionar que no seu trabalho original Parzen não cunhou o termo *kernel*, mas usou o termo função $K(x)$. Posteriormente, alguns autores utilizaram os termos “função de ponderação” e “função *kernel*” [Wegman, 1972; Silverman, 1986; Phillips, 1999].

Parzen, 1962]:

$$\begin{aligned} \text{Bias} [\hat{p}(x)] &= -p''(x) \frac{l^2(N)}{2} \int_{-\infty}^{\infty} u^2 K^2(u) du \\ \text{Variância} [\hat{p}(x)] &= \frac{p(x)}{N \cdot l(N)} \int_{-\infty}^{\infty} \infty K^2(u) du, \end{aligned} \quad (5.48)$$

para valores grandes de N e $p(x)$ é a verdadeira fdp. A função de suavização é escolhida de forma a respeitar

$$\begin{aligned} l(N) &\rightarrow 0, \quad \text{para } N \rightarrow \infty \\ N \cdot l(N) &\rightarrow \infty, \quad \text{para } N \rightarrow \infty, \end{aligned} \quad (5.49)$$

o que garante que o estimador é assintoticamente não-polarizado e a variância do estimador tende a zero [Parzen, 1962].

Logicamente, a escolha de l e $K(\cdot)$ influencia diretamente o desempenho do estimador, uma vez que algumas funções são mais adequadas para certos tipos de dados. Desta forma, uma classe de *kernels* pode ser estabelecida, com suas respectivas funções de suavização, conforme apresentado em [Parzen, 1962, pág. 1068] e [Phillips, 1999, pág. 22].

Entretanto, a função gaussiana normalizada é a mais utilizada como *kernel*, sendo mesmo denominada de *kernel de Parzen*. Com isso, uma escolha razoável para o *kernel* e a respectiva função de suavização é

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \quad (5.50a)$$

$$l(N) = \frac{1}{\sqrt{N}} \quad (5.50b)$$

Este *kernel*, além de respeitar as condições da Equação (5.47), apresenta-se bastante adequado para a utilização em sistemas de comunicação digital, nos quais são considerados ruído gaussiano e sistema linear. Logo, o método de estimação de Parzen torna-se adequado para a estimação da densidade de probabilidade de vários sinais de um sistema de comunicação digital, cuja fdp é geralmente uma combinação de gaussianas [Cavalcante, 2001].

Com isso, para o caso de um sistema de comunicação digital, utilizando o método de Parzen, o *kernel* seria dado pela função gaussiana e a função de suavização a própria variância de cada gaussiana.

Desta maneira, pode-se verificar que o FPC é um caso particular da estimação de Parzen, no qual utiliza-se como medida de similaridade entre a verdadeira densidade e sua estimativa a divergência de Kullback-Leibler. Outros métodos também utilizam a estimação pelo método de Parzen para realizar a equalização. Este é o tópico do item seguinte.

5.3.2 Critérios baseados em métodos de estimação de fdp

Grande parte dos critérios para recuperação de sinais propostos na literatura tiveram como meta inicial a melhoria da velocidade de convergência de abordagens clássicas como o CMA.

Desta forma, métodos de otimização dos parâmetros do equalizador utilizam-se da estimação da fdp para tornar o sinal na saída do filtro equalizador com uma distribuição a mais “próxima” possível da distribuição de probabilidade dos dados na entrada do sistema, ou com pouco erro de estimação. Isto implica em respeitar o Teorema BGR e, conseqüentemente equalizar.

Entretanto, o algoritmo proposto em [Benveniste et al., 1980] busca a equalização dos cumulantes que compõem a fdp do sinal, tornando o processo computacionalmente complexo.

Com o intuito de diminuir esta complexidade, algumas técnicas são propostas na utilização de métricas do erro de estimação que preservam a informação estatística da função de densidade de probabilidade sem necessitar da estimação direta dos cumulantes.

De uma maneira geral, o erro de estimação é medido através da distância \mathbb{L}_p , quando outras medidas, como por exemplo a KLD, não são utilizadas. A distância \mathbb{L}_p é definida como [Devroye, 1987]:

$$\mathbb{L}_p(\hat{p}, p) \triangleq \left[\int_{-\infty}^{\infty} |\hat{p}(u) - p(u)|^p du \right]^{\frac{1}{p}}, \quad 0 < p < \infty. \quad (5.51)$$

A escolha da melhor métrica é feita geralmente de acordo com o tipo de dados disponíveis, pois isto tem grande impacto no tipo de estimativa a ser realizada. Na maior parte dos trabalhos na literatura em estatística, encontram-se valores medidos em relação aos erros \mathbb{L}_1 e \mathbb{L}_2 , mas ainda assim alguns aspectos devem ser observados. Por exemplo, medidas baseadas em erros tipo \mathbb{L}_1 são melhores para estimação de “caudas” de fdp que medidas baseadas em erros do tipo \mathbb{L}_2 . A seguir, alguns critérios que utilizam medidas do tipo \mathbb{L}_2 são apresentados.

► Distância quadrática

Adotando uma medida do tipo \mathbb{L}_2 , em [Santamaría et al., 2002] é proposto um método que utiliza como critério uma medida quadrática entre a fdp da potência do sinal transmitido e a fdp da potência do sinal na saída do equalizador.

O critério resultante é dado por [Santamaría et al., 2002]

$$J_{\text{DQ}}(\mathbf{w}) = \int_{-\infty}^{\infty} [p_{Y^2}(u) - p_{\mathfrak{A}^2}(u)]^2 du, \quad (5.52)$$

em que DQ significa *distância quadrática* e Y^2 e \mathfrak{A}^2 representam, respectivamente, as variáveis da potência dos sinais na saída do equalizador e dos sinais transmitidos.

Utilizando-se o *kernel* de Parzen para estimar a fdp de ambos os sinais, tem-se então

$$\hat{p}_{Y^2}(u) = \frac{1}{N} \sum_{i=0}^{N-1} K_{\sigma_0}(u - |y(n-i)|^2) \quad (5.53a)$$

$$\hat{p}_{\mathfrak{A}^2}(u) = \frac{1}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} K_{\sigma_0}(u - |\mathbf{a}_i|^2), \quad (5.53b)$$

em que $K_{\sigma_0}(u) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{|u|^2}{2\sigma_0^2}\right)$.

Sabendo que para *kernels* gaussianos

$$\int_{-\infty}^{\infty} K_{\sigma_0}(u - \alpha_1) K_{\sigma_0}(u - \alpha_2) du = K_{\sigma_0\sqrt{2}}(\alpha_1 - \alpha_2), \quad (5.54)$$

utilizando-se as estimativas das fdps e substituindo pelas verdadeiras densidades na Equação (5.52), tem-se a seguinte função custo [Santamaría et al., 2002]:

$$\begin{aligned} J_{\text{DQ}}(\mathbf{w}) &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} K_{\sigma} (|y(n-j)|^2 - |y(n-i)|^2) + \frac{1}{\mathfrak{C}^2} \sum_{i=1}^{\mathfrak{C}} \sum_{j=1}^{\mathfrak{C}} K_{\sigma} (|\mathbf{a}_j|^2 - |\mathbf{a}_i|^2) \\ &\quad - \frac{2}{N \cdot \mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} \sum_{j=0}^{N-1} K_{\sigma} (|y(n-j)|^2 - |\mathbf{a}_i|^2), \end{aligned} \quad (5.55)$$

em que, por motivos de simplicidade, denota-se $\sigma_0\sqrt{2}$ como σ . Assim, o gradiente estocástico para o critério DQ é dado, para o caso de $N = 2$, por [Santamaría et al., 2002]

$$\begin{aligned} \nabla J(\mathbf{w}) &= \frac{1}{2} K'_{\sigma} (|y(n)|^2 - |y(n-1)|^2) [y(n)\mathbf{x}^*(n) - y(n-1)\mathbf{x}^*(n-1)] \\ &\quad - \frac{1}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} [K'_{\sigma} (|y(n)|^2 - |\mathbf{a}_i|^2) y(n)\mathbf{x}^*(n) + K'_{\sigma} (|y(n-1)|^2 - |\mathbf{a}_i|^2) y(n-1)\mathbf{x}^*(n-1)], \end{aligned} \quad (5.56)$$

em que $K'(\cdot)$ é a derivada do *kernel* $K(\cdot)$.

► *Sampled pdf fitting (SPF)*

Este critério proposto em [Lázaro et al., 2003, em ICA '2003] simplifica o critério DQ tentando equalizar a fdp da potência dos sinais envolvidos somente através das medidas obtidas nos instantes de amostragem. Desta forma, considerando-se N_s pontos de amostragem, pode-se escrever a função de otimização como:

$$J_{\text{SPF}}(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{N_s} [p_{Y^2}(u_i) - T_i]^2, \quad (5.57)$$

em que cada valor T_i corresponde à fdp da potência do sinal em u_i , ou seja, $T_i = p_{\mathfrak{X}^2}(u_i)$.

Com esta função custo, a meta é igualar a distribuição $p_{\mathfrak{X}^2}(u)$ com $p_{Y^2}(u)$ nos pontos mais representativos.

Ainda é necessário estimar a densidade nos pontos T_i . Isto é realizado com o estimador de Parzen, obtendo-se [Lázaro et al., 2003, em ICA '2003]:

$$T_i = \frac{1}{\mathfrak{C}} \sum_{j=1}^{\mathfrak{C}} K_{\sigma}(u_i - |\mathbf{a}_i|^2), \quad (5.58)$$

o que resulta na seguinte função custo

$$J_{\text{SPF}}(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\frac{1}{N} \sum_{j=0}^{N-1} K_{\sigma}(u_i - |y(n-j)|^2) - T_i \right]. \quad (5.59)$$

A versão do critério SPF foi desenvolvida considerando apenas $N = 1$, ou seja, a estimação de Parzen da Equação (5.46) é realizada apenas com uma amostra. Assim, são obtidas as seguintes estimativas para as amostras da fdp e para a saída do equalizador [Lázaro et al., 2003, em ICA '2003]:

$$\hat{p}_{Y^2}(u) = K_{\sigma_0}(u_i - |y(n-i)|^2) \quad (5.60a)$$

$$T_i = K_{\sigma}(u_i - |\mathbf{a}_i|^2) = K_{\sigma}(0). \quad (5.60b)$$

Com isso, o termo de adaptação estocástica é dada por [Lázaro et al., 2003, em ICA '2003]:

$$\nabla J_{\text{SPF}}(\mathbf{w}) = -\frac{2}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} [K_{\sigma}(u_i - |y(n)|^2) - K_{\sigma}(0)] \cdot K'_{\sigma}(u_i - |y(n)|^2) y(n) \mathbf{x}^*(n). \quad (5.61)$$

► *Matched pdf* (MP)

Ainda na linha de tentar uma simplificação do critério DQ, em [Lázaro et al., 2003, em *ICASSP'2003*] é feita uma abordagem igual à realizada para derivação do critério FP, no que diz respeito a medida de similaridade de funções.

Considerando o critério DQ e expandindo-se o termo quadrático, tem-se:

$$J_{\text{DQ}}(\mathbf{w}) = \int_{-\infty}^{\infty} [p_{Y^2}^2(u) + p_{\mathfrak{A}^2}^2(u) - 2 \cdot p_{Y^2}(u) \cdot p_{\mathfrak{A}^2}(u)] du.$$

A simplificação, a exemplo da derivação do critério FP, consiste em observar que a otimização como medida de similaridade das fdps das potências dos sinais é dada pelo termo cruzado. Logo, a função custo a ser minimizada é [Lázaro et al., 2003, em *ICASSP'2003*]:

$$J_{\text{MP}}(\mathbf{w}) = \int_{-\infty}^{\infty} p_{Y^2}^2(u) \cdot p_{\mathfrak{A}^2}^2(u) du. \quad (5.62)$$

Tomando-se as estimativas de $p_{Y^2}(u)$ e $p_{\mathfrak{A}^2}(u)$ através do método de Parzen, obtém-se a seguinte função custo [Lázaro et al., 2003, em *ICASSP'2003*]:

$$J_{\text{MP}}(\mathbf{w}) = \frac{1}{N \cdot \mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} \sum_{j=0}^{N-1} K_{\sigma} [|y(n-j)|^2 - |\mathbf{a}_i|^2]. \quad (5.63)$$

Da Equação (5.63) resulta a seguinte expressão para o gradiente estocástico, considerando $N = 1$, [Lázaro et al., 2003, em *ICASSP'2003*]:

$$\nabla J_{\text{MP}}(\mathbf{w}) = \frac{1}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} K'_{\sigma} (|y(n)|^2 - |\mathbf{a}_i|^2) \cdot y(n) \cdot \mathbf{x}^*(n). \quad (5.64)$$

Embora alguns trabalhos possam ser listados como pertencentes à classe de abordagens baseadas em estimação da fdp, como por exemplo [Adalı, Liu, & Sönmez, 1997], não há, de acordo com os conhecimentos do autor, critérios com estruturas lineares baseados em abordagens de equalização cega para processamento multiusuário. Nesta direção, a próxima seção é dedicada à proposta de uma família de critérios para separação cega de fontes utilizando o critério FP.

5.4 Família de critérios multiusuário baseados na estimação da fdp

Ao se conceber critérios para processamento multiusuário é necessário ter em mente quais são as condições necessárias para se garantir a separação dos sinais. Estas condições são fornecidas pelo Teorema de Shalvi-Weinstein, às quais adiciona-se a condição para que as fontes recuperadas sejam diferentes entre si. Na Seção 4.5.4 (pág. 106), quando discutidas as motivações para o critério MUK, tais condições foram listadas. A título de clareza estas condições são aqui listadas novamente. Sendo as fontes $a_i(n)$ e as estimativas de fontes $y_i(n)$, tem-se:

C_{RS1}. $a_l(n)$ é i.i.d. e de média zero ($l = 1, \dots, K$);

C_{RS2}. $a_l(n)$ e $a_q(n)$ são estatisticamente independentes para $l \neq q$ e têm a mesma fdp;

C_{RS3}. $|\mathcal{K}[y_l(n)]| = |\mathcal{K}_a|$ ($l = 1, \dots, K$);

C_{RS4}. $\mathbb{E}\{|y_l(n)|^2\} = \sigma_a^2$ ($l = 1, \dots, K$);

C_{RS5}. $\mathbb{E}\{y_l(n)y_q^*(n)\} = 0$, $l \neq q$.

Novamente, vale mencionar que as Condições C_{RS1}-C_{RS4} derivam diretamente do Teorema SW e a Condição C_{RS5} refere-se à necessidade de se obter a estimativa de *todas* as fontes e, para isso, força-se a ortogonalidade entre elas.

Na verdade, as condições anteriores são explicitadas em termo da maximização da *kurtosis* que norteia a derivação do algoritmo de SW.

Entretanto, o Teorema de SW é bem mais abrangente que o contexto de maximização da *kurtosis* mostrado. Embora a maximização da *kurtosis* seja suficiente para garantir a recuperação do sinal, sob a restrição da equalização da variância, **qualquer** momento de ordem superior poderia ser empregado [Lii & Rosenblatt, 1982]. Na verdade, de acordo com o Teorema BGR, **todos** os momentos de ordem superior poderiam ser empregados para obtenção da recuperação dos sinais, embora isto não seja necessário.

Isso significa que qualquer estratégia que garanta a remoção da interferência pode ser utilizada, desde que a equalização de um ou mais cumulantes seja garantida.

Nesta tese, são utilizados critérios que têm por base a estimativa da fdp de um sinal idealmente equalizado, utilizando para tal fim o critério FP. A garantia de que o critério respeita a condição de remoção da interferência é dada pelo fato de que *todos* os momentos são equalizados, uma vez que a estimação de duas fdps é realizada. Desta forma, o critério respeita não só o Teorema SW, mas também o Teorema BGR, garantindo com isto a recuperação do

sinal.

O ponto remanescente sobre as condições de recuperação de sinais trata da garantia de que as fontes obtidas são diferentes entre si. A maneira como esta condição é assegurada é o assunto dos próximos itens da presente seção.

5.4.1 Utilização da descorrelação explícita

Uma vez que a Condição C_{RS5} exige que as fontes apresentem saídas descorrelacionadas, a primeira derivação de um algoritmo multiusuário baseou-se na utilização de um critério auxiliar para garantir a descorrelação explícita, tal como proposto no MU-CMA em [Papadias & Paulraj, 1997] e discutido na Seção 4.5.2.

Como o critério de descorrelação explícita força as diferentes saídas de um sistema de separação a apresentarem descorrelação entre si, as Condições C_{RS1} - C_{RS5} estão compreendidas no critério.

Desta forma, o critério multiusuário aqui proposto, baseado na estimativa da fdp, e que utiliza a descorrelação explícita como critério adicional é dado, para o k -ésimo usuário, por:

$$J_{\text{MU-FPC}}(\mathbf{w}_k) = J_{\text{FPC}}(\mathbf{w}_k) + \gamma \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K |r_{ij}|^2, \quad (5.65)$$

em que γ é o fator de regularização do termo de descorrelação e r_{ij} é o elemento (i, j) da matriz de autocorrelação das saídas dos filtros de separação. O critério dado pela Equação (5.65) é denominado *Multiuser Fitting pdf Criterion* (MU-FPC).

Assim, o gradiente estocástico do MU-FPC, para o k -ésimo filtro, é dado por:

$$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) - \mu \nabla J_{\text{FPC}}(\mathbf{w}_k) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \hat{r}_{ik}(n) \mathbf{p}_i(n), \quad (5.66)$$

em que $\nabla J_{\text{FPC}}(\mathbf{w}_k)$ é dado pela Equação (5.12), $\hat{r}_{ik}(n)$ é o elemento (i, k) da matriz $\mathbf{R}_{yy}(n)$ e $\mathbf{p}_i(n)$ é a i -ésima coluna da matriz $\mathbf{P}(n)$, que podem ser estimadas através de suas médias temporais como

$$\begin{aligned} \mathbf{R}_{yy}(n+1) &= \xi \mathbf{R}_{yy}(n) + (1 - \xi) \mathbf{y}(n) \mathbf{y}^H(n) \\ \mathbf{P}(n+1) &= \xi \mathbf{P}(n) + (1 - \xi) \mathbf{x}(n) \mathbf{y}^H(n), \end{aligned} \quad (5.67)$$

nas quais ξ é um fator de esquecimento.

O algoritmo estocástico do MU-FPC é denominado de *Multiuser Fitting pdf Algorithm* (MU-FPA).

Vale ressaltar que nas Equações (5.65)-(5.67) não se considera a ISI, ou seja, apenas processamento espacial é realizado. Nos casos em que ocorre ISI o critério torna-se [Cavalcante et al., 2002b]:

$$J_{\text{MU-FPC}}(\mathcal{W}_k) = J_{\text{FPC}}(\mathcal{W}_k) + \gamma \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} |r_{ij}(\ell)|^2, \quad (5.68)$$

em que $r_{ij}(\ell) = \mathbb{E} \{y_i(n)y_j^*(n-\ell)\}$ é a correlação cruzada entre os sinais à saída dos filtros espaço-temporais de índices i e j , com intervalo de tempo ℓ e sendo $\frac{\Delta}{2}$ o atraso máximo para o qual os sinais dos diversos usuários devem estar descorrelacionados, conforme o modelo descrito nas Seções 4.3 e 4.5.

Com isso, o MU-FPA para o caso espaço-temporal pode ser dado pelas seguintes equações:

$$\mathcal{W}_k(n+1) = \mathcal{W}_k(n) - \mu \nabla J_{\text{FPC}}(\mathcal{W}_k) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \hat{r}_{ik,\ell}(n) \hat{\mathbf{p}}_{i,\ell}(n) \quad (5.69)$$

$$\mathbf{R}_{y,\ell}(n+1) = \xi \mathbf{R}_{y,\ell}(n) + (1-\xi) \mathbf{y}(n) \mathbf{y}^H(n-\ell) \quad (5.70)$$

$$\mathbf{P}_\ell(n+1) = \xi \mathbf{P}_\ell(n) + (1-\xi) \mathcal{X}(n) \mathbf{y}^H(n-\ell) \quad (5.71)$$

$$\mathbf{y}(n-\ell) = [y_1(n-\ell) \quad \cdots \quad y_K(n-\ell)]^T \quad (5.72)$$

$$\ell = -\frac{\Delta}{2}, \dots, \frac{\Delta}{2}, \quad (5.73)$$

em que $\hat{r}_{ik,\ell}(n)$ é o termo de correlação cruzada entre os i -ésimo e j -ésimo usuários com atraso ℓ e corresponde ao (i, j) -ésimo elemento da matriz $\mathbf{R}_{y,\ell}(n)$, sendo $\hat{\mathbf{p}}(i, \ell)(n)$ a i -ésima coluna da matriz $\mathbf{P}_\ell(n)$

O MU-FPA para os casos espacial e espaço-temporal pode ser resumido nas Tabelas 5.1(a) e 5.1(b).

(a): Processamento espacial com MU-FPA.	(b): Processamento espaço-temporal com MU-FPA.
<p>(1). Inicialize as matrizes \mathbf{R}_{yy}, \mathbf{P} e os vetores \mathbf{w}_k</p> <p>(2). Para $n > 0$</p> <p>(3). Calcule matrizes de correlação</p> $\begin{aligned}\mathbf{R}_{yy}(n+1) &= \xi \mathbf{R}_{yy}(n) \\ &\quad + (1-\xi) \mathbf{y}(n) \mathbf{y}^H(n) \\ \mathbf{P}(n+1) &= \xi \mathbf{P}(n) \\ &\quad + (1-\xi) \mathbf{x}(n) \mathbf{y}^H(n)\end{aligned}$ <p>(4). Para $k = 1 : K$, calcule</p> $\begin{aligned}\mathbf{w}_k(n+1) &= \mathbf{w}_k(n) - \mu \nabla J_{\text{FPC}}(\mathbf{w}_k) \\ &\quad - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \hat{r}_{ik}(n) \mathbf{p}_i(n),\end{aligned}$ <p>em que</p> $\begin{aligned}\nabla J_{\text{FPC}}(\mathbf{w}_k) &= \\ \frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{ y_k(n) - \mathbf{a}_i ^2}{2\sigma_r^2}\right) [y_k(n) - \mathbf{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{ y_k(n) - \mathbf{a}_i ^2}{2\sigma_r^2}\right)} \mathbf{x}(n)\end{aligned}$ <p>e $\hat{r}_{ik}(n)$ é o elemento (i, k) da matriz $\mathbf{R}_{yy}(n)$ e $\mathbf{p}_i(n)$ é a i-ésima coluna de $\mathbf{P}(n)$.</p> <p>(5). Retorne ao passo 4</p> <p>(6). Retorne ao passo 2</p>	<p>(1). Inicialize as matrizes $\mathbf{R}_{y,\ell}$, \mathbf{P}_ℓ e os vetores \mathcal{W}_k</p> <p>(2). Para $n > 0$</p> <p>(3). Para $\ell = -\frac{\Delta}{2} : \frac{\Delta}{2}$, calcule</p> $\begin{aligned}\mathbf{R}_{yy}(n+1) &= \xi \mathbf{R}_{yy}(n) \\ &\quad + (1-\xi) \mathbf{y}(n) \mathbf{y}^H(n) \\ \mathbf{P}(n+1) &= \xi \mathbf{P}(n) \\ &\quad + (1-\xi) \mathbf{x}(n) \mathbf{y}^H(n)\end{aligned}$ <p>(4). Para $k = 1 : K$, calcule</p> $\begin{aligned}\mathcal{W}_k(n+1) &= \mathcal{W}_k(n) - \mu \nabla J_{\text{FPC}}(\mathcal{W}_k) \\ &\quad - \gamma \sum_{\substack{i=1 \\ i \neq k}}^K \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \hat{r}_{ik,\ell}(n) \hat{\mathbf{p}}_{i,\ell}(n),\end{aligned}$ <p>em que</p> $\begin{aligned}\nabla J_{\text{FPC}}(\mathcal{W}_k) &= \\ \frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{ y_k(n) - \mathbf{a}_i ^2}{2\sigma_r^2}\right) [y_k(n) - \mathbf{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{ y_k(n) - \mathbf{a}_i ^2}{2\sigma_r^2}\right)} \mathcal{X}(n)\end{aligned}$ <p>e $\hat{r}_{ik,\ell}(n)$ é o elemento (i, k) da matriz $\mathbf{R}_{y,\ell}(n)$ e $\mathbf{p}_{i,\ell}(n)$ é a i-ésima coluna de $\mathbf{P}_\ell(n)$.</p> <p>(5). Retorne ao passo 4</p> <p>(6). Retorne ao passo 3</p> <p>(7). Retorne ao passo 2</p>

Tabela 5.1: Resumo do *Multiuser Fitting pdf Algorithm* para os casos espacial e espaço-temporal.

Embora a decorrelação explícita implemente a Condição CRS5 , seu desempenho apresenta alguns problemas relativos a sua estrutura. O termo de decorrelação influi bastante no desempenho em regime permanente do algoritmo devido ao fator de regularização γ , o qual apresenta um forte compromisso entre erro em estado permanente e número de usuários perdidos.

A Figura 5.10 ilustra a dependência do termo de decorrelação em relação ao número médio de usuários perdidos. Para avaliar tal situação foram utilizadas 20 simulações independentes nas quais o número de usuários perdidos era contabilizado. Os usuários apresentavam modulações QPSK de módulo unitário. A figura de mérito utilizada foi o erro do módulo constante (CME, *Constant Modulus Error*), definido, para o k -ésimo usuário, como [Cavalcanti, 1999]:

$$\text{CME}_k(n) = (|y_k(n)| - \rho_2)^2. \quad (5.74)$$

Este parâmetro indica quão próximo de um sinal de módulo constante encontra-se a saída de cada um dos filtros do sistema de separação.

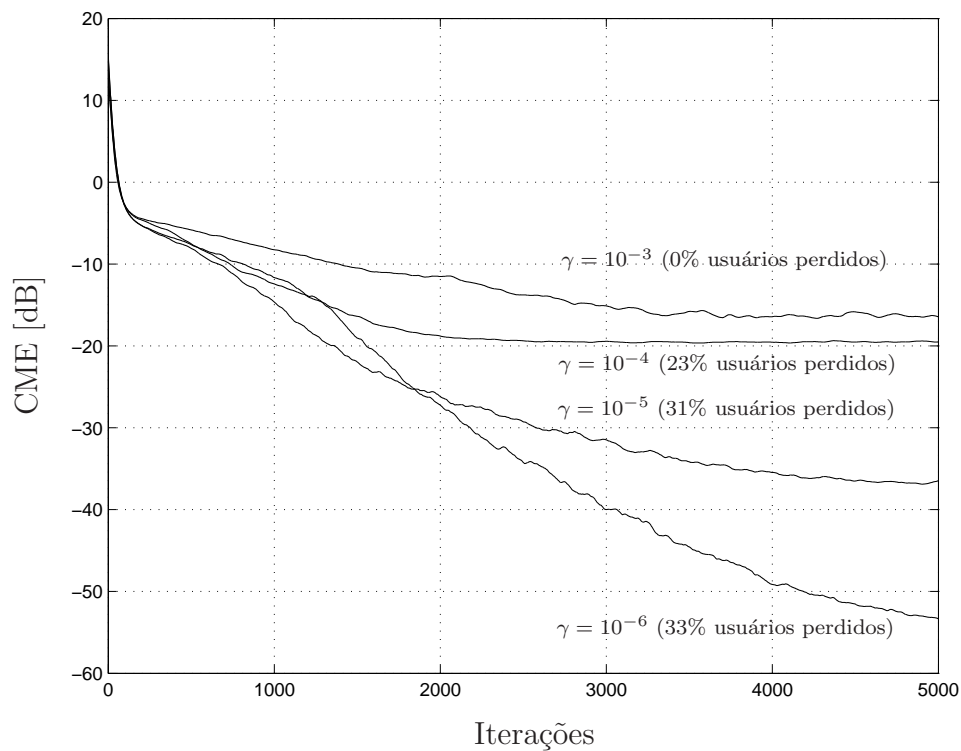


Figura 5.10: Desempenho do MU-FPA em relação ao fator de regularização γ e número de usuários perdidos.

Desta maneira, uma nova estratégia de implementação da Condição C_{RS5} é realizada para evitar o problema de perda de usuários com valores de erro em regime baixo. Este é o assunto da próxima seção.

5.4.2 Critério com restrição

Uma alternativa para respeitar a Condição C_{RS5} é utilizando restrições. Com a inserção de uma restrição sobre as saídas do sistema de separação, o problema recai em como utilizar de forma adaptativa esta restrição.

Em [Papadias, 2000a, 2000b] a proposta do critério MUK, descrito na Seção 4.5.4, introduz a restrição sobre a separação de fontes de tal forma que a matriz de resposta global de um sistema seja ortogonal.

Sob esta restrição, as saídas do sistema de separação serão diferentes uma vez que a adaptação não é permitida para soluções nas quais um ou mais usuários são perdidos.

Novamente, uma vez que o critério FP fornece a mitigação da interferência, é possível deduzir um critério sob a restrição da ortogonalidade da matriz da resposta combinada. Logo, o seguinte critério multiusuário é possível:

$$\left\{ \begin{array}{l} \min_{\mathbf{W}} J_{FP}(\mathbf{W}) = \sum_{i=1}^K D(p_Y(y_i) \|\Phi(y_i)) \\ \text{sujeito a: } \mathbf{G}^H \mathbf{G} = \mathbf{I} \end{array} \right., \quad (5.75)$$

em que

$$\begin{aligned} \mathbf{G} &= \mathbf{W}^H(n) \mathbf{H} = [\mathbf{g}_1 \mid \cdots \mid \mathbf{g}_K] \\ &= \begin{bmatrix} g_{11} & \cdots & g_{K1} \\ \vdots & \ddots & \vdots \\ g_{1K} & \cdots & g_{KK} \end{bmatrix}_{K \times K}. \end{aligned} \quad (5.76)$$

Este novo critério, proposto na forma da Equação (5.75), é denominado ***Multiuser Constrained Fitting pdf Criterion*** (MU-CFPC).

Na verdade, o fato de se escrever a parte de retirada da interferência como sendo a KLD entre a densidade de probabilidade e o modelo paramétrico explicita a estimação da fdp, uma vez que a KLD atinge o mínimo se e somente se as funções (fdp e modelo paramétrico) forem iguais. Embora seja um abuso de notação, não representa uma incoerência, já que a minimização do critério em relação aos parâmetros dos filtros de equalização/separação é dada por parte da equação da KLD, o outro termo, que corresponde à entropia de y , representa somente uma normalização da variável quando a KLD atinge seu mínimo [Bishop, 1995]. Então, através do mesmo procedimento do critério MUK, pode-se utilizar uma ortogonalização de Gram-Schmidt para se garantir a restrição do algoritmo.

Um outro aspecto importante a ser considerado é a introdução de uma notação para os coeficientes da matriz de separação para cada etapa do processamento, tal como realizado

para o algoritmo MUK:

- etapa de equalização – \mathbf{W}^e
- etapa de separação – \mathbf{W} .

Assim, a adaptação na etapa de equalização deste novo critério da Equação (5.75) é baseada no gradiente estocástico do FPA, Equação (5.12), enquanto que a etapa de separação se dá pela ortogonalização da matriz \mathbf{W}^e , por meio do algoritmo de Gram-Schmidt. Isto resulta nas seguintes equações para cada usuário k [Cavalcante et al., 2003]:

$$\mathbf{w}_k^e(n+1) = \mathbf{w}_k(n) - \mu \left(\frac{\sum_{i=1}^{\mathfrak{c}} \exp\left(-\frac{|y_k(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right) [y_k(n) - \mathbf{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{c}} \exp\left(-\frac{|y_k(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n) \right) \quad (5.77a)$$

$$\mathbf{w}_1(n+1) = \frac{\mathbf{w}_1^e(n+1)}{\|\mathbf{w}_1^e(n+1)\|} \quad (5.77b)$$

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1) \mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1)}{\left\| \mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1) \mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1) \right\|}, \quad (5.77c)$$

em que \mathbf{w}_k^e e \mathbf{w}_k são as k -ésimas colunas das matrizes \mathbf{W}^e e \mathbf{W} , respectivamente.

O algoritmo descrito pela Equação (5.77) é denominado *Multiuser Constrained Fitting pdf Algorithm* (MU-CFPA) e pode ser resumido na Tabela 5.2.

(1). Inicialize $\mathbf{W}(0)$ e $\mathbf{W}^e(0)$

(2). Para $n > 0$

(3). Para $k = 1 : K$, calcule

$$\mathbf{w}_k^e(n+1) = \mathbf{w}_k(n) - \mu \left(\frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y_k(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right) [y_k(n) - \mathbf{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y_k(n) - \mathbf{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n) \right)$$

(4). Retorne ao passo 3

(5). Calcule $\mathbf{w}_1(n+1) = \frac{\mathbf{w}_1^e(n+1)}{\|\mathbf{w}_1^e(n+1)\|}$

(6). Para $k = 2 : K$, calcule

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1)}{\left\| \mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} [\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)] \mathbf{w}_l(n+1) \right\|}$$

(7). Retorne ao passo 6

(8). Retorne ao passo 2

Tabela 5.2: Algoritmo MU-CFPA.

A Figura 5.11 ilustra o ganho obtido pela utilização de um processo de ortogonalização ao invés da decorrelação explícita para os algoritmos multiusuário baseados no FPC. As curvas são obtidas através da média de 50 realizações independentes e uma $\text{SNR} = 30$ dB.

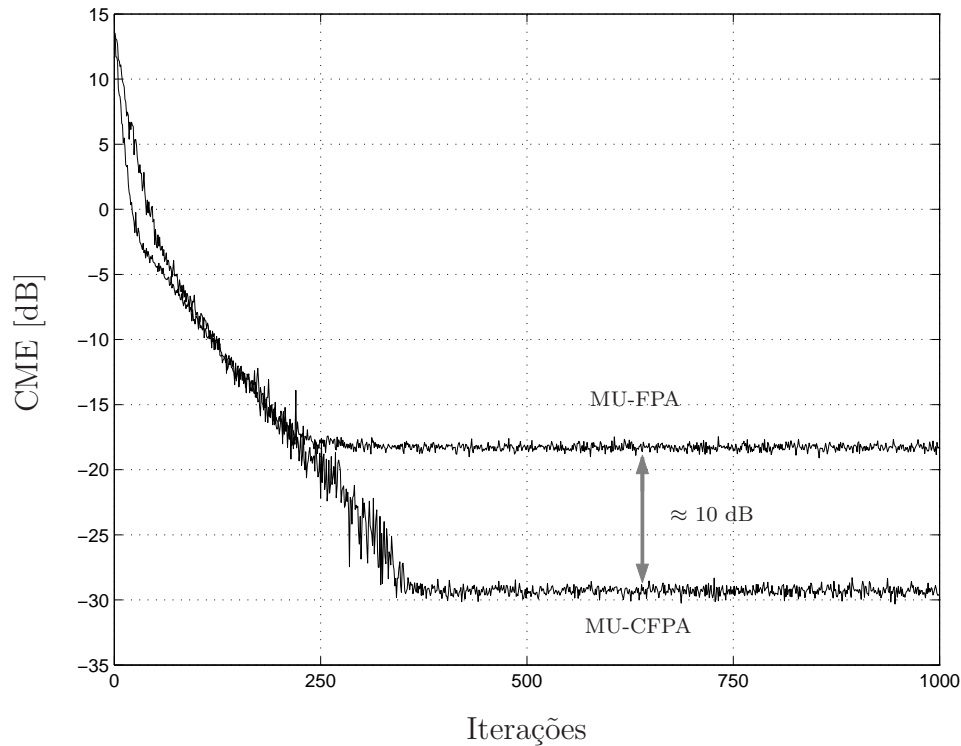


Figura 5.11: Comparação entre os algoritmos MU-FPA e MU-CFPA.

Pode-se observar um ganho em torno de 10 dB somente pelo uso do processamento com restrição e não da decorrelação explícita. Esta diferença é devida à existência do compromisso entre o erro em regime e o número de usuários perdidos.

Embora o MU-CFPA apresente vantagens se comparado com sua versão com decorrelação explícita, ele não possui uma versão para processamento espaço-temporal. Isto porque, no caso de haver ISI, a matriz de separação global é dada por um tensor de ordem 3 (sensores \times usuários \times tempo) e a ortogonalização de um tensor é uma questão bastante complexa, que ainda não apresenta uma versão adaptativa, ao contrário da ortogonalização de Gram-Schmidt para matrizes.

Embora as comparações entre os diversos algoritmos sejam apresentadas no Capítulo 7, vale ilustrar uma comparação entre o MU-CFPA e o MUK, para que alguns aspectos relevantes dos dois algoritmos sejam observados e permitam a análise a ser realizada no Capítulo 6.

Seja um cenário bastante simples, com dois usuários com modulação QPSK. O sistema possui como canal uma matriz de mistura instantânea com coeficientes aleatórios que respeitam a normalização do canal, ou seja, $\mathbf{H}^H \mathbf{H} = \mathbf{I}$. Este cenário é o mesmo descrito em [Papadias, 2000a, 2000b]. Assim, o desempenho dos algoritmos MUK e MU-CFPA podem ser observados,

em termos do CME, na Figura 5.12. As curvas foram obtidas através de 10 simulações independentes.

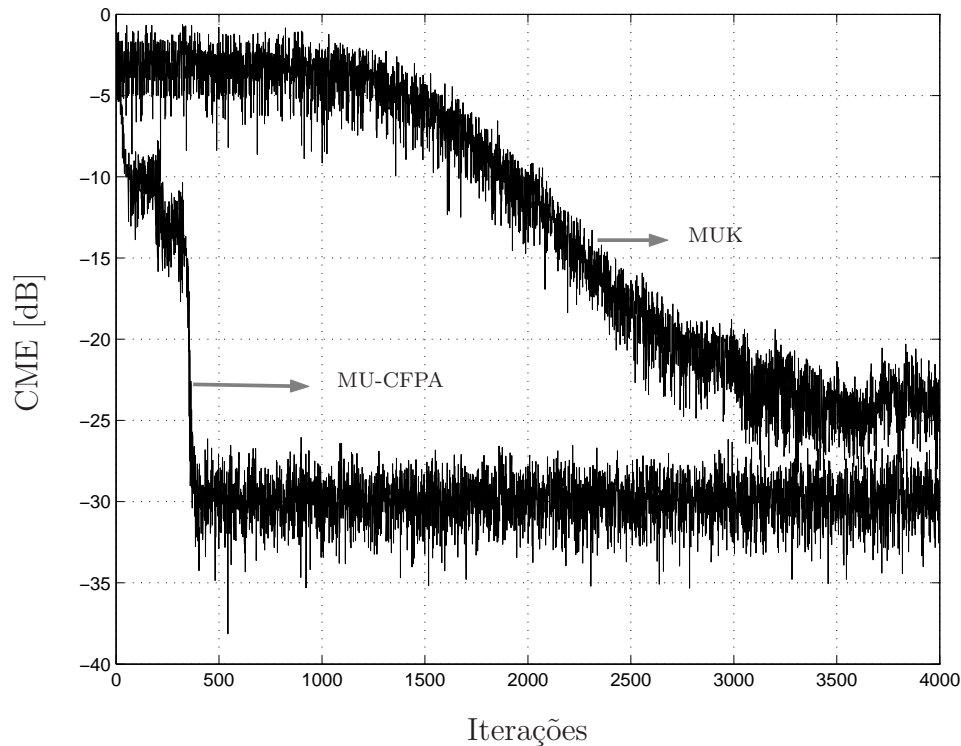


Figura 5.12: Comparação em termos de CME entre os algoritmos MUK e MU-CFPA.

Como pode ser observado, o MU-CFPA converge muito mais rápido que o MUK, embora os dois utilizem o mesmo procedimento de decorrelação entre as fontes. Fica a pergunta de qual diferença entre eles permite um ganho tão considerável em termos de taxa de convergência. A resposta para esta pergunta será dada no Capítulo 6.

Outra abordagem que poderia levar a um novo critério de processamento multiusuário é a *Matched pdf*. Entretanto, uma análise da função custo do critério MP mostrou que seu comportamento dificulta a convergência para um dos pontos de equilíbrio. Esta mesma verificação foi feita através da avaliação de uma função custo FP modificada, na qual não é considerado o logaritmo para compor a métrica de similaridade. Esta função custo é denominada $J_{\text{FPC modif}}$.

A Figura 5.13 ilustra as funções custo para o critério MP e FP modificado. Os parâmetros utilizados para a geração das curvas foram os mesmos dos casos anteriores (SNR = 10 dB, modulação BPSK), sendo ainda $\sigma_r^2 = 0,1$ para o FP modificado e K_σ com variância de 0,1 para o critério MP.

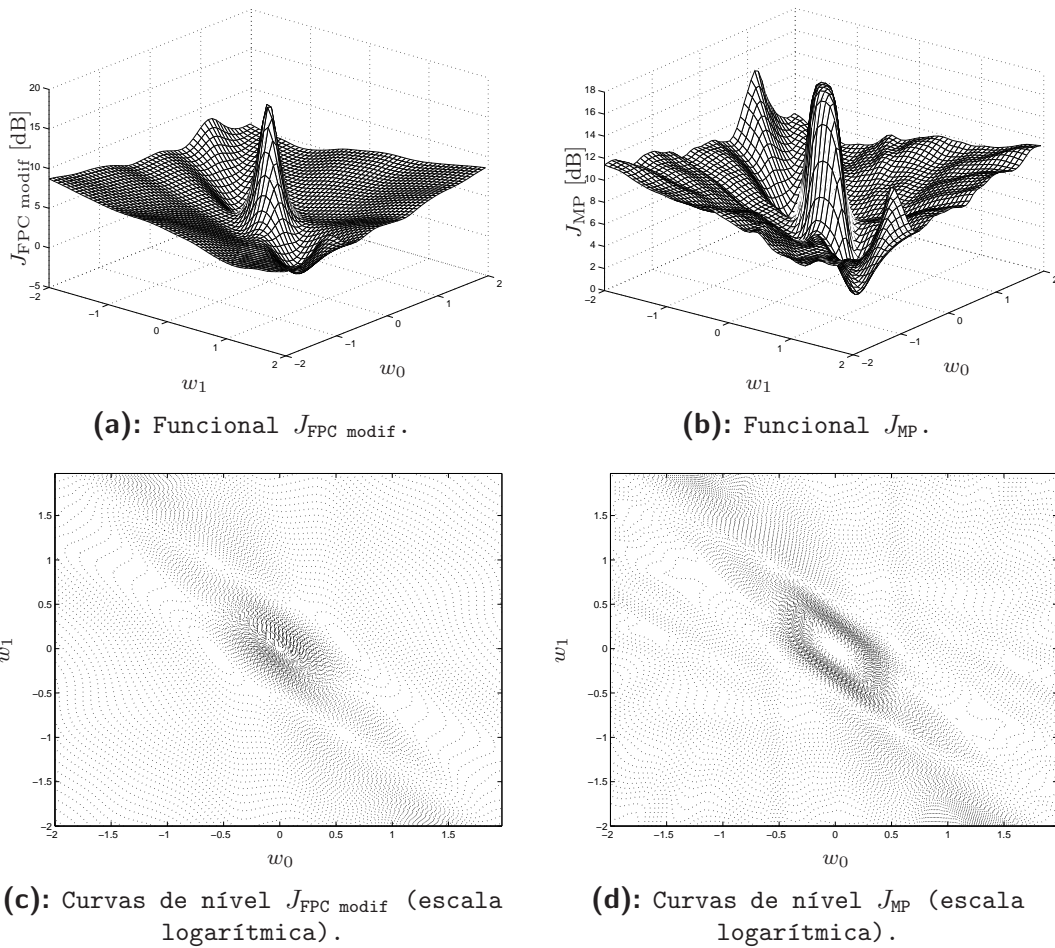


Figura 5.13: Análise das funções de custo e curvas de nível dos critérios MP e FP modificado.

Deve-se observar que os funcionais apresentam uma geometria que dificulta a derivação de um algoritmo baseado no gradiente descendente, visto que os funcionais apresentam vários mínimos locais e uma geometria (inclinação do plano), conforme pode ser observado nas figuras acima, que desfavorece a busca utilizando-se um algoritmo estocástico. Em [Cavalcante, 2001] este fato motivou a utilização do logaritmo para a concepção do critério FP.

Uma outra importante observação, que fica bastante clara a partir do critério MU-FPC na Equação (5.75), é que ele pode ser visto como uma medida de log-verossimilhança entre as saídas da matriz de separação e o modelo paramétrico adotado para os sinais que se deseja obter. Esta observação pode ser facilmente verificada pela comparação entre a função contraste do critério MV para BSS na Equação (3.9) e o critério da etapa de equalização.

No caso do critério MV para BSS, conforme mencionado no Capítulo 3, as estratégias sofrem do problema de um provável desvio quanto ao modelo adotado para as fontes. Considerando o processamento multiusuário, tal indeterminação não ocorre já que as fontes são consideradas como possuindo a mesma fdp. Isto facilita bastante o critério MV e torna necessário apenas uma diferenciação entre as fontes na saída do sistema de separação, conforme expresso pela Condição C_{RS5} .

Visualizando-se o critério como uma medida de MV, pode-se também tratar o MU-CFPA como um critério que maximiza a função de contraste Ψ_{ICA} , definida na Equação (2.43), uma vez que a maximização da função contraste do critério MV corresponde à maximização, em relação ao modelo das fontes, da função contraste de ICA [Cardoso, 1998; Kofidis, 2001].

De forma indireta, esta relação de log-verossimilhança permite uma analogia com uma grande gama de critérios de BSS/ICA, mas uma visão unificada dos critérios de BSS e processamento multiusuário não está no escopo desta tese.

Entretanto, outras relações com critérios importantes em processamento multiusuário podem ser deduzidas. A seção seguinte trata de uma destas relações.

5.5 Critério para minimização da probabilidade de erro

Em sistemas práticos de comunicação, o critério que geralmente rege o projeto e a implementação de um receptor é a taxa de erro de bit. Assim, é interessante encontrar critérios que também minimizem a probabilidade de erro para a concepção do receptor.

Entretanto, a derivação de critérios que levem em conta a minimização da probabilidade de erro são, geralmente, computacionalmente complexos e realizados de forma supervisionada, uma vez que é necessária uma estimação da fdp do sinal transmitido.

Na literatura, alguns trabalhos foram dedicados a avaliar o desempenho de receptores que utilizam a minimização da probabilidade de erro.

Em [Laster, 1997], é proposto um receptor que minimiza a BER, utilizando uma seqüência de 1's para estimar a fdp condicional do sinal recebido através do estimador de Parzen. O tamanho da seqüência de treinamento deve ser suficiente para garantir uma estimativa não-polarizada. Neste caso, o estimador da BER é expresso pela função de erro complementar $Q(\cdot)$ [Verdú, 1998]. Seu resultado é então utilizado por um conjunto de demoduladores atuando em diversidade para recuperar o sinal enviado.

Na mesma linha, [Phillips, 1999] utiliza o estimador de Parzen para construir uma

estimativa adaptativa da BER sem utilizar uma seqüência particular capaz de fornecer a fdp do sinal recebido. Neste caso, a estimativa também é calculada com auxílio da função $Q(\cdot)$.

Similarmente, [Yeh & Barry, 2000] propõem um algoritmo adaptativo capaz de minimizar a BER de maneira direta quando o número de coeficientes não é suficiente. A BER é estimada também no espaço dos parâmetros do equalizador, mas diretamente a partir da função $Q(\cdot)$, sem utilizar um estimador nem uma seqüência de dados específica, como em [Laster, 1997]. Uma versão gradiente-Newton do algoritmo em [Yeh & Barry, 2000] é proposta em [Lamare & Sampaio-Neto, 2002], resultando em ganhos de desempenho em relação a suas versões do tipo LMS.

Aplicações de DS-CDMA são também contempladas pelos algoritmos de minimização da BER. Por exemplo, [Chen et al., 2001] propõem o uso de uma variante do algoritmo de [Yeh & Barry, 2000] aplicado ao problema de detecção multiusuário na presença de multipercursos.

Num trabalho bastante recente [Hjørungnes et al., 2004], é investigada a otimização conjunta do transmissor e receptor para atingir a mínima probabilidade de erro de bit em um sistema MIMO com dispersão temporal. O procedimento é realizado através da consideração de restrições nos sinais no transmissor e receptor para remover a interferência introduzida por um canal MIMO conhecido.

Mas uma questão que se coloca de forma bastante incômoda é a necessidade de supervisão (seqüências de treinamento) para a derivação dos algoritmos de minimização da probabilidade de erro, uma vez que precisa-se de uma estimativa da densidade de probabilidade do sinal recebido. Isto parece sugerir a impossibilidade, ou pelo menos uma grande dificuldade, da utilização de critérios cegos que busquem a minimização da probabilidade de erro em um sistema de múltiplos usuários.

Nesta linha, é interessante analisar o critério MAP (*Maximum a Posteriori*), que também minimiza a probabilidade de erro, para verificar a possibilidade de uma equivalência entre critérios cegos e o critério de minimização da probabilidade de erro.

O critério MAP pode ser escrito, para o k -ésimo usuário, como [Proakis, 1995]:

$$J_{\text{MAP}}(\mathbf{w}_k) = \mathbb{E} \{ \ln [p(\mathbf{a}_i | y)] \}, \quad (5.78)$$

em que $y = \mathbf{w}^H \mathbf{x}$, considerando-se uma medida logarítmica por simplicidade de cálculo [Benedetto et al., 1987]. Desta forma, o critério maximiza a probabilidade de ter sido enviado um sinal \mathbf{a}_i dado que na saída do equalizador foi observado o sinal y .

Através da regra de Bayes, pode-se escrever a probabilidade *a posteriori* como [Proakis,

1995]:

$$p(\mathbf{a}_i | y) = \frac{p(y | \mathbf{a}_i) \cdot p(\mathbf{a}_i)}{\sum_{i=1}^c p(y | \mathbf{a}_i) \cdot p(\mathbf{a}_i)} \quad (5.79)$$

Deve-se notar que o denominador da Equação (5.79) corresponde ao modelo paramétrico considerado no critério FP, ponderado pelas probabilidades dos sinais \mathbf{a}_i , dado na Equação (5.29), desde que a escolha de σ_r^2 seja adequada. Além disso, a Equação (5.79) corresponde à função auxiliar da Equação (5.31), a qual representa uma medida da confiabilidade da decisão de um símbolo, como já mencionado anteriormente.

Desta forma, pode-se escrever o critério MAP como sendo:

$$\begin{aligned} \mathbb{E} \{ \ln [p(\mathbf{a}_i | y)] \} &= \mathbb{E} \left\{ \ln \left[\frac{p(y | \mathbf{a}_i) \cdot p(\mathbf{a}_i)}{\Phi(y)} \right] \right\} \\ J_{\text{MAP}} &= \mathbb{E} \{ \ln [p(y | \mathbf{a}_i)] \} - \underbrace{\mathbb{E} \{ \ln [\Phi(y)] \}}_{J_{\text{FPC}}}. \end{aligned} \quad (5.80)$$

Neste ponto, algumas considerações devem ser feitas para permitir alguma análise. A densidade condicional $p(y | \mathbf{a}_i)$ diz respeito ao modelo assumido do sinal na saída do filtro de separação. Como está sendo considerado um sinal idealmente equalizado na presença de ruído gaussiano, tem-se que:

$$p(y | \mathbf{a}_i) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \exp\left(-\frac{|y - \mathbf{a}_i|^2}{2\sigma_\vartheta^2}\right). \quad (5.81)$$

Com isso, pode-se reescrever a Equação (5.80) como

$$\begin{aligned} J_{\text{MAP}} &= -\frac{1}{2\sigma_\vartheta^2} \mathbb{E} \{ |y - \mathbf{a}_i|^2 \} + \ln \left[\frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \right] + J_{\text{FPC}} \\ J_{\text{FPC}} - J_{\text{MAP}} &= \frac{1}{2\sigma_\vartheta^2} \mathbb{E} \{ |y - \mathbf{a}_i|^2 \} - \ln \left[\frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \right], \end{aligned} \quad (5.82)$$

na qual o termo do lado direito da equação é o critério MMSE, conforme mencionado na Equação (5.27). Assim, pode-se escrever a seguinte relação do critério MAP:

$$J_{\text{MAP}} = J_{\text{FPC}} - J_{\text{MMSE}}. \quad (5.83)$$

A relação entre os critérios mostra algo bastante interessante. Quando há sinal de treinamento disponível, o critério FP é igual ao critério MMSE, segundo a Equação (5.27),

o que anula o valor do critério MAP, mostrando que neste caso a probabilidade de erro foi minimizada. Entretanto, atingir a mínima probabilidade de erro através da minimização do critério MMSE só é factível quando o equalizador possui uma ordem suficiente para que toda a ISI seja eliminada e toda a probabilidade de erro se deva ao ruído gaussiano [Yeh & Barry, 2000]. Na prática isto não é possível, pois tal efeito só é obtido com equalizadores de comprimento duplamente infinito.

Quando não há sinal de treinamento disponível, o critério MMSE não pode ser otimizado e a otimização do critério FP fornece uma estimativa polarizada da probabilidade de erro.

Além disso, como os critérios são definidos como funções positivas, pode-se ainda escrever a seguinte inequação para o FPC e o MAP:

$$J_{\text{FPC}} \geq J_{\text{MAP}}, \quad (5.84)$$

o que mostra que atingir o mínimo para J_{FPC} não necessariamente implica atingir mínima probabilidade de erro.

Assim, minimizar a probabilidade de erro requer a disponibilidade de informação *a priori* da fonte para que o valor atingido seja o menor possível. Na ausência deste, uma estimativa polarizada de tal probabilidade de erro será o melhor resultado obtido.

O impacto desta conclusão para o problema de separação de fontes, utilizando algoritmos que busquem minimizar a probabilidade de erro, faz com que *alguma* informação disponível sobre as fontes seja necessária para que a separação seja factível. De fato, observando a Equação (3.20), a minimização da probabilidade de erro requer alguma informação \mathcal{I} , para que as hipóteses adotadas para o modelo do sistema de fontes e de mistura sejam suficientes para atingir a separação de sinais.

Além disto, é mostrada uma relação de um critério cego baseado na estimativa da fdp do sinal com critérios tipicamente supervisionados, levantando a questão da possibilidade de uma abordagem por teoria da informação do critério MAP.

5.6 Conclusões e síntese

Este capítulo é dedicado à proposta de uma família de critérios de processamento multiusuário baseados na estimativa da densidade de probabilidade dos sinais na saída da matriz de separação.

A proposta do critério de processamento multiusuário é fundamentada sob o Teorema de Shalvi-Weinstein, generalizado para a situação de múltiplas fontes através de condições de

recuperação de sinais. A nova família de critérios multiusuário estende a consideração de Shalvi-Weinstein sobre as estatísticas necessárias para recuperação de fontes e utiliza duas estratégias para garantir a recuperação das diversas fontes envolvidas.

Como a proposta se baseia num critério de equalização previamente proposto, é realizada uma breve revisão das hipóteses envolvidas na proposta deste critério bem como a derivação do algoritmo associado. Na seqüência, é realizada uma análise inédita de tal método envolvendo aspectos de convergência do algoritmo, equivalência com outros critérios e ainda a avaliação de alguns parâmetros inerentes ao critério que impactam diretamente o desempenho dos algoritmos.

Esta parte do trabalho traz um novo aporte de conhecimentos sobre as estruturas de alguns critérios conhecidos de equalização cega, permitindo uma conexão da teoria clássica de filtragem adaptativa com aspectos de teoria da informação.

A apresentação da proposta de processamento multiusuário é de modo a ressaltar suas diferenças com os outros critérios existentes de estimativa de fdp, explicitando as questões relativas às diferentes estratégias de assegurar a identificação de todas as fontes. São também discutidas as diversas abordagens de separação de fontes que permeiam a concepção e a análise do critério proposto.

Outro aspecto, também inédito, discutido nesta tese é a relação de critérios cegos de recuperação da informação com critérios clássicos de detecção de sinais. No caso, é mostrada uma relação do critério de máxima probabilidade *a posteriori* com o critério proposto e com um critério supervisionado (MMSE), ilustrando as limitações de desempenho dos mesmos.

Os resultados teóricos indicam uma confluência de critérios cegos e supervisionados, através de princípios da teoria da informação. E uma vez que o critério FP utiliza tais princípios, parece ter um papel chave na compreensão de aspectos de equivalência entre os vários critérios analisados em filtragem adaptativa e em separação de fontes.

Algumas poucas simulações computacionais comparativas são realizadas neste capítulo, uma vez que serão contempladas posteriormente, quando inseridas no contexto das aplicações. Entretanto, os resultados de comparação de algoritmos que se norteiam pelo Teorema de Shalvi-Weinstein indicam uma grande diferença de desempenho entre a proposta aqui apresentada e o algoritmo MUK descrito na literatura.

A análise e o entendimento dos aspectos conceituais que justificam essas diferenças de desempenho constituem o escopo do próximo capítulo.

– “A estatística é a primeira das ciências inexatas.”
Edmond e Jules Goncourt, 1867

6

Análise da Influência das HOS em Algoritmos Adaptativos para Separação de Fontes

ESTATÍSTICAS de ordem superior (HOS, *Higher Order Statistics*) têm desempenhado um papel fundamental no desenvolvimento de ferramentas e critérios para separação cega de fontes. A partir dos teoremas de Benveniste-Goursat-Rouget e de Shalvi-Weinstein, ficou explícita a necessidade de estatísticas de ordem maiores que dois para recuperação de sinais quando não há disponibilidade do sinal transmitido.

A despeito de sua bem compreendida necessidade para se obter a separação de fontes com distribuições arbitrárias, é pouco conhecido o real impacto do número de HOS envolvidas nos algoritmos adaptativos para separação de fontes. Neste capítulo, pretende-se avaliar este aspecto através de aproximações da função de densidade de probabilidade, de modo a fornecer novas luzes sobre a análise teórica de algoritmos e critérios de separação cega de fontes. Esta análise, que de nosso conhecimento é original, se faz com base numa proposta de aproximação de função de densidade de probabilidade, na forma de expansão em torno de mistura de

gaussianas também apresentada no capítulo.

Sendo assim, o presente capítulo é organizado como descrito a seguir. A Seção 6.1 discute a motivação do presente capítulo. Na Seção 6.2, são discutidas as aproximações de cálculo da função densidade de probabilidade, incluindo dois casos particulares conhecidos na literatura, bem como a derivação da nova expansão em série. De posse deste ferramental matemático, os aspectos da influência quantitativa das HOS em algoritmos adaptativos para BSS são abordados na Seção 6.3. A Seção 6.4 trata da extensão dos resultados obtidos para separação de fontes no problema de desconvolução cega. Finalmente, na Seção 6.5 são listadas as principais conclusões, observações e perspectivas do assunto tratado no presente capítulo.

6.1 Motivação

Pode-se dizer, sem exageros, que a utilização de ferramentas estatísticas possibilitou os maiores avanços nas técnicas de separação cega de fontes, através da utilização da análise em componentes independentes.

Conforme mostrado no Capítulo 2, os mais diversos métodos de separação de fontes são baseados, sempre, na exploração das características estatísticas dos sinais envolvidos, medidas através de ferramentas baseadas em teoria da informação. E, como também mostrado anteriormente, a utilização de HOS é um imperativo quando deseja-se separar fontes com distribuições genéricas.

Esta constatação motivou a investigação qualitativa das estatísticas de ordem superior e seus impactos em algoritmos adaptativos, uma vez que em versões adaptativas apenas uma pequena parte dos dados está disponível a cada iteração para prover a estimação de uma fdp ou a otimização de qualquer outro critério capaz de prover a separação.

Uma importante motivação para tal investigação deve-se à grande diferença de desempenho entre os algoritmos MU-CFPA e MUK, em relação à taxa de convergência, apresentada no Capítulo 5. Sob a abordagem da recuperação dos sinais das múltiplas fontes em duas etapas, equalização e separação, ambos os algoritmos compartilham da mesma estratégia de separação das fontes, realizada através de uma restrição sobre a ortogonalidade da matriz de separação global. Desta maneira, a análise recai sobre a etapa de equalização dos dois critérios.

O algoritmo MUK, conforme já discutido anteriormente, deriva *diretamente* do Teorema de Shalvi-Weinstein. Isto significa que ele procura *maximizar a kurtosis* dos sinais na saída do filtro de separação, de maneira a garantir que a interferência tenha sido removida dos

sinais. A restrição imposta sobre a preservação da variância é realizada através de uma etapa de pré-branqueamento, realizada previamente, garantindo que os sinais a serem tratados pelo MUK apresentam uma matriz de correlação igual à matriz identidade. Sem esta etapa, o algoritmo não funciona adequadamente e a convergência para uma solução do problema de separação não é possível.

No caso do algoritmo MU-CFPA, a etapa de equalização é realizada através da *maximização da similaridade da fdp do sinal* à saída dos filtros de separação com um modelo paramétrico que modela a densidade de probabilidade de um sinal ideal a ser recuperado. Uma vez que a medida de similaridade recai numa estimação paramétrica da fdp do sinal recuperado, o Teorema de Shalvi-Weinstein também se encontra respeitado. Na verdade, o critério de estimação da fdp é mais abrangente, no sentido de que **todos** os momentos dos sinais na entrada e na saída do sistema, inclusive os de ordem dois, devem ser igualados entre si para que as fdps destes sinais sejam iguais.

Desta forma, duas diferenças básicas estão em jogo na comparação dos algoritmos:

- (1) **número de momentos a serem igualados** – o número de estatísticas consideradas nos algoritmos é bem diferente. Enquanto o MUK respeita a exigência do Teorema SW, de utilizar **um** momento de ordem superior, o MU-CFPA utiliza uma abordagem mais próxima daquela ilustrada pelo Teorema BGR, na qual **todos** os momentos são considerados;
- (2) **necessidade de pré-branqueamento** – como o MUK não utiliza a restrição sobre a variância dos sinais, esta é considerada na etapa de pré-branqueamento, ou PCA. O MU-CFPA por sua vez, considera a variância, bem como todos os outros momentos, de uma única vez, guardando com isso uma grande vantagem de não necessitar de dados esferatizados (branqueados e de potência unitária) para que o algoritmo funcione.

De fato, o problema da separação cega de fontes em sua variante mais conhecida, análise por componentes independentes, pode ser abordado sob vários aspectos, tais como os de não-gaussianidade das fontes, maximização da entropia (informação mútua), maximização de funções contraste, e outros. De acordo com o discutido na Seção 3.1.3, uma abordagem unificada reflete-se na busca, ou estimativa, de uma aproximação para as funções de densidade de probabilidade (ou funções de densidade cumulativa) das fontes [Lee et al., 2000].

É através desta abordagem unificada que se torna possível avaliar o impacto do uso de mais HOS para a implementação de algoritmos adaptativos para BSS. Uma vez que a estimativa da fdp é fundamental para tal, a complexidade computacional envolvida deve ser levada em consideração. Isto motiva a análise através de aproximações da densidade de probabilidade, tópico da seção seguinte.

6.2 Aproximações de funções de densidade de probabilidade

A aproximação de funções de densidade de probabilidade é um tópico que foi abordado superficialmente, e de maneira implícita, no Capítulo 5 (Seção 5.3), quando foram comentados alguns aspectos de estimação de fdp.

A estimação de uma fdp é uma aproximação realizada sobre duas possíveis estruturas: paramétrica e não-paramétrica. No caso paramétrico, assumem-se previamente alguns modelos para as distribuições das fontes e busca-se estimar seus *parâmetros* de maneira a representar da melhor maneira possível a densidade dos dados disponíveis. A estimação não-paramétrica se dá a partir de alguma característica dos dados, mas sem levar em consideração um modelo pré-assumido para a distribuição das fontes.

Dentre as estimativas não-paramétricas mais conhecidas estão as estimativas por *kernels* [Wegman, 1972; Zabin & Wright, 1994b], como o estimador de Parzen, descrito em maiores detalhes na Seção 5.3.1, e métodos de estimação através de *séries ortornormais* [Schwartz, 1967].

Embora as estimativas por *kernels* sejam bastante eficientes para os casos de fdps de sinais em sistemas de comunicação digital, particularmente utilizando-se o estimador de Parzen, seu desempenho não é satisfatório para distribuições genéricas [Hyvärinen, 1998]. Além disto, a estimação por séries ortogonais são melhores que estimativas por *kernels* para dados multivariável, uma vez que a convergência através de séries ortogonais permanece constante com o aumento da dimensão dos dados, enquanto que a estimativa por *kernels* apresenta uma taxa de convergência inversamente proporcional à dimensão dos mesmos [Schwartz, 1967; Phillips, 1999].

Como as séries ortogonais apresentam-se definidas através de coeficientes provenientes dos momentos da variável que se deseja estimar [Devroye, 1987], sua utilização é a mais indicada para a análise que se deseja aqui realizar quanto às HOS utilizadas em algoritmos adaptativos.

De uma maneira geral, estimativas utilizando séries ortornormais são geradas a partir da estimativa dos coeficientes de uma expansão ortonormal. A série de Fourier, por exemplo, utiliza senóides complexas como base para sua expansão. No caso de funções de densidade de probabilidade, a decomposição em base de Fourier não é adotada, uma vez que os componentes não contêm informação específica sobre os momentos da distribuição.

Por outro lado, as expansões de Gram-Charlier e Edgeworth¹[Haykin, 1998], explicitam os momentos conforme descritas a seguir.

¹Apesar de ter sido descrita brevemente no Capítulo 3, aqui será novamente apresentada em detalhes.

6.2.1 Expansão de Gram-Charlier

A série ortonormal para a expansão de Gram-Charlier, é obtida a partir da relação entre a função densidade de probabilidade e a função característica.

Seja a função característica, também chamada de função geradora de momentos, de uma v.a. real y com uma função densidade de probabilidade $p_Y(y)^2$, definida como [Papoulis, 1991]:

$$\Omega_Y(\omega) \triangleq \int_{-\infty}^{\infty} p_Y(y) \exp(j\omega y) dy, \quad (6.1)$$

em que $j = \sqrt{-1}$ e $\omega \in \mathbb{R}$. Desta forma, $\Omega_Y(\omega)$ é a transformada de Fourier da fdp $p_Y(y)$, exceto por uma mudança de sinal algébrico no expoente do argumento da função exponencial. De uma forma geral, $\Omega_Y(\omega)$ é um número complexo cujas partes real e imaginária são finitas para qualquer valor de ω [Haykin, 1998].

Se o k -ésimo momento da v.a. y existe, então pode-se expandir $\Omega_Y(\omega)$ em uma série de potências em torno de $\omega = 0$, através de [Papoulis, 1991]:

$$\Omega_Y(\omega) = 1 + \sum_{k=1}^{\infty} \frac{(j\omega)^k}{k!} \cdot \kappa_k(y), \quad (6.2)$$

na qual $\kappa_k(y)$ é o momento central de ordem k da variável y definido como [Papoulis, 1991]:

$$\begin{aligned} \kappa_k(y) &\triangleq \mathbb{E} \{y^k\} \\ &= \int_{-\infty}^{\infty} y^k p_Y(y) dy. \end{aligned} \quad (6.3)$$

A Equação (6.2) é resultante de uma expansão da função exponencial em série de potências, a partir da Equação (6.3), na qual fez-se uma mudança na ordem da integral e do somatório [Haykin, 1998].

Por outro lado, pode-se representar o logaritmo da função característica em série de potências na forma [Papoulis, 1991; Therrien, 1992]:

$$\ln [\Omega_Y(\omega)] = \sum_{k=1}^{\infty} \frac{c_k}{k!} \cdot (j\omega)^k, \quad (6.4)$$

²Por questões de simplicidade de notação será considerada apenas uma v.a. real para a demonstração nesta seção. Entretanto, nas derivações presentes no Apêndice A são consideradas v.a. complexas.

em que c_k é chamado de cumulante de k -ésima ordem ou *semi-invariante* da v.a. y . Os cumulantes de diferentes ordens relacionam-se com os momentos através da recursão

$$c_k = \kappa_k - \sum_{i=1}^{k-1} \binom{k-1}{i-1} c_i \cdot \kappa_{k-i}, \quad (6.5)$$

detalhada no Apêndice A.

Para simplificar, considera-se que y tem média zero e variância unitária, o que implica em $\kappa_1 = 0$ e $\kappa_2 = 1$ e, da Equação (6.4), obtém-se:

$$\ln [\Omega_Y(\omega)] = \frac{1}{2}(j\omega)^2 + \sum_{k=3}^{\infty} \frac{c_k}{k!} (j\omega)^k. \quad (6.6)$$

Define-se, por conveniência, uma função geradora de momentos como referência, dada por $\Omega_0(\omega)$, na forma

$$\ln [\Omega_0(\omega)] = \frac{1}{2}(j\omega)^2 + \sum_{k=3}^{\infty} \frac{c_{k,0}}{k!} (j\omega)^k, \quad (6.7)$$

em que $c_{k,0}$ é o cumulante de k -ésima ordem da distribuição de momentos $\Omega_0(\omega)$.

Subtraindo-se da Equação (6.6) a Equação (6.7) tem-se [Lacoume et al., 1997]:

$$\ln \left[\frac{\Omega_Y(\omega)}{\Omega_0(\omega)} \right] = \ln [\Omega_Y(\omega)] - \ln [\Omega_0(\omega)] = \sum_{k=3}^{\infty} \frac{c_k - c_{k,0}}{k!} (j\omega)^k \quad (6.8)$$

Denotando-se ainda:

$$r(\omega) = \sum_{k=3}^{\infty} \frac{c_k - c_{k,0}}{k!} (j\omega)^k,$$

tomando-se sua exponencial e escrevendo-a sob a forma de série de potências, obtém-se [Lacoume et al., 1997; Haykin, 1998]:

$$\exp [r(\omega)] = 1 + \sum_{l=1}^{\infty} \frac{r^l(\omega)}{l!} \quad (6.9)$$

Com isso, pode-se substituir a Equação (6.9) na Equação(6.8), o que resulta na expressão:

$$\frac{\Omega_Y(\omega)}{\Omega_0(\omega)} = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \cdot \left(\sum_{k=3}^{\infty} \frac{c_k - c_{k,0}}{k!} (j\omega)^k \right), \quad (6.10)$$

com a qual pode-se fazer um agrupamento dos termos com as mesmas potências, obtendo-se novos coeficientes para a expansão da função geradora de momentos.

A partir daí, é possível utilizar uma transformada inversa de Fourier sobre $\Omega_Y(\omega)$ e escrever a seguinte aproximação para a função de densidade de probabilidade da variável y em torno de uma distribuição de referência $p_0(y)$ [Lacoume et al., 1997; Laster, 1997; Haykin, 1998]:

$$p_Y(y) = p_0(y) \left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{b}_i(y) \right), \quad (6.11)$$

em que C_k são os coeficientes da expansão em uma série ortonormal dada por funções matemáticas adequadas $\mathfrak{b}_i(y)$.

Os coeficientes da série ortonormal são definidos em função dos momentos centrais, de ordem k , da variável y . Os 8 primeiros coeficientes da série são dados por [Haykin, 1998]:

$$\begin{aligned} C_1 &= 0 \\ C_2 &= 0 \\ C_3 &= \frac{c_3}{6} \\ C_4 &= \frac{c_4}{24} \\ C_5 &= \frac{c_5}{120} \\ C_6 &= \frac{1}{720} (c_6 + 10c_3^2) \\ C_7 &= \frac{1}{5040} (c_7 + 35c_4c_3) \\ C_8 &= \frac{1}{40320} (c_8 + 56c_5c_3 + 35c_4^2), \end{aligned} \quad (6.12)$$

e podem ser escritos em termo dos seus momentos como [Laster, 1997]

$$\begin{aligned}
C_1 &= \kappa_1 \\
C_2 &= \frac{1}{2}(\kappa_2 - 1) \\
C_3 &= \frac{1}{6}(\kappa_3 - 3\kappa_2) \\
C_4 &= \frac{1}{24}(\kappa_4 - 6\kappa_2 + 3) \\
C_5 &= \frac{1}{120}(\kappa_5 - 10\kappa_3 + 15\kappa_1) \\
C_6 &= \frac{1}{720}(\kappa_6 - 15\kappa_4 + 45\kappa_2 - 15) \\
C_7 &= \frac{1}{5040}(\kappa_7 - 21\kappa_5 + 105\kappa_3 - 105\kappa_1) \\
C_8 &= \frac{1}{40320}(\kappa_8 - 28\kappa_6 + 210\kappa_4 - 420\kappa_2 + 105).
\end{aligned} \tag{6.13}$$

Ao considerar-se a densidade de referência como sendo a distribuição gaussiana (normal) obtém-se a *Série de Gram-Charlier* ou ***Expansão de Gram-Charlier*** dada por :

$$p_Y(y) = p_G(y) \left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{h}_k(y) \right), \tag{6.14}$$

em que $p_G(y)$ é uma distribuição gaussiana normalizada da variável y dada por

$$p_G(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).$$

Os elementos da base ortonormal e as sucessivas derivadas da função de referência relacionam-se pela expressão [Haykin, 1998]:

$$\mathfrak{h}_k(y) = (-1)^k \cdot \frac{1}{p_0(y)} \cdot \frac{d^k p_0(y)}{dy^k}, \tag{6.15}$$

em que no caso da aproximação de Gram-Charlier a densidade de referência é a gaussiana e a base ortonormal é constituída pelos polinômios de Hermite³.

³Alguns poucos trabalhos denominam os polinômios obtidos para a aproximação de Gram-Charlier de polinômios de Tchebycheff-Hermite [Laster, 1997].

Os polinômios de Hermite, até o grau 8, são [Laster, 1997]

$$\begin{aligned}
 \mathfrak{h}_0(y) &= 1 \\
 \mathfrak{h}_1(y) &= y \\
 \mathfrak{h}_2(y) &= y^2 - 1 \\
 \mathfrak{h}_3(y) &= y^3 - 3y \\
 \mathfrak{h}_4(y) &= y^4 - 6y^2 + 3 \\
 \mathfrak{h}_5(y) &= y^5 - 10y^3 + 15y \\
 \mathfrak{h}_6(y) &= y^6 - 15y^4 + 45y^2 - 15 \\
 \mathfrak{h}_7(y) &= y^7 - 21y^5 + 105y^3 - 105y \\
 \mathfrak{h}_8(y) &= y^8 - 28y^6 + 210y^4 - 420y^2 + 105.
 \end{aligned} \tag{6.16}$$

Os polinômios de Hermite apresentam ainda as seguintes propriedades:

$$\mathbf{P}_{\text{PH1.}} \quad \mathfrak{h}_k(y)p_G(y) = \frac{d[\mathfrak{h}_{k-1}(y)p_G(y)]}{dy}$$

$$\mathbf{P}_{\text{PH2.}} \quad \mathfrak{h}_{k+1}(y) = y \cdot \mathfrak{h}_k - k \cdot \mathfrak{h}_{k-1}(y)$$

$$\mathbf{P}_{\text{PH3.}} \quad \int_{-\infty}^{\infty} \mathfrak{h}_k \frac{d^l p_G(y)}{dy^l} dy = \begin{cases} 0, & k \neq l \\ (-1)^l \cdot l! & k = l \end{cases}$$

A Propriedade \mathbf{P}_{PH3} é chamada de propriedade da *biortogonalidade* entre os polinômios de Hermite e a derivada de ordem l da função $p_G(y)$.

Os coeficientes C_k da expansão de Gram-Charlier são obtidos através da multiplicação de ambos os lados da Equação (6.14) por $\mathfrak{h}_k(y)$, integrando-se no intervalo $]-\infty, \infty[$. Devido à Propriedade \mathbf{P}_{PH3} (biortogonalidade), os coeficientes da expansão em série podem ser dados por [Laster, 1997]:

$$\begin{aligned}
 c_k &= \frac{1}{k!} \int_{-\infty}^{\infty} \mathfrak{h}_k(y)p_Y(y)dy \\
 &= \frac{1}{k!} \left[\kappa_k - \frac{k^{[2]}}{2 \cdot 1!} \kappa_{k-2} + \frac{k^{[4]}}{2^2 \cdot 2!} \kappa_{k-4} - \dots \right],
 \end{aligned} \tag{6.17}$$

na qual

$$k^{[m]} = \frac{k!}{(k-m)!} = k \cdot (k-1) \cdots [k-(m-1)], \quad k \geq m. \tag{6.18}$$

6.2.2 Expansão de Edgeworth

A expansão de Gram-Charlier apresenta os termos da série ordenados a partir de seus índices. Entretanto, esta maneira pode não ser a mais adequada de utilizar em alguns casos.

Em situações práticas, é necessário fazer um truncamento da série em um número de termos que represente adequadamente a densidade de probabilidade em questão. Desta forma, é de interesse que a série seja ordenada a partir dos termos que apresentam a maior magnitude, os quais são mais importantes de ser retidos em um truncamento da expansão [Faÿ, Moulines, & Soulier, 2004].

A *Expansão de Edgeworth* é dada pela expansão de Gram-Charlier ordenada pela ordem de magnitude dos seus coeficientes⁴ [Haykin, 1998]. Esta ordem é dada pelo agrupamento dos termos com índices

$$k = (0), (3), (4, 6), (5, 7, 9), (8, 10, 12, 14), \dots \quad (6.19)$$

Desta maneira, a expansão de Edgeworth, em torno de uma distribuição gaussiana, é dada por [Lacoume et al., 1997; Haykin, 1998]:

$$\begin{aligned} p_Y(y) = p_G(y) \cdot & \left(1 + \frac{c_3}{3!} \mathfrak{h}_3(y) + \frac{c_4}{4!} \mathfrak{h}_4(y) + \frac{10c_3^2}{6!} \mathfrak{h}_6(y) + \frac{c_5}{5!} \mathfrak{h}_5(y) \right. \\ & + \frac{35c_3c_4}{7!} \mathfrak{h}_7(y) + \frac{280c_3^3}{9!} \mathfrak{h}_9(y) + \frac{c_6}{6!} \mathfrak{h}_6(y) + \frac{56c_3c_5}{8!} \mathfrak{h}_8(y) \\ & \left. + \frac{35c_4^2}{8!} \mathfrak{h}_8(y) + \frac{2100c_3^2c_4}{10!} \mathfrak{h}_{10}(y) + \frac{15400c_3^4}{12!} \mathfrak{h}_{12}(y) + \dots \right). \end{aligned} \quad (6.20)$$

A utilização da expansão de Edgeworth truncada representa uma redução na complexidade computacional, uma vez que o truncamento preserva os termos de maior magnitude, sendo assim capaz de fazer uma estimativa de $p_Y(y)$ com um menor número de termos da série.

6.2.3 Comentários e considerações

De uma forma geral, a expansão de Edgeworth é mais adequada para estimativa de uma fdp $p_Y(y)$, dado que permite uma redução na complexidade computacional em função do número de termos reduzido estarem associados aos coeficientes de maior magnitude.

⁴Embora apresentem a mesma estrutura, a expansão de Edgeworth e de Gram-Charlier foram desenvolvidas de forma independente [Wegman, 1972].

Esta ferramenta tem sido utilizada em vários trabalhos sobre separação cega de fontes como uma medida da aproximação da gaussianidade ou mesmo da informação mútua do sinal na saída da matriz de separação, uma vez que ambas as estratégias requisitam a fdp do sinal para sua otimização. Em [Comon, 1994], a expansão da fdp como uma série ortonormal através de seus cumulantes é utilizada para realizar a otimização de uma função contraste. Através do uso da estimativa em séries ortonormais, [Lee et al., 2000] promove uma abordagem na qual verifica que vários tipos de estratégias para separação cega de fontes consistem em estimar as características da densidade de probabilidade. Com caráter mais tutorial, a expansão em séries ortonormais, particularmente a expansão de Edgeworth, é brevemente discutida em [Hyvärinen et al., 2001].

Alguns outros métodos de estimação poderiam ser utilizados para prover uma aproximação da fdp. Em [Kay et al., 2001] são propostas algumas distribuições multidimensionais descritas a partir dos coeficientes de reflexão, estimativas de autocorrelação e coeficientes cepstrais. Além destes, [Deco & Brauer, 1995] utilizam uma estimativa da fdp através de cumulantes obtidos na saída de dispositivos não-lineares para criar uma adaptação de uma rede neural capaz de separar fontes independentes. Outro trabalho que apresenta uma linha interessante de estimação de parâmetros de fdps para separação está em [Lin et al., 1997], no qual são utilizados mapas de Kohonen (redes auto-organizadas) para inferir sobre distribuições adequadas para que a separação das fontes seja possível por uma divisão no espaço das distribuições das fontes.

Uma consideração importante a ser destacada é a de uma densidade gaussiana normalizada como função de referência. No caso de desejar-se utilizar uma densidade gaussiana não-normalizada a seguinte estimativa é adotada [Laster, 1997]:

$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{(y - \mu_y)^2}{2\sigma_y^2}\right] \cdot \left(1 + \sum_{k=3}^{\infty} C_k \mathfrak{h}_k\left(\frac{y - \mu_y}{\sigma_y}\right)\right), \quad (6.21)$$

em que μ_y e σ_y^2 são, respectivamente, a média e o desvio padrão da distribuição de referência. Com isto, os coeficientes C_k são expressos em função dos momentos não-centrados, dados por

$$\kappa_k = \mathbb{E}\left\{\left(\frac{y - \mu_y}{\sigma_y}\right)^k\right\}. \quad (6.22)$$

Outra questão de extrema importância: *qualquer densidade de probabilidade poderia ser utilizada como função de referência*. A densidade gaussiana é escolhida porque a manipulação matemática da base ortonormal torna-se mais fácil, uma vez que os polinômios de Hermite apresentam algumas propriedades que facilitam sua utilização. Para alguns tipos de densidade, a base ortonormal associada pode nem mesmo ser formada por polinômios, dificultando mais ainda a representação através de fórmulas matematicamente tratáveis.

Uma vez que a formulação através da densidade gaussiana apresenta a grande vantagem da base ortonormal ser de tratamento matemático mais fácil, sua utilização parece ser sempre a mais indicada. Entretanto, uma ressalva deve ser mencionada: *não é possível aproximar todas as distribuições utilizando a gaussiana como função de referência*. Embora uma grande variedade de fdps sejam passíveis de representação através da expansão em série em torno da gaussiana, vários tipos de fdp de interesse prático, por exemplo as densidades que são somatórios de gaussianas (também chamadas de misturas gaussianas), não apresentam representação através da expansão de Gram-Charlier ou Edgeworth [Laster, 1997]. Esta limitação inviabiliza a utilização direta da expansão em série em torno da distribuição gaussiana para diversos problemas de comunicação digital, nos quais os sinais geralmente apresentam distribuições de probabilidades do tipo misturas de gaussianas.

Para estes tipos de distribuições, as expansões do tipo exponencial são mais indicadas. Uma expansão, obtida a partir da inversão de Fourier da função geradora de momentos, é chamada de **expansão em torno do ponto de sela**⁵ [Lindsey, 1996]. Nesta expansão, a densidade é escrita como uma integral de termos do tipo $\exp[n \cdot \Omega_Y(\omega)]$, em que n é o número de amostras. Assim, pode-se derivar uma expansão em torno de um ponto de sela y_0 , o qual deverá ser um zero de $\frac{\partial \Omega_Y(\omega)}{\partial \omega}$, através do método da descida mais íngreme (*steepest descent*). Isto fornece a seguinte expansão [Lindsey, 1996]:

$$p_Y(y) = \left[\frac{n}{2\pi \frac{\partial^2 \Upsilon_Y(\omega)}{\partial \omega^2}} \right]^{\frac{1}{2}} \cdot \exp(n [\Upsilon_Y(y\omega) - y\omega]) \cdot \left\{ 1 + \frac{1}{n} \left[\frac{\frac{\partial^4 \Upsilon_Y(\omega)}{\partial \omega^4}}{8 \cdot \left(\frac{\partial^2 \Upsilon_Y(\omega)}{\partial \omega^2}\right)^2} - \frac{5 \cdot \left(\frac{\partial^3 \Upsilon_Y(\omega)}{\partial \omega^3}\right)^2}{24 \cdot \left(\frac{\partial^2 \Upsilon_Y(\omega)}{\partial \omega^2}\right)^3} \right] + \dots \right\} \Bigg|_{\omega=\omega_0}, \quad (6.23)$$

em que $\Upsilon_Y(\omega)$ é a segunda função característica (ou função geradora de cumulantes) da v.a. y [Papoulis, 1991].

Outro aspecto sobre a expansão de fdps em séries ortogonais é que as mesmas sofrem do problema de estimação nas “caudas” das distribuição. Nestes pontos, é bastante comum observar que a estimativa da fdp assume valores negativos, desrespeitando assim a premissa de que a fdp é uma função não-negativa. Por isso, em aplicações estatísticas práticas, não são utilizadas expansões em série e sim outros métodos não-paramétricos.

A partir da observação da expansão em torno do ponto de sela, pode-se verificar que as derivadas da função geradora de momentos fornecem um comportamento bastante similar à expansão de Gram-Charlier ou Edgeworth, uma vez que a função característica é definida através da função exponencial.

⁵No inglês encontra-se o termo *saddlepoint expansion*.

Visando fornecer uma ferramenta de análise adequada aos tipos de distribuições encontradas em sistemas de processamento multiusuário, foi desenvolvida neste trabalho uma nova expansão em série, a qual é apresentada na seção seguinte.

6.2.4 Expansão em série em torno de distribuições de misturas gaussianas

Uma vez que a Equação (6.11) é genérica para todas as distribuições de referência desde que as mesmas sejam contínuas, a derivação de uma expansão em série necessita encontrar a base ortonormal adequada para uma nova escolha da distribuição a ser utilizada como referência.

Como as funções de densidade de probabilidade dos sinais na saída do dispositivo de separação em sistemas de processamento multiusuário são misturas gaussianas, dadas por

$$p_{SG}(y) = \sum_{i=1}^{\mathfrak{c}} \frac{1}{\sqrt{2\pi\sigma_{\vartheta}^2}} \exp \left[-\frac{(y - \mathbf{a}_i)^2}{2\sigma_{\vartheta}^2} \right] \cdot \Pr(\mathbf{a}_i), \quad (6.24)$$

em que σ_{ϑ}^2 é a variância de cada uma das gaussianas e SG significa somatório de gaussianas, é natural que a expansão seja realizada em torno de uma fdp do mesmo tipo para que a estimativa da fdp seja realizada sobre um modelo adequado.

Então, ao utilizar-se como função de referência uma mistura de gaussianas como dado na Equação (6.24), é necessário deduzir a base ortonormal que representa as características da fdp que está sendo buscada. Esta base ortonormal deve necessariamente respeitar a relação com a respectiva função de referência:

$$\mathbf{b}_k(y) = (-1)^k \cdot \frac{1}{p_0(y)} \cdot \frac{d^k p_0(y)}{dy^k}. \quad (6.25)$$

Uma questão delicada é relativa as derivadas $\frac{d^k p_0(y)}{dy^k}$, uma vez que para algumas densidades de referência pode-se obter expressões bastante complicadas. Para simplificar a notação, pode-se escrever:

$$p_{SG}(y) = \sum_{i=1}^{\mathfrak{c}} \frac{1}{\sqrt{2\pi\sigma_{\vartheta}^2}} \exp \left[-\frac{(y - \mathbf{a}_i)^2}{2\sigma_{\vartheta}^2} \right] \cdot \Pr(\mathbf{a}_i) = \sum_{i=1}^{\mathfrak{c}} p_0^i(y - \mathbf{a}_i), \quad (6.26)$$

em que $p_0^i(y - \mathbf{a}_i) = \frac{1}{\sqrt{2\pi\sigma_{\vartheta}^2}} \cdot \exp \left[-\frac{(y - \mathbf{a}_i)^2}{2\sigma_{\vartheta}^2} \right]$ é o *kernel* gaussiano centrado em \mathbf{a}_i . Logo, ao tomar-se diretamente as derivadas na Equação (6.25), obtém-se, por exemplo para $k = 1$:

$$\frac{dp_{SG}(y)}{dy} = \sum_{i=1}^{\mathfrak{c}} \frac{\Pr(\mathbf{a}_i)}{\sqrt{2\pi\vartheta^2}} \cdot \frac{-(y - \mathbf{a}_i)}{\sigma_{\vartheta}^2} \cdot \exp \left[-\frac{(y - \mathbf{a}_i)^2}{2\sigma_{\vartheta}^2} \right], \quad (6.27)$$

o que resulta em

$$\mathfrak{b}_1(y) = \frac{y}{\sigma_{\vartheta}^2} - \frac{1}{\sigma_{\vartheta}^2} \frac{\sum_{i=1}^{\mathfrak{c}} \mathfrak{a}_i \cdot p_0^i(y)}{\sum_{i=1}^{\mathfrak{c}} p_0^i(y)}. \quad (6.28)$$

Para evitar a obtenção de expressões ainda mais difíceis, particularmente nos graus mais elevados, aborda-se o problema como uma estimação de Parzen, ou seja, como uma aproximação de fdp por um somatório de gaussianas.

Desta forma, pode-se pensar na seguinte abordagem:

Se a expansão de Gram-Charlier (ou Edgeworth) aproxima fdps em torno de uma densidade de probabilidade gaussiana, aproximar fdps em torno de um somatório de gaussianas corresponde a aproximar fdps em torno de um somatório de expansões de Gram-Charlier (ou Edgeworth).

Pode-se formalizar esta idéia através da propriedade da linearidade dos operadores derivada e somatório. Assim, pode-se escrever, por exemplo, para o $k = 1$ termo, a seguinte equação:

$$\frac{d \sum_{i=1}^{\mathfrak{c}} p_0^i(y)}{dy} = \sum_{i=1}^{\mathfrak{c}} \frac{dp_0^i(y)}{dy}. \quad (6.29)$$

Logo, a Equação (6.29) transforma o problema em um somatório de derivadas para diferentes médias. Outra questão que deve ser considerada na derivação da base ortonormal é a propriedade de que as fdps de sistemas de processamento multiusuário possuem, para cada uma das fontes, média nula. Tomando, novamente, como exemplo o termo para $k = 1$, tem-se então o seguinte polinômio \mathfrak{c}_1 ⁶ relacionado com a expansão em torno de um somatório de gaussianas:

$$\begin{aligned} \mathfrak{c}_1(y) &= \frac{(y - \mathfrak{a}_1)}{\sigma_{\vartheta}^2} + \frac{(y - \mathfrak{a}_2)}{\sigma_{\vartheta}^2} + \dots + \frac{(y - \mathfrak{a}_{\mathfrak{c}})}{\sigma_{\vartheta}^2} \\ \mathfrak{c}_1(y) &= \frac{\mathfrak{c} \cdot y}{\sigma_{\vartheta}^2} - \frac{1}{\sigma_{\vartheta}^2} \sum_{i=1}^{\mathfrak{c}} \mathfrak{a}_i \\ \mathfrak{c}_1(y) &= \frac{\mathfrak{c} \cdot y}{\sigma_{\vartheta}^2}. \end{aligned} \quad (6.30)$$

⁶Deve-se notar que $\mathfrak{c}_k \neq c_k$, embora as fontes sejam similares.

Então, utilizando o mesmo procedimento para os k termos, encontra-se a seguinte recursão para a forma genérica dos polinômios que determinam a base ortonormal:

$$\mathbf{c}_{k+1}(y) = \left(\frac{\mathbf{e} \cdot y}{\sigma_{\vartheta}^2} \right) \cdot \mathbf{c}_k(y) - k \cdot \mathbf{c}_{k-1}(y), \quad (6.31)$$

resultando nos seguintes polinômios até a oitava ordem;

$$\begin{aligned} \mathbf{c}_0(y) &= 1 \\ \mathbf{c}_1(y) &= \mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \\ \mathbf{c}_2(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^2 - 1 \\ \mathbf{c}_3(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^3 - 3 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right) \\ \mathbf{c}_4(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^4 - 6 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^2 + 3 \\ \mathbf{c}_5(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^5 - 10 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^3 + 15 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right) \\ \mathbf{c}_6(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^6 - 15 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^4 + 45 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^2 - 15 \\ \mathbf{c}_7(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^7 - 21 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^5 + 105 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^3 - 105 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right) \\ \mathbf{c}_8(y) &= \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^8 - 28 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^6 + 210 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^4 - 420 \left(\mathbf{e} \cdot \frac{y}{\sigma_{\vartheta}^2} \right)^2 + 105, \end{aligned} \quad (6.32)$$

que são também uma classe de polinômios de Hermite, uma vez que há apenas fatores multiplicativos e as Propriedades P_{PH1}-P_{PH3} continuam válidas.

Assim, a expansão para densidade de probabilidade em torno de uma densidade de referência do tipo somatório de gaussianas é dada por:

$$p_Y(y) = p_{SG}(y) \left(\sum_{k=1}^{\infty} C_k \cdot \mathbf{c}_k(y) \right). \quad (6.33)$$

Uma outra base ortonormal poderia ser derivada se utilizada a seguinte aproximação para a densidade de probabilidade de misturas gaussianas proposta em [Torkkola, 1998]:

$$p_Y(y) = \sum_{k=1}^{\mathbf{c}} (1 - \tanh^2 [w \cdot |y - \mathbf{a}_i|]), \quad (6.34)$$

em que w é um fator que regula a suavidade da estimação, sendo que para $w \rightarrow \infty$ a distribuição tende a ser um somatório de impulsos que é a distribuição das modulações digitais. Neste caso, entretanto, a base ortonormal seria dada em polinômios em função de $\tanh[\cdot]$, cujas propriedades seriam de difícil visualização. A escolha da representação da densidade em sua forma direta do somatório de gaussianas é também justificada pela obtenção de uma classe de polinômios de Hermite, os quais possuem propriedades desejadas para a base ortonormal.

Com esta nova expansão, é possível avaliar a evolução dinâmica de uma estimativa da função de densidade de probabilidade e prover uma análise sobre os cumulantes envolvidos em algoritmos de separação cega de fontes. Este é o assunto da seção seguinte.

6.3 Evolução temporal da estimativa da densidade de probabilidade

Voltando ao ponto da motivação inicial deste capítulo, procura-se avaliar a importância dos cumulantes a serem considerados para se obter a separação de fontes, questão tratada em alguns trabalhos importantes da literatura. Por exemplo, em [Nadal & Parga, 1997] são avaliadas as características e condições dos cumulantes para que as diversas fontes independentes sejam recuperadas; a abordagem empregada é através da utilização dos cumulantes de quarta ordem das saídas de uma rede neural, utilizada como dispositivo de separação. O recente trabalho [Laar, 2004] utiliza as características geométricas dos cumulantes de ordem arbitrária para forçar uma decorrelação temporal dos sinais à saída de um sistema de separação de fontes. Neste caso, um conjunto de equações polinomiais é derivado para formar uma base ortonormal de projeção dos sinais separáveis.

Por outro lado, no contexto de equalização cega, o assunto de escolha de cumulantes vem sendo tratado desde a década de 90, estendendo-se ao caso multicanal, conforme pode ser visto em [Mota, 1992; Tugnait, 1995a, 1995b, 1997b; Tugnait et al., 2000; Tugnait, 2001] e em várias referências citadas nestes trabalhos.

A proposta desta seção vai na direção de responder a uma questão que deriva, basicamente, da diferença entre os Teorema de Shalvi-Weinstein e de Benveniste-Goursat-Rouget:

- ◆ *Qual o ganho que se tem em algoritmos adaptativos quando considera-se mais momentos de ordem superior que somente o de quarta ordem?*

Para tentar responder tal questão, o problema de separação de fontes é abordado segundo a busca por encontrar a distribuição de probabilidade das fontes. Neste contexto, em um sistema de separação de fontes em tempo real, a estimativa da densidade de probabilidade deve ser

realizada para permitir a separação das fontes instantaneamente ou por breves intervalos de tempo. Para permitir uma abordagem temporal sobre a influência dos momentos de ordem superior, é interessante observar a dinâmica dos termos que são utilizados numa expansão em série ortonormal da densidade de probabilidade.

Durante um processo adaptativo de separação de fontes, utilizando a abordagem de estimativa dos cumulantes para construção da fdp, a igualdade entre os cumulantes deve ser realizada a cada instante de tempo. Como exemplo desta abordagem, o algoritmo MUK maximiza o módulo da *kurtosis* de modo a igualar as *kurtosis* na entrada e saída global de um sistema de separação de fontes e garantir, através do Teorema SW, a separação das mesmas.

Através da expansão de Edgeworth, e supondo uma distribuição simétrica, é possível mostrar que o coeficiente da série correspondente ao cumulante de quarta-ordem possui a maior magnitude. Com isso, para cálculo da solução, a utilização somente da *kurtosis* para estimação da fdp é suficientemente representativa. Isso é mais importante quando a *kurtosis* já atingiu um valor próximo do seu valor em estado permanente. Mas o que ocorre durante os instantes iniciais da adaptação, quando o sinal na saída do dispositivo de separação não apresenta características estatísticas próximas àquelas do sinal que se deseja obter?

É fato que nos instantes iniciais da estimação (adaptação), a distribuição dos dados na saída de um sistema de separação apresenta características altamente dependentes da inicialização da matriz de separação. Além disso, os cumulantes vão sendo estimados com base nos dados disponíveis até o instante de tempo atual e a estimativa da função densidade de probabilidade é dada não somente pelos cumulantes envolvidos no processo de otimização mas também pelos termos da base ortonormal associados.

Supondo que o critério de otimização utilize, de forma direta ou indireta, um número S de cumulantes, a estimativa da fdp das fontes será realizada com base nos S cumulantes considerados pelo critério. Por exemplo, o MU-CMA utiliza os momentos de ordem quatro e dois, de forma indireta, e as estimativas das fdps das fontes são construídas somente por tais cumulantes e seus termos associados na expansão em série, conforme pode ser visualizado na Equação (6.33).

Como mostrado, ao utilizar-se a expansão para somatório de gaussianas baseadas naquela de Gram-Charlier⁷, o k -ésimo termo da expansão em série é dado pelo produto do k -ésimo cumulante multiplicado pelo polinômio de Hermite de grau k ⁸.

⁷Neste ponto a escolha da expansão de Gram-Charlier é feita com base na simplicidade de escrever os termos em sua ordem crescente da utilização de cumulantes. No caso de utilizar-se a expansão de Edgeworth, a diferença reside em que o k -ésimo termo da expansão não corresponde ao k -ésimo cumulante multiplicado pelo polinômio de Hermite de grau k , mas isto não invalida a análise apresentada nesta seção, representando somente uma reordenação nos termos da série.

⁸Uma vez que os polinômios c_k são uma classe de polinômios de Hermite, para simplificação da notação será adotada aqui a denominação de polinômio de Hermite indistintamente para os polinômios c_k ou h_k .

Por exemplo, considerando-se o termo para $k = 8$, tem-se então

$$\begin{aligned}
 p_Y^{(8)}(y) &= \mathbf{c}_8(y) \cdot C_8 \\
 &= \left[\left(\mathbf{c} \cdot \frac{y}{\sigma_\vartheta^2} \right)^8 - 28 \left(\mathbf{c} \cdot \frac{y}{\sigma_\vartheta^2} \right)^6 + 210 \left(\mathbf{c} \cdot \frac{y}{\sigma_\vartheta^2} \right)^4 - 420 \left(\mathbf{c} \cdot \frac{y}{\sigma_\vartheta^2} \right)^2 + 105 \right] \cdot \\
 &\quad \cdot \left[\frac{1}{40320} (c_8 + 56c_5c_3 + 35c_4^2) \right], \quad (6.35)
 \end{aligned}$$

em que $p_Y^{(8)}(y)$ representa o oitavo termo da expansão em série de $p_Y(y)$. Durante os instantes iniciais da adaptação, o coeficiente de oitava ordem apresenta um valor provavelmente bastante distinto daquele após a separação, devido ao fato de não ter ainda sido recebido um número importante de dados até então, e também um valor reduzido devido ao numerador presente no seu cálculo. Por outro lado, o polinômio de Hermite associado apresenta potências elevadas de y , o que leva a “amplificar” qualquer informação do real valor de y no seu sinal; também seu número de termos é maior que o de ordens menores, trazendo com isso uma “diversidade” de informação do sinal a ser estimado.

Para um termo k genérico da expansão, a análise é a mesma, uma vez que as potências dos polinômios de Hermite são crescentes com o aumento da ordem k e os cumulantes de ordem k têm aumentado o seu correspondente denominador, contribuindo para um decréscimo no seu valor numérico.

Com isso, algoritmos que utilizam um maior número de HOS apresentam uma maior taxa de convergência que algoritmos com estruturas similares que utilizam menos HOS. Isto devido à capacidade, por parte dos outros termos de ordem mais elevadas, de possibilitarem uma certa multiplicidade da informação, através das várias exponenciações do sinal, fazendo com que a convergência dos cumulantes seja mais rápida.

6.3.1 Comparação MUK \times MU-CFPA

Para comparação, utilizam-se os algoritmos MUK e MU-CFPA devido a duas questões básicas:

- (1) possuem a mesma estrutura de processamento;
- (2) utilizam diferentes números de cumulantes para fornecer a separação.

Neste caso a análise é realizada nos dois extremos: utilização de **uma** estatística de ordem superior no caso do MUK e utilização da **todas** as estatísticas de ordem superior no caso do MU-CFPA.

Seja então um sistema com dois usuários transmitindo sinais BPSK, cujos sinais são coletados em dois sensores no receptor com uma SNR = 30 dB. Desta forma a matriz de

mistura é dada por

$$\mathbf{H} = \begin{bmatrix} 0.7138 & 0.7004 \\ 0.7004 & -0.7138 \end{bmatrix}. \quad (6.36)$$

Vale ressaltar que embora a escolha dos coeficientes de \mathbf{H} seja realizada de maneira aleatória, é imposta a condição de que $\mathbf{H}^H \mathbf{H} = \mathbf{I}$ para que a condição de preservação da variância dos sinais de entrada seja respeitada sem nenhum pré-processamento. Com isto a comparação torna-se mais justa para ambos os métodos. Os parâmetros de simulação são dados na Tabela 6.1.

Parâmetros de simulação para os algoritmos	
MUK	$\mu = 2 \cdot 10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$
MU-CFPA	$\mu = 2 \cdot 10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$ $\sigma_r^2 = 0.1$

Tabela 6.1: Parâmetros de simulação dos algoritmos MUK e MU-CFPA para comparação das estatísticas de ordem superior.

A evolução temporal do CME para ambos os algoritmos é ilustrada na Figura 6.1, a partir da qual pode-se visualizar a significativa diferença em termos de taxa de convergência entre os algoritmos. Um ponto importante a ser mencionado é que o fator de passo do MU-CFPA pode ainda assumir valores mais elevados do que o utilizado nesta simulação, aumentando ainda mais a sua taxa de convergência. A utilização de um fator de passo mais baixo do que aquele máximo suportado pelo algoritmo deve-se ao fator ilustrativo de ressaltar a evolução da estimativa em tempo real dos sinais na saída do sistema de separação.

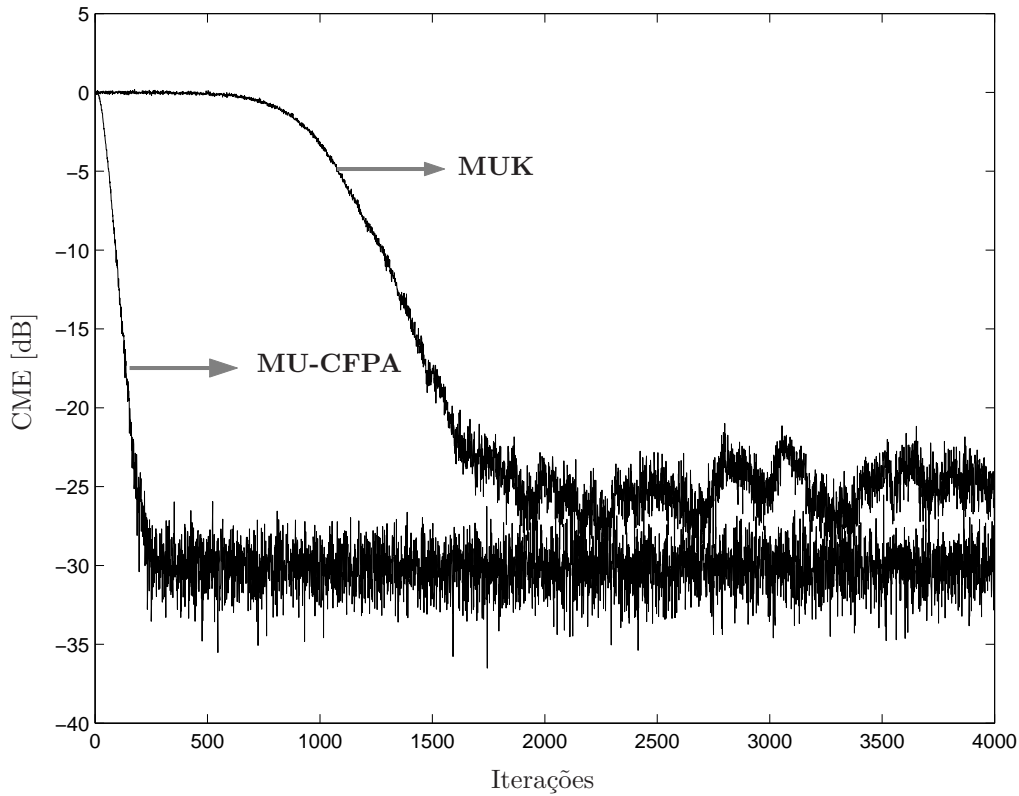


Figura 6.1: Evolução temporal do erro do módulo constante para os algoritmos MUK e MU-CFPA.

A partir dos dados obtidos na saída do sistema de separação para vários intervalos, são calculados os coeficientes da expansão em série em torno de misturas gaussianas e os polinômios \mathbf{c}_k , comparando-se assim a estimativa da fdp em função do número de cumulantes considerados. Vale lembrar que, para o caso BPSK, $\mathfrak{C} = 2$. No caso do MUK, apenas o cumulante de ordem 4 é retido, tornando os coeficientes da série função apenas da kurtosis. Já o MU-CFPA retém todos os cumulantes, provendo uma melhor aproximação. Por questões de simplificação da implementação, foram considerados os termos até a ordem $k = 8$, embora os resultados sejam extensivos a outros valores.

A Figura 6.2 ilustra a evolução dinâmica do algoritmo MUK. Deve-se observar que a estimativa da densidade de probabilidade aproxima-se da densidade dos dados que se deseja a partir do instante próximo à convergência do algoritmo.

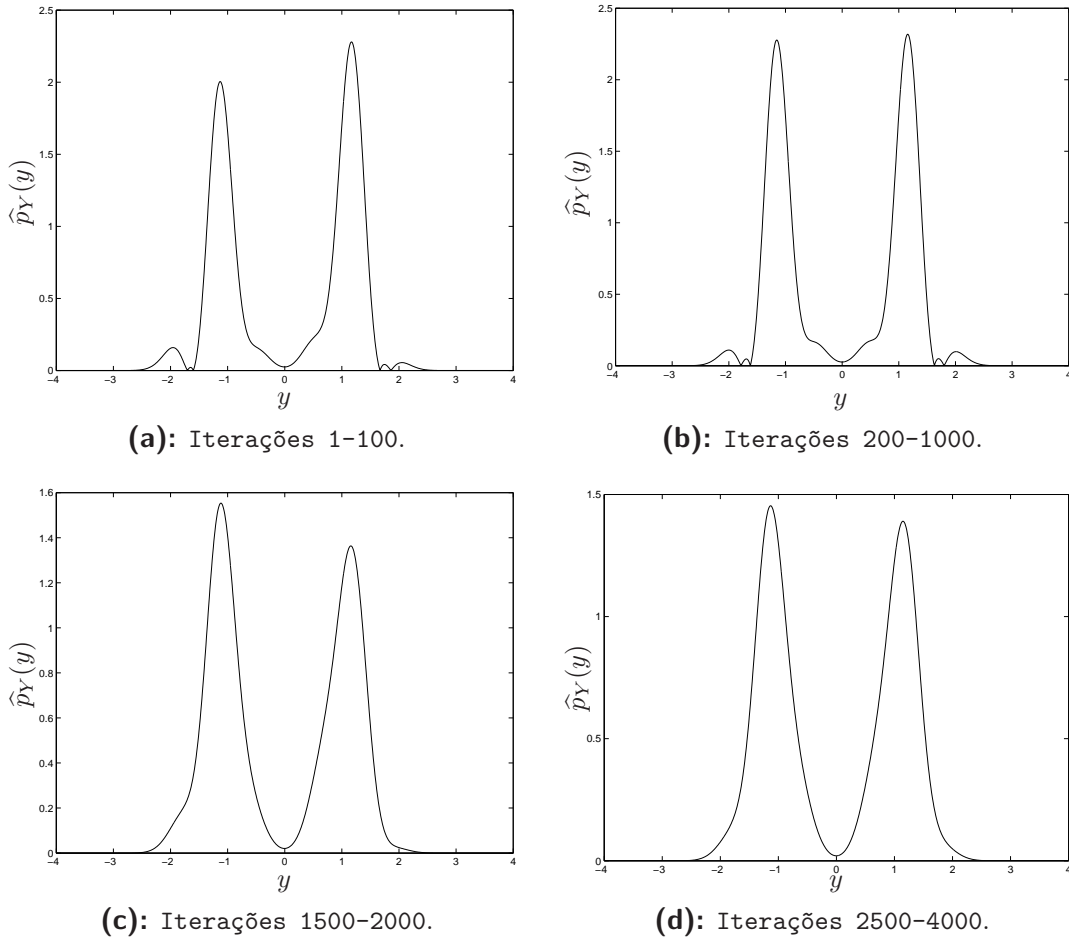


Figura 6.2: Evolução dinâmica da estimativa da função de densidade de probabilidade para o algoritmo MUK em diferentes intervalos temporais.

Deve-se observar que as estimativas das fdps apresentam um pequeno desvio em torno das médias esperadas, neste caso, $+1$ e -1 . Isto se deve ao fato de que a estimativa dos cumulantes é uma estimativa polarizada devido à complexidade da mesma do número de dados disponíveis. Isso poderia ser evitado caso tais estimativas fossem realizadas com blocos de dados, tentando preservar as propriedades estatísticas em cada um dos blocos. Entretanto, nas simulações apresentadas neste capítulo, este procedimento não foi realizado.

A Figura 6.3 apresenta o comportamento dinâmico da estimativa da fdp dos sinais na saída do sistema de separação utilizando o MU-CFPA, ilustrando a importância do uso dos diversos cumulantes para a taxa de convergência do algoritmo adaptativo. Observa-se, a exemplo do caso anterior, uma polarização das médias dos *kernels* das estimativas.

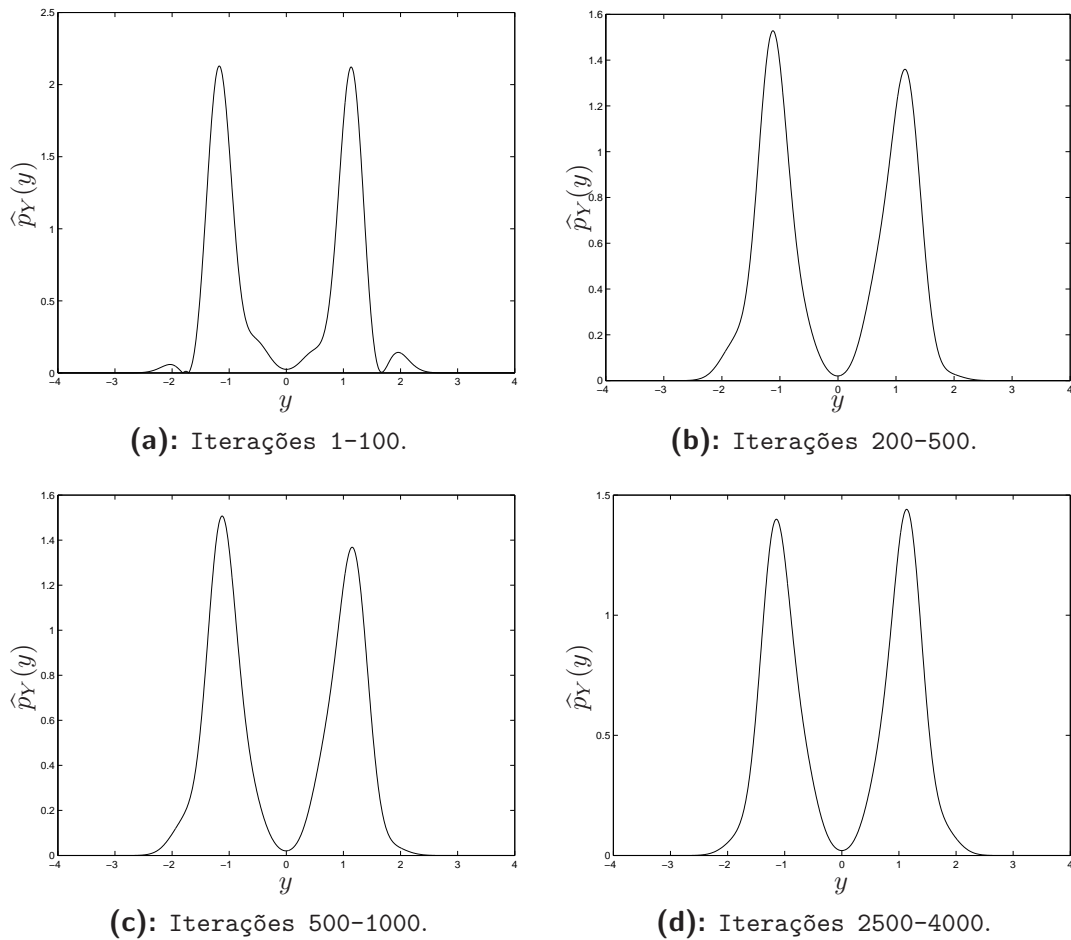


Figura 6.3: Evolução dinâmica da estimativa da função de densidade de probabilidade para o algoritmo MU-CFPA em diferentes intervalos temporais.

Ainda que o caso de simulação apresentado seja um tanto quanto simples, ele é bastante ilustrativo do comportamento observado em uma vasta gama de simulações utilizando misturas convolutivas instantâneas. Para os diversos valores de relação sinal-ruído utilizados e variadas condições de simulação (matriz de mistura, número de sensores, fatores de passo, etc) o comportamento é análogo e as conclusões tiradas são as mesmas que para o exemplo precedente.

Embora as conclusões decorrentes da análise do número de HOS considerados em critérios para BSS apontem para a maior taxa de convergência quando mais momentos são considerados, deve-se mencionar que alguns resultados de simulação indicam que algum ganho de “taxa de convergência local” pode ser possível em algoritmos que retém menos HOS em sua estrutura. Isto pode ser devido à menor complexidade computacional de adaptação quando se utiliza um menor número de HOS, especialmente se aqueles retidos forem relacionados aos de maior

magnitude dados pela expansão de Edgeworth.

Ainda na linha de generalizar os resultados desta seção, deve-se mencionar que foi também observado o mesmo comportamento quando se utilizam modulações mais elaboradas ($\mathfrak{C} > 2$). Entretanto, conforme discutido em [Almeida, 2003], soluções baseadas em estimativas de probabilidades sofrem com mínimos locais quando misturas gaussianas são empregadas. Este fenômeno também foi observado nos casos estudados e o número de mínimos locais aumenta com o aumento do número de *kernels* gaussianos inseridos no modelo, necessitando de mais dados para que as estimativas forneçam bons resultados.

O tipo de análise aqui apresentada fornece alguns elementos importantes que podem ser estendidos para processamento adaptativo cego de uma maneira geral. Este tópico é o assunto da seção seguinte.

6.4 Separação cega de fontes \times desconvolução cega no contexto de HOS

Uma vez que problemas de HOS estão presentes tanto no campo de separação cega de fontes quanto no campo de desconvolução cega vale a pena ressaltar algumas classes de problemas que se aproveitam das características estatísticas de ordem superior em ambos os casos.

Inicialmente, é interessante fazer uma ligação entre a abordagem de separação de fontes e de desconvolução cega para permitir comparações justas. Relembrando que em BSS, para o caso sem ruído e desprezando-se o índice temporal por motivos de simplificação, tem-se

- Sinais nos sensores

$$\mathbf{x} = \mathbf{H}\mathbf{a}$$

- Estimativas das fontes

$$\mathbf{y} = \mathbf{W}^H \mathbf{x},$$

em que

$$\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_K]^T,$$

e K é o número de fontes.

Então, supondo-se independência das fontes presentes no vetor \mathbf{a} , deve-se construir um critério que force a independência estatística das estimativas \mathbf{y} para se obter a separação. Isto consiste na base da análise por componentes independentes.

No caso da desconvolução cega, embora seja um problema SISO, é possível escrever através de equações matriciais, as relações entre os diversos sinais. Seja portanto,

$$\mathbf{x} = \mathcal{H}\mathbf{a},$$

em que \mathbf{a} é o vetor com as amostras temporais de uma dada fonte e \mathbf{x} é o vetor com todas as possibilidades de saída do canal \mathbf{h} , de comprimento L e que gera a matriz de convolução \mathcal{H} , em que $\mathbf{x} = [x(n) \ x(n-1) \ \cdots \ x(n-L-K+1)]^T$. Ao tentar-se considerar todas as possíveis estimativas dos dados em \mathbf{a} , pode-se também escrever:

$$\mathbf{y} = \mathbf{W}^H \mathbf{x},$$

em que

$$\mathbf{y} = [y(n) \ y(n-1) \ \cdots \ y(n-K+1)]^T,$$

e K é o número de sinais e \mathbf{W} é um filtro projetado para recuperar os diferentes atrasos do sinal da fonte.

Assim, como supõe-se que as amostras temporais da fonte são independentes e uma vez que a formulação matemática é a mesma, é possível afirmar que recuperar as K amostras temporais das fontes é possível através da imposição de que as amostras na saída do equalizador sejam independentes.

Esta questão abre um grande leque de possibilidades de tratamento do problema de desconvolução cega, uma vez que a estimativa do atraso de processamento é algo relativamente complexo e de impacto bastante significativo em sistemas reais.

Uma vez que se encontra a “matriz de equalização”, os sinais para todos os atrasos são passíveis de recuperação. A maneira de encontrar o critério capaz de forçar independência é a mesma que para ICA. Uso de funções de contraste, por exemplo, é uma das alternativas possíveis. Este tópico, em particular, tem bastante relação com a questão das HOS, uma vez que as funções contraste são, em geral, definidas em termos de cumulantes [Comon, 1994].

Alguns trabalhos levam em consideração a questão da utilização de igualdade de cumulantes para obtenção da equalização ou identificação cega de canais geralmente através de estratégias não-lineares, uma vez que a utilização de técnicas lineares com algoritmos tipo gradiente estocástico apresentam problemas de convergência associados à questão de mínimos locais [Mo & Shafai, 1994; Chen et al., 1997].

Neste ponto a análise apresentada na seção anterior mostra uma direção na melhoria da convergência através da utilização de mais HOS para construção de critérios de otimização.

De toda forma, deve-se ressaltar que a inclusão de mais estatísticas de ordem superior não necessariamente facilita o critério, uma vez que a estimação direta de cumulantes é custosa e bastante complexa. Todavia, permite uma nova abordagem tanto para o problema de BSS quanto para desconvolução/identificação cega.

6.5 Conclusões e síntese

Neste capítulo, foram apresentadas algumas ferramentas de aproximações de densidade de probabilidade que permitem a análise teórica do uso de cumulantes para separação cega de fontes, notadamente no que se refere à taxa de convergência de algoritmos adaptativos para BSS em tempo real.

A abordagem proposta incluiu a dedução de uma aproximação de fdp por meio de uma expansão, através de base ortonormal, em torno de uma função de referência dada por misturas gaussianas.

A análise teórica foi confirmada por simulações, tendo sido apresentado o caso dos algoritmos MUK e MU-CFPA, considerando-se o caso simples de sinais BPSK e dois sensores.

Em última análise, o capítulo propôs-se a responder, baseando-se em argumentos matemáticos sólidos, sobre o ganho de um algoritmo adaptativo ao se considerar estatísticas de ordem mais elevadas. A compreensão de tal aspecto pode ajudar no projeto de critérios de separação cega de fontes no sentido de melhorar a taxa de convergência que, em muitas aplicações reais, se coloca como um problema bastante acentuado.

Uma contribuição final do capítulo foi a de estender o conceito em pauta para os problemas de desconvolução cega, estudando-os sob a ótica da análise por componentes independentes, o que possibilitaria uma vantagem sobre as abordagens atuais no que se refere aos problemas de estimativa ou obtenção do atraso de processamento.

As questões aqui discutidas, fundamentalmente de caráter mais teórico, fazem parte de uma gama de análises que permitem melhorias nas estratégias de separação cega de fontes. Aspectos mais práticos, relacionados sobretudo às aplicações em processamento espacial e espaço-temporal são tratados no capítulo seguinte.

– “O sábio que não coloca seu saber em prática é uma abelha que não produz mel.”

Ditado persa

7

Aplicações em Sistemas de Mistura Instantânea e Convolutiva

SISTEMAS práticos de comunicação digital buscam continuamente a inclusão de mais usuários de forma a utilizar ao máximo os recursos disponíveis. Esta meta entra em conflito com a capacidade de processamento que os atuais sistemas possuem e pode levar a uma degradação geral de desempenho.

O objetivo deste capítulo é pôr em evidência a importância das ferramentas de separação cega de fontes, em particular dos algoritmos propostos, na melhoria de desempenho em ambientes de processamento multiusuário. Algumas situações típicas de sistemas práticos de múltiplos usuários são consideradas nas simulações e respectivas avaliações de desempenho.

O restante do capítulo é organizado da seguinte forma: a Seção 7.1 é dedicada a avaliar os algoritmos em ambientes cujos sinais são submetidos a interferência de múltiplo acesso majoritariamente; nesta seção casos particulares deste modelo genérico são destacados, fornecendo várias conclusões interessantes sobre os algoritmos propostos. Na Seção 7.2, é mostrada uma aplicação bastante em

evidência no cenário de comunicações móveis e comunicações por pacotes. Na Seção 7.3 um modelo genérico de interferência dos sinais de diversos usuários, incluindo a inter-simbólica, é utilizado para avaliar o desempenho das técnicas de BSS. As conclusões e síntese do capítulo são apresentadas na Seção 7.4.

7.1 Sistemas de mistura instantânea: processamento espacial

Num ambiente multiusuário, quando os sinais estão sendo transmitidos durante um intervalo de tempo suficientemente pequeno em relação às variações do canal, a única interferência de fato atuando é a de múltiplo acesso. Neste caso, o processamento espacial pode ser utilizado para prover a mitigação de tal efeito nos sinais de cada um dos usuários.

Para embasar melhor a avaliação de desempenho em sistemas com processamento multiusuário, considera-se primeiramente alguns aspectos que diferenciam as características de um receptor.

7.1.1 Receptor com sensores simples

O termo *sensores simples*, aqui utilizado, designa o caso no qual o receptor não apresenta uma organização particular dos sensores que o compõem. Desta maneira, o canal entre o k -ésimo usuário (fonte) e o m -ésimo sensor é dado por uma constante complexa que modela o ganho e a fase inseridas em cada um dos sinais. Com isso, a geração dos canais obedece à seguinte regra:

$$[h_{mk}] = a + j \cdot b, \quad (7.1)$$

em que as v.a. a e b são tomadas de uma distribuição gaussiana normalizada e independentes entre si.

Considera-se, então, 4 usuários transmitindo sinais QPSK de potência unitária, os quais são recebidos por 6 sensores nos quais é inserido um ruído aditivo gaussiano com potência dada por $\text{SNR} = 15$ dB. Procurando avaliar uma gama maior de aspectos relativos à dependência do desempenho com a variação do canal, foram utilizadas 100 simulações de Monte Carlo, nas quais a matriz de mistura foi escolhida aleatoriamente para cada uma das tentativas. Os parâmetros de simulação são dados na Tabela 7.1.

Parâmetros de simulação dos algoritmos	
MUK	$\mu = 4.10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$
MU-CFPA	$\mu = 2.10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$ $\sigma_r^2 = 0.1$
MU-CMA	$\mu = 4.10^{-3}$ $\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\varsigma = 0.96$ $\gamma = 10^{-3}$
FMU-CMA	$\mu = 4.10^{-3}$ $\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\mathbf{R}_{xy,k} = \mathbf{I}, \mathbf{d}_{xy,k} = \mathbf{0}_{M \times 1}$ $\varsigma = \zeta = 0.98$ $\gamma_k(0) = 10^{-1}$
DMI	$\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\varsigma = 0.96$

Tabela 7.1: Parâmetros de simulação para processamento multiusuário espacial para 100 canais independentes num sistema de 4 usuários e 6 sensores.

Devido a não garantia de que $\mathbf{H}^H \mathbf{H} = \mathbf{I}$, é realizada uma esferatização dos dados para permitir a utilização do MUK como algoritmo de processamento multiusuário. Esta esferatização é realizada através da utilização de PCA embora outros métodos, tal como a utilização do algoritmo de Schur, possam ser utilizados [Papadias, 2000b].

A Figura 7.1 ilustra o desempenho em termos da evolução do CME para os algoritmos analisados. O algoritmo DMI é colocado somente como uma referência uma vez que há conhecimento total dos sinais enviados por cada usuário. Neste caso também, devido à observação do melhor desempenho do MU-CFPA comparado ao MU-FPA em termos de erro em estado permanente, somente o primeiro é mostrado nas simulações.

Pode-se verificar que o algoritmo MU-CFPA possui a melhor taxa de convergência entre os algoritmos cegos e um desempenho final bastante próximo do DMI, com uma taxa de convergência comparável a este algoritmo. Embora os valores de passo sejam diferentes para a maioria dos algoritmos, aqueles utilizados nas simulações foram selecionados como os valores máximos para os quais os algoritmos convergiam em todas as 100 simulações independentes.

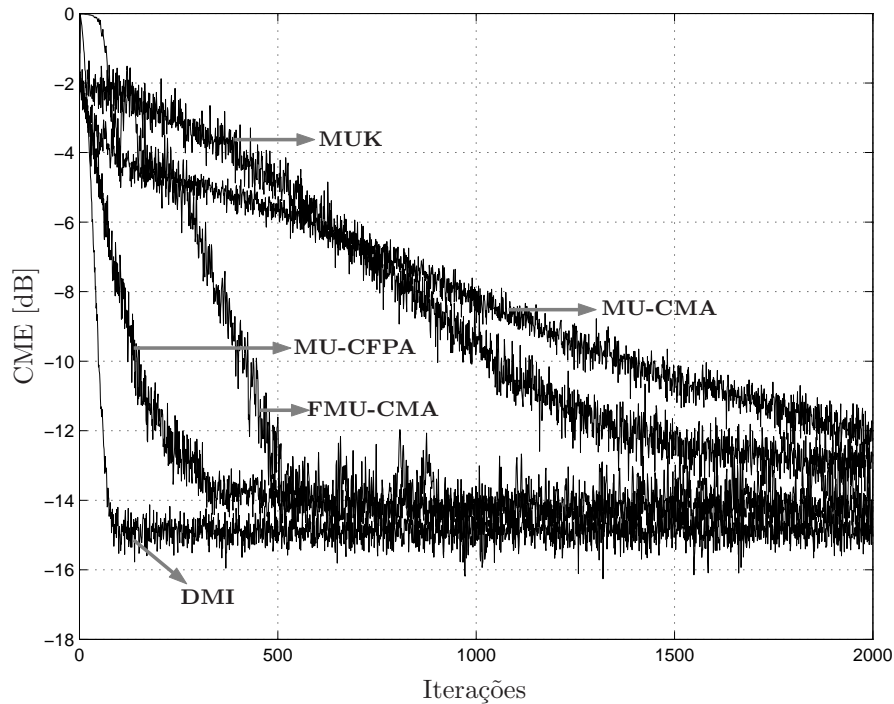


Figura 7.1: Evolução temporal dos erro do módulo constante para processamento espacial com sensores simples em um sistema de 4 usuários e 6 sensores.

Um ponto que deve ser colocado em destaque é o ganho do MU-CFPA em termos de taxa de convergência, mesmo em relação a um algoritmo que utiliza a estrutura recursiva de adaptação como o FMU-CMA. Isto também reforça a questão teórica da influência do número de momentos considerados para a velocidade de adaptação, discutida no Capítulo 6, uma vez que o FMU-CMA utiliza somente as estatísticas de ordem 4, de forma indireta, estimadas de forma recursiva.

Outro ponto interessante de ser avaliado é a obtenção das constelações na saída dos filtros de separação para cada um dos algoritmos. A Figura 7.2 ilustra as constelações para os 10% últimos símbolos obtidos para uma simulação escolhida de forma aleatória de cada um dos algoritmos nas mesmas condições. Os símbolos \bullet , $+$, \circ e \diamond representam as constelações de cada um dos usuários. Vale lembrar que os usuários emitem sinais QPSK de potência unitária.

Pode-se notar que todos os algoritmos cegos, à exceção do MU-CFPA, apresentam a questão da ambiguidade de fase. O MU-CFPA por sua vez preserva a mesma fase do sinal original devido a sua comparação com a densidade de probabilidade do sinal desejado, a qual carrega a informação da fase. Esta questão também está relacionada à utilização de todos os momentos de ordem superior, que retira a ambiguidade relacionada à rotação de fase. De fato, o MU-CFPA sofre do problema de rotação de fase sobre rotações do tipo $l \cdot \frac{c}{2\pi}$, em que $l \in \mathbb{Z}$, as quais só podem ser removidas, em qualquer processamento, com a utilização de modulações

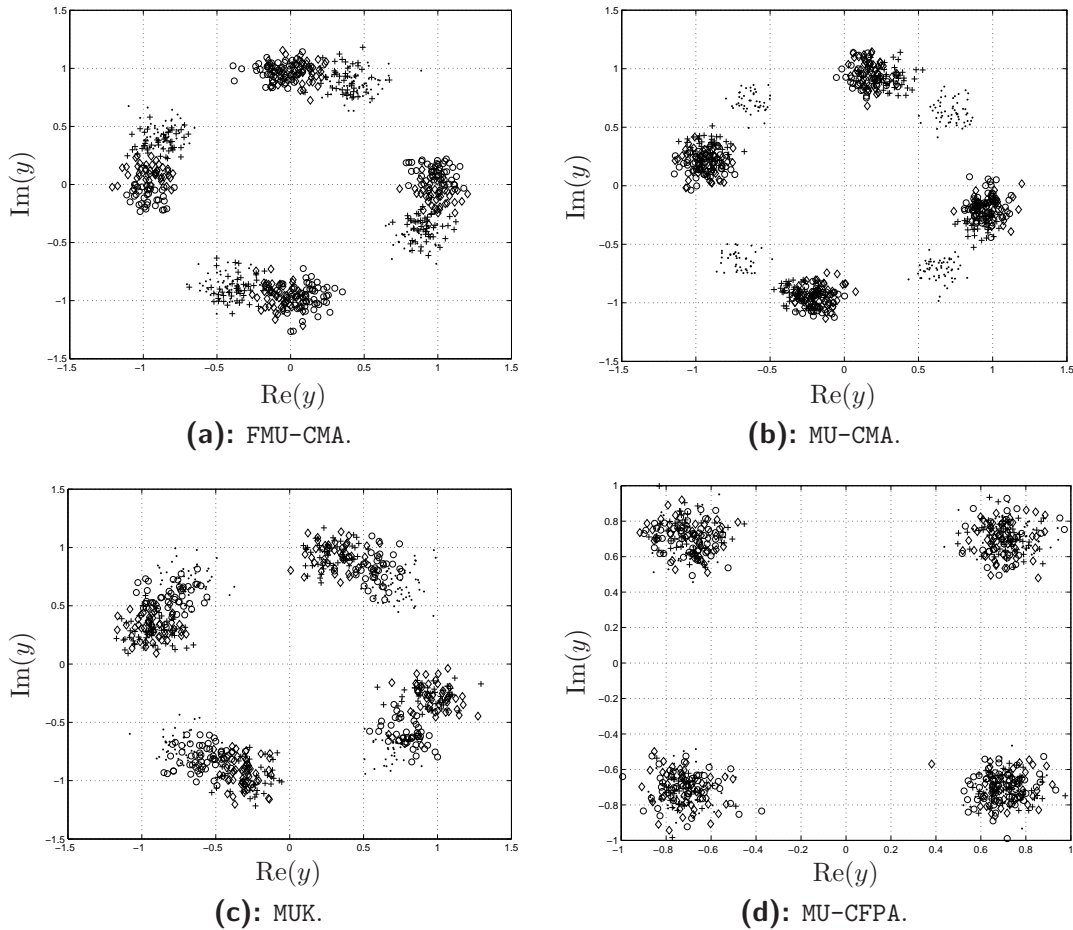


Figura 7.2: Constelações dos 10% símbolos finais para cada um dos algoritmos.

diferenciais codificadas [Proakis, 1995].

Este tipo de contexto é bastante genérico e pode se aplicar a diversos problemas de processamento multiusuário em que os sinais não apresentam interferência inter-simbólica mas o canal pode apresentar um comportamento de variações bastante abruptas em relação aos parâmetros de instantes anteriores. Além disso, alguns casos típicos em comunicação móvel, nos quais os usuários estão sujeitos a ambientes com uma diversidade bastante elevada de posição, são modelados pela mesma abordagem apresentada nesta seção. Quando os sinais em questão não apresentam tal diversidade, ou a apresentam em menor grau, a utilização de receptores que aproveitam as suas características é bastante desejada. Este é o assunto do tópico seguinte.

7.1.2 Receptores com arranjos de antenas

Conforme mencionado no Capítulo 4, antenas adaptativas constituem uma das principais estratégias de recepção para redução da interferência em sistemas de múltiplos usuários.

O arranjo de antenas permite, entre outras coisas, que os sinais sejam combinados de uma forma a explorar a posição espacial do usuário. Para verificar o desempenho de algoritmos cegos para processamento multiusuário é necessário supor que os sinais são provenientes de alguma posição espacial (angular). Nesta aplicação, considera-se a configuração para as posições angulares dos usuários, em função da direção de chegada dos sinais, dada na Tabela 7.2 [Cavalcanti, 1999].

Usuário	DOA (graus)
# 1	1°
# 2	-52°
# 3	29°
# 4	76°

Tabela 7.2: Configuração dos usuários e suas respectivas direções de chegada num sistema que emprega um receptor com um arranjo de antenas.

Utiliza-se então, para detectar os $K = 4$ usuários transmitindo sinais QPSK de potência unitária, um arranjo linear (conforme descrito no Capítulo 4) de $M = 8$ elementos espaçados de meio comprimento de onda e o ruído aditivo é inserido com uma potência calculada a partir de uma $\text{SNR} = 20$ dB.

Uma vez que a utilização do arranjo de antenas fornece uma matriz de mistura geralmente não unitária, o MUK não é utilizado nesta aplicação. Embora um processo de esferatização possa ser utilizado, as características do processamento com antenas adaptativas são desprezadas quando este pré-processamento é realizado. Os parâmetros de simulação para cada um dos algoritmos avaliados são descritos na Tabela 7.3.

A Figura 7.3 mostra a evolução do CME médio para os algoritmos cegos obtidos através de médias em 20 realizações independentes. Novamente, o algoritmo DMI é incluído como referência para os demais.

Parâmetros de simulação dos algoritmos	
MU-CMA	$\mu = 10^{-3}$ $\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\zeta = 0.96$ $\gamma = 10^{-3}$
FMU-CMA	$\mu = 10^{-3}$ $\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\mathbf{R}_{xy,k} = \mathbf{I}, \mathbf{d}_{xy,k} = \mathbf{0}_{M \times 1}$ $\zeta = \zeta = 0.98$ $\gamma_k(0) = 10^{-1}$
MU-CFPA	$\mu = 10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$ $\sigma_r^2 = 0.1$
DMI	$\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\zeta = 0.96$

Tabela 7.3: Parâmetros de simulação para processamento multiusuário em um sistema com 4 usuários empregando um arranjo de antenas com 8 no receptor.

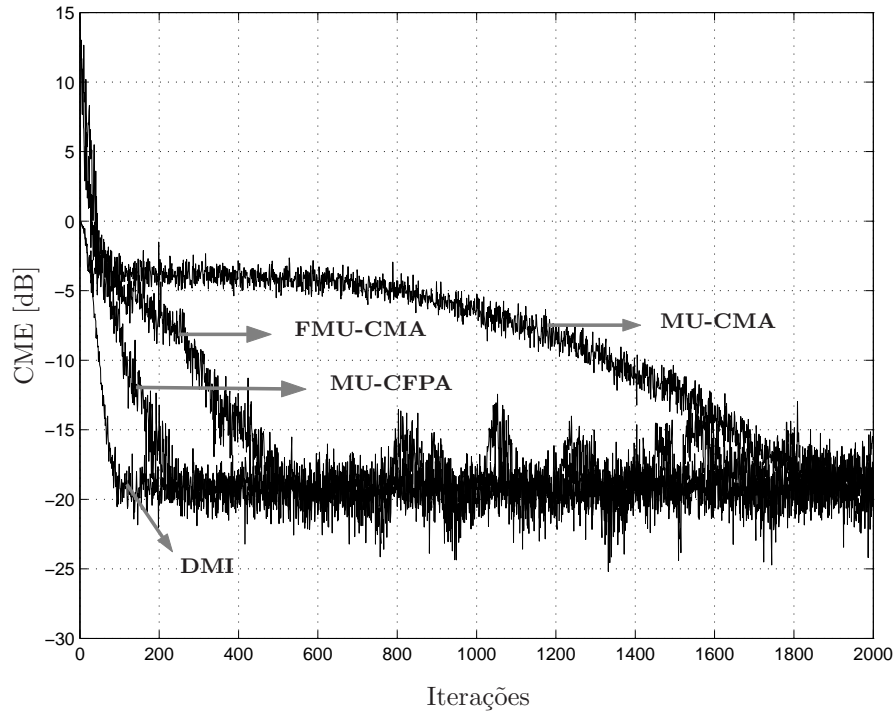


Figura 7.3: Evolução do erro do módulo constante para um sistema de 4 usuários e um arranjo de antenas de 8 elementos.

Observa-se mais uma vez que o MU-CFPA possui um desempenho em termos de taxa de convergência melhor que os outros algoritmos cegos.

Uma métrica importante a ser utilizada quando da utilização de arranjos de antenas no receptor é a obtenção do diagrama de radiação de cada um dos filtros conformadores, que é uma medida da distribuição espacial da energia irradiada pelo arranjo nas diversas direções de chegada. Na verdade, o diagrama de radiação corresponde ao valor absoluto da resposta global do sistema para toda a faixa angular, considerando como canal o vetor de resposta do arranjo. Então, pode-se definir o diagrama de radiação como [Cavalcanti, 1999]:

$$\mathbf{g}_k(\theta) = |\mathbf{w}_k^H \cdot \mathbf{f}(\theta)| \quad \theta \in [0, 2\pi), \quad (7.2)$$

em que \mathbf{w}_k é o filtro conformador para o k -ésimo usuário e $\mathbf{f}(\theta)$ é o vetor de resposta do arranjo.

As Figuras 7.4 a 7.6 ilustram os diagramas de radiação, em coordenadas polares, para os valores finais dos filtros de separação para cada um dos algoritmos. As linhas pontilhadas representam as direções de chegada de cada um dos usuários.

Deve-se notar que os diagramas de radiação ilustram o comportamento mais adequado das estratégias de garantia da decorrelação, baseadas na adaptação do fator de regularização (FMU-CMA) e ortogonalização da resposta global (MU-CFPA), que a decorrelação explícita utilizada pelo MU-CMA. Além disso, deve-se notar que os sinais recuperados pelo MU-CFPA apresentam um fator de escala, ressaltando neste caso a ambiguidade da técnica de BSS quanto a um fator de escala. Embora não colocada de forma explícita, deve-se mencionar que, também neste cenário os algoritmos baseados em recuperação de sinais com módulo constante obtêm as estimativas das fontes com uma rotação de fase, enquanto o MU-CFPA não apresenta este comportamento devido às suas características já descritas anteriormente.

A separação de usuários a partir de informações sobre sua posição angular apresenta a limitação de que os usuários devem estar distantes o suficiente para que a resolução do arranjo seja suficiente para realizar a separação. Isto se reflete no fato do diagrama de radiação necessitar possuir lóbulos bastante estreitos para não inserir ganhos nas direções que contenham sinais provenientes de outros usuários.

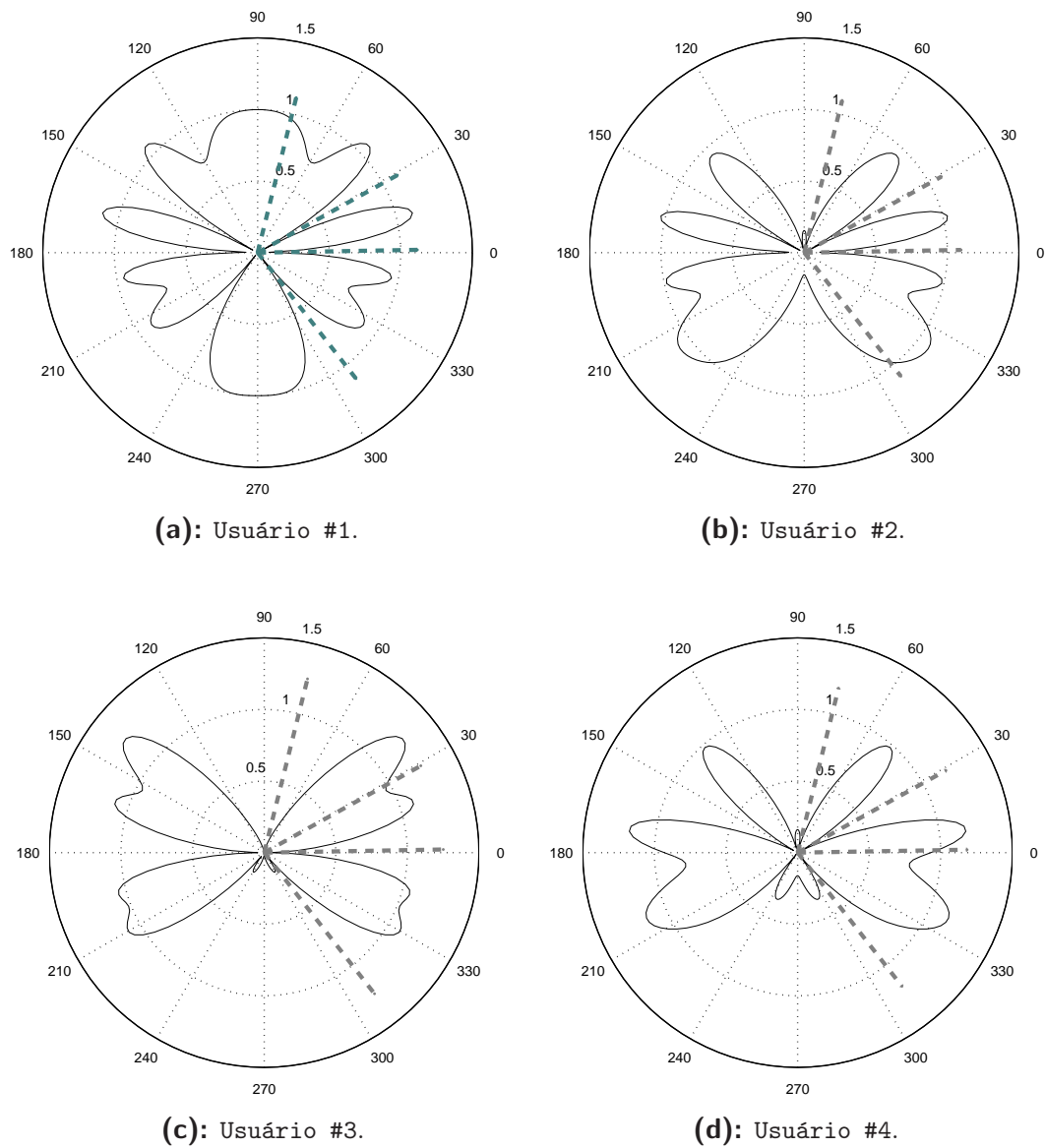


Figura 7.4: Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o MU-CMA.

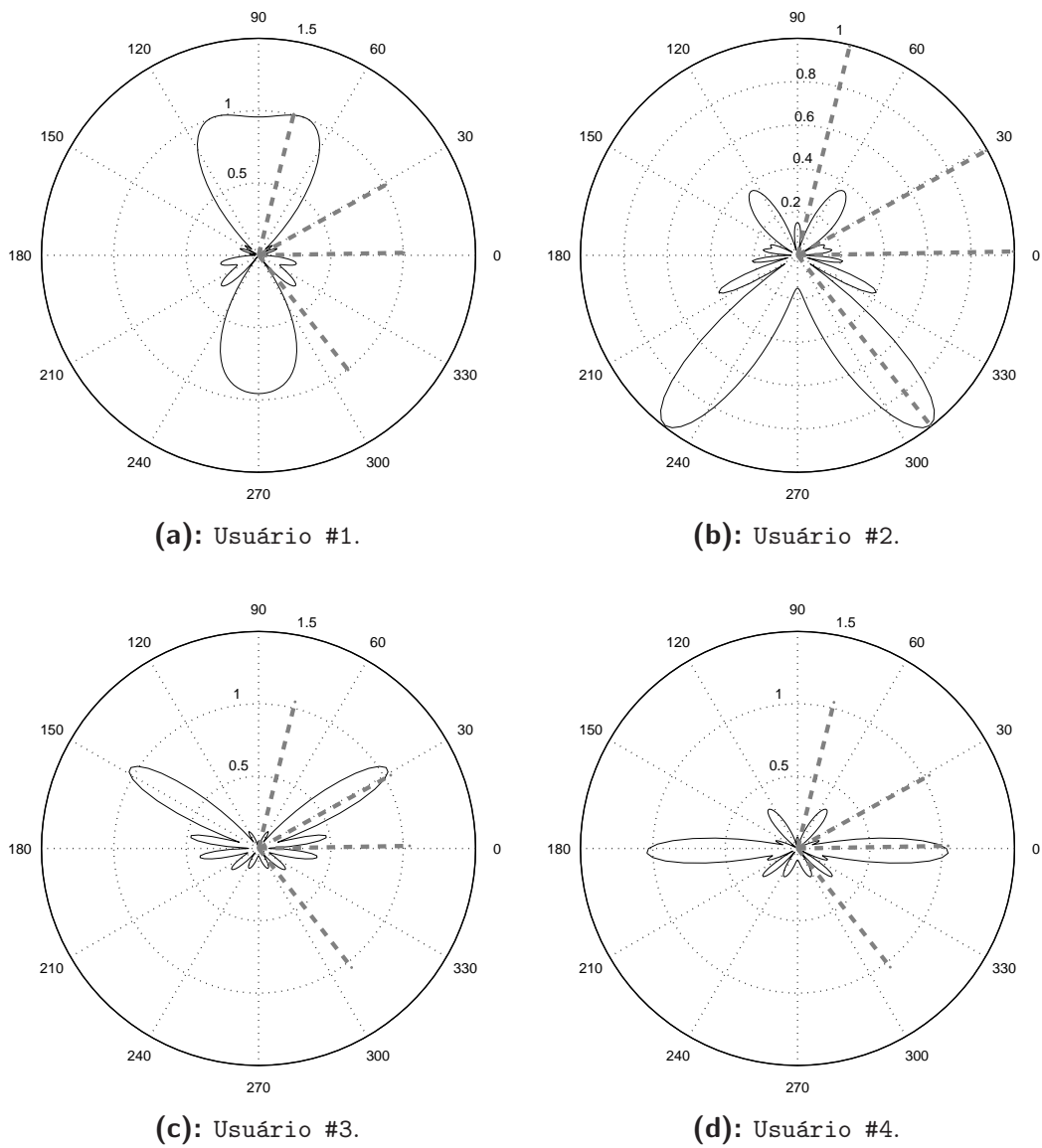


Figura 7.5: Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o FMU-CMA.

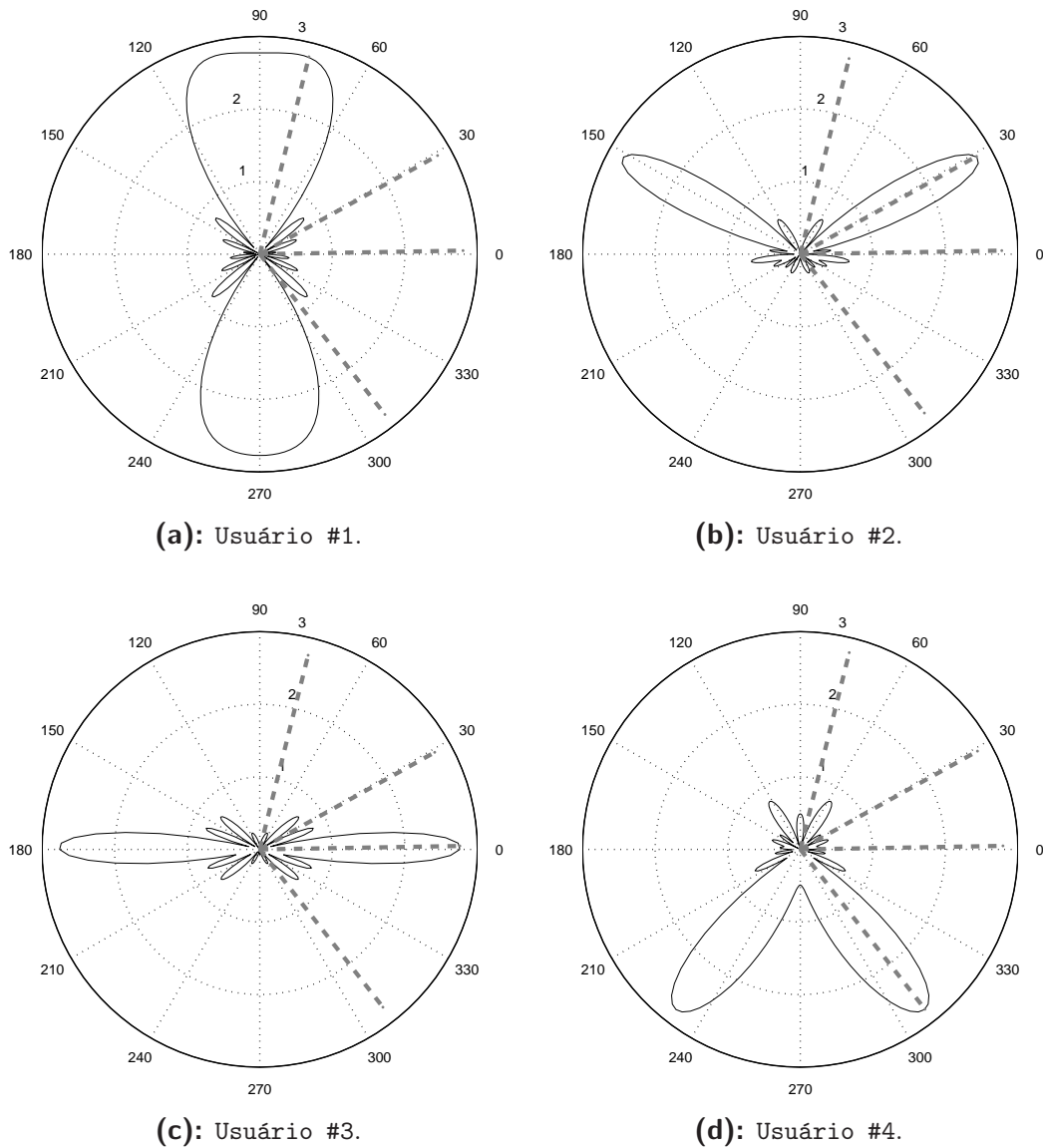


Figura 7.6: Diagramas de radiação para os 4 usuários de um sistema de processamento multiusuário utilizando o MU-CFPA.

7.1.3 Estimação do número de usuários

Até então foi suposto que o número de usuários K é conhecido. Sem este conhecimento, o projeto dos filtros espaciais apresenta problemas pois, se o número de filtros espaciais for menor que o de usuários, haverá perda de informação. No caso de mais filtros que usuários,

múltiplas cópias de um usuário estarão presentes no sistema aumentando a interferência de múltiplo acesso e dificultando a tarefa de separação e identificação das fontes.

Visando uma generalização das estratégias, técnicas de projeção em subespaços podem ser utilizadas para estimar o número de usuários a partir de hipóteses simples com baixa complexidade computacional. Uma vez que o número de sensores deve ser maior ou igual ao número de fontes, uma abordagem baseada na utilização da projeção nos subespaços dos sinais e do ruído é utilizada para determinação do número de usuários ativos no sistema de separação de fontes [Paraschiv-Ionescu et al., 2002].

Relembrando a expressão do sinal recebido, tem-se

$$\mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{v}(n),$$

na qual se considera que o ruído aditivo $\mathbf{v}(n)$ é inserido somente nos sensores.

A matriz de autocorrelação do sinal nos sensores é dada por

$$\begin{aligned} \mathbf{R}_x &= \mathbb{E} \{ \mathbf{x}(n)\mathbf{x}^H(n) \} \\ &= \mathbf{H}\mathbb{E} \{ \mathbf{a}(n)\mathbf{a}^H(n) \} \mathbf{H}^H + \mathbb{E} \{ \mathbf{v}(n)\mathbf{v}^H(n) \}, \end{aligned} \quad (7.3)$$

e utilizando as hipóteses consideradas até então, ou seja, de que os usuários fornecem sinais independentes e o ruído é gaussiano e de espectro branco, tem-se

$$\mathbb{E} \{ \mathbf{a}(n)\mathbf{a}^T(n) \} = \mathbf{R}_a = \mathbf{I}_K \quad (7.4a)$$

$$\mathbb{E} \{ \mathbf{v}(n)\mathbf{v}^T(n) \} = \mathbf{R}_v = \sigma_v^2 \mathbf{I}_M, \quad (7.4b)$$

em que \mathbf{I}_K e \mathbf{I}_M são as matrizes identidade de ordem K e M , respectivamente.

A partir da Equação (7.4), distingue-se os subespaços do sinal e subespaço do ruído na Equação (7.3). Daí, ao ordenar-se em ordem decrescente os autovalores da matriz de autocorrelação, obtém-se a seguinte relação:

$$\underbrace{\lambda_1 \geq \lambda_2 \geq \cdots \lambda_K}_{\text{subespaço de sinal}} > \underbrace{\lambda_{K+1} \approx \cdots \approx \lambda_M}_{\text{subespaço de ruído}} = \lambda_{\text{ruído}}, \quad (7.5)$$

em que λ_i é o i -ésimo autovalor de \mathbf{R}_x em ordem decrescente.

Então, determinar o número de usuários consiste em determinar o *rank* do subespaço de sinal, ou seja, $\text{rank}(\mathbf{H}\mathbf{H}^H)$, suposto completo, que significa, no contexto de recuperação cega de sinais, determinar a multiplicidade da matriz de autocorrelação do sinal recebido, antes de um possível pré-branqueamento (esferatização).

Uma questão relacionada a esta estratégia é a necessidade de estimação da matriz de autocorrelação. Desta forma, uma vez que é desejado que tal estimativa seja realizada em tempo real, algumas amostras são utilizadas para garantir uma boa estimativa da matriz \mathbf{R}_x .

Já que a busca deve ser realizada para encontrar a multiplicidade dos autovalores, a decomposição em valores singulares pode ser utilizada (SVD, *Singular Value Decomposition*). O fato de que a SVD possui um grande número de soluções [Golub & Loan, 1996] permite a utilização de algoritmos que têm uma complexidade computacional de até $\mathcal{O}(K^2)$, o que não coloca uma grande limitação para aplicação em sistemas reais.

Aqui, o procedimento adotado é dado pela seguinte seqüência de passos:

1. cálculo da matriz de autocorrelação através da equação

$$\mathbf{R}_x(n) = \varsigma \mathbf{R}_x(n-1) + (1 - \varsigma) \mathbf{x}(n) \cdot \mathbf{x}^H(n),$$

em que ς é um fator de esquecimento;

2. aplicação da SVD;
3. ordenação em ordem decrescente dos autovalores;
4. partindo do último para o primeiro, os autovalores são comparados e aqueles que diferem uns dos outros de um fator de até 2% são considerados iguais;
5. o número de autovalores diferentes é assumido ser igual a $K + 1$.

Vale ressaltar que a estimação em tempo real dos autovalores é possível quando redes neurais artificiais são empregadas e, neste caso, a SVD é desnecessária [Paraschiv-Ionescu et al., 2002]. Entretanto, o ganho em redução da complexidade não existe pois a complexidade computacional do cálculo da SVD é substituído por uma maior complexidade no dispositivo de separação, o que também implica na utilização de algoritmos com outras características, como por exemplo minimização da entropia, para separação de fontes.

Um ponto de grande apelo prático é relacionado à inserção/retirada de usuários no sistema. Se esta alternativa é considerada, a abordagem mais apropriada é observar a “fronteira” entre o subespaço de sinal e o subespaço de ruído, ou seja, comparar λ_K e λ_{K+1} . Se um usuário é retirado do sistema, $\lambda_K \approx \lambda_{K+1}$, o que implica que o K -ésimo autovalor “migra” para o subespaço de ruído. No caso da inserção, ocorre o contrário e λ_{K+1} “migra” para o subespaço do sinal. Isto torna o método adequado para ser aplicado em sistemas nos quais o número de usuários é variante no tempo, tal como os atuais sistemas de comunicação sem fio. Apesar de mais genérico, esta abordagem aumenta a complexidade computacional uma vez que a estimação do número de usuários deve ser realizada após um pré-determinado número de amostras para (re)avaliação do número de usuários ativos no sistema.

A simulação a seguir ilustra a evolução da estimativa do número de usuários ativos em um sistema a partir da utilização do método descrito anteriormente. Tentando generalizar as conclusões, são utilizados dois tipos de situação:

- receptor com sensores simples – canais escolhidos aleatoriamente;
- receptor com arranjo de antenas – canal dependente das direções de chegada dos sinais dos usuários.

A Figura 7.7 mostra a evolução temporal da estimativa do número de usuários em função do número de iterações utilizadas para estimativa da matriz de autocorrelação. Neste caso utilizou-se $K = 4$, $M = 8$ e $\zeta = 0.97$ (fator de esquecimento) em ambos os casos (antenas adaptativas e sensores simples).

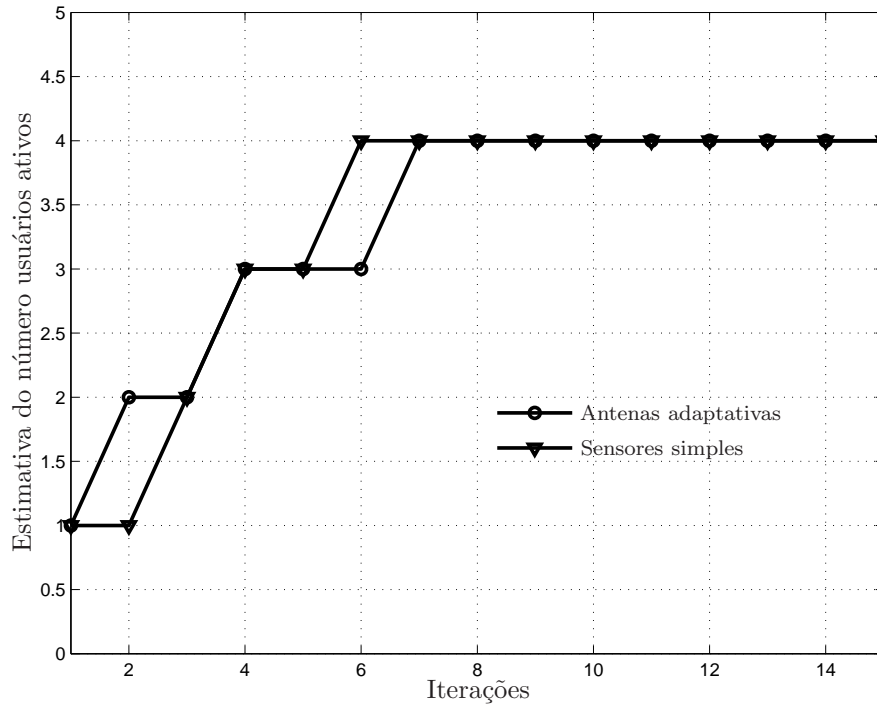


Figura 7.7: Evolução da estimativa do número de usuários ativos no sistema em função do número de iterações utilizadas para estimativa da matriz de autocorrelação.

Claramente, poucas iterações são necessárias para a estimativa correta do número de usuários em ambos os cenários. Isto indica que as técnicas de separação de fontes podem ser empregadas mesmo sem o conhecimento do exato número de usuários no sistema. Na prática uma inicialização com um número máximo de usuários pode ser realizada e, após um certo número de iterações, pode-se reavaliar a estimativa do número de usuários ativos.

Logicamente, a estimativa depende do número de usuários e também do número de sensores presentes no receptor. A Figura 7.8 ilustra o número mínimo de iterações necessárias para estimação de $K = 8$ usuários em função do número de sensores utilizados no receptor, para ambas as configurações de sensores simples e arranjos de antenas.

A partir da observação de que a complexidade de informação é pequena, tal método apresenta-se como candidato potencial para compor métodos de processamento multiusuário que necessitam da estimação em tempo real do número de usuários ativos no sistema.

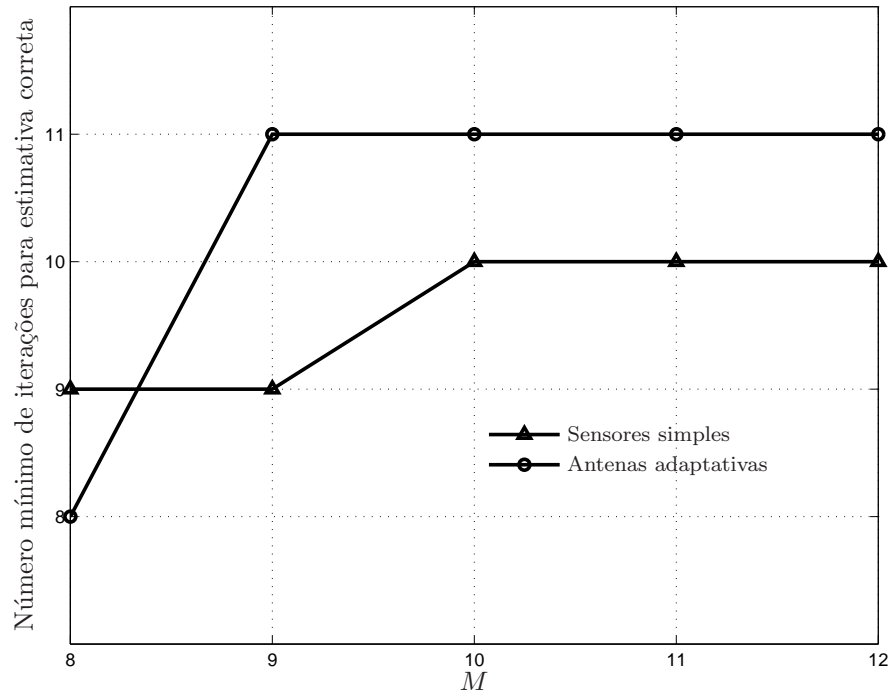


Figura 7.8: Número mínimo de iterações necessárias para a correta estimativa de $K = 8$ usuários em função do número de sensores utilizados.

7.2 Resolução de colisão em redes de pacotes *Slotted ALOHA*

Um outro problema prático no qual pode ser vislumbrada a aplicação de técnicas de separação de fontes é o de *resolução de colisão de pacotes* em redes de comunicação digital.

Em sistemas de transmissão de dados por pacotes, os usuários utilizam *slots* de tempo pré-definidos para enviar dados a uma estação base. Estes pacotes devem ser recuperados na estação base e um *reconhecimento* é enviado de volta ao usuário quando o pacote é recebido sem erros. Entretanto, quando dois ou mais usuários tentam acessar a estação base no mesmo *slot*, os pacotes colidem e há erro de detecção, causando um descarte dos pacotes recebidos.

Este tipo de problema é geralmente solucionado na camada de acesso, chamada de MAC (*Medium Access Control layer*), utilizando-se protocolos que demandam, após um certo tempo, que os usuários que não receberem o reconhecimento de recepção de seu pacote devem retransmiti-lo. Entretanto, este tipo de solução causa uma grande sobrecarga ao sistema uma vez que o número de retransmissão aumenta a probabilidade de novas colisões [Zhao

et al., 2001; Orange, 1998]. Para evitar operações *full-duplex*, o reconhecimento é realizado em uma frequência diferente daquela dos sinais dos usuários.

O sistema ALOHA, desenvolvido por um grupo de pesquisadores da *University of Hawaii* na década de 1970, emprega um satélite repetidor que envia os pacotes recebidos dos vários usuários que acessam o satélite. Neste caso, os usuários podem monitorar as transmissões do satélite e verificar se seus pacotes foram transmitidos com sucesso. Há basicamente dois tipos de sistemas ALOHA: sincronizado ou *slotted* e não-sincronizado ou *unslotted*. Em sistemas não-sincronizados, os usuários iniciam a transmissão de um pacote num instante de tempo arbitrário. No caso sincronizado, os *slots* de tempo possuem instantes de início e fim bem definidos [Proakis, 1995]. Nesta tese, será considerado somente o caso sincronizado que é denominado *Slotted ALOHA* (S-ALOHA).

Técnicas de processamento de sinais podem ser utilizadas para solucionar a questão da colisão dos pacotes sem a necessidade de protocolos de retransmissão, provendo uma diminuição no congestionamento da rede através da detecção e recuperação de pacotes que são recebidos com erros por meio de estratégias de separação de fontes. Técnicas supervisionadas, semi-cegas ou mesmo cegas, inclusive baseadas em ICA, podem ser utilizadas para a resolução de colisão, ver por exemplo [Zhao et al., 2001; Cavalcanti & Romano, 2000; Özgül, 2002] e suas referências. As técnicas baseadas em processamento de sinais têm um apelo bastante significativo quando são consideradas redes *ad hoc*, nas quais não há uma unidade de controle central, dificultando a implementação de protocolos eficientes para a resolução de pacotes.

Então, para avaliar o desempenho de técnicas de separação de fontes em um sistema de rede de pacotes com um cenário de tráfego de vários usuários transmitindo simultaneamente, utiliza-se a medida de vazão em uma rede S-ALOHA

A vazão de sistemas S-ALOHA convencionais¹ é dada por [Proakis, 1995]:

$$\mathcal{T}(\sigma) = \sigma \cdot \exp(-\sigma) \quad (7.6)$$

em que σ é a carga oferecida em pacotes/*slot*. É conhecido na literatura que o máximo valor da vazão na Equação (7.6) é em torno de 0.3684 pacotes/*slot* [Carleial & Hellman, 1975; Proakis, 1995].

Um ponto relevante em sistemas S-ALOHA é a estabilidade. Com um número infinito de usuários, o sistema é estável até uma certa carga oferecida. Desta forma, é possível que o número de retransmissões cresça indefinidamente causando instabilidade no sistema [Carleial & Hellman, 1975].

¹A título de curiosidade, a vazão de sistemas *Unslotted ALOHA* é dada por [Proakis, 1995]:

$$\mathcal{T}(\sigma) = \sigma \cdot \exp(-2\sigma),$$

possuindo uma vazão máxima de 0.184 pacotes/*slot*, valor relativamente baixo que demonstra sua ineficiência.

Em [Ghez et al., 1989], é proposto um sistema S-ALOHA controlado que permite aumentar a vazão sem trazer instabilidade. Tal controle é baseado no monitoramento do nível de pedidos de retransmissões de pacotes. Para este sistema S-ALOHA controlado, é considerado que as chegadas dos pacotes obedecem a uma distribuição de Poisson e é mostrado em [Ghez et al., 1989] que a vazão, para uma carga oferecida de σ pacotes/*slot*, é dada por:

$$\mathcal{T}(\sigma) = \exp(-\sigma) \cdot \sum_{m=1}^{\infty} \frac{\sigma^m}{m!} \mathfrak{L}_m \quad (7.7)$$

na qual $\mathfrak{L}_m = \sum_{k=1}^m k \cdot P(k|m)$ é o número médio de pacotes recebidos com sucesso, dado que m pacotes foram transmitidos, e $P(k|m)$ é a probabilidade de recebimento correto de k pacotes dado que m pacotes foram transmitidos. Pode-se então, generalizar esta abordagem para o caso da utilização de um arranjo de antenas no receptor por [Ward & Compton, 1993]:

$$\mathcal{T}(\sigma) = \exp(-\sigma) \cdot \sum_{m=1}^{M+\ell} \frac{\sigma^m}{m!} \mathfrak{L}_{m|M} \quad (7.8)$$

em que $\mathfrak{L}_{m|M} = \sum_{k=1}^m k \cdot P(k|m, M)$ é o número médio de pacotes com sucesso, dado que m pacotes foram transmitidos, e que o arranjo de antenas tem M elementos. Neste caso, $P(k|m, M)$ é a probabilidade de recepção correta de k pacotes dado que m pacotes foram transmitidos e um arranjo de M elementos é utilizado no receptor. Quando a vazão não apresenta ganhos, ou apresenta ganhos desprezíveis, à medida que são incluídos mais pacotes, ℓ é selecionado para prover mais vazão ao sistema.

O número de termos no somatório na Equação (7.8) depende do número de sinais recuperáveis. Esta questão é tratada de maneira diferente em alguns trabalhos. Em [Ward & Compton, 1993] é considerado que não há completa transmissão de sucesso (recuperação de *todos* os pacotes enviados sem erros) para $K > M$, ou seja, quando há mais pacotes sendo detectados que a capacidade de resolução do arranjo. Neste caso, para $m > M$, tem-se $P(k|m, M)$ ou $\ell = 0$ na Equação (7.8). No citado trabalho, é também proposto um critério MMSE para recuperação dos pacotes que sofreram colisão.

Nesta tese, é utilizada a mesma abordagem que em [Cavalcanti & Romano, 2000], quando medidas de erro de bit são utilizadas para derivar probabilidades de sucesso. Esta consideração permite relaxar a hipótese utilizada em [Ward & Compton, 1993] e permite a avaliação do desempenho dos algoritmos para processamento multiusuário.

Pode-se então utilizar o algoritmo MU-CFPA para adaptação dos filtros conformadores de maneira a identificar e separar os pacotes dos diversos usuários e evitar a retransmissão de pacotes [Cavalcante et al., 2003, em *WWC'2003*].

Para esta avaliação, foram escolhidos dois cenários típicos em comunicações móveis: ambiente de diversidade (DIV) e em linha de visada (LOS, *Line-Of-Sight*). O ambiente de diversidade corresponde àquele denotado anteriormente de “sensores simples”. Neste caso, não há uma dependência da direção de chegada dos sinais e o canal é dado por uma escolha como a da Equação 7.1. Entretanto, é considerado um fator de redução de $\frac{1}{\sqrt{2}}$. No caso de linha de visada, o ambiente corresponde àquele que utiliza antenas adaptativas no receptor, no qual há a dependência da direção de chegada dos usuários.

Para o problema de resolução de pacotes, foram considerados apenas os algoritmos MU-CFPA e FMU-CFPA, por possuírem as melhores taxas de convergência, em torno de 500 símbolos. Estas taxas são necessárias uma vez que os pacotes geralmente apresentam blocos de dados em torno de 500 amostras. Além dos dois algoritmos cegos, o critério MMSE também é utilizado como um valor de referência.

Desta forma, são utilizados sinais QPSK e uma SNR = 20 dB. Quando da consideração do ambiente LOS, um setor de 180° com usuários uniformemente distribuídos é considerado. O cálculo das probabilidades de erro é realizado sobre 200 realizações independentes. Os parâmetros utilizados para cada algoritmo são listados na Tabela 7.4.

Parâmetros de simulação para os algoritmos	
MU-CFPA	$\mu = 2 \cdot 10^{-3}$ $\mathbf{W}(0) = \mathbf{W}^e(0) = \mathbf{I}$ $\sigma_r^2 = 0.1$
FMU-CMA	$\mu = 4 \cdot 10^{-3}$ $\mathbf{W}(0) = \mathbf{R}_y(0) = \mathbf{P}(0) = \mathbf{I}$ $\mathbf{R}_{xy,k} = \mathbf{I}, \mathbf{d}_{xy,k} = \mathbf{0}_{M \times 1}$ $\varsigma = \zeta = 0.97$ $\gamma_k(0) = 10^{-1}$

Tabela 7.4: Parâmetros de simulação para resolução de pacotes em uma rede S-ALOHA.

A Figura 7.9 ilustra a probabilidade de eficiência máxima, ou seja, a probabilidade de K pacotes serem detectados corretamente dado que foram enviados K pacotes.

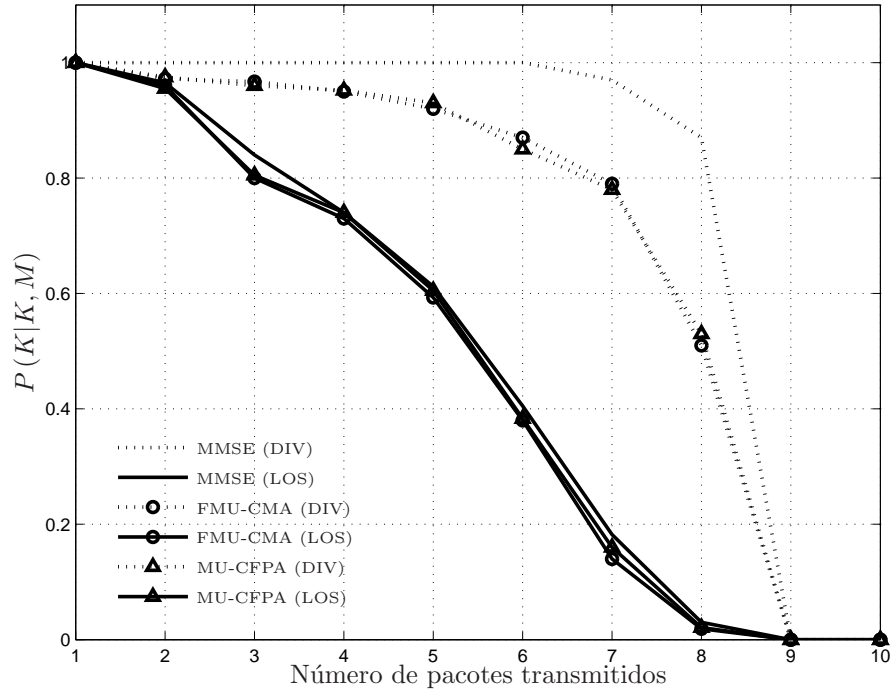


Figura 7.9: Probabilidade de atingir transmissão com 100% de eficiência $P(K|K, M)$ para $M = 8$ antenas.

Pode-se verificar que os algoritmos MU-CFPA e FMU-CMA apresentam praticamente o mesmo desempenho para ambas as situações, LOS e DIV. Além disso, os dois algoritmos também conseguem obter desempenhos próximos o suficiente da solução MMSE para ambos os cenários (LOS e DIV). Nota-se também que, conforme esperado, as probabilidades de acerto tendem a zero quando $K > M$.

A Figura 7.10 mostra a vazão em função da carga oferecida a partir das probabilidades de sucesso calculadas numericamente para cada um dos algoritmos.

Observa-se que o desempenho em termos da vazão para os dois algoritmos cegos é bastante similar, uma vez que as probabilidades de ambos não diferem consideravelmente. Além disso, ambos atingem uma fração bastante significativa da solução MMSE com uma economia das seqüências de treinamento de todos os usuários. Este fato, adicionado à questão de que em redes *ad hoc* torna-se bastante complexo determinar quaisquer seqüências conhecidas *a priori* numa estrutura não-hierárquica, indica a potencialidade de uma das técnicas cegas em questão.

Entretanto, deve ser ressaltado que o algoritmo FMU-CMA utiliza inversões de matrizes para cálculo dos filtros de cada um dos usuários, o que resulta numa complexidade computacional de ordem mais elevada que a do algoritmo MU-CFPA, que trabalha com uma complexidade da ordem do algoritmo LMS. Este ponto é de extrema importância quando se

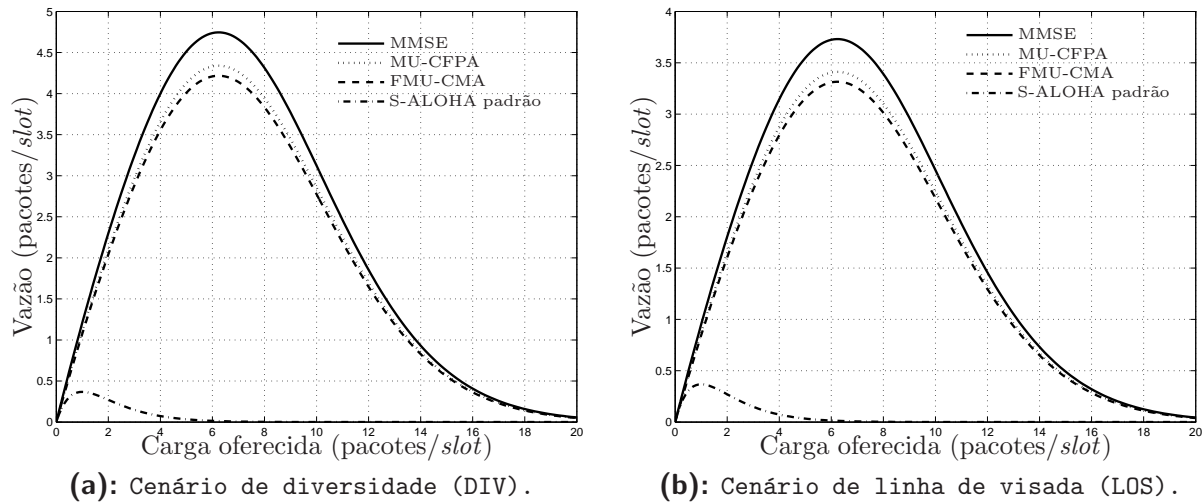


Figura 7.10: Resultados de vazão em função da carga oferecida para os cenários de diversidade (DIV) e linha de visada (LOS).

cogita implementações práticas.

Neste tipo de processamento não foi considerada a interferência inter-simbólica presente em grande número de sistemas práticos. Este tópico é tratado na seção seguinte.

7.3 Sistemas de mistura convolutiva: processamento espaço-temporal

Um caso de relevante interesse na área de comunicações móveis é relativo à supressão de interferência quando os sinais dos usuários estão sujeitos, além de interferência de múltiplo acesso, à interferência inter-simbólica. Conforme discutido no Capítulo 4, esta situação corresponde à do processamento espaço-temporal, quando se utiliza vários sensores para detecção dos sinais. O esquema utilizado para a separação dos sinais é mostrado na Figura 4.4 na página 93.

Neste exemplo são considerados dois usuários com mesma potência que transmitem simultaneamente sinais QPSK de potência unitária. São considerados 2 percursos para cada usuário cujos atrasos relativos, em termos de período de símbolo e direções de chegada, são mostrados na Tabela 7.5. É inserido ruído aditivo em cada um dos sensores sob uma SNR = 30 dB. Além disso, é utilizado um filtro de transmissão dado por um cosseno levantado com fator de excesso de faixa igual a 0.35.

Usuário	Percurso 1		Percurso 2	
	Atraso ($\times T$)	DOA (rad)	Atraso ($\times T$)	DOA (rad)
# 1	0.1	$\frac{2\pi}{5}$	1.1	$-\frac{\pi}{3}$
# 2	0.4	$\frac{\pi}{7}$	1.2	$-\frac{\pi}{6}$

Tabela 7.5: Configuração do sistema para processamento espaço-temporal considerando 2 usuários com 2 percursos cada.

No receptor, são considerados $M = 4$ elementos num arranjo linear com espaçamento de $\frac{\lambda}{2}$ e filtros temporais com $L = 2$ coeficientes para eliminação da ISI, além de $\Delta = 2$ amostras de atraso para decorrelação dos sinais. Outra consideração é de que a recuperação dos sinais para ambos os usuários ocorre para um atraso de decisão $\ell = 0$.

São empregados para efeito de comparação três algoritmos, MU-CMA, MU-FPA e o LMS supervisionado. Os parâmetros de simulação de cada um dos algoritmos encontram-se descritos na Tabela 7.6. Vale ressaltar que o MU-FPA foi utilizado em detrimento do MU-CFPA, pelo fato de que este último não apresenta, ainda, uma versão para processamento espaço-temporal. Outros algoritmos (por exemplo, FMU-CMA e MUK) também não foram incluídos na simulação pela mesma razão.

Parâmetros de simulação para os algoritmos	
MU-CMA	$\mathcal{W}_k(0) = \mathbf{R}_{y,\ell}(0) = \mathbf{P}_\ell(0) = \mathbf{I}$ $\mu = 5 \cdot 10^{-3}$ $\varsigma = 0.99$
MU-FPA	$\mathcal{W}_k(0) = \mathbf{R}_{y,\ell}(0) = \mathbf{P}_\ell(0) = \mathbf{I}$ $\mu = 10^{-3}$ $\varsigma = 0.99$ $\sigma_r^2 = 0.1$
LMS	$\mathcal{W}_k(0) = \mathbf{I}$ $\mu = 10^{-2}$

Tabela 7.6: Parâmetros de simulação para processamento espaço-temporal.

A figura de mérito utilizada é a medida da interferência residual (RI, *Residual Interference*) de cada usuário, a qual é definida como [Brossier, 1997]:

$$\text{RI}_k(n) = \frac{\left| \sum_i |g_{k,i}(n)| - \max_i |g_{k,i}(n)| \right|}{\max_i |g_{k,i}(n)|}, \quad (7.9)$$

em que $g_{k,i}$ é o i -ésimo elemento do vetor de resposta global do k -ésimo usuário dado, para o

caso espaço-temporal, por

$$\mathbf{g}_k(n) = \mathcal{W}_k^H \cdot \mathcal{H}, \quad (7.10)$$

na qual \mathcal{H} é matriz multicanal espaço-temporal dada por

$$\mathcal{H} = [\mathcal{H}_1 \mid \mathcal{H}_2 \mid \cdots \mid \mathcal{H}_K]. \quad (7.11)$$

A Figura 7.11 ilustra a evolução temporal da RI média, ponderado por todos os usuários do sistema, para os três algoritmos em questão, obtidas através de uma média de 20 simulações independentes.

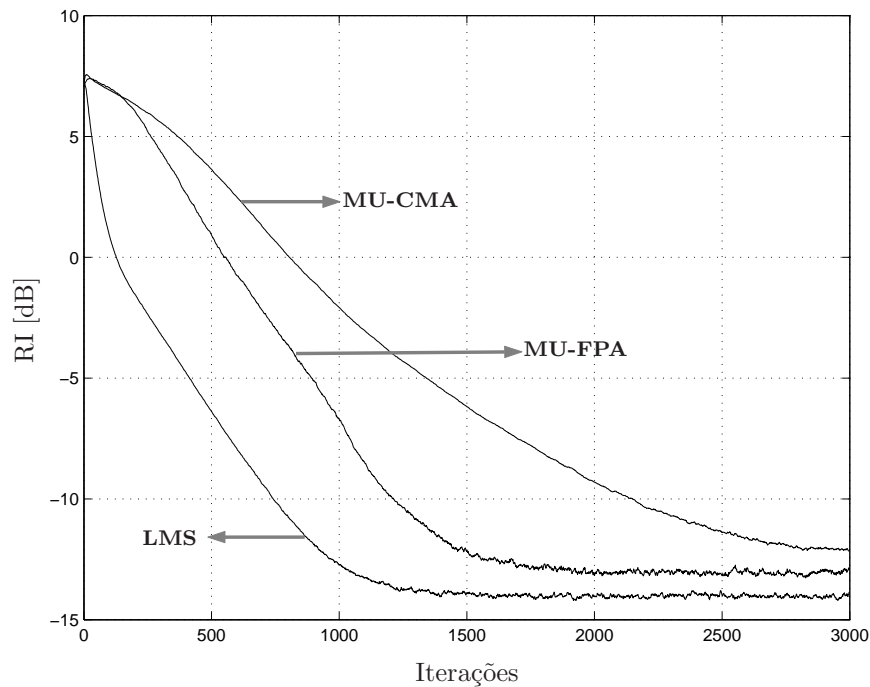


Figura 7.11: Evolução temporal da interferência residual média para os diversos algoritmos utilizados no processamento espaço-temporal.

Nota-se um melhor desempenho por parte do MU-FPA que pelo MU-CMA, em termos de velocidade de convergência e também de interferência residual menor. Além disso, o MU-FPA apresenta uma interferência residual bastante próxima da do LMS (algoritmo supervisionado) com uma taxa de convergência muito similar para os dois algoritmos. Este comportamento permite considerar tal estratégia em situações na qual a disponibilidade da seqüência de treinamento não é possível ou de custo proibitivo.

Para ilustrar o comportamento da dinâmica das estimativas dos sinais dos usuários, são mostradas as evoluções das constelações para cada um dos algoritmos nas Figuras 7.12, 7.13 e 7.14. Os símbolos \bullet e $+$ são os sinais recuperados para o usuário 1 e 2, respectivamente.

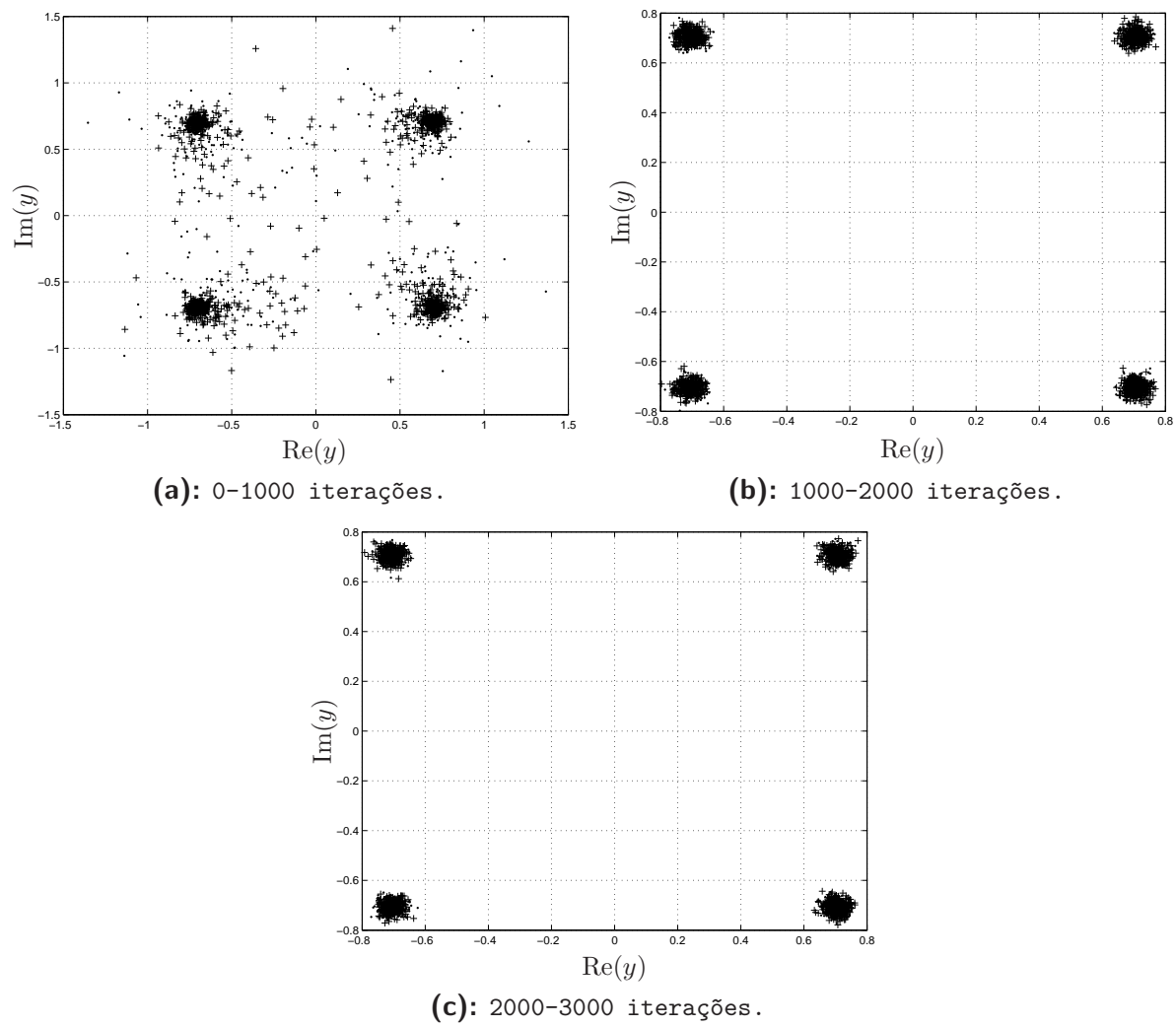


Figura 7.12: Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o LMS.

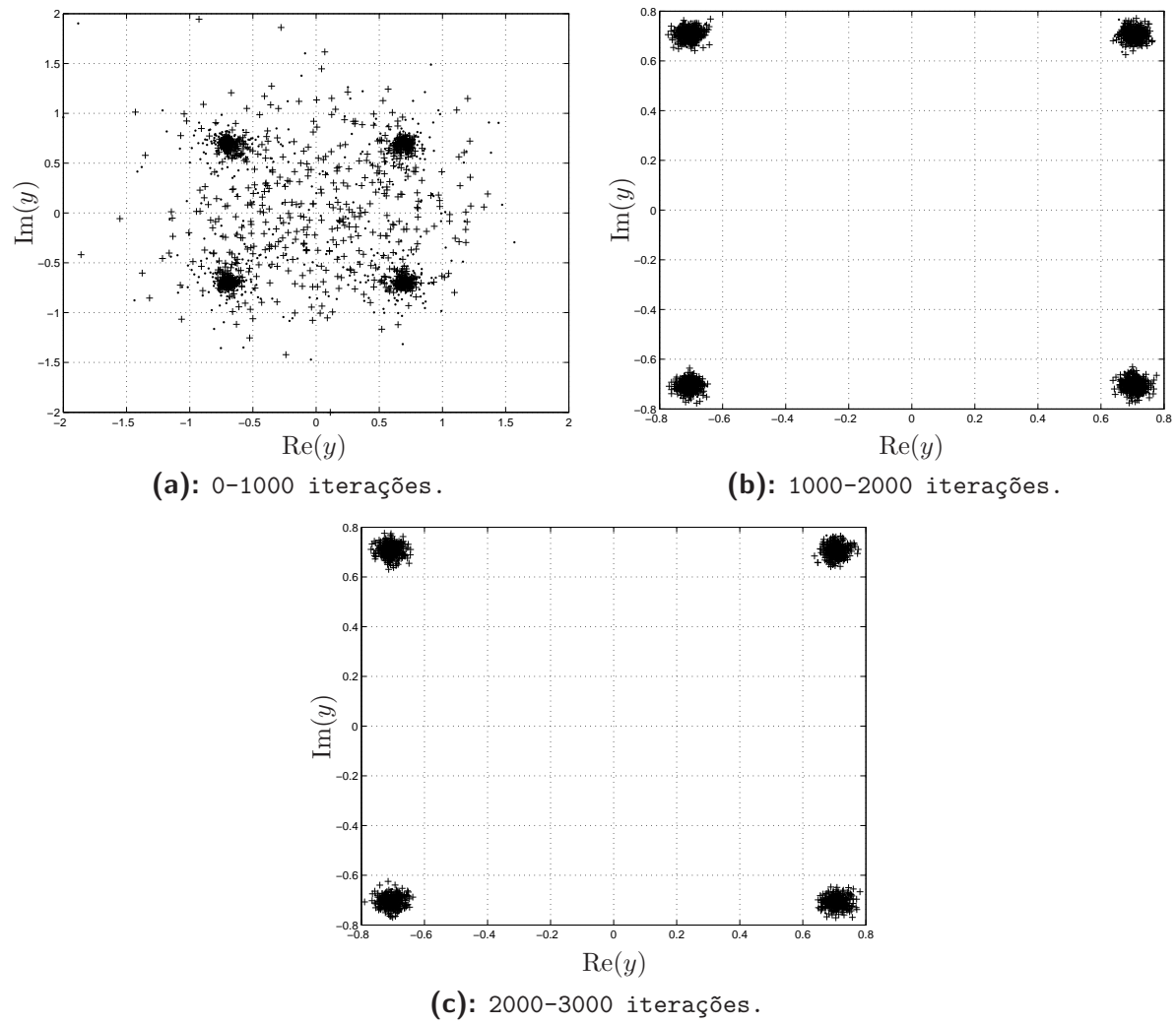


Figura 7.13: Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o MU-FPA.

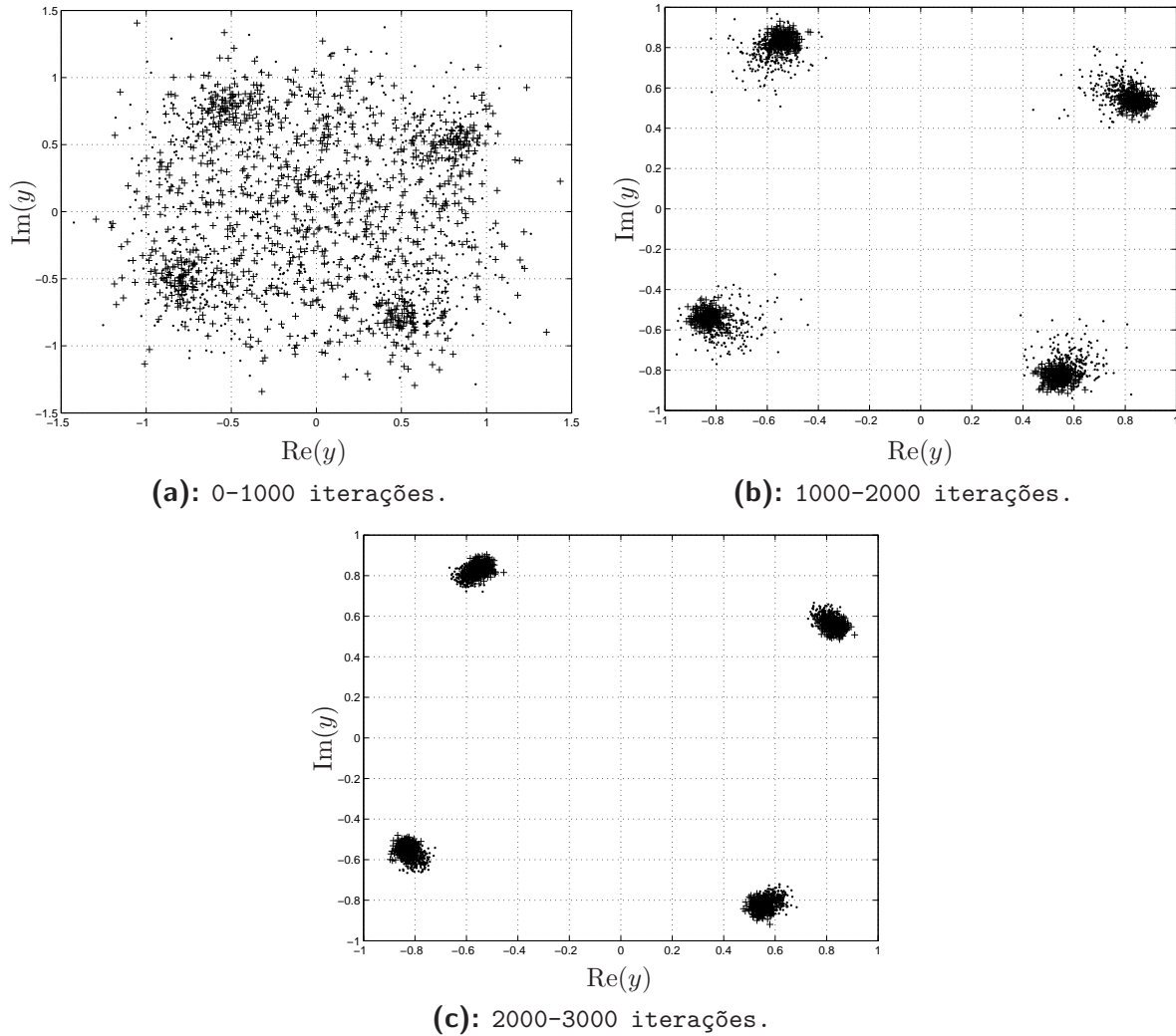


Figura 7.14: Evolução das constelações dos sinais na saída do dispositivo de separação espaço-temporal através da adaptação com o MU-CMA.

Nos gráficos anteriores fica evidente a recuperação do MU-CMA, a menos de uma rotação de fase e também com uma dispersão um pouco maior que a dos outros algoritmos empregados. Entretanto, tal diferença não é relevante quando se imagina que em situações práticas, após a “abertura do olho”, geralmente se chaveia para um modo de equalização de decisão dirigida. Com isso, o maior ganho do algoritmo MU-FPA é relacionado com a maior taxa de convergência.

7.4 Conclusões e síntese

Neste capítulo, foram apresentadas algumas aplicações que encontram, nas ferramentas e estratégias de separação de fontes, um grande suporte para problemas que ocorrem em vários sistemas práticos.

Sob a óptica do uso em sistemas de mistura convolutiva, ambiente natural do desenvolvimento de técnicas baseadas na hipótese de fontes independentes, são avaliados alguns dos principais aspectos dos algoritmos propostos no Capítulo 5.

Dentre algumas das situações consideradas, destaca-se a utilização em sistemas com comportamento aleatório, a estimativa do número de usuários e a utilização de estratégias de BSS em sistemas que empregam arranjos de antenas no receptor. Nestes casos, pode-se ver a robustez da proposta frente a vários aspectos de importância em sistemas reais.

Uma aplicação bastante relevante no campo de sistemas de comunicação digital é também mostrada. A resolução de colisão de pacotes por técnicas de processamento de sinais, particularmente de separação cega de fontes, ilustra a eficácia dos métodos cegos, destacando mais uma vez um melhor desempenho do algoritmo MU-CFPA.

Por fim, o processamento espaço-temporal, quando há interferência gerada por símbolos do mesmo usuário, é introduzida num contexto de recuperação de sinais de comunicação digital. Novamente, pode-se verificar que o algoritmo MU-FPA apresenta um melhor desempenho, principalmente em termos de taxa de convergência, quando comparado com propostas não-supervisionadas já existentes na literatura.

A lista de aplicações investigadas é sugestiva no que se refere às potencialidades das propostas desta tese e serve para solidificar as análises teóricas discutidas nos capítulos anteriores.

– “A vida é a arte de tirar conclusões suficientes de premissas insuficientes.”

Samuel Butler, 1894

8

Conclusões, Sugestões e Perspectivas

ESTE capítulo é dedicado a apresentar as conclusões da tese, bem como sugestões e perspectivas para trabalhos que evoluam a partir do conteúdo aqui apresentado.

Como os capítulos anteriores apresentam seções de conclusão e síntese para cada uma das partes da tese, este capítulo tenta sumarizar e destacar os principais pontos de tais seções.

Outro aspecto abordado neste capítulo é a discussão sobre possíveis linhas de pesquisa para continuidade do atual trabalho, fornecendo algumas argumentações que suportam tais expectativas.

A organização do capítulo é estruturada da seguinte maneira. As conclusões e contribuições são apresentadas, de forma destacada pelas partes e capítulos da tese, na Seção 8.1. Na Seção 8.2 são listadas algumas perspectivas e sugestões de continuidade da pesquisa na linha de trabalho desta tese.

8.1 Conclusões e contribuições

Um dos principais temas desta tese é a contribuição ao problema de separação cega de fontes a partir da proposta e análise de *critérios que utilizam métodos de estimação da função de densidade de probabilidade do sinal na saída do dispositivo de separação*.

Estes métodos são baseados em critérios decorrentes de estratégias propostas inicialmente no contexto da equalização autodidata. A generalização para o problema multivariável é realizada através de condições que permitem garantir a identificação e a separação de todas as fontes. Entretanto, devido à natureza do critério de otimização utilizado, a proposta é aplicável a sistemas que apresentam sinais pertencentes a um alfabeto discreto finito.

Além de fornecer uma nova classe de soluções para o problema de separação de fontes, a proposta permite uma análise comparativa com alguns outros critérios existentes na literatura, fornecendo um ponto chave para a obtenção de relações entre critérios importantes e de aplicação prática bastante considerável. Foi possível ainda contribuir com a análise teórica no que se refere à concepção de algoritmos adaptativos para separação cega de fontes.

De uma maneira geral, pôde-se observar que a técnica apresentada provê uma redução do nível de interferência inserida por usuários que compartilham os mesmos recursos, bem como uma solução adaptativa que apresenta uma taxa de convergência comparável àquelas supervisionadas, ressaltando a factibilidade de aplicação das técnicas de separação cega de fontes para o processamento de sistemas multi-usuário.

É interessante, entretanto, verificar as conclusões e contribuições apresentadas capítulo a capítulo nesta tese.

Parte I

No **Capítulo 2** são revisitados os conceitos de separação cega de fontes destacando-se a consideração de independência das fontes. A descrição das principais ferramentas matemáticas, entre as quais se destacam aquelas que avaliam a não-gaussianidade das fontes, são apresentadas para fundamentar a formulação de critérios para separação cega de fontes.

O **Capítulo 3** é dedicado à descrição de algumas das principais estratégias e técnicas de separação de fontes, ressaltando as hipóteses e considerações adotadas em cada caso. Embora o número de estratégias apresentadas esteja longe do número existente de soluções, seus fundamentos possibilitam a extensão dos conceitos a outras também passíveis de serem inseridas no contexto de recuperação autodidata de sinais. Algumas aplicações em cenários clássicos do problema em questão são apresentadas, ilustrando também a potencialidade de

alguns métodos (critérios e algoritmos) existentes na literatura.

No **Capítulo 4** são abordados sistemas de múltiplos usuários que possuem características pré-determinadas para o sistema de combinação dos sinais, bem como características particulares das fontes, de modo que torna-se possível uma simplificação na estratégia de separação de fontes. Ainda neste capítulo, são discutidos alguns modelos de receptores tipicamente empregados em sistemas de recuperação de sinais de usuários submetidos ao efeito de um canal de transmissão, o qual executa a tarefa do sistema de mistura. A modelagem destes canais, bem como a estrutura de seus receptores associados, é apresentada para os casos de interferência de múltiplo acesso e interferência inter-simbólica, compondo, respectivamente, as chamadas estruturas de processamento espacial e espaço-temporal. O capítulo é finalizado com a apresentação de alguns critérios e algoritmos projetados para a remoção não-supervisionada da interferência em sistemas de processamento multi-usuário, discutindo um pouco o estado da arte destas técnicas.

A primeira parte desta tese é mais do que uma revisão bibliográfica. Ela tenta colocar de forma unificada dois campos que são, freqüentemente, apresentados de forma disjunta nos diversos trabalhos. De fato, embora o processamento multi-usuário possa ser considerado como um caso particular (devido às características das fontes) do problema de separação de fontes, as linhas de ação dos trabalhos existentes não vêm explorando tal consideração. A tentativa de colocar os dois problemas sob a mesma óptica vem ao encontro de buscar ferramentas de análise de separação de fontes que possibilitem um aporte de conhecimento aos problemas práticos de comunicação digital.

Parte II

No que se refere a resultados originais, as contribuições mais evidentes deste trabalho estão concentradas na segunda parte, conforme se observa a seguir.

No **Capítulo 5**, uma família de critérios para processamento multi-usuário é proposta, tendo por base um método para equalização cega que leva em conta a estimação da densidade de probabilidade do sinal na saída do equalizador. A análise do método em relação às questões de convergência e de suas equivalências com outras técnicas é realizada, de modo a fundamentar o entendimento da proposta. Uma relação com o critério de decisão dirigida, ressaltando algumas características de teoria da informação não discutidas até o momento na literatura, é também apresentada, bem como uma dedução de que a solução ótima (solução de Wiener) pode ser obtida a partir da proposta. As análises são realizadas através da utilização de probabilidades condicionais que fornecem as características desejadas do sinal a ser recuperado. A partir daí, é possível verificar uma relação entre os critérios de minimização do erro médio quadrático, critério de máxima *a posteriori* (que minimiza a taxa de erro de bit) e o critério baseado na estimação da densidade de probabilidade. Os critérios de

processamento multi-usuário são então concebidos utilizando duas estratégias distintas de garantia da recuperação das diferentes fontes envolvidas: a utilização de um critério auxiliar de decorrelação explícita e a utilização de filtragem com restrições para forçar a ortogonalidade da resposta global do sistema. Os critérios desta família apresentam a característica de serem relacionados à máxima log-verossimilhança e possuem a capacidade de recuperação de fase dos sinais, com uma complexidade computacional equivalente à dos algoritmos LMS para cada usuário.

O **Capítulo 6** discute e analisa as características de algoritmos adaptativos de separação cega de fontes quanto à utilização dos momentos de ordem superior dos sinais, quando da concepção do critério. A análise é conduzida através da utilização de expansões em séries ortonormais da função densidade de probabilidade. Estas expansões utilizam os cumulantes como coeficientes da base ortonormal e permitem avaliar a necessidade e utilidade de tais cumulantes para a correta estimação da densidade de probabilidade. Como as densidades de sinais típicos em sistemas de processamento multi-usuário são dadas por misturas gaussianas, foi necessário a dedução de uma expansão em série para tal distribuição. Esta nova distribuição é utilizada para avaliar em tempo real a diferença entre os critérios e respectivos algoritmos que utilizam diferentes números de momentos de ordem superior na sua concepção. Com isso, é possível observar e avaliar o comportamento dinâmico quando da obtenção da solução para separação de fontes. Esta análise é de caráter bastante inovador uma vez que, na atual literatura, os critérios e algoritmos para separação cega de fontes não levam em consideração o efeito do número de cumulantes na taxa de convergência de algoritmos adaptativos. Desta maneira, as conclusões da análise do uso das estatísticas de ordem superior permitem novas considerações para a concepção de métodos de separação cega de fontes. Neste ponto, fica evidente a diferença entre métodos baseados em separação cega de fontes e aqueles de equalização cega, seja em relação à generalidade da proposta, como também à capacidade de rastreamento de diferentes tipos de distribuição envolvidas em sistemas com múltiplas fontes. Seguindo esta linha, são ainda avaliadas as características comuns entre o problema de separação cega de fontes e o de desconvolução cega e, conseqüentemente, o impacto do uso das estatísticas no processo adaptativo também em desconvolução.

Algumas aplicações são apresentadas no **Capítulo 7** visando comparar as propostas com os outros métodos existentes para processamento multi-usuário. As comparações são conduzidas de maneira a explorar os diferentes cenários possíveis de sistemas multi-usuário. Destaca-se aqui a aplicação de arranjos de antenas, o problema de resolução de pacotes em redes S-ALOHA e a utilização de processamento espaço-temporal para remoção de interferência de múltiplo acesso e inter-simbólica. Um ponto inovador é a proposição de uma técnica de estimativa do número de usuários baseado em métodos de subespaço a partir da decomposição em valores singulares da matriz de autocorrelação dos sinais do receptor.

8.2 Sugestões e perspectivas para trabalhos futuros

Devido ao fato de que o problema de separação cega de fontes congrega várias áreas da engenharia, estatística e matemática, as sugestões e perspectivas surgem em diversas direções que convergem para a busca de estratégias mais robustas, novas ferramentas de análise e generalização dos modelos existentes. Do ponto de vista dos métodos propostos, também algumas perspectivas podem ser listadas.

1. A utilização de outras medidas de similaridade entre funções de densidade de probabilidade como métrica a ser minimizada para obtenção das estimativas dos sinais dos diversos usuários é uma linha natural a ser seguida. Uma das medidas alternativas que devem ser investigadas é a medida de Hellinger, a qual deriva da medida \mathbb{L}_p [Devroye, 1987]. Outras medidas, principalmente aquelas que levam em conta entropias diferenciais baseadas na definição de entropia dada por Renyi [Renyi, 1987], são de grande interesse e podem prover avanços na área da análise dos métodos e suas equivalências com outros critérios já existentes e não reportados aqui. Ainda visando melhorar os métodos tratados nesta tese, um algoritmo que utilize a adaptação do parâmetro σ_r^2 é de grande interesse. Esta adaptação pode ser inicialmente investigada pelos métodos de otimização da variância do estimador de Parzen mencionadas na literatura, mas que utilizam técnicas empíricas para encontrar o melhor valor para o parâmetro de variância do *kernel*.
2. Uma perspectiva de continuidade no sentido de generalizar os problemas passíveis de tratamento com os métodos aqui propostos é a utilização de estimativas não-paramétricas das densidades de probabilidade dos sinais na saída do dispositivo de separação. Esta estimativa pode ser realizada através da utilização de *kernels*, particularmente o de Parzen, permitindo uma extensão dos métodos apresentados na tese de forma bastante eficiente para tratamento de problemas cujas densidades de probabilidade das fontes não são conhecidas *a priori*. Dentre elas podem-se citar processamento de sinais biomédicos e processamento de voz, bem como problemas de comunicação digital em que coexistam vários tipos de sinais diferentes, como se imagina ser o caso nos sistemas de quarta geração de comunicações móveis. Outras estimativas não-paramétricas poderiam ser realizadas através de expansões em série a partir dos cumulantes dos dados, em torno de uma função de referência adequada.
3. Considerando a perspectiva descrita anteriormente, a derivação de novas expansões em série que sejam adequadas aos problemas a serem tratados é uma linha de pesquisa ainda bastante interessante. A consideração de funções de densidade de probabilidade de referência que sejam adequadas às características dos sinais em questão não é uma tarefa elementar e a derivação da base ortonormal pode ser de difícil tratamento matemático. Entretanto, tais expansões podem fornecer grandes

avanços na análise e na concepção de métodos adaptativos de separação cega de fontes.

4. Na linha da avaliação das estatísticas de ordem superior, uma abordagem através das funções de Donoho [Donoho, 1981], definidas em função de cumulantes de ordem arbitrária para minimização da entropia, pode representar um novo ganho de conhecimento nesta área ainda incipiente. Isto vem no sentido de determinar quais são aqueles cumulantes e/ou momentos mais representativos de serem retidos num critério de otimização para maximizar a relação complexidade \times taxa de convergência. Alguns trabalhos, como por exemplo [Mboup & Regalia, 2000], já consideraram as funções de Donoho para compreensão do critério de maximização de cumulantes. Entretanto, a análise realizada até o momento na literatura é relacionada à demonstração dos pontos estacionários das soluções existentes. Uma abordagem baseada na importância dos cumulantes pode inclusive ajudar a deduzir algumas relações entre critérios de separação de fontes.
5. Uma meta que parece ser possível, dados alguns resultados recentes na literatura, é a demonstração de forma direta da equivalência da solução do critério proposto na tese com outros critérios já bastante conhecidos, como o do módulo constante e o de Shalvi-Weinstein.
6. De fato, uma importante contribuição que pode vir a ser somada às técnicas atuais, é a derivação de algoritmos de separação de fontes, para aplicação em sistemas que apresentam dispersão temporal. Critérios com restrição, tais como o MUK e o MU-CFPA, necessitam, para atuar em tais sistemas, da derivação de ferramentas para manipulação de tensores de ordem maior ou igual a três e suas operações (ortogonalização, produto interno, etc), uma vez que os tensores que modelam os sistemas com interferência de múltiplos usuários e interferência inter-simbólica possuem mais dimensões que apenas a dimensionalidade das fontes e dos sensores. Além disso, poderiam ser incluídas algumas outras variáveis, como por exemplo código e/ou frequência, para tornar ainda mais genérico o modelo de mistura e das fontes. Uma decorrência natural desta derivação, se alcançada, é a aplicação em sistemas de mistura convolutiva (dependência temporal das fontes) e em particular em resolução de pacotes em redes S-ALOHA, nos quais há a presença de interferência inter-simbólica.
7. Um aspecto que não foi abordado, e que também não se encontra referenciado a contento, refere-se ao sincronismo das fontes. Métodos que sejam capazes de recuperar o sincronismo, ou mais ainda, tratar sinais que se apresentam combinados de forma assíncrona, podem ser de interesse prático como por exemplo em detecção de sinais acústicos provenientes de instrumentos musicais em um concerto.
8. Finalmente, uma perspectiva e anseio de continuidade de trabalho, reside na investigação de métodos de geometria diferencial para processamento de sinais. A geometria diferencial estuda a geometria de uma função custo a ser minimizada

quando há restrições envolvidas no espaço de Riemann, observando assim algumas características dos tensores de ordem qualquer, para avaliação das questões de convergência e sucesso na otimização. Esta teoria matemática vem sendo aplicada a problemas de engenharia, em particular processamento de sinais, nos contextos de estimação bayesiana, filtragem com restrição, processamento multidimensional e codificação espaço-temporal. Então, a aplicação dos conceitos desta poderosa ferramenta de análise pode incrementar bastante a compreensão de várias características dos critérios de separação de fontes.

Sendo assim, esta tese, desenvolvida sobre a motivação de prover uma ampla revisão, e ao mesmo tempo pessoal, da literatura e, a partir deste texto tutorial, fornecer resultados inovadores de cunho teórico e com aplicações em processamento multi-usuário, abre também diversas perspectivas de trabalho. Entende-se assim estar contribuindo com o tema de separação de fontes, sobretudo para a comunidade brasileira onde pesquisadores poderão, talvez, encontrar neste texto ferramentas adequadas a seus temas de trabalho ou frentes teóricas que lhe abram novos interesses.

PARTE **III**

Apêndices

– “A matemática pode ser definida como a ciência na qual não se sabe jamais sobre o que fala nem se o que se diz é verdade.”

Bertrand Russel, 1920

A

Definições Matemáticas

CONCEITOS e definições matemáticas que possuem grande importância no decorrer da tese são descritos em mais detalhes neste apêndice.

O apêndice é dividido em duas seções. Na Seção A.1 são descritos os principais aspectos relativos aos cumulantes e momentos de uma distribuição de probabilidade qualquer. Aspectos relativos à entropia de variáveis aleatórias são exploradas na Seção A.2.

A.1 Cumulantes e momentos

A.1.1 História

Os cumulantes foram inicialmente introduzidos pelo astrônomo, contador, matemático e estatístico dinamarquês Thorvald N. Thiele (1838-1910) que os denominou *semi-invariantes*.

O termo *cumulante* surgiu pela primeira vez em 1931 no artigo “The Derivation of the

Pattern Formulæ of Two-Way Partitions from Those of Simpler Patterns”, Proceedings of the London Mathematical Society, Series 2, vol. 33, pp. 195-208, publicado pelo geneticista e estatístico Sir Ronald Fisher e o estatístico John Wishart, epônimo da distribuição de Wishart.

O historiador Stephen Stigler comenta que o termo cumulante foi sugerido a Fisher numa carta de Harold Hotelling. Em um outro artigo publicado em 1929, Fisher chamou-os de funções de momentos cumulativos.

A.1.2 Cumulantes e momentos de distribuições de probabilidade

Dada uma distribuição de probabilidade $p_Y(y)$, os momentos são obtidos a partir da *função característica*, também chamada de *função geradora de momentos*, definida, para uma variável real y , como

$$\begin{aligned}\Omega_Y(\omega) &\triangleq \int_{-\infty}^{\infty} p_Y(y) \exp(j\omega y) dy \\ &\triangleq \mathbb{E} \{ \exp(j\omega y) \}.\end{aligned}\tag{A.1}$$

Expandindo-se $\Omega_Y(\omega)$ em uma série de potências em torno da origem obtém-se [Papoulis, 1991]:

$$\Omega_Y(\omega) = \sum_{k=0}^{\infty} \frac{\kappa_k}{k!} (j\omega)^k,\tag{A.2}$$

em que κ_k é o momento centrado de ordem k .

O cumulante de ordem k é definido como [Papoulis, 1991]

$$c_k = \frac{\partial^k \Upsilon_Y(\omega)}{\partial \omega^k},\tag{A.3}$$

em que

$$\Upsilon_Y(\omega) = \ln [\Omega_Y(\omega)],\tag{A.4}$$

é a *função geradora de cumulantes*.

Para o caso de variáveis complexas, a função característica é dada por [Amblard et al., 1996a]:

$$\begin{aligned}\Omega_{Y,Y^*}(\omega, \omega^*) &\triangleq \int_{-\infty}^{\infty} p_{Y,Y^*}(y, y^*) \exp \left[j \left(\frac{\omega y^* + \omega^* y}{2} \right) \right] dy dy^* \\ &\triangleq E \left\{ \exp \left[j \left(\frac{\omega y^* + \omega^* y}{2} \right) \right] \right\}.\end{aligned}\tag{A.5}$$

Neste caso, a função geradora de cumulantes é escrita como:

$$\Upsilon_Y(\omega) \triangleq \ln [\Omega_{Y,Y^*}(\omega, \omega^*)] \quad (\text{A.6})$$

A.1.3 Algumas propriedades

Invariância e equivariância

O cumulante de ordem um é equivariante enquanto todos os demais são invariantes a deslocamentos. Então, para um cumulante de ordem k da variável Y , denotado por $c_k(Y)$, tem-se

$$\begin{aligned} c_1(Y + \alpha) &= c_1(Y) + \alpha \\ c_k(Y + \alpha) &= c_k(Y), \end{aligned} \quad (\text{A.7})$$

para α uma constante qualquer.

Homogeneidade

O cumulante de ordem k é homogêneo¹ de grau k , ou seja, para o caso real tem-se:

$$c_k(\alpha Y) = \alpha^k \cdot c_k(Y). \quad (\text{A.8})$$

Considerando-se o caso complexo, o k -ésimo cumulante é definido como

$$c_k(Y, Y^*) = c_k(\underbrace{Y, \dots, Y}_{s \text{ termos}}, \underbrace{Y^*, \dots, Y^*}_{q \text{ termos}}) \quad \forall s + q = k. \quad (\text{A.9})$$

Então, de acordo com a Equação (A.9), a propriedade da homogeneidade para variáveis complexas é dada por [Lacoume et al., 1997; Amblard et al., 1996b]:

$$c_k(\alpha Y, \alpha Y^*) = (\alpha)^s \cdot (\alpha^*)^q \cdot c_k(Y, Y^*). \quad (\text{A.10})$$

Desta maneira, para os cumulantes de ordem par, pode-se definir $s = q$ que fornece a homogeneidade como

$$c_k(\alpha Y) = |\alpha|^k \cdot c_k(Y). \quad (\text{A.11})$$

¹Esta propriedade é algumas vezes denominada de *multilinearidade*.

Aditividade

Se X e Y são variáveis aleatórias independentes então vale a seguinte relação:

$$c_k(X + Y) = c_k(X) + c_k(Y). \quad (\text{A.12})$$

A.1.4 Cumulantes e momentos

Os cumulantes são relacionados com os momentos através da seguinte recursão [Nikias & Petropulu, 1993]:

$$c_k = \kappa_k - \sum_{i=1}^{k-1} \binom{k-1}{i-1} c_i \cdot \kappa_{k-i}. \quad (\text{A.13})$$

Desta forma, o k -ésimo momento é um polinômio de grau k dos k primeiros cumulantes, dados, para o caso em que $k = 6$, na seguinte forma:

$$\begin{aligned} \kappa_1 &= c_1 \\ \kappa_2 &= c_2 + c_1^2 \\ \kappa_3 &= c_3 + 3c_2c_1 + c_1^3 \\ \kappa_4 &= c_4 + 4c_3c_1 + 3c_2^2 + 6c_2c_1^2 + c_1^4 \\ \kappa_5 &= c_5 + 5c_4c_1 + 10c_3c_2 + 10c_3c_1^2 + 15c_2^2c_1 + 10c_2c_1^3 \\ \kappa_6 &= c_6 + 6c_5c_1 + 15c_4c_2 + 15c_4c_1^2 + 10c_3^2 + 60c_3c_2c_1 + 20c_3c_1^3 + 15c_2^3 + 45c_2^2c_1^2 + 15c_2c_1^4 + c_1^6. \end{aligned} \quad (\text{A.14})$$

No caso de ser uma distribuição de média nula, basta anular na Equação (A.14) os termos dos polinômios nos quais c_1 aparece.

Os polinômios da Equação (A.14) possuem uma interpretação combinatorial na qual os coeficientes “contam” as partições de conjuntos. Uma fórmula geral dos polinômios é dada por

$$\kappa_k = \sum_{\aleph} \prod_{B \in \aleph} \kappa_{|B|}, \quad (\text{A.15})$$

em que \aleph contém toda a lista de partições de um conjunto de tamanho k , e $B \in \aleph$ significa que B é um dos “blocos” nos quais o conjunto é particionado, sendo $|B|$ o tamanho do conjunto B .

Com isso, cada monômio é dado por uma constante multiplicando um produto de cumulantes nos quais a soma dos índices é k , por exemplo no termo $c_3c_2^2c_1$ a soma dos índices é $3 + 2 \cdot 2 + 1 = 8$, indicando que este termo aparece no polinômio do momento de oitava ordem.

A.1.5 Cumulantes conjuntos

O cumulante conjunto de várias variáveis aleatórias Y_1, \dots, Y_k é dado por [Nikias & Petropulu, 1993]:

$$c(Y_1, \dots, Y_k) = \sum_{\aleph} \prod_{B \in \aleph} (|B| - 1)! \cdot (-1)^{|B|-1} \prod_{i \in B} \mathbb{E} \{Y_i\}, \quad (\text{A.16})$$

em que \aleph é o conjunto com todas as combinações da seqüência $\{1, \dots, k\}$ e B é o conjunto com todas as combinações em bloco do conjunto \aleph . Por exemplo,

$$\begin{aligned} c(X, Y, Z) = \mathbb{E} \{XYZ\} - \mathbb{E} \{XY\} \mathbb{E} \{Z\} - \mathbb{E} \{XZ\} \mathbb{E} \{Y\} \\ - \mathbb{E} \{YZ\} \mathbb{E} \{X\} + 2 \cdot \mathbb{E} \{X\} \mathbb{E} \{Y\} \mathbb{E} \{Z\}. \end{aligned}$$

Se as variáveis forem independentes, o cumulante conjunto delas é nulo e se as k variáveis forem todas iguais, o cumulante conjunto é dado por $c_k(Y)$.

O significado combinatorial da expressão dos momentos em termos dos cumulantes mostra-se mais elegante, conforme mostrado abaixo [Nikias & Petropulu, 1993]:

$$\mathbb{E} \{Y_1 \cdots Y_k\} = \sum_{\aleph} \prod_{B \in \aleph} c(Y_B), \quad (\text{A.17})$$

em que $c(Y_B)$ é o cumulante conjunto associado às variáveis aleatórias Y_1, \dots, Y_k , cujos índices são incluídos no bloco B . Por exemplo,

$$\mathbb{E} \{XYZ\} = c(X, Y, Z) + c(X, Y)c(Z) + c(X, Z)c(Y) + c(Y, Z)c(X) + c(X)c(Y)c(Z).$$

A.1.6 Cumulantes condicionais

A lei de média total, que afirma que $\mathbb{E} \{Y\} = \mathbb{E} \{\mathbb{E} \{Y|X\}\}$ e a lei de variância total, na qual $\text{var}(Y) = \mathbb{E} \{\text{var}(Y|X)\} + \text{var}(\mathbb{E} \{Y|X\})$, são naturalmente generalizadas para os cumulantes condicionais. Em geral tem-se:

$$c(Y_1, \dots, Y_k) = \sum_{\aleph} c(c(Y_{\aleph_1}|X), \dots, c(Y_{\aleph_b}|X)), \quad (\text{A.18})$$

em que o somatório é tomado sobre todas as partições \aleph do conjunto $\{1, \dots, k\}$ dos índices, $\aleph_1, \dots, \aleph_b$ são todos os blocos da partição de \aleph e $c(Y_{\aleph_k})$ indica o cumulante conjunto das variáveis aleatórias cujos índices estão naquele bloco da partição.

A.2 Entropia de variáveis aleatórias

A.2.1 Definição de entropia

Seja uma variável aleatória Y multidimensional, contínua, real e centrada (média nula) com uma função de densidade de probabilidade $p_Y(\mathbf{y})$. Define-se por *entropia* a seguinte quantidade:

$$\begin{aligned} \mathcal{H}(y) &= -\mathbb{E} \{ \ln [p_Y(\mathbf{y})] \} \\ &= - \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln [p_Y(\mathbf{y})] dy. \end{aligned} \quad (\text{A.19})$$

A.2.2 Distribuições com máxima entropia

É de grande interesse em processamento de sinais encontrar as distribuições que possuem máxima entropia. Desta maneira é interessante solucionar o seguinte problema [Cover & Thomas, 1991]:

Maximizar $\mathcal{H}(y)$ sob todas as distribuições $p_Y(y)$ que satisfazem

1. $p_Y(y) \geq 0$, com a igualdade válida somente fora do domínio S da variável;
2. $\int_S p_Y(y) dy = 1$;
3. $\int_S p_Y(y) f_i(y) dy = \kappa_i$, para $1 \leq i \leq k$;

em que κ_i é o momento centrado de i -ésima e $f_i(y)$ é uma função que faz $p_Y(y)$ respeitar a restrição.

Para resolver o problema acima, é necessário utilizar os multiplicadores de Lagrange. Assim, pode-se escrever o seguinte Lagrangiano [Cover & Thomas, 1991]:

$$J(p_Y(y)) = - \int_S p_Y(y) \cdot \ln [p_Y(y)] dy + \beta_0 \cdot \left(\int_S p_Y(y) dy \right) + \sum_{i=1}^k \beta_i \left(\int_S p_Y(y) f_i(y) dy = \kappa_i \right), \quad (\text{A.20})$$

em que β_0, \dots, β_i são os multiplicadores de Lagrange.

Derivando-se a Equação (A.20) em relação à distribuição $p_Y(y)$ tem-se então:

$$\frac{\partial J(p_Y(y))}{\partial p_Y(y)} = -\ln [p_Y(y)] - 1 + \beta_0 + \sum_{i=1}^k \beta_i \cdot p_Y(y), \quad (\text{A.21})$$

em que β_0, \dots, β_i são escolhidos de tal forma que $p_Y(y)$ satisfaça as restrições.

Então quais são as distribuições que maximizam² a Equação (A.21)?

A resposta depende das restrições impostas. A título de exemplo considera-se dois casos:

1. Suporte fixo ($S = [a, b]$)

Neste caso, não há nenhuma restrição quanto aos momentos, logo os multiplicadores $\beta_1, \dots, \beta_k = 0$ uma vez que não há necessidade de restrição. Deste modo, igualando-se a Equação (A.21) a zero tem-se:

$$\begin{aligned} -\ln [p_Y(y)] - 1 + \beta_0 &= 0 \\ \ln [p_Y(y)] &= \beta_0 - 1 \\ p_Y(y) &= \exp [\beta_0 - 1]. \end{aligned} \tag{A.22}$$

Resolvendo a integral sobre o suporte determinado, tem-se:

$$\begin{aligned} \int_a^b p_Y(y) dy &= 1 \\ \int_a^b \exp [\beta_0 - 1] dy &= 1 \\ \exp [\beta_0 - 1] \cdot (b - a) &= 1 \\ p_Y(y) = \exp [\beta_0 - 1] &= \frac{1}{b - a}. \end{aligned} \tag{A.23}$$

Assim, sob a restrição de um suporte fixo, a distribuição com máxima entropia é a *distribuição uniforme*.

2. Média e variância fixas

Sob estas restrições, $\beta_0, \beta_1, \beta_2 \neq 0$ e $S =] - \infty, \infty [$. Assim tem-se a seguinte solução para a distribuição ao tomar-se $\frac{\partial J(p_Y(y))}{\partial p_Y(y)} = 0$:

$$p_Y(y) = \exp [\beta_0 + \beta_1 y + \beta_2 y^2 - 1]. \tag{A.24}$$

Logo, necessita-se encontrar os valores de β_0, β_1 e β_2 através do seguinte sistema de

²A rigor deve-se tomar a segunda derivada da Equação (A.20) para mostrar que é um valor de máximo.

equações:

$$\begin{aligned}
 \int_{-\infty}^{\infty} \exp [\beta_0 + \beta_1 y + \beta_2 y^2 - 1] dy &= 1 \\
 \int_{-\infty}^{\infty} y \cdot \exp [\beta_0 + \beta_1 y + \beta_2 y^2 - 1] dy &= \kappa_1 \\
 \int_{-\infty}^{\infty} y^2 \cdot \exp [\beta_0 + \beta_1 y + \beta_2 y^2 - 1] dy &= \sigma^2 = \kappa_2.
 \end{aligned} \tag{A.25}$$

A solução do sistema na Equação (A.25) fornece os seguintes valores para os multiplicadores de Lagrange:

$$\begin{aligned}
 \beta_0 &= -\ln \left[\sqrt{2\pi\sigma} \right] \\
 \beta_1 &= \kappa_1 \\
 \beta_2 &= -\frac{1}{2\sigma^2}.
 \end{aligned} \tag{A.26}$$

Desta maneira, substituindo-se os valores da Equação (A.26) na Equação (A.24) obtém-se

$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{(y - \kappa_1)^2}{2\sigma^2} \right]. \tag{A.27}$$

Logo, para a restrição de média e variância fixas, a *distribuição gaussiana* apresenta a máxima entropia.

É possível mostrar esta propriedade da distribuição gaussiana sob outra abordagem, conforme descrito na seção a seguir.

A.2.3 Entropia de uma variável gaussiana: abordagem alternativa

Seja Y uma variável aleatória gaussiana multidimensional e de média nula cuja densidade é escrita como:

$$p_G(\mathbf{y}) = \frac{1}{\sqrt[2]{2\pi} \cdot |\det(\mathbf{R}_y)|^{\frac{1}{2}}} \cdot \exp \left(-\frac{1}{2} \mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y} \right), \tag{A.28}$$

em que n é a dimensão do vetor \mathbf{y} e \mathbf{R}_y é a matriz de autocorrelação de \mathbf{y} .

Ao tomar-se o logaritmo natural da Equação (A.28), obtém-se

$$\ln [p_G(\mathbf{y})] = -\frac{n}{2} \cdot \ln[2\pi] - \frac{1}{2} \ln [|\det(\mathbf{R}_y)|] - \frac{1}{2} \mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y}. \tag{A.29}$$

Como pode-se escrever [Picinbono & Barret, 1990]

$$\mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y} = \text{tr}(\mathbf{y} \mathbf{y}^T \mathbf{R}_y^{-1}),$$

em que $\text{tr}(\cdot)$ é o traço da matriz, tem-se que

$$\mathbb{E} \{ \mathbf{y}^T \mathbf{R}_y^{-1} \mathbf{y} \} = \mathbb{E} \{ \text{tr}(\mathbf{y} \mathbf{y}^T \mathbf{R}_y^{-1}) \} = \text{tr}(\mathbf{I}) = n. \quad (\text{A.30})$$

Então, substituindo os resultados obtidos nas Equações (A.29) e (A.30) na Equação (A.19) tem-se:

$$\mathcal{H}_G(\mathbf{y}) = \frac{n}{2} \cdot \{\ln[2\pi] + 1\} + \frac{1}{2} \cdot \ln[|\det(\mathbf{R}_y)|], \quad (\text{A.31})$$

em que $\mathcal{H}_G(\mathbf{y})$ é a entropia da distribuição gaussiana de média nula.

Um aspecto importante a ser demonstrado é que a distribuição gaussiana apresenta a maior entropia entre todas as distribuições. Para tal, considera-se uma função de densidade de probabilidade qualquer sobre a variável representada por $p_Y(\mathbf{y})$.

A média da v.a. $\ln[p_G(\mathbf{y})]$ é a mesma tanto considerando-a com uma distribuição qualquer $p_Y(\mathbf{y})$ como no caso particular de uma distribuição gaussiana $p_G(\mathbf{y})$. Isto porque a matriz de autocorrelação para as duas distribuições é a mesma, ou seja, a restrição é de que a distribuição tenha uma variância definida [Picinbono & Barret, 1990].

Daí, pode-se escrever

$$\int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln[p_G(\mathbf{y})] d\mathbf{y} = \int_{-\infty}^{\infty} p_G(\mathbf{y}) \cdot \ln[p_G(\mathbf{y})] d\mathbf{y} = -\mathcal{H}_G(\mathbf{y}). \quad (\text{A.32})$$

A partir da definição da divergência de Kulback-Leibler pode-se escrever:

$$\begin{aligned} \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln \left[\frac{p_Y(\mathbf{y})}{p_G(\mathbf{y})} \right] d\mathbf{y} &= \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln[p_Y(\mathbf{y})] d\mathbf{y} - \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln[p_G(\mathbf{y})] d\mathbf{y} \\ \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln \left[\frac{p_Y(\mathbf{y})}{p_G(\mathbf{y})} \right] d\mathbf{y} &= -\mathcal{H}_Y(\mathbf{y}) - \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln[p_G(\mathbf{y})] d\mathbf{y} \\ \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln \left[\frac{p_G(\mathbf{y})}{p_Y(\mathbf{y})} \right] d\mathbf{y} &= \mathcal{H}_Y(\mathbf{y}) + \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln[p_G(\mathbf{y})] d\mathbf{y}, \end{aligned} \quad (\text{A.33})$$

e substituindo-se na Equação (A.32) obtém-se a seguinte relação

$$\mathcal{H}_Y(\mathbf{y}) - \mathcal{H}_G(\mathbf{y}) = \int_{-\infty}^{\infty} p_Y(\mathbf{y}) \cdot \ln \left[\frac{p_G(\mathbf{y})}{p_Y(\mathbf{y})} \right] d\mathbf{y}. \quad (\text{A.34})$$

Ao utilizar-se então a desigualdade $\ln[x] \leq x - 1$, a igualdade só ocorre quando $x = 1$, obtendo-se

$$\mathcal{H}_Y(\mathbf{y}) - \mathcal{H}_G(\mathbf{y}) \leq 0, \tag{A.35}$$

obtendo-se a igualdade somente quando $p_Y(\mathbf{y}) = p_G(\mathbf{y})$.

Desta maneira mostra-se que a entropia da variável gaussiana é máxima.

– “A matemática é como o amor. Uma idéia simples mas que pode, às vezes, se complicar.”

R. Drabek

B

Demonstrações e Derivações

NESTE apêndice são derivados ou provados alguns resultados importantes utilizados nesta tese.

A divisão do apêndice é dada na seguinte estrutura. Na Seção B.1 são destacadas as condições necessárias para a recuperação de sinais baseadas no teorema de Shalvi-Weinstein. O teorema de Benveniste-Goursat-Ruget é provado na Seção B.2 num formato original nesta tese e o método de ortogonalização de Gram-Schmidt é apresentado em detalhes na Seção B.3.

B.1 Recuperação de sinais baseado no Teorema de Shalvi-Weinstein

O Teorema de Shalvi-Weinstein pode ser enunciado da seguinte maneira:

Teorema SW: *Seja um sinal de distribuição não-gaussiana, pertencente a um alfabeto finito, com amostras i.i.d. e de média nula, transmitido em um canal*

linear. A equalização linear é obtida através da igualdade entre a kurtosis e a variância dos sinais transmitido e recuperado.

O Teorema SW original pode então ser estendido para um conjunto de condições necessárias à recuperação dos sinais num ambiente multi-usuário:

- C_{RS1}**. $a_l(n)$ é i.i.d. e de média zero ($l = 1, \dots, K$);
- C_{RS2}**. $a_l(n)$ e $a_q(n)$ são estatisticamente independentes para $l \neq q$ e têm a mesma fdp;
- C_{RS3}**. $|\mathcal{K}[y_l(n)]| = |\mathcal{K}_a|$ ($l = 1, \dots, K$);
- C_{RS4}**. $\mathbb{E}\{|y_l(n)|^2\} = \sigma_a^2$ ($l = 1, \dots, K$);
- C_{RS5}**. $\mathbb{E}\{y_l(n)y_q^*(n)\} = 0$, $l \neq q$,

em que \mathcal{K}_a e σ_a^2 são, respectivamente, a kurtosis e a variância da seqüência transmitida e $\mathcal{K}[\cdot]$ é o operador kurtosis. Neste caso, a Condição C_{RS5} é inserida para permitir o tratamento de múltiplas fontes e garantir que os sinais recuperados sejam diferentes entre si.

Prova:

Seja \mathbf{G} a matriz de resposta global de um sistema de recuperação de sinais dado por $\mathbf{G} = \mathbf{W}^H \mathbf{H}$, em que \mathbf{W} e \mathbf{H} são as matrizes de separação e de mistura, respectivamente.

Tomando os símbolos das K fontes representados pelo vetor $\mathbf{a} = [a_1 \ \dots \ a_K]^T$, pode-se escrever a variância e a kurtosis do k -ésimo sinal na saída da matriz de separação em função daquelas das fontes como

$$\mathbb{E}\{|y_k(n)|^2\} = \sigma_a^2 \cdot \sum_{i=1}^M |g_{ik}|^2, \quad k = 1, \dots, K \quad (\text{B.1})$$

$$\mathcal{K}[y_k(n)] = \mathcal{K}_a \cdot \sum_{i=1}^M |g_{ik}|^4, \quad k = 1, \dots, K, \quad (\text{B.2})$$

em que K é o número de fontes e M é o número de sensores.

Além disso, sabe-se que

$$\sum_{i=1}^M |g_{ik}|^4 \leq \left(\sum_{i=1}^M |g_{ik}|^2 \right)^2, \quad k = 1, \dots, K, \quad (\text{B.3})$$

a qual só mantém a igualdade quando \mathbf{g}_k , que é a k -ésima coluna de \mathbf{G} , for um vetor com apenas um elemento não-nulo com magnitude unitária.

De fato, para se respeitar as Condições C_{RS4} e C_{RS3} , deve-se ter

$$\sum_{i=1}^M |g_{ik}|^4 = 1, \quad k = 1, \dots, K, \quad (B.4)$$

e

$$\sum_{i=1}^M |g_{ik}|^2 = 1, \quad k = 1, \dots, K, \quad (B.5)$$

o que só pode ser possível se \mathbf{g}_k for da forma

$$\mathbf{g}_k = [0 \ \cdots \ 0 \ \exp(j\phi_k) \ 0 \ \cdots \ 0]^T, \quad k = 1, \dots, K, \quad (B.6)$$

em que $j = \sqrt{-1}$ e $\phi_k \in [0, 2\pi)$ é uma fase arbitrária da recuperação do sinal da fonte k .

De acordo com a Condição C_{RS5} deve-se ter também

$$\mathbf{g}_k^H \mathbf{g}_l = 0, \quad k \neq l, \quad k = 1, \dots, K. \quad (B.7)$$

Desta forma, para que todas as fontes sejam recuperadas sem interferência, as colunas da matriz \mathbf{G} , que correspondem aos vetores \mathbf{g}_k , devem ter a forma dada na Equação (B.6), mas com o elemento não-nulo em diferentes posições para cada uma das colunas de \mathbf{G} ■.

B.2 Teorema de Benveniste-Goursat-Ruget

Teorema BGR: *Seja um sinal cuja distribuição de probabilidade é não-gaussiana, transmitido através de um canal linear. O sinal na saída de um equalizador linear será aquele da entrada do sistema, se as distribuições da entrada e da saída do sistema forem iguais.*

Prova:

Seja \mathbf{g} o vetor de resposta global dado por

$$\mathbf{g} = \mathcal{H} \cdot \mathbf{w} \quad (B.8)$$

em que \mathcal{H} é a matriz de convolução do canal e \mathbf{w} é o vetor representando a resposta do do equalizador. Então, pode-se obter a seguinte relação:

$$y(n) = \sum_{i=0}^{L+M-2} g_i \cdot a(n-i), \quad (B.9)$$

em que L e M são respectivamente os comprimentos do canal e do equalizador e $a(n)$ é o símbolo transmitido no instante n .

Pode-se então calcular o cumulante de ordem k para o sinal na saída do sistema obtendo-se

$$c_k(Y) = c_k\left(\sum_{i=0}^{L+M-2} g_i \cdot a(n-i)\right), \quad (\text{B.10})$$

e denotando $\alpha = \sum_{i=0}^{L+M-2} g_i$ por simplicidade de notação e evocando as propriedades da homogeneidade e da aditividade dos cumulantes mostradas no Apêndice A, pode-se escrever:

$$c_k(Y) = \sum_{i=0}^{L+M-2} \alpha^k \cdot c_k(a(n-i)). \quad (\text{B.11})$$

Assumindo-se estacionariedade da fonte, tem-se que $c_k(a(n-i)) = c_k(A)$, o que resulta em

$$c_k(Y) = \sum_{i=0}^{L+M-2} \alpha^k \cdot c_k(A). \quad (\text{B.12})$$

Deste modo ao igualar-se todos os cumulantes da entrada aos da saída do sistema, correspondendo a igualar as distribuições [Prasad & Menicucci, 2004], deve-se ter, simultaneamente:

$$\sum_{i=0}^{L+M-2} (g_i)^k = 1 \quad \forall k \in \mathbb{N}. \quad (\text{B.13})$$

A expressão (B.13) é satisfeita $\forall k \in \mathbb{N}$ se a resposta global \mathbf{g} tem $L + M - 2$ coeficientes nulos e um elemento com módulo unitário e fase $\frac{2\pi \cdot m}{k}$, para $m \in \mathbb{Z}$.

Com isso, igualar todos os cumulantes dos sinais na entrada e na saída corresponde a igualar as densidades de probabilidade, uma vez que esta é determinada unicamente por todos os cumulantes ■.

Unicidade da solução:

De fato, como os sinais de entrada $a(n)$ são limitados, ou seja, $\sum_{i=1}^{L+M-1} |g_i|^2 < \infty$, pode-se escrever que:

$$\sum_{i=1}^{L+M-1} |g_i|^n = \sum_{i=1}^{L+M-1} |g_i|^q = 1, \quad (\text{B.14})$$

em que $n \neq q$ são diferentes ordens de estatísticas superiores.

Ao escrever-se a soma da diferença das potências na Equação (B.14), e assumindo por conveniência que $n > q$, tem-se:

$$\sum_{i=1}^{L+M-1} |g_i|^n - |g_i|^q = 0. \quad (\text{B.15})$$

Para a igualdade na Equação (B.15) ser verdadeira, deve-se ter que

$$|g_i|^n = |g_i|^q, \quad (\text{B.16})$$

Logo, uma vez que assumiu-se que $n \neq q$, é necessário ter

$$|g_i| = 1 \quad \text{ou} \quad g_i = 0, \quad (\text{B.17})$$

para assegurar a Equação (B.16), uma vez que somente termos positivos cuja soma elevada a qualquer potência deve ser igual à 1.

Além disso, a seguinte condição tem de ser satisfeita:

$$\sum_{i=1}^{L+M-1} |g_i| = 1. \quad (\text{B.18})$$

A Equação (B.18) mostra que apenas um elemento não-nulo com módulo unitário respeita todas as restrições. Isto mostra a unicidade da solução quando $p_Y(y) = p_A(a)$ ■.

B.3 Ortogonalização de Gram-Schmidt

Em álgebra linear, o procedimento de Gram-Schmidt é um método de ortogonalização de um conjunto de vetores em um espaço do produto interno. Um dos mais comuns espaços no qual se trabalha é o *espaço euclidiano* \mathbb{R}^n .

Neste contexto, ortogonalização significa: ter-se um conjunto de K vetores representados por $\mathbf{v}_1, \dots, \mathbf{v}_K$ que são linearmente independentes, e obter, a partir deles, um conjunto de K vetores mutuamente ortogonais $\mathbf{u}_1, \dots, \mathbf{u}_K$ que geram o mesmo subespaço dos vetores $\mathbf{v}_1, \dots, \mathbf{v}_K$.

Então, denotando o produto interno (escalar) de dois vetores como sendo $\mathbf{v}^H \mathbf{u}$, pode-se descrever a ortogonalização de Gram-Schmidt através do seguinte procedimento [Haykin, 1989; Golub & Loan, 1996]:

$$\begin{aligned}
 \mathbf{u}_1 &= \mathbf{v}_1 \\
 \mathbf{u}_2 &= \mathbf{v}_2 - [\mathbf{v}_2^H \mathbf{u}_1] \mathbf{u}_1 \\
 \mathbf{u}_3 &= \mathbf{v}_3 - [\mathbf{v}_3^H \mathbf{u}_1] \mathbf{u}_1 - [\mathbf{v}_3^H \mathbf{u}_2] \mathbf{u}_2 \\
 &\vdots \quad \dots \quad \vdots \\
 \mathbf{u}_K &= \mathbf{v}_K - [\mathbf{v}_K^H \mathbf{u}_1] \mathbf{u}_1 - [\mathbf{v}_K^H \mathbf{u}_2] \mathbf{u}_2 - \dots - [\mathbf{v}_K^H \mathbf{u}_{K-1}] \mathbf{u}_{K-1}.
 \end{aligned} \tag{B.19}$$

A Equação(B.19) pode ser resumida através da seguinte recursão [Golub & Loan, 1996]:

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} [\mathbf{u}_i^H \mathbf{v}_k] \mathbf{u}_i. \tag{B.20}$$

Pode-se verificar a validade das fórmulas iniciando o cálculo do produto escalar $\mathbf{u}_1^H \mathbf{u}_2$ e verificando que o mesmo é zero utilizando o valor de \mathbf{u}_2 dado na Equação (B.19). Depois utilizar o resultado e calcular o produto $\mathbf{u}_3^H \mathbf{u}_1$ e verificar a ortogonalidade. O restante da verificação é realizada por indução.

Do ponto de vista geométrico este método realiza o seguinte procedimento: para calcular \mathbf{u}_k ele projeta \mathbf{v}_k ortogonalmente no subespaço U gerado por $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$, o qual é o mesmo que o subespaço gerado pelos vetores $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$. Define-se então \mathbf{u}_k igual a diferença entre \mathbf{v}_k e sua projeção, garantindo ser ortogonal a todos os vetores no subespaço U .

Quando se deseja obter vetores *ortonormais*, deve-se dividir os vetores \mathbf{u}_k pela sua norma.

Vale ressaltar que embora a ortogonalização de Gram-Schmidt seja bastante eficiente, para implementações práticas a transformação de Householder é geralmente preferida por ser numericamente mais estável [Haykin, 1996].

Referências

- ADALI, T., LIU, X., & SÖNMEZ, M. K. (1997). Conditional Distribution Learning with Neural Networks and Its Application to Channel Equalization. *IEEE Transactions on Signal Processing*, 45(4), 1051-1064. Citado na(s) pág(s): 147
- ALMEIDA, L. B. (2003). MISEP – Linear and Nonlinear ICA Based on Mutual Information. *Journal of Machine Learning Research*, 4, 1297-1318. Citado na(s) pág(s): 187
- AMARA, R. (2001). *Égalisation de Canaux Linéaires et Non Linéaires, Approche Bayésienne*. Tese de doutorado, Université Paris XI Orsay, France. Citado na(s) pág(s): 122
- AMBLARD, P.-O., GAETA, M., & LACOUME, J.-L. (1996a). Statistics for Complex Variables and Signals - Part I: Variables. *Signal Processing*, 53, 1-13. Citado na(s) pág(s): 228
- AMBLARD, P.-O., GAETA, M., & LACOUME, J.-L. (1996b). Statistics for Complex Variables and Signals - Part II: Signals. *Signal Processing*, 53, 15-25. Citado na(s) pág(s): 229
- ATTUX, R. R. F. (2001). *Sobre Dinâmica Caótica e Convergência em Algoritmos de Equalização Autodidata*. Dissertação de mestrado, Universidade Estadual de Campinas (UNICAMP) – DECOM, Campinas - SP, Brasil. Citado na(s) pág(s): 126
- BARBAROSSA, S., & SCAGLIONE, A. (1997). Blind Equalization Using Cost Function Matched to the Signal Constellation. In *Proceedings of IEEE 31st Asilomar Conference on Signals, Systems and Computer* (p. 550-554). Pacific Grove (CA) - USA. Citado na(s) pág(s): 123
- BELL, A. J., & SEJNOWSKI, T. J. (1995). An Information-Maximisation Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, Vol. 7(No. 6), pp. 1129-1159. Citado na(s) pág(s): 40, 44, 45
- BENEDETTO, S., BIGLIERI, E., & CASTELLANI, V. (1987). *Digital Transmission Theory*.

- Prentice Hall International. Citado na(s) pág(s): 160
- BENVENISTE, A., GOURSAT, M., & RUGET, G. (1980). Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications. *IEEE Transactions on Automatic Control*, Vol. AC-25(No. 3), 385-399. Citado na(s) pág(s): 40, 45, 140, 144
- BISHOP, C. M. (1995). *Neural Networks for Pattern Recognition*. UK: Oxford University Press. Citado na(s) pág(s): 153
- BROSSIER, J.-M. (1997). *Signal et Communication Numérique: égalisation et synchronisation*. HERMES. Citado na(s) pág(s): 135, 211
- BUGALLO, M. F., MÍGUEZ, J., & CASTEDO, L. (2000). Semiblind Linear Multiuser Interference Cancellation: A Maximum Likelihood Approach. *Signal Processing*, 81, 2041-2057. Citado na(s) pág(s): 113
- BUGALLO, M. F., MÍGUEZ, J., & CASTEDO, L. (2001). A Maximum Likelihood Approach to Blind Multiuser Interference Cancellation. *IEEE Transactions on Signal Processing*, 49(6), 1228-1239. Citado na(s) pág(s): 113
- CAO, X.-R., & LIU, R. WEN. (1996). General Approach to Blind Source Separation. *IEEE Transactions on Signal Processing*, 44(3), 562-571. Citado na(s) pág(s): 27
- CARDOSO, J.-F. (1989). Source Separation Using Higher Order Moments. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'89)* (p. 2109-2112). Glasgow, Scotland. Citado na(s) pág(s): 17
- CARDOSO, J.-F. (1997). Infomax and Maximum Likelihood for Blind Source Separation. *IEEE Signal Processing Letters*, 4(4), 112-114. Citado na(s) pág(s): 48
- CARDOSO, J.-F. (1998). Blind Signal Separation: Statistical Principles. *Proceedings of the IEEE*, 86(10), 2009-2025. Citado na(s) pág(s): 48, 49, 53, 63, 159
- CARDOSO, J.-F. (2000). Entropic Contrasts for Source Separation: Geometry and Stability. In *Unsupervised Adaptive Filtering* (Vol. 1, p. 139-190). John Wiley & Sons. Citado na(s) pág(s): 38, 55, 56
- CARDOSO, J.-F., & LAHELD, B. H. (1996). Equivariant Adaptive Source Separation. *IEEE Transactions on Signal Processing*, 44(12), 3017-3030. Citado na(s) pág(s): 55, 56, 69
- CARLEIAL, A. B., & HELLMAN, M. E. (1975). Bistable Behavior of ALOHA-Type Systems. *IEEE Transactions on Communications*, 23, 410-440. Citado na(s) pág(s): 206

- CASTEDO, L., ESCUDERO, C. J., & DAPENA, A. (1997). A Blind Signal Separation Method for Multiuser Communications. *IEEE Transactions on Signal Processing*, 45(5), 1343-1348. Citado na(s) pág(s): 112
- CAVALCANTE, C. C. (2001). *Predição Neural e Estimação de Função Densidade de Probabilidade Aplicadas à Equalização Cega*. Dissertação de mestrado, Departamento de Engenharia Elétrica, Universidade Federal do Ceará (UFC), Fortaleza-CE. Citado na(s) pág(s): 33, 44, 60, 118, 119, 120, 135, 137, 143, 158
- CAVALCANTE, C. C., CAVALCANTI, F. R. P., & MOTA, J. C. M. (2002a). A PDF Estimation-Based Blind Criterion for Adaptive Equalization. In *Proceedings of IEEE Int. Symposium on Telecommunications (ITS 2002)*. Natal, Brazil. Citado na(s) pág(s): 124, 125, 135, 137
- CAVALCANTE, C. C., CAVALCANTI, F. R. P., & MOTA, J. C. M. (2002b). Adaptive Blind Multiuser Separation Criterion Based on Log-Likelihood Maximisation. *IEE Electronics Letters*, 38(20), 1231-1233. Citado na(s) pág(s): 150
- CAVALCANTE, C. C., CAVALCANTI, F. R. P., MOTA, J. C. M., & FREITAS, W. D. C., JR. (2003). Collision Resolution in Slotted ALOHA Smart Antenna System Using PDF Estimation-Based Blind Criteria. In *Proceedings of World Wireless Communications 2003 (WWC'03)*. San Francisco, USA. Citado na(s) pág(s): 207
- CAVALCANTE, C. C., CAVALCANTI, F. R. P., MOTA, J. C. M., & ROMANO, J. M. T. (2003). A Constrained Version of Fitting PDF Algorithm for Blind Source Separation. In *Proceeding of IEEE Signal Processing Advances for Wireless Communications (SPAWC 2003)*. Rome, Italy. Citado na(s) pág(s): 154
- CAVALCANTI, F. R. P. (1999). *Antenas Inteligentes & Processamento Espaço-Temporal para Sistemas de Comunicação Sem Fio*. Tese de doutorado, Universidade Estadual de Campinas (UNICAMP), Campinas, SP - Brasil. Citado na(s) pág(s): 8, 9, 83, 86, 87, 89, 91, 92, 94, 100, 102, 103, 104, 105, 112, 152, 196, 198
- CAVALCANTI, F. R. P., BRANDÃO, A. L., & ROMANO, J. M. T. (1999). The Generalized Constant Modulus Algorithm Applied to Multiuser Space-Time Equalization. In *Proceedings of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'99)* (p. 94-97). Annapolis, MD - USA. Citado na(s) pág(s): 105, 112, 113
- CAVALCANTI, F. R. P., & ROMANO, J. M. T. (1999). Blind Multiuser Detection in Space Division Multiple Access Systems. *Annales des Télécommunications*(7-8), 411-419. Citado na(s) pág(s): 89, 100, 102, 103, 104

- CAVALCANTI, F. R. P., & ROMANO, J. M. T. (2000). Using Adaptive Array for Collision Resolution in Slotted ALOHA Packet Radio Systems. *Journal of Communications and Networks*, 2(4), 344-350. Citado na(s) pág(s): 206, 207
- CAVALCANTI, F. R. P., ROMANO, J. M. T., & BRANDÃO, A. L. (1999). Lest-Squares CMA with Decorrelation for Fast Blind Multiuser Signal Separation. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'99)* (Vol. V, p. 2527-2530). Phoenix, AZ - USA. Citado na(s) pág(s): 112
- CHEN, S., SAMINGAN, A. K., MULGREW, B., & HANZO, L. (2001). Adaptive Minimum-BER Linear Multiuser Detection for DS-CDMA Signals in Multipath Channels. *IEEE Transactions on Signal Processing*, 49(6), 1240-1247. Citado na(s) pág(s): 160
- CHEN, S., WU, Y., & MCCLAUGHLIN, S. (1997). Genetic Algorithm Optimization for Blind Channel Identification with Higher Order Cumulant Fitting. *IEEE Transactions on Evolutionary Computation*, 1(4), 259-265. Citado na(s) pág(s): 188
- CHI, C.-Y., CHEN, C.-H., & CHEN, C.-Y. (2002). Blind MAI and ISI Suppression for DS/CDMA Systems HOS-Based Inverse Filter Criteria. *IEEE Transactions on Signal Processing*, 50(6), 1368-1381. Citado na(s) pág(s): 113
- CICHOCKI, A., & AMARI, S. ICHI. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons. Citado na(s) pág(s): 7, 22, 25, 36, 54, 61, 63, 65, 66, 71, 73, 75, 76, 80, 107
- COMON, P. (1989). Separation of Sources Using Higher-Order Cumulants. In *SPIE Conference on Advanced Algorithms and Architectures for Signal Processing* (p. 170-181). San Diego, California, USA. Citado na(s) pág(s): 17
- COMON, P. (1994). Independent Component Analysis: A New Concept? *Signal Processing*, 36(3), 287-314. Citado na(s) pág(s): 17, 27, 34, 37, 38, 39, 52, 175, 188
- COPPI, R., & BOLASCO, S. (Eds.). (1989). *Multiway Data Analysis*. Netherlands: Elsevier Science Publishers B.V. (North-Holland). Citado na(s) pág(s): 24
- COVER, T. M., & THOMAS, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons. Citado na(s) pág(s): 232
- DECO, G., & BRAUER, W. (1995). Nonlinear Higher-Order Statistical Decorrelation by Volume-Conserving Neural Architectures. *Neural Networks*, 8(4), 525-535. Citado na(s) pág(s): 175
- DECO, G., & OBRADOVIC, D. (1996). *An Information-Theoretic Approach to Neural*

- Computing*. Springer-Verlag. Citado na(s) pág(s): 32
- DELFOSSÉ, N., & LOUBATON, P. (1995). Adaptive Blind Separation of Independent Sources: A Deflation Approach. *Signal Processing*, 45, 59-83. Citado na(s) pág(s): 54, 57, 58
- DENEIRE, L. (1998). *Estimation Aveugle de Canal et Accès Multiple par Répartition Spatiale*. Tese de doutorado, École National des Télécommunications - ENST Paris, Paris, France. Citado na(s) pág(s): 107, 109
- DENEIRE, L., & SLOCK, D. T. M. (1999a). A Deterministic Schur Method for Multichannel Blind Identification. In *Proceedings of 2nd IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'99)*. Annapolis, MD - USA. Citado na(s) pág(s): 109
- DENEIRE, L., & SLOCK, D. T. M. (1999b). A Schur Method for Multiuser Multichannel Blind Identification. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'99)* (p. 2905-2908). Phoenix, AZ - USA. Citado na(s) pág(s): 109
- DEVROYE, L. (1987). *A Course in Density Estimation* (Vol. 14). Birkhäuser. Citado na(s) pág(s): 141, 144, 168, 221
- DING, Z., & LI, Y. G. (2001). *Blind Equalization and Identification*. New York, USA: Marcel Dekker. Citado na(s) pág(s): 133
- DINIZ, P. S. R. (1997). *Adaptive Filtering: Algorithms and Practical Implementation*. Kluwer Academic. Citado na(s) pág(s): 99, 100
- DONOHO, D. (1981). On Minimum Entropy Deconvolution. In *Applied Time Series Analysis II* (p. 565-608). Academic Press. Citado na(s) pág(s): 40, 60, 222
- DUDA, R., & HART, P. (1973). *Pattern Classification and Scene Analysis*. New York, USA: Wiley. Citado na(s) pág(s): 138
- FAÏ, G., MOULINES, E., & SOULIER, P. (2004). Edgeworth Expansions for Linear Statistics of Possibly Long-Range-Dependent Linear Processes. *Statistics & Probability Letters*, 66, 275-288. Citado na(s) pág(s): 174
- FREITAS, W. D. C., JR. (2002). *Adaptação de Enlace em Comunicações Móveis com Aplicação ao Sistema EGPRS*. Dissertação de mestrado, Universidade Federal do Ceará (UFC), Fortaleza - CE. Citado na(s) pág(s): 98
- FREITAS, W. D. C., JR., CAVALCANTI, F. R. P., CAVALCANTE, C. C., ZANATTA, D., FILHO, & ALMEIDA, A. L. F. DE. (2002). Link Performance Evaluation for EGPRS

- with Multipl Antennas. In *Proceedings of XIII International Symposium on Personal and Indoor Mobile Radio Communications (PIMRC2002)*. Lisboa, Portugal. Citado na(s) pág(s): 98
- GAMBOA, F., & GASSIAT, E. (1997). Source Separation when the Input Sources Are Discrete or Have Constant Modulus. *IEEE Transactions on Signal Processing*, 45(12), 3062-3072. Citado na(s) pág(s): 98
- GHEZ, S., VERDÚ, S., & SCHWARTZ, S. (1989). Optimal Decentralized Control in the Random Access Multipacket Channel. *IEEE Transactions on Automatic Control*, 34(11), 1153-1163. Citado na(s) pág(s): 207
- GIANNAKIS, G. B., HUA, Y., STOICA, P., & TONG, L. (Eds.). (2001). *Signal Processing Advances in Wireless & Mobile Communications* (Vol. 2: Trends in Single- and Multi-User Systems). Upple Sader River, NJ: Prentice-Hall. Citado na(s) pág(s): 9
- GIBSON, J. D. (1993). *Principles of Digital and Analog Communications* (2nd ed.). New Jersey, USA: Prentice-Hall. Citado na(s) pág(s): 31
- GODARA, L. C. (1997a). Application of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations. *Proceedings of the IEEE*, 85(8), 1195-1245. Citado na(s) pág(s): 87
- GODARA, L. C. (1997b). Applications of Antenna Arrays to Mobile Communications, Part I: Performance Improvement, Feasibility, and System Considerations. *Proceedings of the IEEE*, 85(7), 1031-1060. Citado na(s) pág(s): 82
- GODARD, D. N. (1980). Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems. *IEEE Transactions on Communications*, Vol. COM-28(No. 11), 1867-1875. Citado na(s) pág(s): 101
- GOLUB, G. H., & LOAN, C. F. V. (1996). *Matrix Computations* (3rd ed.). Baltimore, USA: The Johns Hopkins University Press. Citado na(s) pág(s): 24, 107, 203, 242
- GOOCH, R. P., & LUNDELL, J. (1986). The CM Array: An Adaptive Beamformer for Constant Modulus Signals. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'86)* (p. 2523-2526). Tokyo, Japan. Citado na(s) pág(s): 111
- GOROKHOV, A., & LOUBATON, P. (1997). Subspace-Based Techniques for Blind Separation of Convulsive Mixtures with Temporally Correlated Sources. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 44(9), 813-820. Citado na(s) pág(s): 110

- GOROKHOV, A., & LOUBATON, P. (1999). Blind Identification of MIMO-FIR Systems: A Generalized Linear Prediction Approach. *Signal Processing*, 73, 105-124. Citado na(s) pág(s): 110
- GRELLIER, O., & COMON, P. (1998). Blind Separation of Discrete Sources. *IEEE Signal Processing Letters*, 5(8), 212-214. Citado na(s) pág(s): 98
- HAGENAUER, J., & HOEHER, P. (1989). A Viterbi Algorithm with Soft-Decision Outputs and its Applications. In *Proceedings of globecom '89* (p. 1680-1686). Dallas, TX - USA. Citado na(s) pág(s): 136
- HAYKIN, S. (1989). *Modern Filters*. New York - USA: Macmillan. Citado na(s) pág(s): 109, 242
- HAYKIN, S. (Ed.). (1994). *Blind Deconvolution*. NJ, USA: Prentice-Hall, Englewood Cliffs. Citado na(s) pág(s): 36, 60, 62, 119, 120
- HAYKIN, S. (1996). *Adaptive Filter Theory* (3rd ed.). Prentice-Hall. Citado na(s) pág(s): 62, 99, 100, 103, 109, 120, 126, 133, 135, 138, 242
- HAYKIN, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall. Citado na(s) pág(s): 31, 32, 33, 34, 44, 52, 60, 79, 168, 169, 170, 171, 172, 174
- HAYKIN, S. (Ed.). (2000a). *Unsupervised Adaptive Filtering* (Vol. I: Source Separation). John Wiley & Sons. Citado na(s) pág(s): 17, 19, 22, 24, 42, 45, 56, 63, 75, 80, 95
- HAYKIN, S. (Ed.). (2000b). *Unsupervised Adaptive Filtering* (Vol. II: Blind Deconvolution). John Wiley & Sons. Citado na(s) pág(s): 94
- HÉRAULT, J., & JUTTEN, C. (1994). *Réseaux Neuronaux et Traitement du Signal*. HERMES. Citado na(s) pág(s): 16, 54, 61
- HÉRAULT, J., JUTTEN, C., & ANS, B. (1985). Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *Actes du Xéme Colloque GRETSI* (p. 1017-1022). Nice, France. Citado na(s) pág(s): 4, 16, 54
- HJØRUNGNES, A., DINIZ, P. S. R., & CAMPOS, M. L. R. (2004). Jointly Minimum BER Transmitter and Receiver FIR MIMO Filters for Binary Signal Sources. *IEEE Transactions on Signal Processing*, 52(4), 1021-1036. Citado na(s) pág(s): 160
- HYVÄRINEN, A. (1997). One-unit Contrast Functions for Independent Component Analysis: A Statistical Analysis. In *Neural Networks for Signal Processing VII (Proceedings of IEEE Workshop on Neural Networks for Signal Processing)* (p. 388-397). Amelia Island,

- Florida. Citado na(s) pág(s): 41, 59
- HYVÄRINEN, A. (1998). New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit. In *Advances in Neural Information Processing Systems 10 (Proceedings of NIPS'97)* (p. 273-279). MIT Press, Cambridge. Citado na(s) pág(s): 168
- HYVÄRINEN, A. (1999). Survey on Independent Component Analysis. *Neural Computing Surveys*, 2, 94-128. Citado na(s) pág(s): 41, 42, 45, 54, 55, 56, 58, 59, 62
- HYVÄRINEN, A., & OJA, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5), 411-430. Citado na(s) pág(s): 17, 22, 26, 27, 30, 34, 35, 36, 42, 58
- HYVÄRINEN, A., OJA, E., & KARHUNEN, J. (2001). *Independent Component Analysis*. John Wiley & Sons. Citado na(s) pág(s): 16, 17, 22, 24, 35, 36, 52, 59, 62, 80, 175
- JUTTEN, C., & HÉRAULT, J. (1991). Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture. *Signal Processing*, 24(1), 1-10. Citado na(s) pág(s): 16
- KAY, S. M., NUTTAL, A. H., & BAGGENSTOSS, P. M. (2001). Multidimensional Probability Density Function Approximations for Detection, Classification, and Model Order Selection. *IEEE Transactions on Signal Processing*, 49(10), 2240-2252. Citado na(s) pág(s): 175
- KNUTH, K. H. (1999). A Bayesian Approach to Source Separation. In *Proceedings of International Symposium on Independent Component Analysis and Blind Signal Separation (ICA'99)* (p. 283-288). Aussois, France. Citado na(s) pág(s): 50
- KOFIDIS, E. (2001). *Blind Source Separation: Fundamentals and Recent Advances*. Mini-curso no XIX Simpósio Brasileiro de Telecomunicações (SBrT 2001), Fortaleza-CE, Setembro. Citado na(s) pág(s): 18, 22, 23, 24, 26, 27, 32, 33, 34, 37, 38, 42, 48, 49, 51, 52, 53, 55, 58, 59, 60, 61, 62, 63, 159
- KÜHN, V. (1999). Evaluating the Performance of Turbo Codes and Turbo-Coded Modulation in a DS-CDMA Environment. *IEEE Journal on Selected Areas in Communications*, 17(12), 2138-2147. Citado na(s) pág(s): 136
- KULLBACK, S., & LEIBLER, R. A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22, 79-86. Citado na(s) pág(s): 32
- LAAR, J. VAN DE. (2004). On MIMO Instantaneous Blind Identification Based on the Exploitation of the Time Structure of Signals Using Arbitrary-Order Cumulants.

- In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP 2004)*. Montreal, Quebec. Citado na(s) pág(s): 180
- LACOUME, J.-L., AMBLARD, P.-O., & COMON, P. (1997). *Statistiques d'Ordre Supérieur pour le Traitement du Signal*. Paris: Masson. Citado na(s) pág(s): 24, 26, 39, 51, 52, 170, 171, 174, 229
- LAMARE, R. C. DE, & SAMPAIO-NETO, R. (2002). Adaptive Channel Equalisation Using Minimum BER Gradient-Newton Algorithms. In *Proceedings of IEEE Int. Symposium on Telecommunications (ITS 2002)*. Natal, RN - Brazil. Citado na(s) pág(s): 160
- LAMBOTHARAN, S., & CHAMBERS, J. A. (1999). On the Surface Characteristics of a Mixed Constant Modulus and Cross-Correlation Criterion for the Blind Equalization of a MIMO Channel. *Signal Processing*, 74, 209-216. Citado na(s) pág(s): 112
- LAMBOTHARAN, S., CHAMBERS, J. A., & CONSTANTINIDES, A. G. (1999). Adaptive Blind Retrieval Techniques for Multiuser DS-CDMA Signals. *IEE Electronics Letters*, 35(9), 693-695. Citado na(s) pág(s): 112
- LASTER, J. D. (1997). *Robust GMSK Demodulation Using Demodulator Diversity and BER Estimation*. Tese de doutorado, Faculty of the Virginia Polytechnic Institute and State University, Blacksburg, Virginia. Citado na(s) pág(s): 143, 159, 160, 171, 172, 173, 175, 176
- LATHAUWER, L. DE. (1997). *Signal Processing Based on Multilinear Algebra*. Tese de doutorado, Katholieke Universiteit Leuven, Belgium. Citado na(s) pág(s): 6, 17
- LÁZARO, M., SANTAMARÍA, I., ERDOGMUS, D., HILD, K. E., II, & PRÍNCIPE, J. C. (2003). Matched PDF-Based Blind Equalization. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP 2003)* (Vol. IV, p. 297-300). Hong Kong, China. Citado na(s) pág(s): 147
- LÁZARO, M., SANTAMARÍA, I., PANTALEÓN, C., ERDOGMUS, D., HILD, K. E., II, & PRÍNCIPE, J. C. (2003). Blind Equalization by Sampled PDF Fitting. In *Proceedings of the 4th Int. Symposium on Independent Component Analysis and Blind Source Separation (ICA 2003)* (p. 1041-1046). Nara, Japan. Citado na(s) pág(s): 146
- LEARY, J. (1997). Least-Squares Multi-User CMArray: A New Algorithm for Blind Adaptive Beamforming. In *Proceedings of IEEE 31st Asilomar Conference on Signals, Systems and Computer* (p. 902-905). Pacific Grove (CA) - USA. Citado na(s) pág(s): 112
- LEE, E. A., & MESSERSCHMITT, D. G. (1993). *Digital Communication* (2nd ed.). Kluwer Academic Publishers. Citado na(s) pág(s): 96

- LEE, T.-W., GIROLAMI, M., BELL, A. J., & SEJNOWSKI, T. J. (2000). A Unifying Information-Theoretic Framework for Independent Component Analysis. *Computers and Mathematics with Applications*, 39(11), 1-21. Citado na(s) pág(s): 52, 167, 175
- LI, T.-H. (1992). Blind Identification and Deconvolution of Linear Systems Driven by Binary Random Sequences. *IEEE Transactions on Information Theory*, 38(1), 26-38. Citado na(s) pág(s): 98
- LI, T.-H. (1995). Blind Deconvolution of Linear Systems with Multilevel Nonstationary Inputs. *The Annals of Statistics*, 23(2), 690-704. Citado na(s) pág(s): 98
- LI, T.-H., & MBAREK, K. (1997). A Blind Equalizer for Nonstationary Discrete-Valued Signals. *IEEE Transactions on Signal Processing*, 45(1), 247-254. Citado na(s) pág(s): 98
- LI, T., & SIDIROPOULOS, N. D. (2000). Blind Digital Signal Separation Using Successive Interference Cancellation Iterative Least Squares. *IEEE Transactions on Signal Processing*, 48(11), 3146-3152. Citado na(s) pág(s): 112
- LI, Y. G., & LIU, K. J. R. (1998). Adaptive Blind Source Separation and Equalization for Multiple-Input/Multiple-Output Systems. *IEEE Transactions on Information Theory*, 44(7), 2864-2876. Citado na(s) pág(s): 112
- LIBERTI, J. C., JR., & RAPPAPORT, T. S. (1999). *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Prentice-Hall. Citado na(s) pág(s): 86, 87, 111
- LII, K.-S., & ROSENBLATT, M. (1982). Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non-Gaussian Linear Processes. *The Annals of Statistics*, 10(4), 1195-1208. Citado na(s) pág(s): 3, 148
- LIN, J. K., GRIER, D. G., & COWAN, J. D. (1997). Faithful Representation of Separable Distributions. *Neural Computation*, 9, 1305-1320. Citado na(s) pág(s): 175
- LINDSEY, J. K. (1996). *Parametric Statistical Inference*. Oxford University Press. Citado na(s) pág(s): 51, 141, 176
- MACCHI, O., & EWEDA, E. (1984). Convergence Analysis of Self-Adaptive Equalizers. *IEEE Transactions on Information Theory*, IT-30(2), 161-176. Citado na(s) pág(s): 135
- MACCHI, O., & MOREAU, E. (1999). Adaptive Unsupervised Separation of Discrete Sources. *Signal Processing*, 73(1-2), 49-66. Citado na(s) pág(s): 98
- MBOUP, M., & REGALIA, P. A. (2000). A Gradient Search Interpretation of the

- Super-Exponential Algorithm. *IEEE Transactions on Information Theory*, 46(7), 2731-2734. Citado na(s) pág(s): 60, 222
- MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall. Citado na(s) pág(s): 51
- MÍGUEZ, J., & CASTEDO, L. (1998). A Constant Modulus Blind Adaptive Receiver for Multiuser Interference Suppression. *Signal Processing*, 71, 15-27. Citado na(s) pág(s): 112
- MO, S., & SHAFAI, B. (1994). Blind Equalization Using Higher Order Cumulants and Neural Network. *IEEE Transactions on Signal Processing*, Vol. 42(No. 11), 3209-3217. Citado na(s) pág(s): 188
- MOHAMMAD-DJAFARI, A. (1999). A Bayesian Approach to Source Separation. In *Proceedings of 19th Int. Workshop on Bayesian and Maximum Entropy Methods (MaxEnt'99)*. Boise, Idaho, USA. Citado na(s) pág(s): 49
- MONTALVÃO, J. R., FILHO. (2000). *Égalisation et Identification de Canaux de Communication Numérique: Une Approche par Reconnaissance des Formes et Mélange de Gaussiennes*. Tese de doutorado, Université Paris XI, France. Citado na(s) pág(s): 33, 60, 120
- MOTA, J. C. M. (1992). *Equalização Adaptativa em Sistemas Rádio-Digitais: das Técnicas em Freqüência à Teoria da Desconvolução Autodidata*. Tese de doutorado, DECOM/FEEC/UNICAMP, Brasil. Citado na(s) pág(s): 180
- MUTHIAC, R., & HULLE, M. M. V. (2003). A Comparative Survey on Adaptive Neural Network Algorithms for Independent Component Analysis. *Romanian Reports in Physics*, 55(1), 43-67. Citado na(s) pág(s): 76, 80
- NADAL, J.-P., & PARGA, N. (1997). Redundancy Reduction and Independent Component Analysis: Conditions on Cumulants and Adaptive Approaches. *Neural Computation*, 9, 1421-1456. Citado na(s) pág(s): 180
- NEVES, A. O. (2001). *Uma Abordagem Unificada para Algoritmos de Equalização Autodidata*. Dissertação de mestrado, Universidade Estadual de Campinas (UNICAMP) – DECOM, Campinas - SP, Brasil. Citado na(s) pág(s): 113
- NIKIAS, C. L., & PETROPULU, A. P. (1993). *Higher-Order Spectra Analysis*. Prentice Hall. Citado na(s) pág(s): 230, 231
- ORANGE, M. D. (1998). *Packetised Wireless Communication Systems in Interference Limited Environments*. Tese de doutorado, University of Auckland, New Zealand. Citado na(s) pág(s): 206

- ÖZGÜL, B. (2002). *Blind Collision Resolution for Wireless Multiple Access Using Independent Component Analysis*. Dissertação de mestrado, Boğaziçi University, Istanbul, Turkey. Citado na(s) pág(s): 206
- PAPADIAS, C. B. (1995). *Methods for Blind Equalization and Identification of Linear Channels*. Tese de doutorado, École National Supérieure des Télécommunications (ENST), France. Citado na(s) pág(s): 103
- PAPADIAS, C. B. (2000a). Blind Separation of Independent Sources Based on Multiuser Kurtosis Optimization Criteria. In *Unsupervised Adaptive Filtering* (Vol. 2, p. 147-179). John-Wiley & Sons. Citado na(s) pág(s): 35, 37, 58, 94, 106, 107, 153, 156
- PAPADIAS, C. B. (2000b). Globally Convergent Blind Source Separation Based on a Multiuser Kurtosis Maximization Criterion. *IEEE Transactions on Signal Processing*, 48(12), 3508-3519. Citado na(s) pág(s): 37, 58, 106, 107, 153, 156, 193
- PAPADIAS, C. B. (2000c). New Unsupervised Processing Techniques for Wideband Multiple Transmitter/Multiple Receiver Systems. In *Conference Record of the 34th Asilomar Conference on Signals, Systems and Computers* (Vol. 1, p. 759-763). Pacific Grove, CA - USA. Citado na(s) pág(s): 110
- PAPADIAS, C. B. (2004). Unsupervised Receiver Processing Techniques for Linear Space-Time Equalization of Wideband Multiple Input / Multiple Output Channels. *IEEE Transactions on Signal Processing*, 52(2), 472-482. Citado na(s) pág(s): 109, 110
- PAPADIAS, C. B., & HUANG, H. (2001). Linear Space-Time Multiuser Detection for Multipath CDMA Channels. *IEEE Journal on Selected Areas in Communications*, 19(2), 254-265. Citado na(s) pág(s): 112
- PAPADIAS, C. B., & PAULRAJ, A. J. (1997). A Constant Modulus Algorithm for Multiuser Signal Separation in Presence of Delay Spread Using Antenna Array. *IEEE Signal Processing Letters*, 4(6), 178-181. Citado na(s) pág(s): 98, 101, 102, 106, 149
- PAPOULIS, A. (1991). *Probability, Random Variables and Stochastic Processes* (3rd ed.). McGraw-Hill International. Citado na(s) pág(s): 24, 26, 29, 35, 40, 46, 47, 120, 132, 169, 176, 228
- PARASCHIV-IONESCU, A., JUTTEN, C., & BOUVIER, G. (2002). Source Separation Based Processing for Integrated Hall Sensor Arrays. *IEEE Sensors Journal*, 2(6). Citado na(s) pág(s): 202, 203
- PARZEN, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3), 1066-1076. Citado na(s) pág(s): 142, 143

- PAULRAJ, A. J., & PAPADIAS, C. B. (1997). Space-Time Processing for Wireless Communications. *IEEE Signal Processing Magazine*, 49-83. Citado na(s) pág(s): 94
- PETRUS, P. (1997). *Novel Adaptive Array Algorithms and Their Impact on Cellular System Capacity*. Tese de doutorado, Virginia Polytechnic Institut and State University, Blacksburg, Virginia - USA. Citado na(s) pág(s): 85, 86, 87, 88, 89, 91, 111
- PHILLIPS, K. A. (1999). *Probability Density Function Estimation Applied to Minimum Bit Error Rate Adaptive Filtering*. Dissertação de mestrado, Virginia Polytechnic Institut and State University. Citado na(s) pág(s): 141, 142, 143, 159, 168
- PICINBONO, B. (1993). *Random Signals and Systems*. Englewood Cliffs, New Jersey: Prentice Hall International Editions. Citado na(s) pág(s): 26, 35, 40, 47, 49
- PICINBONO, B., & BARRET, M. (1990). Nouvelle Présentation de la Méthode du Maximum d'Entropie. *Traitement du Signal, Vol. 7*(No. 2), pp. 153-158. Citado na(s) pág(s): 34, 235
- POOR, H. V., & VERDÚ, S. (1997). Probability of Error in MMSE Multiuser Detection. *IEEE Transactions on Information Theory*, 47(3), 858-871. Citado na(s) pág(s): 120
- PRASAD, S., & MENICUCCI, N. C. (2004). Fisher information with respect to cumulants. *IEEE Transactions on Information Theory*, 50(4), 638-642. Citado na(s) pág(s): 240
- PROAKIS, J. G. (1995). *Digital Communications* (3rd ed.). McGraw-Hill. Citado na(s) pág(s): 8, 82, 84, 89, 96, 160, 161, 195, 206
- REGALIA, P. A. (1999). On the Equivalence Between the Godard and Shalvi-Weinstein Schemes of Blind Equalization. *Signal Processing*, 73(1-2), 185-190. Citado na(s) pág(s): 60
- RENYI, A. (1987). *A Diary on Information Theory*. Wiley. Citado na(s) pág(s): 221
- REYNOLDS, D., WANG, K., & POOR, H. V. (2002). Blind Adaptive Space-Time Multiuser Detection with Multiple Transmitter and Receiver Antennas. *IEEE Transactions on Signal Processing*, 50(6), 1261-1276. Citado na(s) pág(s): 113
- RISTANIEMI, T. (2000). *Synchronization and blind signal processing in cdma systems*. Tese de doutorado, University of Jyväskylä, Faculty of Information Technology, Department of Mathematical Information Technology, Jyväskylä, Finland. Citado na(s) pág(s): 113
- RISTANIEMI, T., & JOUTSENSALO, J. (1999). On the Performance of Blind Source Separation in CDMA Downlink. In *Proc. of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)* (p. 437-441). Aussois, France. Citado na(s)

pág(s): 113

- RISTANIEMI, T., & JOUTSENSALO, J. (2002). Advanced ICA-based Receivers for Block Fading DS-CDMA Channels. *Signal Processing*, 82, 417-431. Citado na(s) pág(s): 113
- ROWE, D. B. (2002). A Bayesian Approach to Blind Source Separation. *Journal of Interdisciplinary Mathematics*, 5(1), 49-76. Citado na(s) pág(s): 50
- SALA-ALVAREZ, J., & VÁZQUEZ-GRAU, G. (1994). Separation of Digital Communication Signals Through Joint Space-Time Decorrelation. In *Proceedings of IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP'94)* (Vol. IV, p. 145-148). Adelaide, Australia. Citado na(s) pág(s): 113
- SALA-ALVAREZ, J., & VÁZQUEZ-GRAU, G. (1997). Statistical Reference Criteria for Adaptive Signal Processing in Digital Communications. *IEEE Transactions on Signal Processing*, Vol. 45(No. 1), 14-31. Citado na(s) pág(s): 133, 137
- SAMARDZIJA, D., PAPADIAS, C. B., & VALENZUELA, R. (2002). Experimental Evaluation of Channel Deconvolution for Wireless Multiple-Transmitter/Multiple-Receiver Systems. *IEE Electronics Letters*, 38(20), 1214-1216. Citado na(s) pág(s): 113
- SANTAMARÍA, I., PANTALEÓN, C., VIELVA, L., & PRÍNCIPE, J. C. (2002). Adaptive Blind Equalization Through Quadratic PDF Matching. In *Proceedings of XI European Signal Processing Conference (EUSIPCO 2002)* (Vol. II, p. 289-292). Toulouse, France. Citado na(s) pág(s): 144, 145
- SATO, Y. (1975). A Method of Self-recovering Equalization for Multi-level Amplitude Modulation. *IEEE Transactions on Communications*, Vol. 23, 679-682. Citado na(s) pág(s): 2
- SCHRAUDOLPH, N. N. (1995). *Optimization of Entropy with Neural Networks*. Tese de doutorado, University of California, San Diego, USA. Citado na(s) pág(s): 137
- SCHWARTZ, S. C. (1967). Estimation of Probability Density by an Orthogonal Series. *The Annals of Mathematical Statistics*, 38(4), 1261-1265. Citado na(s) pág(s): 168
- SHALVI, O., & WEINSTEIN, E. (1990). New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels). *IEEE Transactions on Information Theory*, Vol. 36(No. 2), 312-321. Citado na(s) pág(s): 3, 57, 106
- SHALVI, O., & WEINSTEIN, E. (1993). Super-Exponential Methods for Blind Deconvolution. *IEEE Transactions on Information Theory*, Vol. 39(No. 2), 504-519. Citado na(s) pág(s): 60

- SHANNON, C. E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, Vol. 27, 379-423, 623-656. Citado na(s) pág(s): 3, 31
- SHYNK, J. J., & GOOCH, R. P. (1993). Convergence Properties of the Multistage CMA Adaptive Beamformer. In *Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers* (Vol. 1, p. 622-626). Pacific Groove, CA - USA. Citado na(s) pág(s): 112
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Bristol, Great Britain: Chapman and Hall. Citado na(s) pág(s): 141, 142
- SMALL, C. G., & MCLEISH, D. L. (1994). *Hilbert Space Methods in Probability & Statistical Inference*. USA: John Willey & Sons. Citado na(s) pág(s): 121
- SMARAGDIS, P. J. (1997). *Information-Theoretic Approaches to Source Separation*. Dissertação de mestrado, Massachusetts Institut of Technology, Boston, USA. Citado na(s) pág(s): 42, 53
- SUYAMA, R. (2003). *Sobre o Critério do Módulo Constante para Equalização Não-Supervisionada e suas Relações com a Teoria de Wiener*. Dissertação de mestrado, Universidade Estadual de Campinas (UNICAMP) – DECOM, Campinas - SP. Citado na(s) pág(s): 130, 139
- TANG, Z., YANG, Z., & YAO, Y. (1999). Blind Multiuser Detector Based on LMK Criterion. *IEE Electronics Letters*, 35(4), 267-268. Citado na(s) pág(s): 113
- THERRIEN, C. W. (1992). *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall International. Citado na(s) pág(s): 40, 47, 131, 169
- TONG, L., LIU, R. WEN, SOON, V. C., & HUANG, Y.-F. (1991). Indeterminacy and Identifiability of Blind Identification. *IEEE Transactions on Circuits and Systems*, 38(5), 499-509. Citado na(s) pág(s): 26
- TORKKOLA, K. (1998). Blind Signal Separation in Communications: Making Use of Known Signal Distributions. In *Proceedings of 1998 ieee dsp workshop*. Bryce Canion, UT - USA. Citado na(s) pág(s): 179
- TOUZNI, A. (1998). *Performance et Robustesse en Égalisation Aveugle Spatio-Temporelle*. Tese de doutorado, Université de Cergy-Pontoise, France. Citado na(s) pág(s): 34
- TOUZNI, A., FIJAKOW, I., LARIMORE, M. G., & TREICHLER, J. R. (2001). A Globally Convergent Approach for Blind MIMO Adaptive Deconvolution. *IEEE Transactions on Signal Processing*, 49(6), 1166-1178. Citado na(s) pág(s): 112

- TREICHLER, J. R., & AGEE, B. G. (1983). A New Approach to Multipath Correction of Constant Modulus Signals. *IEEE Transactions on Adaptive Speech and Signal Processing, Vol. ASSP-31* (No. 4), 349-472. Citado na(s) pág(s): 101
- TREICHLER, J. R., LARIMORE, M. G., & HARP, J. C. (1998). Practical Blind Demodulators for High-Order QAM Signals. *Proceedings of the IEEE, 86*(10), 1907-1926. Citado na(s) pág(s): 111
- TUGNAIT, J. K. (1995a). Blind Estimation and Equalization of Digital Communication FIR Channels Using Cumulant Matching. *IEEE Transactions on Communications, 43*(2, Part III), 1240-1245. Citado na(s) pág(s): 180
- TUGNAIT, J. K. (1995b). Parameter Identifiability of Multichannel ARMA Models of Linear Non-Gaussian Signals Via Cumulant Matching. *IEEE Transactions on Signal Processing, 43*(12), 3067-3069. Citado na(s) pág(s): 180
- TUGNAIT, J. K. (1997a). Blind Spatio-Temporal Equalization and Impulse Response Estimation for MIMO Channels Using a Godard Cost Function. *IEEE Transactions on Signal Processing, 45*(1), 268-271. Citado na(s) pág(s): 112
- TUGNAIT, J. K. (1997b). Identification and Deconvolution of Multichannel Linear Non-Gaussian Processes Using Higher-Order Statistics and Inverse Filter Criteria. *IEEE Transactions on Signal Processing, 45*(3), 658-672. Citado na(s) pág(s): 180
- TUGNAIT, J. K. (1999). Adaptive Blind Separation of Convolutional Mixtures of Independent Linear Signals. *Signal Processing, 73*, 139-152. Citado na(s) pág(s): 113
- TUGNAIT, J. K. (2001). Channel Estimation and Equalization Using Higher-Order Statistics. In G. B. GIANNAKIS, Y. HUA, P. STOICA, & L. TONG (Eds.), *Signal Processing Advances in Wireless & Mobile Communications* (Vol. 1, p. 1-39). Prentice Hall. Citado na(s) pág(s): 180
- TUGNAIT, J. K., TONG, L., & DING, Z. (2000). Single-User Channels Estimation and Equalization. *IEEE Signal Processing Magazine, 17*(3), 16-28. Citado na(s) pág(s): 180
- VERDÚ, S. (1998). *Multiuser Detection*. Cambridge Press. Citado na(s) pág(s): 9, 84, 90, 159
- WARD, J., & COMPTON, R. T., JR. (1993). High Throughput Slotted ALOHA Packet Radio Networks with Adaptive Arrays. *IEEE Transactions on Communications, 41*(3), 460-470. Citado na(s) pág(s): 207
- WEGMAN, E. J. (1972). Nonparametric Probability Density Estimation: I. A Summary of Available Methods. *Technometrics, 14*(3), 533-546. Citado na(s) pág(s): 141, 142, 168, 174

- WINTERS, J. H. (1998). Smart Antennas for Wireless Systems. *IEEE Personal Communications Magazine*(2), 23-27. Citado na(s) pág(s): 88
- YANG, H. H. (1998). On-line Blind Equalization Via On-line Blind Separation. *Signal Processing*, 68, 271-281. Citado na(s) pág(s): 113
- YANG, H. H., & AMARI, S. ICHI. (1997). Adaptive On-Line Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information. *Neural Computation*, 9, 1457-1482. Citado na(s) pág(s): 45
- YEH, C.-C., & BARRY, J. R. (2000). Adaptive Minimum Bit-Error Rate Equalization for Binary Signaling. *IEEE Transactions on Communications*, 48(7), 1226-1235. Citado na(s) pág(s): 160, 162
- YELLIN, D., & PORAT, B. (1993). Blind Identification of FIR Systems Excited by Discrete-Alphabet Inputs. *IEEE Transactions on Signal Processing*, 41(3), 1331-1339. Citado na(s) pág(s): 98
- ZABIN, S. M., & WRIGHT, G. A. (1994a). Nonparametric Density Estimation and Detection in Impulsive Interference Channels - Part I: Detectors. *IEEE Transactions on Communications*, 42(2/3/4), 1698-1711. Citado na(s) pág(s): 141
- ZABIN, S. M., & WRIGHT, G. A. (1994b). Nonparametric Density Estimation and Detection in Impulsive Interference Channels - Part I: Estimators. *IEEE Transactions on Communications*, 42(2/3/4), 1684-1697. Citado na(s) pág(s): 141, 168
- ZANATTA, D., FILHO. (2002). *Algoritmos de Processamento Espacial para Enlaces de Comunicação sem Fio*. Dissertação de mestrado, Universidade Estadual de Campinas (UNICAMP) – DECOM, Campinas - SP, Brasil. Citado na(s) pág(s): 94
- ZHAO, Q., BAO, J. Q., & TONG, L. (2001). Signal Processing Based Collision Resolution in Slotted ALOHA Wireless Ad Hoc Networks. In G. B. GIANNAKIS, Y. HUA, P. STOICA, & L. TONG (Eds.), *Signal Processing Advances in Wireless & Mobile Communications* (Vol. 1, p. 315-356). Prentice Hall. Citado na(s) pág(s): 206

Índice Remissivo

A

- Acesso múltiplo
 - por divisão em código, 83
 - por divisão em frequência, 82
 - por divisão no espaço, 83
 - por divisão no tempo, 83
- Alfabeto, 97
 - finito, 97
- Algoritmo
 - AMUSE, 26
 - deflation, 57
 - DMI
 - espaço-temporal, 100
 - espacial, 100
 - EASI, 56
 - FastICA, 58
 - FMU-CMA, 103
 - resumo, 105
 - Jutten-Hérault, 53
 - MU-CMA, 101
 - espaço-temporal, 103
 - espacial, 102
 - MU-FPA, 150
 - MUK, 106
 - resumo, 108
- Antenas adaptativas, 85
 - hipóteses, 85
 - resposta relativa, 87
 - vetor de resposta do arranjo, 87
- Aplicação de BSS
 - extração de imagens, 77

- sinais biomédicos, 71
- sinais de comunicação digital, 68
- sinais de voz, 65
- sinais eletroencefalográficos, 74
- Aplicações de BSS, 65
- Aprendizado
 - anti-Hebbiano, 60
- Aproximações
 - de funções de densidade de probabilidade, 168

C

- Caracterização
 - fontes de sistema multiusuário, 96
 - resposta do canal, 87
- Caraterísticas
 - algoritmos de separação de fontes, 64
- Condições
 - recuperação de sinais, 106, 148, 238
- Constelação, 96
- Contraste
 - discriminante, 38
- Critério
 - baseados em estimação da fdp, 144
 - de Donoho, 60
 - InfoMax, 44
 - máxima a posteriori, 50
 - MaxEnt, 44
 - minimização da probabilidade de erro, 159

MU-FPC, 149
 multiusuário
 família baseada na estimação da fdp,
 148
 Cumulante, 170

D

Decomposição
 de Pitágoras, 33
 Descorrelação
 estatística, 29
 explícita, 103
 Dispersão
 da constelação, 101
 da distribuição, 36
 intraclasse, 120
 Divergência
 assimétrica, 32
 de Kullback-Leibler, 32
 propriedades, 32

E

Entropia, 31
 condicional, 31
 cruzada, 32
 diferencial, *ver* Entropia, 32
 Equivalência
 critérios cegos e FPC, 131
 critérios supervisionados e FPC, 131
 Equivariância, 55
 Erro
 do módulo constante, 152
 Estimação
 fdp, 140
 método de Parzen, 142
 tipos, 140
 Estrutura
 processamento espaço-temporal, 93
 processamento espacial, 90
 Expansão
 de Edgeworth, 51, 174

de Gram-Charlier, 169, 172
 para misturas gaussianas, 177

F

Fator
 de regularização, 101
 de esquecimento, 100
 de regularização
 adaptativo, 104
 fdp, *ver* Função densidade de probabilidade
 Fitting pdf Algorithm, 123
 Fitting pdf Criterion, 122
 Fonte discreta, 96
 FPA, *ver* Fitting pdf Algorithm
 FPC, *ver* Fitting pdf Criterion
 Função
 de referência, 32
 logística sigmóide, 53
 Função densidade de probabilidade, 31
 condicional, 31
 conjunta, 31
 gaussiana
 generalizada, 36
 marginal, 33
 Funções
 de contraste, 37
 condições, 37
 escore, 53

G

Gradiente
 natural, 55
 relativo, 55, 56

I

ICA, *ver* Independent Component Analysis
 Independência estatística, 28
 Informação mútua, 31
 Interferência
 residual, 211

K

Kurtosis, 35
de funções de densidade de probabilidade, 35

M

Múltiplo acesso, 82
Matriz
de convolução do canal espaço-temporal, 94
de convolução do canal mono-usuário, 119
de mistura, *ver* de mistura convolutiva
de mistura convolutiva, 21
de separação, 22
pseudo-inversa, 109
Modelo
sub-determinado, 22

N

Near-far, 101
Negentropia, 34
aproximações da, 41

O

Ortogonalização
de Gram-Schmidt, 108

P

PCA, *ver* Principal Component Analysis
Polinômios
de Hermite, 172
propriedades, 173
Principal Component Analysis, 24
Processamento
espaço-temporal, 91
espacial, 88
multiusuário, 81
Propriedade
convergência

FPC, 123

R

Rede neural artificial
Jutten-Hérault, 54
Relação
sinal-ruído, 65

S

Separação de fontes, 15, 43
áreas de aplicações, 18
abordagem Bayesiana, 46
critério
universal, 51
esquema de processamento, 21
estatísticas necessárias, 23
hipóteses usuais, 21
histórico, 16
indeterminação
escalonamento, 22
ordem, 22
máxima verossimilhança, 46
Solução
ótima, 99
processamento espaço-temporal, 100
processamento espacial, 99
de Wiener, 99

T

Teorema
Benveniste-Goursat-Rouget, 140, 237
Transformação
de Mahalanobis, 25

Índice de Autores

Symbols

Özgül, B., 206, 254

A

Adalı, T., 147, 243
Agee, B. G., 101, 258
Almeida, A. L. F. de, 247
Almeida, L. B., 187, 243
Amara, R., 122, 243
Amari, S. ichi, 7, 22, 25, 36, 45, 54, 61, 63, 65, 66, 71, 73, 75, 76, 80, 107, 246, 259
Amblard, P.-O., 228, 229, 243, 251
Ans, B., 249
Attux, R. R. F., 126, 243

B

Baggenstoss, P. M., 250
Bao, J. Q., 259
Barbarossa, S., 123, 243
Barret, M., 34, 235, 255
Barry, J. R., 160, 162, 259
Bell, A. J., 40, 44, 45, 243, 252
Benedetto, S., 160, 243
Benveniste, A., 40, 45, 140, 144, 244
Biglieri, E., 243
Bishop, C. M., 153, 244
Bolasco, S., 24, 246
Bouvier, G., 254
Brandão, A. L., 245, 246

Brauer, W., 175, 246
Brossier, J.-M., 135, 211, 244
Bugallo, M. F., 113, 244

C

Campos, M. L. R., 249
Cao, X.-R., 27, 244
Cardoso, J.-F., 17, 38, 48, 49, 53, 55, 56, 63, 69, 159, 244
Carleial, A. B., 206, 244
Castedo, L., 112, 244, 245, 253
Castellani, V., 243
Cavalcante, C. C., 33, 44, 60, 118–120, 124, 125, 135, 137, 143, 150, 154, 158, 207, 245, 247
Cavalcanti, F. R. P., 8, 9, 83, 86, 87, 89, 91, 92, 94, 100, 102–105, 112, 113, 152, 196, 198, 206, 207, 245–247
Chambers, J. A., 112, 251
Chen, C.-H., 246
Chen, C.-Y., 246
Chen, S., 160, 188, 246
Chi, C.-Y., 113, 246
Cichocki, A., 7, 22, 25, 36, 54, 61, 63, 65, 66, 71, 73, 75, 76, 80, 107, 246
Comon, P., 17, 27, 34, 37–39, 52, 98, 175, 188, 246, 249, 251
Compton, R. T., Jr., 207, 258
Constatinides, A. G., 251
Coppi, R., 24, 246

Cover, T. M., 232, 246

Cowan, J. D., 252

D

Dapena, A., 245

Deco, G., 32, 175, 246

Delfosse, N., 54, 57, 58, 247

Deneire, L., 107, 109, 247

Devroye, L., 141, 144, 168, 221, 247

Ding, Z., 133, 247, 258

Diniz, P. S. R., 99, 100, 247, 249

Donoho, D., 40, 60, 222, 247

Duda, R., 138, 247

E

Erdogmus, D., 251

Escudero, C. J., 245

Eweda, E., 135, 252

F

Faÿ, G., 174, 247

Fijakow, I., 257

Freitas, W. d. C., Jr., 98, 245, 247

G

Gaeta, M., 243

Gamboa, F., 98, 248

Gassiat, E., 98, 248

Ghez, S., 207, 248

Giannakis, G. B., 9, 248

Gibson, J. D., 31, 248

Girolami, M., 252

Godara, L. C., 82, 87, 248

Godard, D. N., 101, 248

Golub, G. H., 24, 107, 203, 242, 248

Gooch, R. P., 111, 112, 248, 257

Gorokhov, A., 110, 248, 249

Goursat, M., 244

Grellier, O., 98, 249

Grier, D. G., 252

H

Hagenauer, J., 136, 249

Hanzo, L., 246

Harp, J. C., 258

Hart, P., 138, 247

Haykin, S., 17, 19, 22, 24, 31–34, 36, 42, 44, 45, 52, 56, 60, 62, 63, 75, 79, 80, 94, 95, 99, 100, 103, 109, 119, 120, 126, 133, 135, 138, 168–172, 174, 242, 249

Hellman, M. E., 206, 244

Hild, K. E., II, 251

Hjørungnes, A., 160, 249

Hoehner, P., 136, 249

Hua, Y., 248

Huang, H., 112, 254

Huang, Y.-F., 257

Hulle, M. M. V., 76, 80, 253

Hyvärinen, A., 16, 17, 22, 24, 26, 27, 30, 34–36, 41, 42, 45, 52, 54–56, 58, 59, 62, 80, 168, 175, 249, 250

Hérault, J., 4, 16, 54, 61, 249, 250

J

Joutsensalo, J., 113, 255, 256

Jutten, C., 16, 54, 61, 249, 250, 254

K

Karhunen, J., 80, 250

Kay, S. M., 175, 250

Knuth, K. H., 50, 250

Kofidis, E., 18, 22–24, 26, 27, 32–34, 37, 38, 42, 48, 49, 51–53, 55, 58–63, 159, 250

Kullback, S., 32, 250

Kühn, V., 136, 250

L

Laar, J. van de, 180, 250

Lacoume, J.-L., 24, 26, 39, 51, 52, 170, 171, 174, 229, 243, 251

Laheld, B. H., 55, 56, 69, 244
Lamare, R. C. de, 160, 251
Lambotharan, S., 112, 251
Larimore, M. G., 257, 258
Laster, J. D., 142, 159, 160, 171–173, 175, 176, 251
Lathauwer, L. de, 6, 17, 251
Leary, J., 112, 251
Lee, E. A., 96, 251
Lee, T.-W., 52, 167, 175, 252
Leibler, R. A., 32, 250
Li, T., 112, 252
Li, T.-H., 98, 252
Li, Y. G., 112, 133, 247, 252
Liberti, J. C., Jr., 86, 87, 111, 252
Lii, K.-S., 3, 148, 252
Lin, J. K., 175, 252
Lindsey, J. K., 51, 141, 176, 252
Liu, K. J. R., 112, 252
Liu, R. wen, 27, 244, 257
Liu, X., 147, 243
Loan, C. F. V., 24, 107, 203, 242, 248
Loubaton, P., 54, 57, 58, 110, 247–249
Lundell, J., 111, 248
Lázaro, M., 146, 147, 251

M

Macchi, O., 98, 135, 252
Mbarek, K., 98, 252
Mboup, M., 60, 222, 252
McCullagh, P., 51, 253
McLaughlin, S., 246
McLeish, D. L., 121, 257
Menicucci, N. C., 240, 255
Messerschmitt, D. G., 96, 251
Mo, S., 188, 253
Mohammad-Djafari, A., 49, 253
Montalvão, J. R., Filho, 33, 60, 120, 253
Moreau, E., 98, 252
Mota, J. C. M., 180, 245, 253
Moulines, E., 174, 247

Mulgrew, B., 246
Mutihac, R., 76, 80, 253
Míguez, J., 112, 244, 253

N

Nadal, J.-P., 180, 253
Neves, A. O., 113, 253
Nikias, C. L., 230, 231, 253
Nuttal, A. H., 250

O

Obradovic, D., 32, 246
Oja, E., 17, 22, 26, 27, 30, 34–36, 42, 58, 80, 250
Orange, M. D., 206, 253

P

Pantaleón, C., 251, 256
Papadias, C. B., 35, 37, 58, 94, 98, 101, 103, 106, 107, 109, 110, 112, 149, 153, 156, 193, 254–256
Papoulis, A., 24, 26, 29, 35, 40, 46, 47, 120, 132, 169, 176, 228, 254
Paraschiv-Ionescu, A., 202, 203, 254
Parga, N., 180, 253
Parzen, E., 142, 143, 254
Paulraj, A. J., 94, 98, 101, 106, 149, 254, 255
Petropulu, A. P., 230, 231, 253
Petrus, P., 85–89, 91, 111, 255
Phillips, K. A., 141–143, 159, 168, 255
Picinbono, B., 26, 34, 35, 40, 47, 49, 235, 255
Poor, H. V., 120, 255
Porat, B., 98, 259
Prasad, S., 240, 255
Proakis, J. G., 8, 82, 84, 89, 96, 160, 195, 206, 255
Príncipe, J. C., 251, 256

R

Rappaport, T. S., 86, 87, 111, 252
Regalia, P. A., 60, 222, 252, 255
Renyi, A., 221, 255
Reynolds, D., 113, 255
Ristaniemi, T., 113, 255, 256
Romano, J. M. T., 89, 100, 102–104, 206,
207, 245, 246
Rosenblatt, M., 3, 148, 252
Rowe, D. B., 50, 256
Ruget, G., 244

S

Sala-Alvarez, J., 113, 132, 133, 137, 256
Samardzija, D., 113, 256
Samingan, A. K., 246
Sampaio-Neto, R., 160, 251
Santamaría, I., 144, 145, 251, 256
Sato, Y., 2, 256
Scaglione, A., 123, 243
Schraudolph, N. N., 137, 256
Schwartz, S., 248
Schwartz, S. C., 168, 256
Sejnowski, T. J., 40, 44, 45, 243, 252
Shafai, B., 188, 253
Shalvi, O., 3, 57, 60, 106, 256
Shannon, C. E., 3, 31, 257
Shynk, J. J., 112, 257
Sidiropoulos, N. D., 112, 252
Silverman, B. W., 141, 142, 257
Slock, D. T. M., 109, 247
Small, C. G., 121, 257
Smaragdis, P. J., 42, 53, 257
Soon, V. C., 257
Soulier, P., 174, 247
Stoica, P., 248
Suyama, R., 130, 139, 257
Sönmez, M. K., 147, 243

T

Tang, Z., 113, 257

Therrien, C. W., 40, 47, 131, 169, 257
Thomas, J. A., 232, 246
Tong, L., 26, 248, 257–259
Torkkola, K., 179, 257
Touzni, A., 34, 112, 257
Treichler, J. R., 101, 111, 257, 258
Tugnait, J. K., 112, 113, 180, 258

V

Valenzuela, R., 256
Verdú, S., 9, 84, 90, 120, 159, 248, 255, 258
Vielva, L., 256
Vázquez-Grau, G., 113, 133, 137, 256

W

Wang, K., 255
Ward, J., 207, 258
Wegman, E. J., 141, 142, 168, 174, 258
Weinstein, E., 3, 57, 60, 106, 256
Winters, J. H., 88, 259
Wright, G. A., 141, 168, 259
Wu, Y., 246

Y

Yang, H. H., 45, 113, 259
Yang, Z., 257
Yao, Y., 257
Yeh, C.-C., 160, 162, 259
Yellin, D., 98, 259

Z

Zabin, S. M., 141, 168, 259
Zanatta, D., Filho, 94, 247, 259
Zhao, Q., 205, 206, 259