



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
CURSO DE DE GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

SEPARAÇÃO CEGA DE FONTES: MÉTODOS E
APLICAÇÕES

LUCAS NOGUEIRA RIBEIRO

Fortaleza, Ceará

1 de dezembro de 2014

LUCAS NOGUEIRA RIBEIRO

SEPARAÇÃO CEGA DE FONTES: MÉTODOS E APLICAÇÕES

Projeto de Final de Curso apresentado à Coordenação do Curso de Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará como requisito parcial para a obtenção do diploma de Engenheiro de Teleinformática.

Orientador: Prof. Dr. João César Moura Mota

UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
CURSO DE DE GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

Fortaleza, Ceará

1 de dezembro de 2014

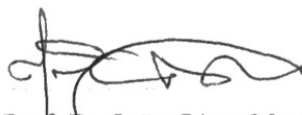
LUCAS NOGUEIRA RIBEIRO

SEPARAÇÃO CEGA DE FONTES: MÉTODOS E APLICAÇÕES

Lucas Nogueira Ribeiro

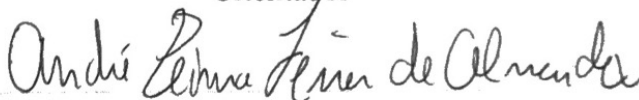
Lucas Nogueira Ribeiro

Este Trabalho foi julgado adequado para a obtenção do título de Engenheiro de Teleinformática e aprovado, em sua forma final, pela Coordenação do Curso de de Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará em 1 de dezembro de 2014, com nota 10, na cidade de Fortaleza, Ceará, pela banca examinadora assim constituída:



Prof. Dr. João César Moura Mota

Orientador



Prof. Dr. André Lima Férrer de Almeida



Prof. Dr. Antônio Macílio Pereira de Lucena

Fortaleza, Ceará

1 de dezembro de 2014

Dedico este trabalho à minha família.

*Ideias são como coelhos.
Você consegue um par, aprende como lidar com elas e logo tem uma dúzia.
John Steinbeck*

AGRADECIMENTOS

Agradeço primeiramente a Deus, meu Criador, por ter me sustentado durante esses 5 anos de graduação,

À minha família, principalmente aos meus pais, João e Ana Maria, pelo apoio constante e por sempre acreditarem em mim,

À minha namorada, Priscilla, pelo seu amor, cuidado, compreensão e companheirismo,

Ao professor João César por ter me orientado durante toda a graduação, me ensinando a arte de processamento de sinais,

À CAPES por ter financiado meu intercâmbio na França por meio do programa BRAFITEC, onde tive meu primeiro contato com o problema de separação de fontes,

Aos meus amigos de graduação pelos momentos inesquecíveis durante a graduação, e à Universidade Federal do Ceará por ser uma instituição pública de excelência, gerando e difundindo conhecimento no nosso estado do Ceará.

RESUMO

A separação cega de fontes é uma técnica de processamento estatístico de sinais que visa recuperar um conjunto de sinais fontes misturados e/ou contaminados a partir de um conjunto de sinais observados sem conhecimento dos integrantes do sistema de mistura. É possível modelar vários problemas de engenharia como um problema de separação cega de fontes. Neste trabalho, alguns métodos de separação cega de fonte como JADE, FastICA e RobustICA foram estudados e comparados por meio de simulações computacionais. Finalmente, utilizou-se o algoritmo RobustICA para extrair o eletrocardiograma fetal do eletrocardiograma materno a fim de ilustrar uma aplicação prática.

Palavras-chaves: ICA, tensores, separação de fontes.

ABSTRACT

Blind source separation is a statistical signal processing technique which seeks recovering a set of mixed source signals from a set of observed signals without knowing the mixing system. Many engineering problems can be modeled as a blind source source separation problem. In this work, some blind source separation methods such as JADE, FastICA and RobustICA were studied and compared by means of computer simulation. Finally, the RobustICA algorithm was used to extract fetal electrocardiogram from the maternal electrocardiogram recordings in order to illustrate a practical application.

Keywords: ICA, tensors, source separation.

SUMÁRIO

Lista de Figuras	10
Lista de Algoritmos	11
Lista de Acrônimos	12
1 Introdução	13
1.1 Formalização do problema	14
1.2 Estratégias de separação de fontes	16
1.3 Aplicações	17
2 Ferramentas matemáticas	20
2.1 Introdução	20
2.2 Tensores	20
2.3 Momentos e cumulantes	22
2.3.1 Propriedades	26
2.3.2 Estimacão de momentos e cumulantes	28
2.4 Teoria da Informacão	29
2.5 Otimizacão numérica	30
3 Métodos e algoritmos	33
3.1 Introducão	33
3.2 Contrastes	33
3.2.1 Extração de fontes	34
3.2.2 Separacão de fontes	35
3.3 Métodos algébricos	40
3.3.1 Maximizacão do traço	41
3.3.2 Diagonalizacão conjunta de matrizes	42
3.4 Algoritmos iterativos	44
3.4.1 FastICA	45
3.4.2 RobustICA	47
3.5 Novos métodos e perspectivas	48
3.5.1 Decomposicão não-negativa de matrizes	49
3.5.2 Block Term Decomposition	51
3.5.3 ICA via matriz de “carrelacão”	53

3.6	Conclusão	55
4	Resultados e discussão	57
4.1	Introdução	57
4.2	Experimentos com sinais sintéticos	57
4.3	Experimentos práticos	65
5	Conclusão	69
	Referências Bibliográficas	70

LISTA DE FIGURAS

1.1	Modelos de composição de fontes.	15
1.2	Ilustração do <i>cocktail party problem</i> , no qual várias fontes distintas são captadas pelos receptores.	18
2.1	Decomposição PARAFAC para um tensor de terceira ordem com dimensões $(I_1 \times I_2 \times I_3)$	22
2.2	Densidades de probabilidade de Laplace, Wigner e normal.	25
3.1	Ilustração da composição de N fontes e extração de uma fonte através do filtro \mathbf{f}	34
3.2	Ilustração da composição de N fontes e separação delas por meio de um filtro \mathbf{B}	36
3.3	Efeito do branqueamento nas observações.	39
3.4	Block Term Decomposition de um tensor \mathcal{T}	52
4.1	Histograma das fontes sintéticas.	59
4.2	Experimento com sinais sintéticos utilizando o algoritmo JADE.	60
4.3	Curvas de SNR vs. ISR.	61
4.4	Curvas SNR vs MSE.	62
4.5	Curvas de número de amostras vs. ISR.	62
4.6	Curvas de número de amostras vs. MSE.	63
4.7	Curvas de número de fontes vs. ISR.	63
4.8	Curvas de número de fontes vs. MSE.	64
4.9	Eletrocardiograma ilustrando as onda P e T juntamente com complexo QRS. Sinal ECG obtido do MIT-BIH <i>normal sinus rhythm database</i> , paciente 16272.	65
4.10	Eletrocardiogramas abdominais da mãe.	67
4.11	Componentes independentes dos ECGs maternos, fornecendo o ECG fetal.	68

LISTA DE ALGORITMOS

1	Algoritmo de busca linear.	32
2	Algoritmo para SCF utilizando métodos algébricos.	44
3	Algoritmo FastICA.	47
4	Algoritmo RobustICA.	48
5	Algoritmo NMF-ALS.	51
6	Algoritmo BSS-BTD.	53

LISTA DE ACRÔNIMOS

ALS *Alternating Least-Squares.*

BTD *Block Term Decomposition.*

ECG Eletrocardiograma.

EEG Eletroencefalograma.

EOS Estatísticas de Ordem Superior.

MSE *Mean Squared Error.*

EVD *Eigenvalue Decomposition.*

ICA *Independent Component Analysis.*

IM Informação Mútua.

MIMO *Multiple-Input Multiple-Output.*

MISO *Multiple-Input Single-Output.*

ML *Maximum Likelihood.*

NMF *Non-negative Matrix Factorization.*

PARAFAC *Parallel Factors.*

PCA *Principal Component Analysis.*

SCF Separação Cega de Fontes.

SVD *Singular Value Decomposition.*

1. INTRODUÇÃO

A separação cega de fontes (SCF) é uma técnica de processamento estatístico de sinais que visa recuperar um conjunto de sinais fontes misturados e/ou contaminados a partir de um conjunto de sinais observados sem conhecimento dos integrantes do sistema de mistura. Em vários campos da ciência e da engenharia, é comum observar composição de fontes de diversas naturezas. Em um ambiente com vários interlocutores falando ao mesmo tempo, grava-se através de alguns microfones a voz misturada de todos eles. Através de técnicas de separação cega de fontes, consegue-se separar a fala de cada interlocutor, levando à identificação e à inteligibilidade de cada fala. Na engenharia biomédica, pode-se modelar os sinais das atividades atrial e ventricular como fontes distintas e separá-las aplicando técnicas de separação de fontes em eletrocardiogramas (ECG), favorecendo o diagnóstico médico de doenças cardíacas. Sistemas de comunicação sem fio multiusuário podem ser modelados como um problema de separação de fontes, no qual se deseja separar os sinais interferentes do sinal desejado, permitindo a obtenção de uma melhor qualidade e segurança do uso de informações transmitidas, como por exemplo na telefonia celular. A maioria dos métodos de separação de fontes está concentrada em modelos lineares e observações homogêneas, ou seja, utilização de sensores semelhantes. No entanto, há uma necessidade do desenvolvimento de novos resultados teóricos em modelos não-lineares e modelos convolucionais não-lineares buscando atender novas situações práticas e a melhoria dos serviços existentes.

De forma geral, propõe-se através deste trabalho explorar as bases teóricas da SCF desde as primeiras abordagens algébricas nascidas nos anos 80 até novos métodos recentemente propostos. Especificamente, este trabalho visa apresentar um estudo sobre as ferramentas matemáticas utilizadas nas técnicas de separação de fontes, tais como tensores e estatísticas de ordem superior. Apresentar as principais estratégias de separação de fontes assim como os principais algoritmos, assim como compará-los através de simulações computacionais, também constituem alguns objetivos específicos deste trabalho. Como será detalhado no Capítulo 4, realizaram-se experimentos de separação de fontes com sinais sintéticos utilizando alguns algoritmos estudados, possibilitando uma análise de desempenho entre estes algoritmos. Também foi realizado um experimento com sinais biomédicos a fim de ilustrar uma aplicação prática desta técnica de processamento de sinais.

A fim de cumprir essas propostas, esta monografia foi dividida da seguinte forma:

- No restante do Capítulo 1, serão introduzidos alguns conceitos preliminares em relação a SCF.

- As ferramentas matemáticas necessárias para o desenvolvimento da monografia serão apresentadas no Capítulo 2.
- Métodos e algoritmos serão investigados no Capítulo 3.
- Através de simulações computacionais, a performance dos métodos de SCF serão discutidas e comparadas no Capítulo 4.
- As conclusões e os destaques finais do que foi realizado serão apresentados no Capítulo 5.

1.1 Formalização do problema

Considera-se um conjunto de N sinais não-observados estacionários com média nula, chamados de fontes, representados por $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$. O problema da separação cega de fontes consiste em recuperar essas fontes a partir de um conjunto de M sinais observados, representados por $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$. As fontes e as observações relacionam-se da seguinte forma:

$$\mathbf{x}(t) = \mathcal{A} \{\mathbf{s}(t)\}, \quad (1.1)$$

no qual $\mathcal{A} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ é uma aplicação desconhecida responsável por “misturar” as fontes.

Os primeiros trabalhos envolvendo separação cega de fontes consideraram o caso em que as observações consistem em combinações lineares das fontes e o sistema de composição não possui memória (COMON; JUTTEN, 2010):

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + \dots + a_{1N}s_N(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + \dots + a_{2N}s_N(t) \\ \vdots \\ x_M(t) = a_{M1}s_1(t) + a_{M2}s_2(t) + \dots + a_{MN}s_N(t) \end{cases},$$

no qual a_{mn} para $m = 1, \dots, M$ e $n = 1, \dots, N$ são os coeficientes de composição que relacionam o m -ésimo sinal de observação à n -ésima fonte. Este sistema de equações pode ser representado utilizando a seguinte notação matricial:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (1.2)$$

no qual $\mathbf{A} \in \mathbb{R}^{M \times N}$ representa a matriz de composição desconhecida. Este modelo de composição de sinais é conhecido como **modelo instantâneo linear**, pois, nesse caso, a aplicação \mathcal{A} é um operador linear \mathbf{A} sem memória, como ilustrado na Figura 1.1(a).

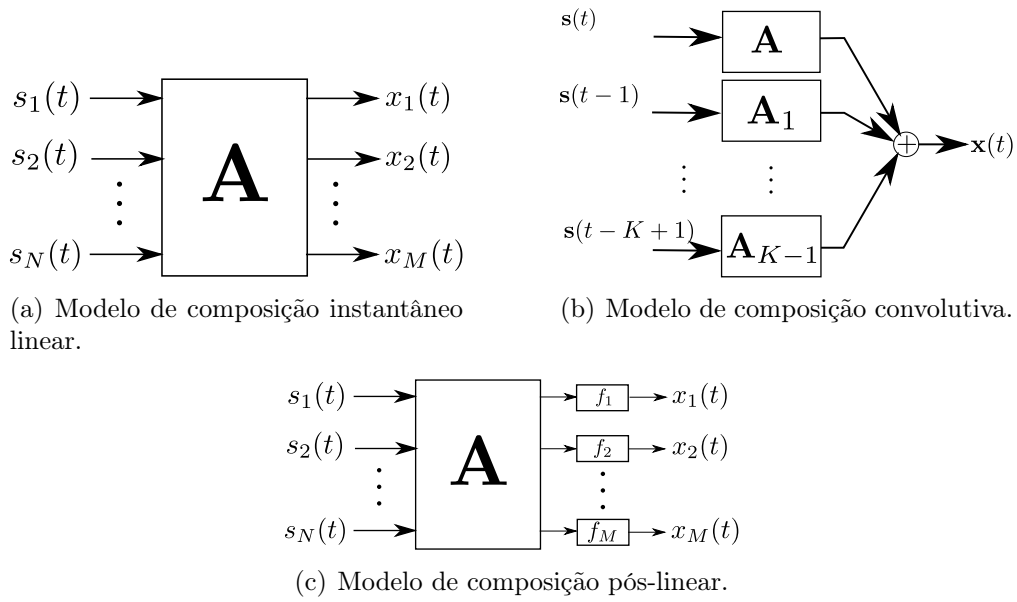


Figura 1.1: Modelos de composição de fontes.

Uma extensão natural consiste em considerar um modelo de composição dinâmico, ilustrado na Figura 1.1(b), também conhecido como **modelo convolutivo**:

$$x_m(t) = \sum_{n=1}^N \sum_{k=0}^{K-1} a_{mn}(k) s_n(t-k), \quad m = 1, \dots, M. \quad (1.3)$$

Pode-se representar a equação (1.3) de forma matricial:

$$\mathbf{x}(t) = \sum_{k=0}^{K-1} \mathbf{A}_k \mathbf{s}(t-k), \quad (1.4)$$

no qual $\mathbf{A}_k = [a_{mn}(k)]_{mn}$ é a matriz de composição convolutiva com $m = 1, \dots, M$ e $n = 1, \dots, N$. Esse modelo é mais adaptado para casos em que ocorre o fenômeno de reverberação, frequente em problemas de acústica (COMON; JUTTEN, 2010). Nota-se que o modelo linear instantâneo é um caso especial do modelo convolutivo quando $K = 1$.

O artigo de (TALEB; JUTTEN, 1999) foi um dos primeiros trabalhos a investigar o **modelo não-linear** de composição de fontes, no qual a aplicação \mathcal{A} na equação (1.1) é não-linear. Neste artigo, os autores consideraram um modelo não-linear particular, conhecido como *post-linear mixture*, formado por um módulo linear, seguido por um módulo não-linear, como ilustrado na Figura 1.1(c). A relação entre as fontes e as observações é dada pela equação

$$x_m(t) = f_m \left(\sum_{n=1}^N a_{mn} s_n(t) \right), \quad m = 1, \dots, M, \quad (1.5)$$

no qual $f_m(\cdot)$ é uma função inversível, derivável e não-linear.

Neste trabalho, será investigado principalmente o sistema linear instantâneo, devido à sua vasta gama de aplicações e simplicidade do modelo e notação. Ruídos e interferências não serão considerados nas equações dos modelos de composição de fontes. Apesar disso, na prática a presença deles é inevitável, podendo reduzir o desempenho dos métodos de SCF.

1.2 Estratégias de separação de fontes

A fim de estimar o vetor $\mathbf{s}(t)$, supõe-se ter conhecimento de \mathcal{A} . Se a aplicação \mathcal{A} é inversível, pode-se estimar as fontes simplesmente calculando $\mathcal{A}^{-1}(\mathbf{x}(t))$. Como consequência da inversibilidade de \mathcal{A} , afirma-se que $M \geq N$. Caso o problema seja sub-determinado, ou seja, $M < N$, então seria impossível inverter \mathcal{A} . Na prática, não se tem conhecimento da aplicação \mathcal{A} . Quando se deseja recuperar as fontes $\mathbf{s}(t)$ conhecendo apenas $\mathbf{x}(t)$, trata-se de um **problema inverso** (STARCK; FADILI, 2009). Se este problema não possui uma única solução estável, ele é **mal-posto**. É fácil ver que a SCF é um problema inverso mal-posto, pois existem infinitas soluções para a equação (1.1).

Uma forma de solucionar um problema inverso mal-posto é a **regularização**, que consiste em usar informações *a priori* sobre as fontes para reduzir o espaço de soluções possíveis. Uma condição *a priori* comumente usada em problemas de SCF é a independência estatística entre as fontes (COMON, 1994). Dessa forma, dentre as possíveis soluções, busca-se uma aplicação \mathcal{H} que aproxima \mathcal{A}^{-1} , de forma que a independência entre os elementos de $\hat{\mathbf{s}}(t) = \mathcal{H}\mathbf{x}(t)$ seja máxima. Um método bastante utilizado em separação fontes, conhecido como *Independent Component Analysis* (ICA), é baseado nesse princípio de regularização.

A independência estatística entre as fontes pode ser uma condição muito restritiva em alguns problemas. Porém, nesses casos, é possível que elas sejam ao menos decorrelacionadas (PEDERSEN *et al.*, 2007). Todavia, a decorrelação entre as fontes por si própria não é suficiente para a separação das fontes em muitos casos. Juntamente com a decorrelação, alguns trabalhos supõem: sistema de composição com fase mínima (GERVEN; COMPERNOLLE, 1995), não-estacionaridade das fontes (WEINSTEIN; FEDER; OPPENHEIM, 1993), ciclo-estacionaridade (SHAMSUNDER; GIANNAKIS, 1997), fontes coloridas (GOROKHOV; LOUBATON, 1997), dentre outras condições.

Novas estratégias de separação de fontes têm sido estudadas nos últimos anos. Em (KIM; LEE; LEE, 2006), os autores propõem uma extensão do modelo (1.1), que consiste em considerar vetores aleatórios, diferentemente do modelo clássico composto de variáveis aleatórias. Dessa forma, cada fonte é representada por um vetor, dando nome ao método *Independent Vector Analysis* (IVA). Nesse modelo, os elementos de fontes distintas são independentes, enquanto os elementos da mesma fonte são altamente dependentes. Por-

tanto, a SCF é realizada maximizando a independência entre as fontes vetoriais. Outras novas estratégias utilizam **tensores**, uma estrutura matemática multidimensional que generaliza as matrizes. No artigo (CICHOCKI *et al.*, 2014), o autor comenta algumas decomposições tensoriais e compara a performance delas no problema de SCF. A decomposição *Block Term Decomposition* (BTD) (LATHAUWER, 2008) é bem adaptada para separar sinais que podem ser expressos como exponenciais oscilatórias (LATHAUWER, 2011), sendo interessante em aplicações de processamento de sinais biomédicos. Uma das grandes vantagens das ferramentas tensoriais advém do fato de suas decomposições serem essencialmente únicas ¹ com condições não tão restritivas, se comparadas às decomposições matriciais. Na Seção 3.5, a decomposição BTD e outras técnicas de SCF serão apresentadas mais detalhadamente.

1.3 Aplicações

A SCF por ser um problema geral possui vários campos de aplicação, tais como engenharia biomédica, sistemas de telecomunicações sem fio, processamento de imagens, acústica, dentre outros. Em razão dessa pluridisciplinaridade, a SCF tem recebido bastante atenção de pesquisadores, que desejam desenvolver algoritmos que ofereçam uma melhor qualidade na estimação das fontes. Alguns exemplos de aplicação da SCF serão discutidos nesta seção.

Uma das primeiras aplicações da SCF data de 1981 no campo da neurociência, especificamente no problema de decodificação motora de vertebrados (ROLL, 1981; COMON; JUTTEN, 2010). Essa decodificação é feita por meio da análise da contração de fibras musculares, coordenada pelo cérebro por meio de duas terminações nervosas: a primária e a secundária. É possível observar experimentalmente alguns padrões nos sinais bioelétricos $f_1(t)$ e $f_2(t)$ que transitam na terminação primária e na secundária, respectivamente. Observou-se que esses padrões variam de acordo com a posição angular $p(t)$ do alongamento muscular e a velocidade angular $v(t) = \frac{dp(t)}{dt}$. Visando determinar uma relação entre os sinais das terminações nervosas e os parâmetros $p(t)$ e $v(t)$, os autores propuseram o seguinte modelo:

$$\begin{cases} f_1(t) = a_{11}v(t) + a_{12}p(t) \\ f_2(t) = a_{21}v(t) + a_{22}p(t) \end{cases}$$

no qual a_{ij} , $v(t)$ e $p(t)$ com $i, j = 1, 2$ são desconhecidos. A solução deste problema consiste em considerar $v(t)$ e $p(t)$ independentes, e, assim, calcular uma transformação que estime $v(t)$ e $p(t)$ a partir de $f_1(t)$ e $f_2(t)$. Quando se trabalha com sinais de origem biológica, é importante verificar se as suposições feitas são biologicamente válidas. Apesar

¹Quando uma decomposição apresenta indeterminações como escala ou permutação, diz-se que a decomposição é essencialmente única.

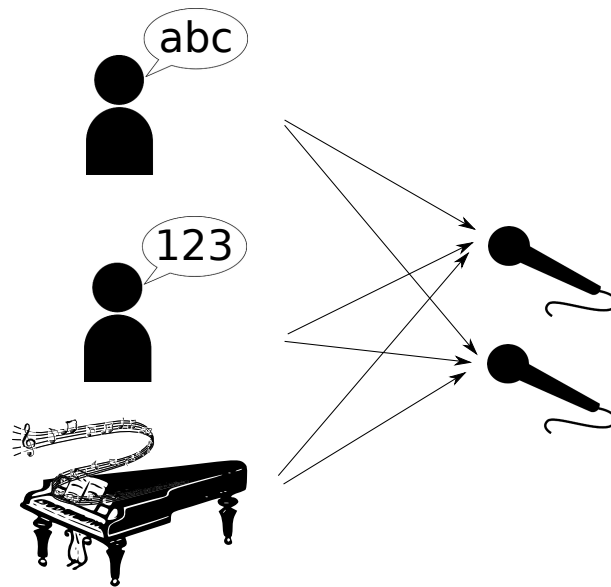


Figura 1.2: Ilustração do *cocktail party problem*, no qual várias fontes distintas são captadas pelos receptores.

da relação temporal entre $v(t)$ e $p(t)$, os autores apresentaram argumentos que indicam a independência mútua entre essas fontes, permitindo a separação delas.

Apesar das primeiras ideias de separação de fontes terem surgido na década de 80, o problema prático que melhor ilustra a ideia da SCF foi discutido na década de 50 por Colin Cherry (CHERRY, 1953). Ao estudar reconhecimento de fala utilizando dois microfones, Cherry considerou o cenário em que várias pessoas falam ao mesmo tempo, chamando-o de *cocktail party problem*, como ilustrado na Figura 1.2. Do ponto de vista de processamento de sinais, esse é um problema difícil, pois devido à grande quantidade de classes de sinais acústicos, é difícil propor um modelo de separação de áudio geral. As duas classes de sinais acústicos mais analisadas são os sons musicais e a fala. Em (SMITH; LUKASIAK; BURNETT, 2006), concluiu-se que as amostras de uma fala tornam-se mais independentes à medida que a quantidade de amostras aumenta. Por outro lado, em (ABRARD; DEVILLE, 2003), observou-se que sinais musicais são altamente correlacionados considerando poucas amostras. A validade da hipótese de independência para sinais acústicos foi discutida em (PUIGT; VINCENT; DEVILLE, 2009) utilizando algumas medidas de independência. Concluiu-se experimentalmente, em sintonia com (SMITH; LUKASIAK; BURNETT, 2006; ABRARD; DEVILLE, 2003), que para sinais acústicos com pequena duração, essa hipótese não é válida no domínio do tempo. Todavia, ela se verifica para sinais de longa duração. Para o caso em que a independência não é válida, pode-se usar métodos como *sparse component analysis* (SCA) (BOFILL; ZIBULEVSKY, 2001) e *non-negative matrix factorization* (NMF) (WANG; PLUMBLEY, 2005).

Uma área que cresceu bastante graças ao avanço das técnicas de processamento de sinais é a engenharia biomédica. Em 1924, o médico alemão Hans Berger conseguiu medir a atividade elétrica cerebral posicionando eletrodos na cabeça dos seus pacientes.

A representação temporal dessa atividade foi chamada de eletroencefalograma (EEG). Hoje o EEG é bastante utilizado pois oferece uma visão geral da atividade cerebral a um baixo custo operacional. Além disso, algoritmos de processamento de sinais permitem a interpretação de EEG, ajudando neurologistas a darem um diagnóstico mais preciso. Na década de 60, equipamentos de medição de campos magnéticos foram criados, usados para medir a atividade magnética cerebral, criando, assim, o exame chamado magnetoencefalograma (MEG). Técnicas de SCF são usadas em sinais EEG e MEG para detecção e remoção de artefatos, análise de resposta a eventos e remoção de ruídos (COMON; JUTTEN, 2010). Também é possível analisar a atividade elétrica do coração por meio de eletrodos posicionados na superfície do corpo, resultando no ECG. Este exame é bastante utilizado na prática médica, pois assim como o EEG, seu custo operacional é baixo e existem algoritmos de processamento de sinais eficientes que permitem a extração de várias informações sobre o estado de saúde do paciente. Utiliza-se a SCF em ECG basicamente para remoção de ruído e separação de diferentes fontes bioelétricas do coração, como a atividade atrial e a atividade ventricular (SÖRNMO; LAGUNA, 2005).

Sistemas de telecomunicações sem fio no cenário multiusuário geralmente são representados por modelos *multiple-input multiple-output* (MIMO) (ZARZOSO; NANDI, 2004). Geralmente se utiliza um modelo de composição convolutiva para representar a transmissão de símbolos por um canal MIMO. A estimação desses símbolos pode ser realizada por meio da SCF, que, nesse contexto, é chamada de equalização cega. Uma vantagem dos métodos cegos é a estimação das fontes sem a necessidade de sinais de referência (sequência piloto), aumentando a eficiência no uso da banda disponível. Vários métodos de equalização cega foram propostos baseados em diferentes estratégias, como: *constant modulus* (CM) (GOOCH; LUNDELL, 1986), não-estacionaridade (AGEE; SCHELL; GARDNER, 1990), não-Gaussianidade (DOGAN; MENDEL, 1994) e alfabeto discreto (TALWAR; VIBERG; PAULRAJ, 1993).

Nesse capítulo, alguns conceitos básicos do problema de separação cega de fontes foram introduzidos. Os três modelos de separação de fontes foram discutidos, juntamente com algumas estratégias e campos de aplicação. No próximo capítulo, as ferramentas matemáticas necessárias para a compreensão desses métodos serão apresentadas.

2. FERRAMENTAS MATEMÁTICAS

2.1 Introdução

Em processamento de sinais, é comum o desenvolvimento de modelos matemáticos que expliquem o comportamento de sinais e sistemas. Neste Capítulo serão apresentadas algumas ferramentas matemáticas que possibilitarão o desenvolvimento de modelos para a SCF. As estatísticas de ordem superior (EOS), que serão apresentadas na Seção 2.3, são fundamentais para os métodos de SCF e podem ser naturalmente representadas através de tensores. Por isso, introduz-se o conceito de tensor e algumas notações na Seção 2.2. Da mesma forma, a Teoria da Informação fornece um *insight* dos sinais que ajuda no desenvolvimento de métodos de separação de fontes, sendo apresentados na Seção 2.4. Frequentemente é preciso otimizar funções objetivos para se encontrar a separação de fontes ótima. Dessa forma, adota-se alguns algoritmos de otimização numérica, apresentados na Seção 2.5.

2.2 Tensores

Tensores constituem uma ferramenta matemática introduzida no século XIX no contexto do cálculo diferencial, sendo bastante utilizada na Física para representar leis independentemente do sistema de coordenadas (COMON, 2014). Um tensor é definido como um operador multilinear que transforma um espaço linear em um outro espaço também linear. Da mesma forma, ele também pode ser interpretado como um *array* multidimensional de dados. A ordem N de um tensor corresponde ao número de índices dos elementos desse *array*. Portanto, uma matriz $\mathbf{A} \in \mathbb{R}^{I \times J}$ é um tensor de segunda ordem. A Álgebra Linear, fundamentada em vetores e matrizes, é um campo de estudos desenvolvido, bastante utilizada em diversas técnicas na Engenharia. Por outro lado, a Álgebra Multilinear, campo da matemática que estuda os tensores de ordem superior ($N > 2$) ainda está em desenvolvimento (KOLDA; BADER, 2009; CICHOCKI *et al.*, 2014), sendo cada vez mais utilizado em diversas aplicações, como a SCF.

Será visto adiante que EOS são naturalmente representadas por tensores. Visto que vários métodos de SCF utilizam EOS, é necessário introduzir algumas ferramentas tensoriais. O produto externo entre dois tensores de ordem N e M $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ e $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_M}$ resulta em um tensor $\mathcal{C} = \mathcal{A} \circ \mathcal{B}$ de ordem $(N + M)$ definido da seguinte forma:

$$c_{i_1 i_2 \dots i_n j_1 j_2 \dots j_m} = a_{i_1 i_2 \dots i_n} b_{j_1 j_2 \dots j_m}, \quad (2.1)$$

no qual $c_{i_1 i_2 \dots i_n j_1 j_2 \dots j_m}$, $a_{i_1 i_2 \dots i_n}$ e $b_{j_1 j_2 \dots j_m}$ representam, respectivamente, os elementos de \mathcal{C} , \mathcal{A} e \mathcal{B} . Pode-se multiplicar o tensor \mathcal{A} por uma matriz por meio do produto no modo- n . Seja uma matriz $\mathbf{U} \in \mathbb{R}^{J \times I_n}$, então o produto no modo- n , representado por $\mathcal{A} \times_n \mathbf{U}$, é definido como

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 \dots i_{n-1} j_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_N} u_{j i_n}. \quad (2.2)$$

Nota-se que o tensor resultante possui dimensões $(I_1 \times I_2 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N)$.

Utilizando a notação “:” para denotar variação de todos os elementos do índice representado, pode-se definir alguns sub-tensores. *Slices* de um tensor são obtidos fixando todos os índices exceto dois, resultando em matrizes. Por exemplo, os slices de um tensor de terceira ordem $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ são definidos como $\mathbf{X}_{i:} \in \mathbb{R}^{I_2 \times I_3}$, $\mathbf{X}_{:j} \in \mathbb{R}^{I_1 \times I_3}$ e $\mathbf{X}_{::k} \in \mathbb{R}^{I_1 \times I_2}$. *Fibers* são definidos fixando todos os índices exceto um, resultando em vetores. Utilizando o mesmo exemplo do tensor de terceira ordem, seus *fibers* são representados por $\mathbf{x}_{:jk} \in \mathbb{R}^{I_1 \times 1}$, $\mathbf{x}_{i:k} \in \mathbb{R}^{I_2 \times 1}$ e $\mathbf{x}_{ij:} \in \mathbb{R}^{I_3 \times 1}$.

Tensores possuem diversas definições de *rank* (KOLDA; BADER, 2009). Um tensor de *rank* unitário de ordem N é aquele definido pelo produto externo de N vetores. Por exemplo, \mathcal{X} possui *rank* unitário se ele puder ser escrito da forma $\mathcal{X} = \mathbf{a}_1 \circ \mathbf{a}_2 \circ \mathbf{a}_3$, no qual $\mathbf{a}_1 \in \mathbb{R}^{I \times 1}$, $\mathbf{a}_2 \in \mathbb{R}^{J \times 1}$ e $\mathbf{a}_3 \in \mathbb{R}^{K \times 1}$. O *rank* de um tensor é definido como o menor número R de tensores de *rank* unitário que, somados, resultam nesse tensor. É importante enfatizar que o esse *rank* não é limitado pelas dimensões do tensor, ao contrário do caso matricial. Existem alguns resultados que conseguem limitar o *rank* em algumas dimensões específicas dos tensores. Uma outra característica interessante é que o *rank* de um tensor pode ser diferente se calculado em \mathbb{R} ou em \mathbb{C} .

Em várias aplicações, busca-se decompor tensores a fim de observar algumas características latentes dos dados armazenados nessas estruturas. A decomposição *PARAllel FACTors* de \mathcal{A} , um tensor de ordem N , é definida como a combinação linear de tensores de *rank* unitários:

$$\mathcal{A} = \sum_{r=1}^R \lambda_r \mathbf{p}_r^{(1)} \circ \mathbf{p}_r^{(2)} \circ \dots \circ \mathbf{p}_r^{(N)}, \quad (2.3)$$

no qual $\mathbf{p}_r^{(n)} \in \mathbb{R}^{I_n \times 1}$ para $n = 1, 2, \dots, N$, $\lambda_r \in \mathbb{R}$ e R é o *rank* de \mathcal{A} . Pode-se escrever o modelo (2.3) utilizando o produto de modo- n (KOLDA; BADER, 2009):

$$\mathcal{A} = \mathcal{D} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \dots \times_N \mathbf{P}^{(N)}, \quad (2.4)$$

no qual \mathcal{D} é um tensor diagonal formado por $\{\lambda_1, \lambda_2, \dots, \lambda_R\}$ e $\mathbf{P}^{(n)} = [\mathbf{p}_1^{(n)}, \mathbf{p}_2^{(n)}, \dots, \mathbf{p}_R^{(n)}]_{I_n \times R}$. Essa decomposição é ilustrada na Figura 2.1 para um tensor de terceira ordem.

A decomposição PARAFAC de um tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ pode ser calculada resolvendo-

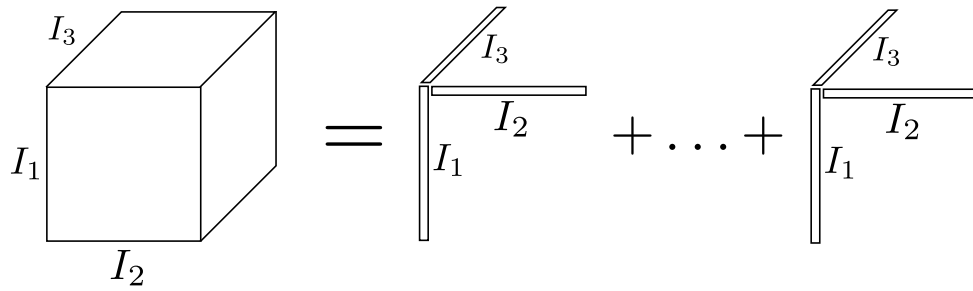


Figura 2.1: Decomposição PARAFAC para um tensor de terceira ordem com dimensões $(I_1 \times I_2 \times I_3)$.

se o seguinte problema de otimização:

$$\min_{\mathbf{P}^{(1)}, \dots, \mathbf{P}^{(N)}} \left\| \mathcal{T} - \mathcal{D} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \dots \times_N \mathbf{P}^{(N)} \right\|_F, \quad (2.5)$$

no qual $\| \cdot \|_F$ é a norma de Frobenius. Nota-se que (2.5) é um problema de otimização não-linear. Um dos métodos mais simples para resolver esse problema consiste em dividi-lo em vários sub-problemas lineares e resolvê-los de forma alternada. Essa abordagem é conhecida como *Alternating Least-Squares* (ALS). Na Subseção 3.5.1, o ALS será melhor detalhado no contexto de decomposição matricial. Apesar de ser conceitualmente simples e possuir baixo custo computacional, o ALS não é muito atrativo em relação a convergência, pois ela não é garantida. Todavia, outros métodos mais robustos foram propostos (TOMASI; BRO, 2006), apresentando melhores resultados de convergência.

2.3 Momentos e cumulantes

Seja x uma variável aleatória real com uma função distribuição cumulada $F_X(u) = Pr(x \leq u)$. A sua função densidade de probabilidade é dada por

$$f_x(u) = \frac{dF_x(u)}{du},$$

no qual $f_x(u) > 0 \forall u$ e $\int_{-\infty}^{\infty} f_x(u) du = 1$. A **primeira função característica** de x é (LACOUME *et al.*, 1997)

$$\Phi_x(v) = \mathbb{E} [e^{-jvx}] \quad (2.6)$$

$$= \int_{-\infty}^{\infty} e^{-jvx} f_x(u) du \quad (2.7)$$

no qual $j^2 = -1$ e $\mathbb{E}[\cdot]$ é o operador esperança. Observa-se que a primeira função característica de x é a transformada de Fourier da sua função densidade de probabilidade. A **segunda função característica** de x é definida como o logaritmo natural da primeira

função característica:

$$\Psi_x(v) = \ln(\Phi_x(v)). \quad (2.8)$$

O **momento** de r -ésima ordem de x é definido como (LACOUME *et al.*, 1997)

$$\mu_{x(r)} = \mathbb{E}[x^r] \quad (2.9)$$

e o momento de r -ésima ordem centrado

$$\mu'_{x(r)} = \mathbb{E}[(x - \mu_{x(1)})^r], \quad (2.10)$$

no qual $\mu_{x(1)}$, o momento de primeira ordem, é também chamado de média de x . Para verificar a relação entre a primeira função característica (2.6) e os momentos de x , primeiro calcula-se a expansão de Taylor de $e^{-jvx} = \sum_{r=0}^{\infty} \frac{(-jv)^r}{r!} x^r$ centrada em zero e a substitui em (2.6):

$$\mathbb{E}[e^{-jvx}] = \sum_{r=0}^{\infty} \frac{(-jv)^r}{r!} \mathbb{E}[x^r]. \quad (2.11)$$

Em seguida, calcula-se a expansão de Taylor de $\Phi_x(v)$:

$$\Phi_x(v) = \sum_{r=0}^{\infty} \frac{d^r \Phi_x(v)}{dv^r} \Big|_{v=0} \frac{v^r}{r!}. \quad (2.12)$$

Como as equações (2.11) e (2.12) são equivalentes, então tem-se que

$$\mu_{x(r)} = \mathbb{E}[x^r] = j^r \frac{d^r \Phi_x(v)}{dv^r} \Big|_{v=0}. \quad (2.13)$$

De forma semelhante, define-se o **cumulante** de r -ésima ordem de x como

$$\kappa_{x(r)} = \text{Cum}[x, x, \dots, x] = j^r \frac{d^r \Psi_x(v)}{dv^r} \Big|_{v=0}. \quad (2.14)$$

As equações (2.13) e (2.14) relacionam os momentos e os cumulantes de x com as suas duas funções características. Dessa forma, pode-se afirmar que uma variável aleatória é completamente caracterizada por suas funções características (LACOUME *et al.*, 1997).

Uma relação direta entre os cumulantes e os momentos de x pode ser encontrada desenvolvendo o valor esperado implícito em (2.8). Assim, obtém-se as seguintes relações:

$$\begin{aligned} \kappa_{x(1)} &= \mu_{x(1)}, \\ \kappa_{x(2)} &= \mu_{x(2)} - \mu_{x(1)}^2, \\ \kappa_{x(3)} &= \mu_{x(3)} - 3\mu_{x(1)}\mu_{x(2)} + 2\mu_{x(1)}^3, \\ \kappa_{x(4)} &= \mu_{x(4)} - 4\mu_{x(3)}\mu_{x(1)} - 3\mu_{x(2)}^2 + 12\mu_{x(2)}\mu_{x(1)}^2 - 6\mu_{x(1)}^4. \end{aligned}$$

Quando a variável aleatória é centrada, ou seja $\mu_{x(1)} = 0$, então as relações acima tornam-se:

$$\kappa_{x(1)} = 0, \quad (2.15)$$

$$\kappa_{x(2)} = \mathbb{E}[x^2], \quad (2.16)$$

$$\kappa_{x(3)} = \mathbb{E}[x^3], \quad (2.17)$$

$$\kappa_{x(4)} = \mathbb{E}[x^4] - 3\mathbb{E}[x^2]^2. \quad (2.18)$$

Identifica-se rapidamente nessas relações a variância $\text{var}(x) = \kappa_{x(2)}$. A **assimetria** é definida como o cumulante de 3^a ordem centrado normalizado pela 1,5-ésima potência do cumulante de 2^a ordem

$$K_{x(3)} = \frac{\mathbb{E}[x^3]}{\mathbb{E}[x^2]^{3/2}}. \quad (2.19)$$

A assimetria é nula se a função densidade de probabilidade de x possui um eixo de simetria.

Quando x possui uma distribuição Gaussiana, a sua segunda função característica é dada por

$$\Psi_x(v) = j\mu_{x(1)}v - \frac{1}{2}\mu_{x(2)}v^2. \quad (2.20)$$

Observa-se que os seus cumulantes de ordem superior a 2 são nulos, assim uma variável Gaussiana é completamente caracterizadas pela suas estatísticas de segunda ordem (LACOUME *et al.*, 1997). Dessa forma, possuir cumulante de 4^a ordem nulo é um requisito para que uma variável seja Gaussiana. Para verificar isso, utiliza-se normalmente o cumulante de 4^a ordem centrado e normalizado, a **curtose**:

$$K_{x(4)} = \frac{\mathbb{E}[x^4]}{\mathbb{E}[x^2]^2} - 3. \quad (2.21)$$

Uma distribuição de probabilidade é **sub-Gaussiana** quando ela possui curtose negativa, sendo “plana” ao redor da média. Por outro lado, quando a curtose é positiva, a distribuição de probabilidade é dita **super-Gaussiana**, apresentando uma distribuição com um pico agudo ao redor da média. Na Figura 2.2, observa-se uma distribuição Laplaciana com curtose = 3, sendo então super-Gaussiana. Também observa-se uma distribuição de Wigner com curtose = -1, sendo sub-Gaussiana.

Agora considera-se o caso multidimensional com o vetor aleatório de ordem N

$$\mathbf{x} = [x_{i_1}, x_{i_2}, \dots, x_{i_N}]^T,$$

no qual x_{i_n} , para $n = 1, 2, \dots, N$ são variáveis aleatórias centradas. A sua **primeira**

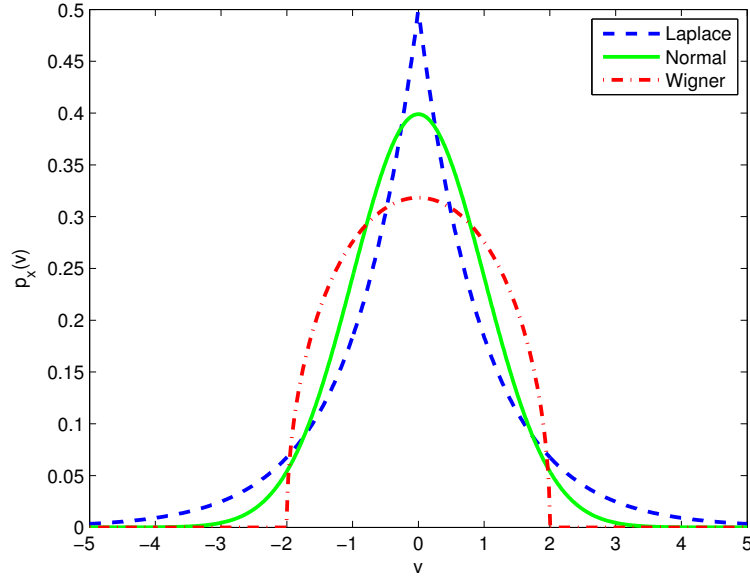


Figura 2.2: Densidades de probabilidade de Laplace, Wigner e normal.

função característica conjunta é dada pela relação (LACOUME *et al.*, 1997)

$$\Phi_{\mathbf{x}}(\mathbf{v}) = \mathbb{E} \left[e^{-j\mathbf{v}^T \mathbf{x}} \right] \quad (2.22)$$

$$= \int_{\mathbb{R}^N} e^{-j\mathbf{v}^T \mathbf{u}} f_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \quad (2.23)$$

no qual $\mathbf{v} = [v_{i_1}, v_{i_2}, \dots, v_{i_N}]^T$ e $f_{\mathbf{x}}(\mathbf{u})$ é a função densidade de probabilidade conjunta de \mathbf{x} . De forma análoga ao caso unidimensional, observa-se que a primeira função característica conjunta de \mathbf{x} é a transformada de Fourier multidimensional da sua função densidade de probabilidade conjunta. A **segunda equação característica** é definida como

$$\Psi_{\mathbf{x}}(\mathbf{v}) = \ln \Phi_{\mathbf{x}}(\mathbf{v}). \quad (2.24)$$

Expandindo a série de Taylor de $e^{-j\mathbf{v}^T \mathbf{x}}$ ao redor de $\mathbf{v} = \mathbf{0}$, e substituindo-a em (2.22), obtém-se de forma semelhante ao caso univariável os momentos

$$\mu_{\mathbf{x}i_1 i_2 \dots i_N} = j^N \left. \frac{\partial^N \Phi_{\mathbf{x}}(\mathbf{v})}{\partial v_{i_1} \partial v_{i_2} \dots \partial v_{i_N}} \right|_{\mathbf{v}=\mathbf{0}} = \mathbb{E} [x_{i_1} x_{i_2} \dots x_{i_N}]. \quad (2.25)$$

De forma análoga, os cumulantes são definidos como

$$\kappa_{\mathbf{x}i_1 i_2 \dots i_N} = j^N \left. \frac{\partial^N \Psi_{\mathbf{x}}(\mathbf{v})}{\partial v_{i_1} \partial v_{i_2} \dots \partial v_{i_N}} \right|_{\mathbf{v}=\mathbf{0}} = \text{Cum} [x_{i_1} x_{i_2} \dots x_{i_N}]. \quad (2.26)$$

2.3.1 Propriedades

Momentos e cumulantes podem ser naturalmente representados por tensores (McCULLAGH, 1987). Define-se o tensor de momentos de N -ésima ordem de \mathbf{x} como

$$\mathcal{M}_{\mathbf{x}}^{(N)} = \mu_{\mathbf{x}i_1i_2\dots i_N} = \mathbb{E}[x_1x_2\dots x_N], \quad (2.27)$$

no qual $\mathcal{M}_{\mathbf{x}}^{(N)} \in \mathbb{R}^{N \times N \times \dots \times N}$. Da mesma forma, define-se o tensor de cumulantes de N -ésima ordem de \mathbf{x} como

$$\mathcal{C}_{\mathbf{x}}^{(N)} = \kappa_{\mathbf{x}i_1i_2\dots i_N} = \text{Cum}[x_1, x_2, \dots, x_N]. \quad (2.28)$$

Essa representação tensorial permite a observação de algumas propriedades dos momentos e cumulantes. Primeiramente, mostra-se (NIKIAS; MENDEL, 1993; LATHAUWER; MOOR; VANDEWALLE, 2000) que os esses tensores são simétricos, ou seja

$$(\mathcal{M}_{\mathbf{x}}^{(N)})_{i_1i_2\dots i_N} = (\mathcal{M}_{\mathbf{x}}^{(N)})_{P(i_1i_2\dots i_N)} \quad (2.29)$$

$$(\mathcal{C}_{\mathbf{x}}^{(N)})_{i_1i_2\dots i_N} = (\mathcal{C}_{\mathbf{x}}^{(N)})_{P(i_1i_2\dots i_N)} \quad (2.30)$$

no qual P é uma permutação de índices.

Uma segunda propriedade diz respeito à multilinearidade dos momentos e cumulantes. A fim de ilustração, considera-se o seguinte exemplo. Sejam dois vetores aleatórios $\mathbf{x} \in \mathbb{R}^{N \times 1}$ e $\mathbf{y} \in \mathbb{R}^{M \times 1}$ relacionados pela transformação linear $\mathbf{y} = \mathbf{A}\mathbf{x}$, no qual $\mathbf{A} = [a_{ij}]_{M \times N}$ é um operador linear. Então os momentos e cumulantes de \mathbf{y} são funções lineares de cada a_{ij} , sendo assim estatísticas **multilineares** em a_{ij} . Nesse caso, a primeira função característica conjunta de \mathbf{y} é

$$\begin{aligned} \Phi_{\mathbf{y}}(\mathbf{v}) &= \mathbb{E} \left[e^{-j\mathbf{v}^T \mathbf{y}} \right] \\ &= \mathbb{E} \left[e^{-j\mathbf{v}^T (\mathbf{A}\mathbf{x})} \right] \\ &= \mathbb{E} \left[e^{-j\mathbf{x}^T (\mathbf{A}^T \mathbf{v})} \right] \\ &= \Phi_{\mathbf{x}}(\mathbf{A}^T \mathbf{v}). \end{aligned} \quad (2.31)$$

Os cumulantes de terceira ordem de \mathbf{y} são obtidos substituindo a equação (2.31) em (2.26):

$$\text{Cum}[y_i, y_j, y_k] = \sum_{m,n,p} a_{im}a_{jn}a_{kp} \text{Cum}[x_m, x_n, x_p]. \quad (2.32)$$

Identifica-se a equação (2.32) como três produtos modo- n (2.2) consecutivos. Utilizando

a notação tensorial, reescreve-se a equação (2.32) como:

$$\mathcal{C}_{\mathbf{y}}^{(3)} = \mathcal{C}_{\mathbf{x}}^{(3)} \times_1 \mathbf{A} \times_2 \mathbf{A} \times_3 \mathbf{A}. \quad (2.33)$$

Adotando-se o mesmo raciocínio acima, pode-se generalizar essa propriedade para o caso de ordem N para momentos e cumulantes (LATHAUWER; MOOR; VANDEWALLE, 2000):

$$\mathcal{M}_{\mathbf{y}}^{(N)} = \mathcal{M}_{\mathbf{x}}^{(N)} \times_1 \mathbf{A} \times_2 \mathbf{A} \dots \times_N \mathbf{A} \quad (2.34)$$

$$\mathcal{C}_{\mathbf{y}}^{(N)} = \mathcal{C}_{\mathbf{x}}^{(N)} \times_1 \mathbf{A} \times_2 \mathbf{A} \dots \times_N \mathbf{A}. \quad (2.35)$$

A terceira propriedade é relacionada à independência das observações. Seja $\mathbf{z} = [\mathbf{x}, \mathbf{y}]^T$ um vetor aleatório formado pela concatenação de outros dois vetores aleatórios, \mathbf{x} e \mathbf{y} . Se \mathbf{x} e \mathbf{y} são estatisticamente independentes, então a função densidade de probabilidade conjunta de \mathbf{z} pode ser escrita da seguinte forma:

$$f_{\mathbf{z}}(\mathbf{u}, \mathbf{v}) = f_{\mathbf{x}, \mathbf{y}}(\mathbf{u}, \mathbf{v}) = f_{\mathbf{x}}(\mathbf{u})f_{\mathbf{y}}(\mathbf{v}). \quad (2.36)$$

Substituindo a relação (2.36) na definição de primeira equação característica conjunta (2.22), tem-se

$$\Phi_{\mathbf{z}}(\mathbf{u}, \mathbf{v}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{x}, \mathbf{y}}(\mathbf{u}, \mathbf{v}) e^{-j(\mathbf{x}^T \mathbf{u} + \mathbf{y}^T \mathbf{v})} \mathbf{d}\mathbf{u} \mathbf{d}\mathbf{v} \quad (2.37)$$

$$= \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{u}) e^{-j\mathbf{x}^T \mathbf{u}} \mathbf{d}\mathbf{u} \int_{-\infty}^{\infty} f_{\mathbf{y}}(\mathbf{v}) e^{-j\mathbf{y}^T \mathbf{v}} \mathbf{d}\mathbf{v} \quad (2.38)$$

$$= \Phi_{\mathbf{x}}(\mathbf{u}) \Phi_{\mathbf{y}}(\mathbf{v}) \quad (2.39)$$

Aplicando o logaritmo natural na equação (2.39), tem-se

$$\Psi_{\mathbf{z}}(\mathbf{u}, \mathbf{v}) = \Psi_{\mathbf{x}}(\mathbf{u}) + \Psi_{\mathbf{y}}(\mathbf{v}). \quad (2.40)$$

Dessa forma, os cumulantes de \mathbf{z} são

$$\kappa_{\mathbf{z}i_1, \dots, i_n, j_1, \dots, j_m} = \text{Cum} [x_{i_1}, \dots, x_{i_n}, y_{j_1}, \dots, y_{j_m}] \quad (2.41)$$

$$= j^r \frac{\partial^r}{\partial u_{i_1} \dots \partial u_{i_n} \partial v_{j_1} \dots \partial v_{j_m}} [\Psi_{\mathbf{x}}(\mathbf{u}) + \Psi_{\mathbf{y}}(\mathbf{v})] \Big|_{\mathbf{v}=\mathbf{0}}, \quad (2.42)$$

no qual $r = n + m$. O cumulante (2.41) é nulo sempre que um dos i_k e um dos j_k são não-nulos simultaneamente, zerando a derivada parcial dessa equação (MCCULLAGH, 1987; LACOUME *et al.*, 1997). Dessa forma, os cumulantes cruzados de dois vetores independentes são nulos.

2.3.2 Estimação de momentos e cumulantes

Em situações práticas, deseja-se calcular as estatísticas de sinais a partir de suas amostras utilizando estimadores. O **viés** de um estimador $\hat{\theta}$ de uma variável aleatória θ é definido como $B(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$. O viés mede a distância entre a média da coleção de estimativas e o valor a ser estimado. O estimador $\hat{\theta}$ será não-enviesado quando $\mathbb{E}[\hat{\theta}] = \theta$. Como o viés de um estimador está ligado à qualidade deste, sempre se busca um estimador com o viés mínimo.

Seja x uma variável aleatória unidimensional centrada e x_n , $1 \leq n \leq N$ a n -ésima realização de x . Um estimador não-enviesado do momento de r -ésima ordem de x é dado por

$$\hat{\mu}_{(r)} = \frac{1}{N} \sum_{n=1}^N x_n. \quad (2.43)$$

Observa-se facilmente que esse estimador é não-enviesado, pois $\mathbb{E}[\hat{\mu}_{(r)}] = \mu_{(r)}$.

Em relação à estimação de cumulantes de uma variável aleatória centrada, utiliza-se as relações teóricas (2.15)-(2.18). Um estimador do cumulante de quarta ordem centrado é

$$\hat{\kappa}_{(4)} = \hat{\mu}_{(4)} - 3\hat{\mu}_{(2)}^2 \quad (2.44)$$

$$= \frac{1}{N} \sum_{i=1}^N x_i^4 - \frac{3}{N^2} \sum_{i,j=1}^N x_i^2 x_j^2 \quad (2.45)$$

Desenvolvendo a equação (2.45) e aplicando o operador esperança, obtém-se

$$\mathbb{E}[\hat{\kappa}_{(4)}] = \kappa_{(4)} - 3 \frac{\kappa_{(4)} + 2\mu_{(2)}^2}{N}. \quad (2.46)$$

Constata-se facilmente que esse estimador é enviesado. Entretanto quando $N \rightarrow \infty$, ele se torna não-enviesado. No caso de poucas amostras, o viés é considerável. Um estimador de cumulante não-enviesado, chamado de **k-estatística** (MCCULLAGH, 1987; LACOUME *et al.*, 1997), é definido como

$$\hat{k}_{(4)} = \frac{N+2}{N(N-1)} \sum_{i=1}^N x_i^4 - \frac{3}{N(N-1)} \sum_{i,j=1}^N x_i^2 x_j^2. \quad (2.47)$$

Uma vez calculado $\hat{k}_{(4)}$, pode-se obter a curtose através da seguinte equação:

$$\hat{K}_{(4)} = \hat{k}_{(4)} / \hat{k}_{(2)}^2. \quad (2.48)$$

Em (MCCULLAGH, 1987), o viés de (2.48) é discutido juntamente com o seu comportamento diante de variáveis Gaussianas. Nesse caso, o valor esperado da curtose é nulo. Apesar da simplicidade e da eficiência dos métodos de separação de fontes baseados em

cumulantes de ordem superior (Capítulo 3), os geralmente os estimadores de cumulantes são sensíveis a *outliers* (HUBER, 1985), indicando uma sensibilidade desse tipo de algoritmo.

2.4 Teoria da Informação

Uma outra forma de se descrever um conjunto de dados é por meio da sua quantidade de informação (COMON; JUTTEN, 2010). Essa abordagem, também conhecida como Teoria da Informação, fundamenta-se na medida de informação por meio da **entropia**, definida como a média de informação de uma variável aleatória. Seja x uma variável aleatória com valores discretos $\{x_1, x_2, \dots\}$ com função densidade de probabilidade $f_x(\cdot)$. A sua entropia é

$$H(x) = - \sum_i f_x(x_i) \log f_x(x_i). \quad (2.49)$$

Quanto maior a entropia de uma variável aleatória, maior será incerteza contida nela. Considerando agora x uma variável aleatória contínua com função densidade de probabilidade $p_x(\cdot)$, define-se a **entropia diferencial**

$$S(x) = - \int p_x(u) \log p_x(u) du. \quad (2.50)$$

Seja um vetor aleatório $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ com função densidade de probabilidade multivariada $p_{\mathbf{x}}(\cdot)$, então a sua entropia diferencial é definida como

$$S(\mathbf{x}) = - \int p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (2.51)$$

A **informação mútua** (IM) de um vetor aleatório \mathbf{x} é uma medida de independência mútua das componentes deste vetor (HYVÄRINEN; KARHUNEN; OJA, 2004). Ela é definida como

$$I(\mathbf{x}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \left[\frac{p_{\mathbf{x}}(\mathbf{u})}{\prod_i p_{x_i}(u_i)} \right] d\mathbf{u}, \quad (2.52)$$

no qual $\mathbf{u} = [u_1, u_2, \dots, u_N]^T$, $p_{\mathbf{x}}(\cdot)$ é a função densidade de probabilidade conjunta do vetor aleatório \mathbf{x} e $p_{x_i}(\cdot)$ é a função densidade de probabilidade marginal das suas componentes. Se as componentes de \mathbf{x} forem mutuamente independentes, então $p_{\mathbf{x}}(\mathbf{u}) = \prod_i p_{x_i}(u_i)$ e $I(\mathbf{x}) = 0$. A divergência de Kullback-Leibler, uma medida de “distância” entre duas distribuições de probabilidade, é definida como

$$K(p_{\mathbf{a}}|p_{\mathbf{b}}) = \int p_{\mathbf{a}}(\mathbf{u}) \log \left[\frac{p_{\mathbf{a}}(\mathbf{u})}{p_{\mathbf{b}}(\mathbf{u})} \right] d\mathbf{u}, \quad (2.53)$$

no qual $p_{\mathbf{a}}(\cdot)$ e $p_{\mathbf{b}}(\cdot)$ denotam a função densidade de probabilidade conjunta dos vetores

\mathbf{a} e \mathbf{b} , respectivamente. Nota-se, que a IM pode ser escrita como uma divergência de Kullback-Leibler:

$$I(\mathbf{x}) = K(p_{\mathbf{x}} | \prod_i p_{x_i}). \quad (2.54)$$

Um outro conceito importante neste trabalho é a **negentropia**. Ela mede a diferença de entropia diferencial entre um vetor aleatório \mathbf{x} e outro \mathbf{x}_g , com distribuição de probabilidade conjunta Gaussiana:

$$J(\mathbf{x}) = S(\mathbf{x}_g) - S(\mathbf{x}). \quad (2.55)$$

Essa medida é feita em relação a um vetor Gaussiano pois para uma variável aleatória de variância unitária cuja densidade de probabilidade é definida em todo espaço real, a densidade Gaussiana proporciona entropia diferencial máxima (HYVÄRINEN; KARHUNEN; OJA, 2004). Dessa forma, pode-se utilizar essas variáveis como referência em entropia. Se $J(\mathbf{x}) = \mathbf{0}$, então o vetor \mathbf{x} é Gaussiano. Em (COMON, 1994), Pierre Comon demonstra a seguinte relação entre informação mútua e a negentropia:

$$I(\mathbf{x}) = J(\mathbf{x}) - \sum_i J(x_i). \quad (2.56)$$

Essa relação será útil no próximo Capítulo ao se derivar métodos de separação de fontes.

2.5 Otimização numérica

Os métodos de SCF são derivados de problemas de otimização que visam minimizar ou maximizar uma certa função objetivo $f : \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}$. Admite-se nesta seção, sem perda de generalidade, que se busca o ponto ótimo $\mathbf{x}^* \in \mathbb{R}^{N \times 1}$ tal que

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} f(\mathbf{x}). \quad (2.57)$$

Uma estratégia para resolver esse problema de otimização, chamada de busca linear, consiste em buscar o ponto ótimo em uma direção \mathbf{p}_k de forma iterativa (NOCEDAL; WRIGHT, 2006). Essa busca é feita através da seguinte equação:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mu_k \mathbf{p}_k, \quad (2.58)$$

no qual μ_k é um fator de passo tal que

$$\min_{\mu_k} f(\mathbf{x}_k + \mu_k \mathbf{p}_k). \quad (2.59)$$

É preciso escolher uma direção \mathbf{p}_k que garanta o decréscimo na função f . Os diferentes

métodos baseados nessa estratégia diferem na direção \mathbf{p}_k adotada e na forma de calcular μ_k .

A primeira direção a ser considerada é a descida mais íngreme (*steepest descent*) $\mathbf{p}_k = -\nabla f(\mathbf{x})$. Essa direção é simples de se calcular, pois ela requer apenas o cálculo do gradiente $\nabla f(\mathbf{x})$. Por outro lado, ela pode convergir para um mínimo local e ser bastante lenta em um ponto de sela. O fator de passo μ_k deve ser escolhido de tal forma que obedeça as condições de Wolfe (NOCEDAL; WRIGHT, 2006), garantindo decréscimo em $f(\mathbf{x}_{k+1})$. Porém, é possível calcular μ_k de forma ótima a cada iteração, assim como é feito no algoritmo RobustICA (Subseção 3.4.2).

Uma alternativa ao *steepest descent* é a direção de Newton. Seja a aproximação de segunda ordem de $f(\mathbf{x}_k + \mathbf{p})$ pela expansão de Taylor:

$$m_k(\mathbf{p}) \approx f_k + \mathbf{p}^T \nabla f_k + \frac{1}{2} \mathbf{p}^T \nabla^2 f_k \mathbf{p}. \quad (2.60)$$

Assume-se que $\nabla^2 f_k$ é positiva definida. A direção de Newton é obtida calculando-se o vetor que minimiza $m_k(\mathbf{p})$. Calculando o gradiente de $m_k(\mathbf{p})$, tem-se:

$$\nabla_{\mathbf{p}} m_k(\mathbf{p}) = \frac{\partial m_k(\mathbf{p})}{\partial \mathbf{p}} \quad (2.61)$$

$$= \nabla f_k + \nabla^2 f_k \mathbf{p}. \quad (2.62)$$

Igualando o gradiente (2.62) ao vetor nulo, tem-se:

$$\mathbf{p}_k^N = -\nabla^2 f_k^{-1} \nabla f_k. \quad (2.63)$$

Normalmente, no método de Newton, faz-se $\mu_k = 1$ para todo k . A vantagem do método de Newton é que ele geralmente apresenta taxa de convergência quadrática. Em contrapartida, é preciso calcular o Hessiano $\nabla^2 f_k$, que pode apresentar um alto custo computacional.

Quando o Hessiano não é uma matriz positiva definida, pode-se utilizar uma aproximação \mathbf{B}_k do Hessiano, constituindo os métodos quasi-Newton. A direção quasi-Newton é $\mathbf{p}_k^{QN} = -\mathbf{B}_k^{-1} \nabla f_k$. Normalmente se impõe algumas restrições como: \mathbf{B}_{k+1} deve ser simétrico, $(\mathbf{B}_{k+1} - \mathbf{B}_k)$ deve possuir *rank* pequeno e $\mathbf{B}_{k+1} \mathbf{s}_k = \mathbf{y}_k$, no qual $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ e $\mathbf{y}_k = \nabla f_{k+1} - \nabla f_k$. Um algoritmo do tipo quasi-Newton popular é o algoritmo Broyden-Fletcher-Goldfarb-Shanno (BFGS), em que a aproximação do Hessiano é atualizada da seguinte forma:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k}. \quad (2.64)$$

Existem algumas variantes do algoritmo BFGS, como a versão de memória limitada, L-BFGS, adaptada para problemas com muitas variáveis. Os métodos estudados nessa

Subseção podem ser resumidos de acordo com o Algoritmo 1. Pode-se usar como critério de convergência a seguinte relação:

$$\|\mathbf{p}_k\| < \epsilon,$$

no qual ϵ é um valor pequeno.

Entrada: Função f a ser minimizada e valor inicial \mathbf{x}_0 .

Saída: Valor após convergência \mathbf{x}^*

1. Inicialização

- $k \leftarrow 0$.

2. Repita até a convergência:

- Calcular direção \mathbf{p}_k .
- Escolher ou calcular o fator de passo μ_k .
- Atualizar utilizando a equação (2.58).
- $k \leftarrow k + 1$.

3. Retornar $\mathbf{x}^* \leftarrow \mathbf{x}_k$.

Algoritmo 1: Algoritmo de busca linear.

Conclusão

Nesse capítulo algumas ferramentas matemáticas fundamentais para o estudo de métodos de separação cega de fontes foram introduzidas. Duas medidas de “Gaussianidade” foram definidas. A curtose, baseada no cumulante de quarta ordem, e a negentropia, derivada de uma abordagem de teoria de informação. Nos próximos capítulos, essas medidas serão utilizadas em métodos matemáticos de separação de fontes juntamente com os algoritmos de otimização numérica aqui expostos.

3. MÉTODOS E ALGORITMOS

3.1 Introdução

Como apresentado no Capítulo 1, a SCF consiste em recuperar fontes não-observadas a partir de um conjunto de observações provenientes da composição dessas fontes. Como não se conhece nem as fontes e nem o sistema de composição, é preciso assumir características e modelos que tentem descrevê-los. Uma vez que os modelos foram estabelecidos, propõe-se uma estrutura de processamento e um critério de otimização, que, ao ser otimizado, resulte na separação das fontes. Finalmente se desenvolve um algoritmo que implemente a otimização desse critério.

Neste Capítulo, serão apresentados modelos, estruturas, critérios de otimização, métodos e algoritmos de SCF. Em relação às fontes, será assumido que elas são mutuamente independentes, não-Gaussianas e assumem valores reais. Apesar dessas condições serem matematicamente fortes, elas são plausíveis em situações práticas, como será visto no Capítulo 4. Além disso, a independência estatística apresenta propriedades interessantes para os métodos de SCF. Assume-se que as fontes serão misturadas por um sistema instantâneo linear. Dessa forma, as fontes podem ser recuperadas por meio de dois tipos de estruturas de filtragem: *multiple-input single-output* (MISO) e *multiple-input multiple-output* (MIMO). A primeira estrutura é associada à extração de fontes, enquanto a outra é relacionada à separação de fontes. Também é assumido que o número de sensores M é igual ao número de fontes N , implicando que a matriz de composição pode ser inversível. Os critérios de otimização para essas estruturas vão se fundamentar na suposição de fontes independentes, explorando essa propriedade por meio da Teoria da Informação (Seção 2.4) e de estatísticas de ordem superior (Seção 2.3). Após a definição de contrastes, serão apresentados métodos algébricos e iterativos de SCF e por último será feito um rápido estudo bibliográfico sobre novas tendências nesse campo de estudo.

3.2 Contrastes

Em um problema de SCF, dado um conjunto de observações, procura-se saber se o sistema de composição pode ser identificado e se as fontes podem ser separadas corretamente. **Funções de contraste** oferecem uma resposta para essas duas questões. Essas funções são definidas de tal forma que os seus máximos globais correspondam à separação de fontes. Para isso, é preciso que elas obedeçam a algumas condições que garantam a consistência das soluções. Essas condições serão discutidas na Subseção a seguir.

3.2.1 Extração de fontes

A primeira estrutura de SCF a ser considerada neste trabalho consiste em um filtro que extrai uma única fonte do conjunto de observações por meio de uma filtragem MISO. Seja o modelo de composição instantâneo e linear (1.2). Define-se o filtro extrator $\mathbf{f} \in \mathbb{R}^{M \times 1}$, tal que a sua saída $y(t) = \mathbf{f}^T \mathbf{x}(t)$ corresponda a uma das fontes $s_n(t) \forall n$, como ilustrado na Figura 3.1. Costuma-se a relacionar a saída $y(t)$ diretamente às fontes da seguinte forma:

$$y(t) = \mathbf{f}^T \mathbf{x}(t) = \mathbf{f}^T \mathbf{A} \mathbf{s}(t) = \mathbf{g}^T \mathbf{s}(t),$$

no qual $\mathbf{g} = \mathbf{A}^T \mathbf{f}$ é chamado de **filtro global**. Quando um filtro possui apenas um único valor não-nulo, ele é chamado de filtro trivial. Portanto, para que o filtro global obtenha apenas uma única fonte, é preciso que ele seja trivial.

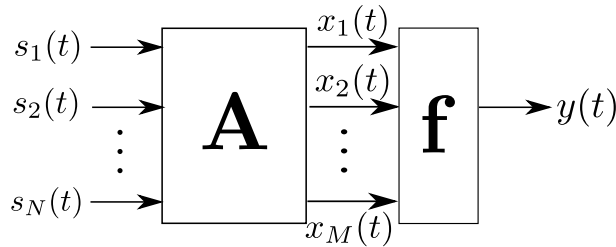


Figura 3.1: Ilustração da composição de N fontes e extração de uma fonte através do filtro \mathbf{f} .

O filtro extrator \mathbf{f} ótimo é obtido através da otimização de uma função contraste $\Upsilon(\cdot)$, que deve obedecer as seguintes propriedades (COMON; JUTTEN, 2010):

1. Para todo filtro trivial \mathbf{g} , existe uma fonte $s_i(t)$ tal que $\Upsilon(\mathbf{g}^T \mathbf{s}) = \Upsilon(s_i(t))$. Assim, a função $\Upsilon(\cdot)$ é **invariante** aos filtros triviais.
2. O efeito da composição, representado por \mathbf{A} em $\mathbf{g} = \mathbf{A}^T \mathbf{f}$, diminui a função contraste $\Upsilon(\mathbf{g}^T \mathbf{s}) \leq \Upsilon(\mathbf{f}^T \mathbf{s})$. Essa é a propriedade da **dominância**.
3. O máximo de $\Upsilon(\cdot)$ deve ser atingido somente para filtros com a estrutura esperada:

$$\Upsilon(\mathbf{g}^T \mathbf{s}) = \max_{\mathbf{f}} \Upsilon(\mathbf{f}^T \mathbf{s}) \Rightarrow \mathbf{g} \text{ é trivial.}$$

A propriedade da **discriminação** impede que valores espúrios sejam tidos como solução do problema.

Na Seção seguinte, serão apresentados vários contrastes e métodos que também podem ser utilizados para a extração de fontes. Agora será apresentado uma técnica para extração consecutiva de fontes conhecida como **deflação**.

Suponha que uma fonte $y_1(t) = \mathbf{f}^T \mathbf{x}(t)$ foi corretamente estimada através de um método de SCF. Para se extrair as outras $(N - 1)$ fontes, subtrai-se a contribuição de

$y_1(t)$ das observações $\mathbf{x}(t)$ da seguinte forma:

$$\mathbf{x}'_1(t) = \mathbf{x}(t) - y_1(t)\mathbf{h}_1(t), \quad (3.1)$$

no qual $\mathbf{h}_1(t) = \mathbb{E}[y_1(t)\mathbf{x}(t)]/\mathbb{E}[y_1^2(t)]$ é o vetor de regressão que minimiza o erro quadrático médio de (3.1) em relação a $\mathbf{h}_1(t)$ (COMON; JUTTEN, 2010). O tamanho do vetor $\mathbf{x}'_1(t)$ pode ser reduzido, pois uma fonte foi extraída, reduzindo o *rank* da sua matriz de covariância. Isso pode ser feito por meio da sua *Singular Value Decomposition* (SVD), retendo somente os $(M - 1)$ valores e vetores singulares. O novo vetor de observação é dado por

$$\mathbf{x}_2(t) = \mathbf{P}_1\mathbf{x}'_1(t),$$

no qual \mathbf{P}_1 é a matriz esquerda de vetores singulares da matriz de covariância de $\mathbf{x}'_1(t)$ com o vetor singular menos significativo subtraído. Repete-se esse processo até a dimensão de $\mathbf{x}_m(t)$ ser unitária. Claramente, se todas as fontes forem sequencialmente extraídas, logo a separação de fontes foi realizada. De fato, essa é uma estratégia de algoritmos de separação de fontes. Com isso, enfatiza-se a diferença entre extração e separação de fontes.

3.2.2 Separação de fontes

Doravante, a notação temporal será ignorada a fim de não sobrecarregar a notação. Uma alternativa à deflação é a estimação conjunta de todas as fontes, realizada por meio de um filtro separador $\mathbf{B} \in \mathbb{R}^{N \times M}$ com saída

$$\mathbf{y} = \mathbf{B}\mathbf{x}. \quad (3.2)$$

Esse filtro possui M entradas e N saídas, sendo, assim, um filtro MIMO, como mostrado na Figura 3.2. Substituindo (3.2) em (1.2), tem-se

$$\mathbf{y} = \mathbf{B}\mathbf{A}\mathbf{s} = \mathbf{G}\mathbf{s},$$

no qual $\mathbf{G} = \mathbf{B}\mathbf{A}$ é o filtro global. A separação de fontes realizada por meio de \mathbf{B} será bem-sucedida se o filtro global \mathbf{G} for um filtro trivial (COMON; JUTTEN, 2010), que no contexto MIMO significa que é possível decompô-lo em uma matriz de permutação \mathbf{P} e matriz de escalas Λ , tal que $\mathbf{G} = \mathbf{P}\Lambda$. Portanto, diz-se que a separação de fontes ocorreu quando \mathbf{s} e \mathbf{y} forem diferentes apenas por fatores de escala e de permutação.

O filtro separador pode ser encontrado através de estimação de máxima verossimilhança¹ (ML) (CARDOSO, 1998). Explorando a hipótese de fontes independentes, será

¹Esse método consiste em buscar os parâmetros de um certo modelo que maximizem a sua verossimilhança (CARDOSO, 1998).

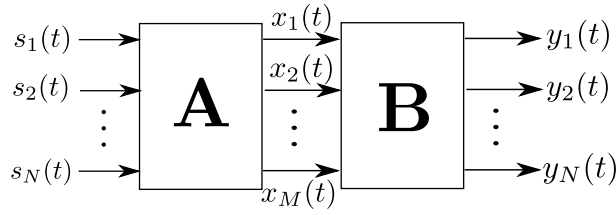


Figura 3.2: Ilustração da composição de N fontes e separação delas por meio de um filtro \mathbf{B} .

mostrado que a estimação ML, é equivalente a um método baseado na Teoria da Informação. Finalmente, será feito o *link* entre esses resultados de Teoria da Informação com a ICA (COMON, 1994).

Considere que T realizações independentes do vetor observação \mathbf{x} distribuídas igualmente de acordo com a função densidade de probabilidade $p_{\mathbf{x}}(\mathbf{x}_t)$, sejam armazenadas em uma matriz $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$. Como as realizações são independentes, então $p_{\mathbf{x}}(\mathbf{X}) = \prod_{t=1}^T p(\mathbf{x}_t)$. Parametriza-se essa probabilidade por meio da verossimilhança, definida como

$$p_{\mathbf{x}|\mathbf{A}}(\mathbf{A}, \mathbf{x}) = |\det \mathbf{A}|^{-1} p_{\mathbf{s}}(\mathbf{A}^{-1} \mathbf{x}), \quad (3.3)$$

no qual $p_{\mathbf{s}}(\cdot)$ é a função densidade de probabilidade **suposta** das fontes. A log-verossimilhança é definida como

$$\begin{aligned} \mathcal{L}_T(\mathbf{A}, \mathbf{x}) &= \frac{1}{T} \log \prod_{t=1}^T p_{\mathbf{x}|\mathbf{A}}(\mathbf{x}_t) \\ &= \frac{1}{T} \sum_{t=1}^T \log p_{\mathbf{x}|\mathbf{A}}(\mathbf{x}_t). \end{aligned} \quad (3.4)$$

Observa-se que a equação (3.4) é a média amostral de $\log p_{\mathbf{x}|\mathbf{A}}(\mathbf{x})$, então, pela Lei dos Grandes Números, essa equação converge para o seu valor esperado (CARDOSO, 1998):

$$\mathcal{L}_{\infty}(\mathbf{A}, \mathbf{x}) = \int p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}|\mathbf{A}}(\mathbf{x}) d\mathbf{x}. \quad (3.5)$$

Fazendo $p_{\mathbf{x}|\mathbf{A}}(\mathbf{x}) = \frac{p_{\mathbf{x}|\mathbf{A}}(\mathbf{x})}{p_{\mathbf{x}}(\mathbf{x})} p_{\mathbf{x}}(\mathbf{x})$, tem-se que:

$$\begin{aligned} \mathcal{L}_{\infty}(\mathbf{A}, \mathbf{x}) &= \int p_{\mathbf{x}}(\mathbf{x}) \log \left[\frac{p_{\mathbf{x}|\mathbf{A}}(\mathbf{x})}{p_{\mathbf{x}}(\mathbf{x})} p_{\mathbf{x}}(\mathbf{x}) \right] d\mathbf{x} \\ &= - \int p(\mathbf{x}) \log \left[\frac{p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{x}|\mathbf{A}}(\mathbf{x})} \right] d\mathbf{x} + \int p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\ &= -K [p_{\mathbf{x}}|p_{\mathbf{x}|\mathbf{A}}] - S(\mathbf{x}), \end{aligned} \quad (3.6)$$

no qual $K[\cdot|\cdot]$ é a divergência de Kullback-Leibler entre duas função densidade de probabilidade e $S(\cdot)$ é a entropia diferencial de uma função densidade de probabilidade.

Pela equação (3.2), pode-se afirmar que $p_{\mathbf{y}}(\mathbf{y}) = |\det \mathbf{B}|^{-1} p_{\mathbf{x}}(\mathbf{B}^{-1} \mathbf{y})$. Considerando a equação (3.3) e a propriedade da invariância da divergência de Kullback-Leibler, tem-se que (CARDOSO, 1998)

$$K [p_{\mathbf{x}} | p_{\mathbf{x}|\mathbf{A}}] = K [p_{\mathbf{y}} | p_{\mathbf{s}}], \quad (3.7)$$

no qual $p_{\mathbf{y}}(\cdot)$ é a função densidade de probabilidade conjunta das saídas do filtro separador. Portanto, maximizar a log-verossimilhança \mathcal{L}_{∞} implica em minimizar o desvio entre as função densidade de probabilidade das saídas do filtro e das fontes, medido por $K [p_{\mathbf{y}} | p_{\mathbf{s}}]$. Isso é um resultado intuitivo, pois é natural que fontes estimadas corretamente sejam distribuídas da mesma forma que as fontes. Baseado nesse resultado, propõe-se então o contraste de máxima verossimilhança:

$$\Upsilon_{ML}(\mathbf{y}) = K [p_{\mathbf{y}} | p_{\mathbf{s}}]. \quad (3.8)$$

Minimizar esse contraste implica em procurar uma matriz inversível $\mathbf{B} \approx \mathbf{A}^{-1}$ tal que a distribuição de $\mathbf{A}^{-1} \mathbf{x}$ seja próxima das função densidade de probabilidade supostas das fontes, maximizando a ML do modelo.

Seja $\tilde{\mathbf{y}}$ um vetor aleatório com a mesma dimensão de \mathbf{y} . As componentes desse vetor são mutuamente independentes e distribuídas da mesma forma que as componentes correspondentes em \mathbf{y} . Prova-se (COMON, 1994) que para \mathbf{s} independente, a equação (3.7) pode ser decomposta em

$$K [p_{\mathbf{y}} | p_{\mathbf{s}}] = K [p_{\mathbf{y}} | p_{\tilde{\mathbf{y}}}] + K [p_{\tilde{\mathbf{y}}} | p_{\mathbf{s}}]. \quad (3.9)$$

Como os vetores $\tilde{\mathbf{y}}$ e \mathbf{s} são independentes, então o segundo termo de (3.9) pode ser escrito como

$$\begin{aligned} K [p_{\tilde{\mathbf{y}}} | p_{\mathbf{s}}] &= \int p_{\tilde{\mathbf{y}}}(\mathbf{u}) \log \left[\frac{p_{\tilde{\mathbf{y}}}(\mathbf{u})}{p_{\mathbf{s}}(\mathbf{u})} \right] d\mathbf{u} \\ &= \int p_{\tilde{\mathbf{y}}}(\mathbf{u}) \sum_{n=1}^N \log \left[\frac{p_{\tilde{y}_n}(\mathbf{u})}{p_{s_n}(\mathbf{u})} \right] d\mathbf{u} \\ &= \sum_{n=1}^N K [p_{\tilde{y}_n} | p_{s_n}] \end{aligned}$$

Então a equação (3.9) pode ser reescrita como

$$K [p_{\mathbf{y}} | p_{\mathbf{s}}] = I(\mathbf{y}) + \sum_{n=1}^N K [p_{\tilde{y}_n} | p_{s_n}], \quad (3.10)$$

no qual $I(\mathbf{y})$ é a informação mútua (2.52) de \mathbf{y} , medindo o seu grau de independência. Por outro lado, o segundo termo mede o desvio marginal das distribuições das componentes

de \mathbf{y} e \mathbf{s} . Como $p_{\tilde{\mathbf{y}}}(\mathbf{u}) = \prod_{n=1}^N p_{\tilde{y}_n}(u_n)$, então a informação mútua de \mathbf{y} pode ser reescrita como

$$\begin{aligned} I(\mathbf{y}) &= \int p_{\mathbf{y}}(\mathbf{u}) \log \left[\frac{p_{\mathbf{y}}(\mathbf{u})}{\prod_{n=1}^N p_{\tilde{y}_n}(u_n)} \right] d\mathbf{u} \\ &= \int p_{\mathbf{y}}(\mathbf{u}) \log p_{\mathbf{y}}(\mathbf{u}) d\mathbf{u} - \sum_{n=1}^N \int p_{\mathbf{y}}(\mathbf{u}) \log p_{\tilde{y}_n}(u_n) d\mathbf{u} \\ &= \sum_{n=1}^N S(y_n) - S(\mathbf{y}). \end{aligned} \quad (3.11)$$

Conclui-se pela equação (3.11) que minimizar a informação mútua equivale a minimizar a entropia marginal e maximizar a entropia conjunta de \mathbf{y} .

A estimação ML é bastante indicada quando se tem conhecimento *a priori* da densidade de probabilidade das fontes, pois além de explorar a hipótese de independência das fontes, procura-se minimizar os desvios entre as distribuições das fontes e das saídas do filtro. Entretanto, muitas vezes não se conhece as densidades das fontes, sendo isso uma restrição do contraste ML. Uma alternativa consiste em utilizar apenas a informação mútua como contraste:

$$\Upsilon_{IM}(\mathbf{y}) = I(\mathbf{y}). \quad (3.12)$$

Uma operação muito importante na separação cega de fontes se chama **branqueamento**, que consiste em cancelar as covariâncias cruzadas de um vetor aleatório, diagonalizando a sua matriz de covariância. Suponha que a matriz de composição admita uma SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, no qual $\mathbf{U} \in \mathbb{R}^{M \times M}$ e $\mathbf{V} \in \mathbb{R}^{N \times N}$ são matrizes ortonormais e $\mathbf{\Sigma} \in \mathbb{R}^{M \times N}$ é uma matriz diagonal. Então considerando que $\mathbb{E}[\mathbf{s}\mathbf{s}^T] = \mathbf{I}$, então a matriz de covariância do vetor \mathbf{x} é:

$$\begin{aligned} \mathbf{C}_{\mathbf{x}} &= \mathbb{E}[\mathbf{x}\mathbf{x}^T] \\ &= \mathbb{E}[\mathbf{A}\mathbf{s}\mathbf{s}^T\mathbf{A}^T] \\ &= \mathbb{E}[\mathbf{A}\mathbf{A}^T] \\ &= \mathbb{E}[\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T] \\ &= \mathbb{E}[\mathbf{U}\mathbf{\Sigma}\mathbf{U}^T\mathbf{\Sigma}^T] \\ &= \mathbf{W}\mathbf{W}^T, \end{aligned} \quad (3.13)$$

no qual $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}$. Essa matriz pode ser obtida de várias formas, como por exemplo, decomposição de Cholesky ou *Principal Component Analysis* (PCA). Agora a versão

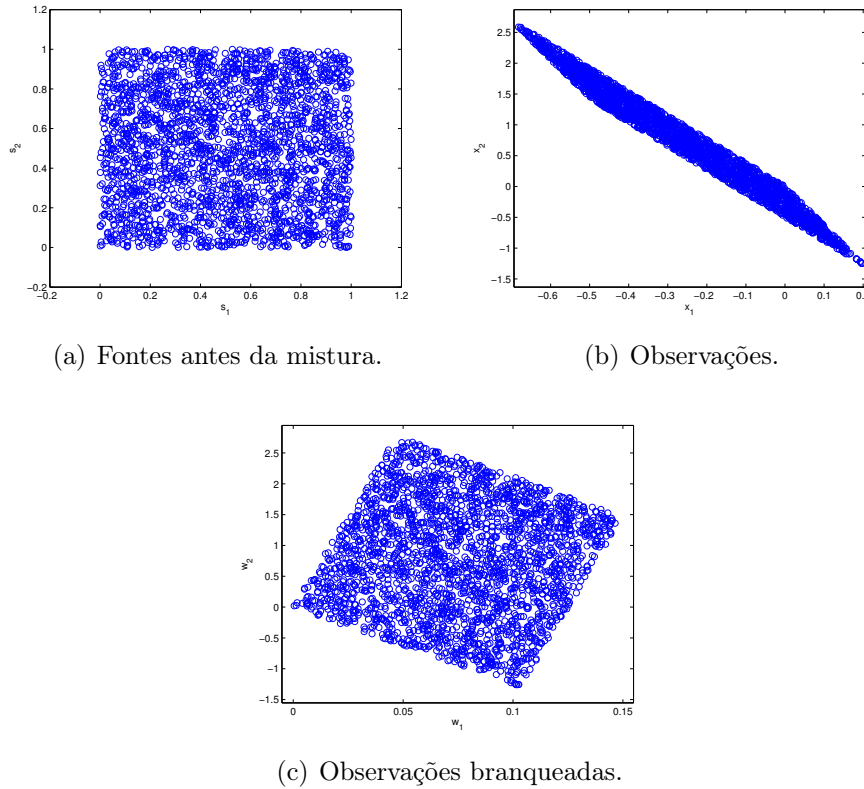


Figura 3.3: Efeito do branqueamento nas observações.

“branca” de \mathbf{x} é definida como

$$\begin{aligned}
 \tilde{\mathbf{x}} &= \mathbf{W}^T \mathbf{x} \\
 &= (\mathbf{U}\boldsymbol{\Sigma}) (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T) \mathbf{s} \\
 &= \mathbf{V}^T \mathbf{s}.
 \end{aligned} \tag{3.14}$$

Dessa forma, o vetor $\tilde{\mathbf{x}}$ possui matriz de covariância diagonal.

A fim de ilustração, mostra-se na Figura 3.3(a) a relação entre duas fontes uniformemente distribuídas. Após misturá-las através de um sistema instantâneo linear, obtém-se a Figura 3.3(b). A relação entre os sinais de observação brancos obtidos por meio da SVD são mostrados na Figura 3.3(c). Nota-se que as observações brancas são as fontes originais **rotacionadas**. Conclui-se então através desse exemplo e da equação (3.14) que o branqueamento reduz a composição de fontes em uma rotação do vetor \mathbf{s} , visto que este vetor e $\tilde{\mathbf{x}}$ possuem matriz de covariância igual à matriz identidade. Dessa forma, o filtro extrator \mathbf{B} pode ser estimado como uma rotação que maximiza a independência da sua saída \mathbf{y} .

O contraste baseado na informação mútua (3.12) pode ser simplificado ao se considerar um vetor de observação branco. A informação mútua pode ser representada em termos de entropia, como mostrado na equação (3.11), no qual o segundo termo desta equação

é constante ao ser processado por transformações ortonormais (CARDOSO, 1998), como \mathbf{B} , por exemplo. Visto que o segundo termo de (3.11) é constante quando \mathbf{x} é branco, a informação mútua pode ser contabilizada apenas em termos de do contraste de entropia marginal (EM)

$$\Upsilon_{EM}(\mathbf{y}) = \sum_{n=1}^N S(y_n). \quad (3.15)$$

Os contrastes apresentados até agora necessitam de uma boa estimação das funções de densidade de probabilidade. Isso pode ser consideravelmente custoso na prática. Por isso, em (COMON, 1994; CARDOSO, 1999) foram propostas aproximações dessas funções por estatísticas de ordem superior por meio da expansão de Edgeworth. A equação (2.56), que relaciona a informação mútua com a negentropia, pode ser aproximada da seguinte forma:

$$I(\mathbf{y}) \approx J(\mathbf{y}) - \frac{1}{48} \sum_{i=1}^N 4\kappa_{y_i(3)}^2 + \kappa_{y_i(4)}^2 + 7\kappa_{y_i(3)}^4 - 6\kappa_{y_i(3)}^2 \kappa_{y_i(4)}, \quad (3.16)$$

no qual $\kappa_{y_i(n)}$ é o cumulante marginal de ordem n da variável aleatória y_i . Como a negentropia é invariante sob transformadas inversíveis (COMON; JUTTEN, 2010), então a aproximação da informação mútua pode ocorrer apenas em termos dos cumulantes. Em vista disso, em (COMON, 1994) foi proposto o contraste

$$\Upsilon_{HOS}(\mathbf{y}) = \sum_{i=1}^N 4\kappa_{y_i(3)}^2 + \kappa_{y_i(4)}^2 + 7\kappa_{y_i(3)}^4 - 6\kappa_{y_i(3)}^2 \kappa_{y_i(4)}. \quad (3.17)$$

A maximização dos cumulantes marginais de \mathbf{y} resulta na maximização da independência devido à relação (3.16). Os cumulantes podem ser estimados através das técnicas apresentadas na subseção (2.3.2).

Até agora foi mostrado alguns contrastes que indicam a qualidade da separação de fontes. Nas seções seguintes serão apresentados métodos que otimizam esses critérios, resultando na SCF.

3.3 Métodos algébricos

Esta seção visa apresentar alguns métodos algébricos para resolver o problema da SCF através de cumulantes, relacionados com independência e não-Gaussianidade de vetores aleatórios. Como visto no Capítulo 2, os momentos de ordem superior podem ser armazenados em uma estrutura matemática chamada tensor. Também sabe-se que vetores aleatórios independentes possuem cumulantes cruzados nulos, de tal forma que o seu tensor de cumulantes é diagonal. A diagonalização do tensor de cumulantes de segunda ordem, a matriz de covariância, implica na decorrelação das variáveis, enquanto a diagonalização do tensor de cumulantes de ordem superior acarreta a independência das

variáveis. Uma forma de se diagonalizar os cumulantes de segunda ordem se dá pelos métodos de branqueamento comentados na Seção anterior. Como pretende-se buscar as componentes independentes, logo deve-se diagonalizar os tensores de cumulantes de ordem superior. Uma estratégia bastante utilizada para esse fim, consiste em diagonalizar o tensor de cumulantes do vetor de observação branco $\tilde{\mathbf{x}}$, pois o número de parâmetros a serem estimados é diminuído, visto que necessita-se então estimar apenas uma rotação, como ilustrado pela equação (3.14). O branqueamento também melhora o condicionamento numérico do problema (COMON; JUTTEN, 2010). Porém tudo isso é feito adicionando-se um certo custo computacional na resolução do problema.

Dado a relação entre $\tilde{\mathbf{x}}$ e \mathbf{s} pela equação (3.14), seus tensores de cumulantes de ordem K relacionam-se através da propriedade da multilinearidade (2.35) da seguinte forma:

$$\mathcal{C}_{\tilde{\mathbf{x}}} = \mathcal{C}_{\mathbf{s}} \times_1 \mathbf{V}^T \times_2 \mathbf{V}^T \dots \times_K \mathbf{V}^T, \quad (3.18)$$

No qual $\mathcal{C}_{\mathbf{s}} \in \mathbb{R}^{\overbrace{N \times N \times \dots \times N}^{K \text{ vezes}}}$ e $\mathcal{C}_{\tilde{\mathbf{x}}} \in \mathbb{R}^{\overbrace{N \times N \times \dots \times N}^{K \text{ vezes}}}$. Como $\mathcal{C}_{\mathbf{s}}$ é diagonal, então a matriz \mathbf{V}^T **diagonaliza** $\mathcal{C}_{\tilde{\mathbf{x}}}$, resultando em um vetor independente, ou seja, realizando a separação de fontes. A diagonalização exata de um tensor qualquer é um problema difícil a ser tratado. Dessa forma, adota-se a diagonalização aproximada de tensores, buscando uma matriz \mathbf{V} que otimize um certo critério resultando na separação de fontes. Nas Subseções seguintes serão apresentadas algumas abordagens para se aproximar o tensor diagonal $\mathcal{C}_{\mathbf{s}}$.

3.3.1 Maximização do traço

De acordo com o contraste (3.17), busca-se maximizar os cumulantes marginais do tensor de cumulantes, localizados na diagonal dele. Portanto, maximizar o traço desse tensor é equivalente à maximização do contraste. Os métodos *CO*ntrast *MA*ximization 1 (**COM1**) e *CO*ntrast *MA*ximization 2 (**COM2**) utilizam essa ideia e foram propostos respectivamente em (COMON; MOREAU, 1997) e (COMON, 1994). Na maioria das situações, esses métodos apresentam uma performance equivalente, porém a performance do COM1 decresce bastante quando há uma fonte Gaussiana (COMON; CHEVALIER; CAPDEVIELLE, 1997).

A equação (3.18) pode ser reescrita de uma forma que facilite a visualização da operação de diagonalização. Sabendo que $\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T \mathbf{I}$ e multiplicando os tensores dos dois lados por \mathbf{V} em cada modo, obtém-se:

$$\mathcal{C}_{\mathbf{s}} = \mathcal{C}_{\tilde{\mathbf{x}}} \times_1 \mathbf{V} \times_2 \mathbf{V} \dots \times_K \mathbf{V}. \quad (3.19)$$

Os métodos COM1 e COM2 buscam a matriz \mathbf{V} que maximizem o traço do tensor de

cumulantes. Dessa forma, os critérios otimizados por esses métodos são respectivamente

$$\Phi_{COM1}(\mathbf{V}) = \sum_{n=1}^K c_{nn\dots n}^{(s)} \quad (3.20)$$

e

$$\Phi_{COM2}(\mathbf{V}) = \sum_{n=1}^K |c_{nn\dots n}^{(s)}|^2. \quad (3.21)$$

O método de diagonalização baseado em matrizes de Jacobi é bastante conhecido no caso matricial. A sua generalização para tensores é bastante semelhante ao caso matricial, por isso esse método será primeiramente apresentado no caso matricial. Considere $\mathbf{D} = \mathbf{C}\mathbf{A}\mathbf{C}^T$, no qual \mathbf{A} é a matriz que se deseja diagonalizar, \mathbf{C} é uma matriz ortogonal e \mathbf{D} é a versão diagonal de \mathbf{A} . A matriz \mathbf{C} pode ser decomposta em um produtório de matrizes de Jacobi (GOLUB; LOAN, 2012), representadas por

$$\mathbf{Q}_{pq}(a, b) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & a_{pp} & \dots & b_{pq} & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & a_{qp} & \dots & b_{qq} & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}. \quad (3.22)$$

Normalmente se parametriza as variáveis a e b em função de um ângulo θ da seguinte forma

$$\begin{aligned} a_{pp} &= \cos \theta \\ a_{qp} &= \cos \theta \\ b_{pq} &= -\text{sen} \theta \\ b_{qq} &= \text{sen} \theta. \end{aligned}$$

O método de diagonalização busca essas matrizes de forma que uma função custo associada à diagonalização de \mathbf{A} seja minimizada. A mesma ideia é utilizada no caso tensorial. Nesse caso, a matriz \mathbf{V} na equação (3.19) é um produtório de matrizes de Jacobi obtidas através da maximização dos critérios (3.20) ou (3.21).

3.3.2 Diagonalização conjunta de matrizes

Uma outra família de métodos algébricos consiste na diagonalização conjunta dos *slices* do tensor de cumulantes. Será mostrado que isso também leva à estimação da matriz \mathbf{V} .

Inicialmente, identifica-se a equação (3.18) como um modelo PARAFAC (2.4). De fato, a diagonalização por meio do PARAFAC é uma abordagem equivalente àquela da seção anterior (COMON; JUTTEN, 2010). Portanto, a representação em elementos da equação (3.18) pode ser escrita da seguinte forma:

$$(\mathcal{C}_{\bar{\mathbf{x}}})_{i_1 i_2 \dots i_n} = \sum_{r=1}^R (\mathcal{C}_{\mathbf{s}})_{i_1 i_2 \dots i_n} v_{i_1 r} v_{i_2 r} \dots v_{i_n r}. \quad (3.23)$$

Os *slices* $(\mathcal{C}_{\bar{\mathbf{x}}})_{::i_3 \dots i_n} \in \mathbb{R}^{N \times N}$ possuem a seguinte estrutura:

$$\begin{aligned} (\mathcal{C}_{\bar{\mathbf{x}}})_{::i_3 \dots i_n} &= \sum_{r=1}^R [(\mathcal{C}_{\mathbf{s}})_{::i_3 \dots i_n} v_{i_3 r} v_{i_4 r} \dots v_{i_n r}] \mathbf{v}_r \circ \mathbf{v}_r \\ &= \sum_{r=1}^R [(\mathcal{C}_{\mathbf{s}})_{::i_3 \dots i_n} v_{i_3 r} v_{i_4 r} \dots v_{i_n r}] \mathbf{v}_r \mathbf{v}_r^T, \end{aligned} \quad (3.24)$$

Reagrupando os vetores \mathbf{v}_r para $r = 1, \dots, R$ na forma de matriz, tem-se $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_R]_{N \times N}$. Também define-se a matriz diagonal $\mathbf{D}_{i_3 \dots i_n} = (\mathcal{C}_{\mathbf{s}})_{::i_3 \dots i_n} v_{i_3 r} v_{i_4 r} \dots v_{i_n r}$. Assim, a equação (3.24) é escrita como

$$(\mathcal{C}_{\bar{\mathbf{x}}})_{::i_3 \dots i_n} = \mathbf{V} \mathbf{D}_{i_3 \dots i_n} \mathbf{V}^T. \quad (3.25)$$

Como \mathbf{V} é ortogonal, então a equação (3.25) é uma *Eigenvalue Decomposition* (EVD). Observa-se que a matriz \mathbf{V} é a mesma para todos os slices. Quando há ruído, entretanto, isso pode não ocorrer. Dessa forma, uma diagonalização conjunta dos slices de $\mathcal{C}_{\bar{\mathbf{x}}}$ resulta em uma boa aproximação de \mathbf{V} . Nota-se que seria necessário diagonalizar N^{K-2} *slices*. Em (CARDOSO; SOULOUMIAC, 1993), foi proposto diagonalizar apenas as automatrizes mais significativas do tensor $\mathcal{C}_{\bar{\mathbf{x}}}$, diminuindo, assim, a quantidade de matrizes a serem diagonalizadas. Esse método foi chamado de *Joint Approximate Diagonalization of*

Eigenmatrices (JADE).

Entrada: Vetor de observação \mathbf{x}

Saída: Vetor de fontes independentes \mathbf{y}

1. Pré-Branqueamento
 - Calcular $\tilde{\mathbf{x}} \leftarrow \mathbf{W}^T \mathbf{x}$.
2. Calcular tensor de cumulantes de $\tilde{\mathbf{x}}$ utilizando as técnicas da Seção 2.3.2.
3. Obter \mathbf{V} ótimo de acordo com algum algoritmo (JADE, COM1, COM2, etc).
4. Calcular estimação do filtro de composição $\mathbf{B}^{-1} = \mathbf{A} \leftarrow \mathbf{WV}$.
5. Obter fontes independentes $\mathbf{y} \leftarrow \mathbf{B}\tilde{\mathbf{x}}$.

Algoritmo 2: Algoritmo para SCF utilizando métodos algébricos.

Resume-se o procedimento de separação cega de fontes utilizando métodos algébricos no Algoritmo 2.

3.4 Algoritmos iterativos

Na Seção anterior foram apresentados métodos que realizam a SCF por meio de soluções algébricas, que decompõem a matriz de rotação desconhecida em matrizes de Jacobi, e as calcula. É difícil analisar a convergência desse tipo de algoritmo devido à forma que a rotação ótima é calculada (HYVÄRINEN; KARHUNEN; OJA, 2004). Além disso, esses métodos não rastreiam variações rápidas no sistema de composição, pois os cálculos apresentados podem apresentar uma certa latência devido à quantidade de amostras necessárias para se estimar os tensores de cumulantes.

Uma abordagem alternativa que evita esses problemas é o uso de soluções iterativas baseados em algoritmos de otimização. Procura-se otimizar funções de contraste utilizando, por exemplo, os métodos vistos na Seção 2.5, permitindo uma análise de convergência. Por outro lado, a performance desses algoritmos depende fortemente de alguns parâmetros como: fator de passo e inicialização. Em alguns métodos, o fator de passo ótimo é calculado em cada iteração, e a inicialização ótima pode ser obtida através do contexto de aplicação. Serão apresentados três algoritmos iterativos para SCF:

- FastICA (HYVARINEN, 1999), algoritmo bastante popular devido à sua simplicidade e eficiência.
- RobustICA (ZARZOSO; COMON, 2010), solução relativamente nova baseada no cálculo ótimo do fator de passo.

É importante salientar que será assumido que os vetores de observação foram branqueados, adotando-se, doravante, apenas a notação $\tilde{\mathbf{x}}$.

3.4.1 FastICA

FastICA é um algoritmo eficiente e popular para SCF proposto por Aapo Hyvärinen (HYVARINEN, 1999). Basicamente, ele maximiza a independência das componentes estimadas maximizando a informação mútua por meio da negentropia dessas componentes, relacionadas pela equação (2.56). Hyvärinen propôs em (HYVÄRINEN, 1998) a seguinte expansão da negentropia:

$$J(y_i) \approx c \{ \mathbb{E} [G(y_i)] - \mathbb{E} [G(x_g)] \}^2, \quad (3.26)$$

no qual $G(\cdot)$ é uma função não quadrática, c é uma constante qualquer, e x_g é uma variável aleatória Gaussiana de média nula e variância unitária. Como a aproximação (3.26) se refere somente a um elemento de \mathbf{y} , então pode-se considerar essa equação como uma função objetivo para a extração de uma fonte, levando ao modelo MISO. Observa-se que maximizar $\mathbb{E} [G(y_i)]$ acarreta diretamente na maximização de $J(y_i)$, sendo portanto a função objetivo desse algoritmo. Portanto, busca-se o vetor \mathbf{w} , com saída $y_i = \mathbf{w}^T \tilde{\mathbf{x}}$, que maximize a função objetivo

$$\Upsilon_{\text{FAST}}(\mathbf{w}) = \mathbb{E} [G(y_i)] \quad (3.27)$$

com a restrição $\|\mathbf{w}\| = 1$, resolvendo problemas de escala e valores muito grandes para y . A performance do algoritmo depende da escolha da função $G(\cdot)$. Busca-se funções que tornem não tornem o algoritmo computacionalmente complexo. Uma função robusta e adaptada a fontes super-Gaussianas² proposta em (HYVARINEN, 1999) é

$$G(u) = -\log \cosh(\alpha u), \quad (3.28)$$

no qual $1 \leq \alpha \leq 2$. Para uso futuro, a sua derivada é $g(u) = \tanh(\alpha u)$.

Inicialmente, o algoritmo supõe que as observações foram branqueadas. De acordo com as condições de Karush-Kuhn-Tucker (NOCEDAL; WRIGHT, 2006), o ponto ótimo de (3.27), sob a restrição $\|\mathbf{w}\| = 1$, é:

$$\mathbb{E} [\tilde{\mathbf{x}} g(\mathbf{w}^T \tilde{\mathbf{x}})] - \beta \mathbf{w} = 0, \quad (3.29)$$

no qual $\beta = \mathbb{E} [\mathbf{w}_0^T \tilde{\mathbf{x}} g(\mathbf{w}^{*T} \tilde{\mathbf{x}})]$ com \mathbf{w}^* sendo o valor de \mathbf{w} no ponto ótimo. Aplicando o

²A maior parte de sinais independentes na prática são, de fato, super-Gaussianas (BELL; SEJNOWSKI, 1995)

método de Newton para resolver (3.29), obtém-se a equação de atualização do filtro:

$$\mathbf{w}^+ = \mathbb{E} [\tilde{\mathbf{x}}g(\tilde{\mathbf{x}}^T \mathbf{w}(k))] - \mathbb{E} [g'(\tilde{\mathbf{x}}^T \mathbf{w}(k))] \mathbf{w}(k) \quad (3.30)$$

$$\mathbf{w}(k+1) = \mathbf{w}^+ / \|\mathbf{w}^+\|. \quad (3.31)$$

Diz-se que o algoritmo convergiu quando $|1 - \mathbf{w}(k)^T \mathbf{w}(k+1)| < \epsilon$, no qual ϵ é um valor pequeno.

Essa abordagem baseada na maximização da negentropia foi uma das primeiras a serem propostas para o FastICA. Outras abordagens podem ser encontradas em (BINGHAM; HYVÄRINEN, 2000; HYVARINEN, 1997; HYVARINEN, 1999), dentre elas há uma abordagem bastante popular baseada na maximização do cumulante de quarta ordem normalizado, a curtose (2.21). De fato, maximizar a curtose é equivalente a maximizar

$$\Upsilon_{\text{CURT}}(\mathbf{w}) = \mathbb{E}[y^4] \quad (3.32)$$

quando as variáveis são reais. Em (COMON, 1994), mostra-se que (3.32) é um contraste. Procura-se um vetor \mathbf{w} que maximize o critério $|\Upsilon_{\text{CURT}}(\mathbf{w})|$, porém é preciso garantir que essa maximização não resulte em valores de y muito grandes assim como na abordagem baseada na negentropia. Para tanto, impõe-se a restrição $\|\mathbf{w}\| = 1$. Resolvendo esse problema de otimização por meio dos multiplicadores de Lagrange, tem-se que os pontos estacionários do critério (3.32) é dado por

$$\mathbb{E} [|\mathbf{w}^T \tilde{\mathbf{x}}|^2 \mathbf{x} \mathbf{x}^T] \mathbf{w} = \lambda \mathbf{w}, \quad (3.33)$$

no qual o multiplicador λ é uma constante definida de acordo com (HYVÄRINEN; KARHUNEN; OJA, 2004). Utiliza-se o método de Newton para calcular o \mathbf{w} ótimo, obtendo-se as seguintes equações de atualização do filtro

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \frac{1}{3} \mathbb{E} [\tilde{\mathbf{x}} (\mathbf{w}^T \tilde{\mathbf{x}})^3]. \quad (3.34)$$

O FastICA é bastante popular devido à sua convergência cúbica, graças ao fato de não haver necessidade de se escolher um fator de passo e não necessitar pela sua capacidade de ser implementado de forma paralela e não necessitar de muitos recursos de memória

(HYVÄRINEN; KARHUNEN; OJA, 2004).

Entrada: Vetor de observação $\tilde{\mathbf{x}}$

Saída: Uma fonte independente y_n

1. Inicialização

- Pré-branqueamento $\tilde{\mathbf{x}} \leftarrow \mathbf{W}^T \mathbf{x}$.
- $k \leftarrow 0$.
- Inicializar $\mathbf{w}(k)$ de acordo com algum critério.

2. Repita até a convergência:

- Calcular equação (3.30).
- Calcular equação (3.31).
- $k \leftarrow k + 1$.

3. Obter fonte independentes $y_n \rightarrow \mathbf{w}(k)^T \tilde{\mathbf{x}}$.

Algoritmo 3: Algoritmo FastICA.

3.4.2 RobustICA

O FastICA encontra o filtro ótimo utilizando o método de Newton. Em (ZARZOSO; COMON, 2010) foi proposto uma alternativa que otimiza o seguinte critério, baseado na curtose:

$$\Upsilon_{\text{ROBUST}}(\mathbf{w}) = \frac{\mathbb{E}\{|y|^4\} - 2\mathbb{E}\{|y|^2\}^2 - |\mathbb{E}\{y^2\}|^2}{\mathbb{E}\{|y|^2\}^2}. \quad (3.35)$$

A ideia do RobustICA é a otimização de (3.35) por meio de um método de busca linear. Nele o fator de passo ótimo μ_{opt} na direção \mathbf{g} (gradiente de Υ_{ROBUST} , por exemplo) é calculado da seguinte forma:

$$\mu_{\text{opt}} = \arg \max_{\mu} |\Upsilon_{\text{ROBUST}}(\mathbf{w} + \mu \mathbf{g})|. \quad (3.36)$$

Normalmente métodos baseados em busca linear podem ser complexos do ponto de vista computacional. Entretanto, se o critério de otimização puder ser expresso como um polinômio em μ então o ponto ótimo global pode ser obtido analiticamente pelas raízes desse polinômio. O RobustICA explora essa propriedade utilizando o contraste (3.35), que pode ser expresso como um polinômio, através das equações mostradas no apêndice de (ZARZOSO; COMON, 2010). Dessa forma, o fator de passo ótimo μ_{opt} como sendo a raiz desse polinômio que maximiza (3.36). Baseado nessa ideia, o RobustICA é resumido no Algoritmo 4.

Quando se conhece o sinal algébrico da curtose das fontes desejadas, pode-se modificar

o critério (3.36) para que o algoritmo extraia primeiramente a fonte desejada da seguinte forma:

$$\mu_{\text{opt}} = \arg \max_{\mu} \quad \epsilon \Upsilon_{\text{ROBUST}}(\mathbf{w} + \mu \mathbf{g}), \quad (3.37)$$

no qual $\epsilon \in \{-1, 1\}$ indica o sinal algébrico da fonte desejada.

Entrada: Vetor de observação $\tilde{\mathbf{x}}$

Saída: Uma fonte independente y_n

1. Inicialização

- $k \leftarrow 0$.
- Inicializar $\mathbf{w}(k)$ de acordo com algum critério.

2. Repita até a convergência:

- Calcular direção $\mathbf{g}(k)$.
- Obter polinômio em μ (detalhado em (ZARZOSO; COMON, 2010)).
- Selecionar raiz μ_{opt} do polinômio que maximize o contraste Υ_{CURT}
- $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k) + \mu_{\text{opt}} \mathbf{g}(k)$
- $\mathbf{w}(k+1) \leftarrow \mathbf{w}(k+1) / \|\mathbf{w}(k+1)\|$
- $k \leftarrow k+1$.

3. Obter fonte independentes $y_n \rightarrow \mathbf{w}(k)^T \tilde{\mathbf{x}}$.

Algoritmo 4: Algoritmo RobustICA.

Através de experimentos, mostrou-se em (ZARZOSO; COMON, 2010) que esse algoritmo é robusto à escolha da inicialização e a pontos de sela, especialmente quando a quantidade de amostras é bastante reduzida. Como esse método não requer o pré-branqueamento e o fator de passo ótimo é obtido analiticamente, esse método possui uma melhor eficiência em relação à complexidade computacional se comparado ao popular FastICA. Entretanto, assim como todos os métodos baseados em cumulantes de ordem superior, o RobustICA é sensível a *outliers*. Até o momento, não existem trabalhos que explorem outros critérios de otimização nem modelos convolutivos com a ideia do RobustICA.

3.5 Novos métodos e perspectivas

A separação fontes de forma não-supervisionada é possível quando se tem conhecimento *a priori* das características das fontes, caso contrário, estimação correta das fontes torna-se um problema muito complexo ou até impossível. Felizmente, dependendo do contexto da aplicação, pode-se assumir várias características ou restrições sobre as fontes. A

independência estatística é uma restrição matematicamente forte mas plausível em várias aplicações. Como dito anteriormente, um dos primeiros trabalhos ao supor independência entre as fontes foi (HÉRAULT; JUTTEN; ANS, 1985). Porém foi em (COMON, 1994) que formalizou-se o ICA e adaptou-se a noção de estatística contraste no contexto de separação de fontes. Nos anos seguintes, vários métodos algébricos e iterativos baseados no ICA foram propostos. Paralelamente, técnicas de SCF baseadas em estatísticas de segunda ordem foram desenvolvidas. Por exemplo, o método *Second-Order Blind Identification* (BELOUCHRANI *et al.*, 1997) supõe que as fontes são coloridas³ e possuem espectro distintos. Dessa forma, é possível estimar a matriz de composição através da diagonalização conjunta das matrizes de covariância das observações com *lags* não-nulos. Outros trabalhos exploram características das fontes como estacionaridade (SHAMSUNDER; GIANNAKIS, 1997) e parcimônia no tempo ou frequência (O'GRADY; PEARLMUTTER; RICKARD, 2005). Em (GERVEN; COMPERNOLLE, 1995), duas fontes foram separadas por decorrelação assumindo que o sistema de composição é um filtro FIR com fase mínima, ou seja, o filtro e o seu sistema inverso são causais e estáveis. Nos últimos anos, a comunidade científica propôs diversos novos métodos de separação de fontes baseados em outras características e modelos das fontes. Em (HYVÄRINEN, 2013; CHABRIEL *et al.*, 2014) os autores comentam algumas novas estratégias em relação a ICA e a separação de fontes por meio de diagonalização conjunta.

Nessa seção serão apresentados algumas dessas novas técnicas que possuem um grande potencial prático, atraindo cada vez mais interesse da comunidade científica. Primeiramente será apresentado a ideia de métodos que exploram a não-negatividade de fontes através de ferramentas matriciais. Em seguida será apresentado uma técnica de separação de fontes baseada numa decomposição tensorial chamada *Block Term Decomposition* (BTD) e finalmente uma nova abordagem de separação de fontes baseada em uma estatística chamada carrelação será apresentada.

3.5.1 Decomposição não-negativa de matrizes

Frequentemente em processamento de sinais trabalha-se com sinais que possuem sentido apenas quando seus valores são não-negativos. Uma imagem digital, por exemplo, pode ser considerada como um sinal multidimensional com pixels não-negativos. Em Economia, algumas variáveis como volumes e preços são naturalmente não-negativas. Dessa forma, é possível realizar a separação de fontes não-negativas sem assumir outras informações *a priori*, como independência.

Uma forma de se explorar a não-negatividade de fontes se dá por meio da *non-negative matrix factorization* (PAATERO; TAPPER, 1994; WANG; PLUMBLEY, 2005). Para utilizar essa decomposição no contexto de SCF, reescreve-se o modelo instantâneo linear

³Sinais correlacionados no tempo.

como

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (3.38)$$

no qual $\mathbf{X} = [\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(T-1)]_{M \times T}$ é a matriz de observação com T amostras temporais, $\mathbf{S} = [\mathbf{s}(0), \mathbf{s}(1), \dots, \mathbf{s}(T-1)]_{N \times T}$ é a matriz das fontes com T amostras temporais e $\mathbf{A} \in \mathbb{R}^{M \times N}$ é a matriz de composição. Os elementos dos fatores \mathbf{A} e \mathbf{S} são todos **não-nulos**. Dependendo da aplicação, esse modelo pode ter sua estrutura modificada ou até possuir outras restrições sobre as fontes, como independência. Por exemplo, existem trabalhos que estendem o modelo (3.38) para o caso multilinear a fim de analisar sinais de EEG (LEE *et al.*, 2007; WANG *et al.*, 2008).

Em (LEE; SEUNG, 2001), o autor propôs algoritmo iterativo para calcular a decomposição (3.38). Definiu-se a seguinte função custo:

$$\Phi_{\text{NMF}}(\mathbf{A}, \mathbf{S}) = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2. \quad (3.39)$$

Mostra-se que essa função é convexa em relação a cada variável mas não em relação às duas simultaneamente. Então busca-se resolver o seguinte problema de otimização:

$$\mathbf{A}^*, \mathbf{S}^* = \arg \min_{\mathbf{A}, \mathbf{S}} \Phi_{\text{NMF}}. \quad (3.40)$$

A estratégia adotada para resolver esse problema é o *Alternating Least-Squares* (ALS), no qual \mathbf{A} e \mathbf{S} são calculados de uma forma alternada, impondo a restrição de elementos não-nulos em cada iteração. Sabe-se que as soluções do problema (3.40) em relação a \mathbf{A} e \mathbf{S} , respectivamente, são

$$\hat{\mathbf{A}} = \mathbf{X}\hat{\mathbf{S}}^\dagger \quad (3.41)$$

e

$$\hat{\mathbf{S}} = \hat{\mathbf{A}}^\dagger \mathbf{X}. \quad (3.42)$$

O procedimento de separação cega de fontes utilizando a NMF é resumido no Algoritmo

5.

Entrada: Matriz de observação \mathbf{X} .

Saída: Estimação da Matriz de composição $\hat{\mathbf{A}}$ e da matriz de fontes $\hat{\mathbf{S}}$.

1. Inicialização

- $\hat{\mathbf{A}} \leftarrow \mathbf{I}$.
- $\hat{\mathbf{S}} \leftarrow \mathbf{X}$.

2. Repita até a convergência:

- Impor a restrição de não-negatividade nas matrizes $\hat{\mathbf{A}}$ e $\hat{\mathbf{S}}$.
- Atualizar $\hat{\mathbf{A}}$ de acordo com a equação (3.41).
- Atualizar $\hat{\mathbf{S}}$ de acordo com a equação (3.42).

Algoritmo 5: Algoritmo NMF-ALS.

Apesar de conceitualmente simples, não há garantias de que esse algoritmo convirja para o mínimo global nem para um ponto estacionário. Outros algoritmos podem ser encontrados em (PAATERO, 1997; CICHOCKI *et al.*, 2009).

3.5.2 Block Term Decomposition

Como visto na Seção 3.3, as ferramentas multilineares são bem adaptadas à SCF. Foi visto que a separação de fontes pode ser realizada por meio da diagonalização do tensor de cumulantes de ordem superior. Por outro lado, algoritmo BSS-BTD (LATHAUWER, 2011; LATHAUWER, 2008), realiza a SCF assumindo que as fontes podem ser modeladas da seguinte forma:

$$s_n(t) = \sum_{l_n=1}^{L_n} c_{l_n,t} z_{l_n,t}^t, \quad 0 \leq t \leq (T-1), 1 \leq n \leq N, \quad (3.43)$$

no qual L_n é o número de termos exponenciais, $c_{l_n,t}$ é o coeficiente do l_n -ésimo termo nos elementos da n -ésima fonte e $z_{l_n,t}^t$ é a base da função exponencial associada ao l_n -ésimo termo de $s_n(t)$. Apesar desse modelo ser bastante restritivo, ele é bem adaptado a sinais oscilatórios, que são naturais em processamento de sinais biomédicos (HUNYADI *et al.*, 2014).

Nesta subseção é considerado o modelo instantâneo linear na forma matricial (3.38) sem restrições de não-negatividade. Mostra-se em (LATHAUWER, 2011) uma forma de construir um tensor de terceira ordem $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ a partir da matriz de observação \mathbf{X}

de forma que a decomposição *Block Term Decomposition* (BTD)

$$\mathcal{T} = \sum_{n=1}^N \mathbf{E}_n \circ \mathbf{c}_n, \quad n = 1, \dots, N, \quad (3.44)$$

forneça a separação de fontes, no qual $\mathbf{E}_n \in \mathbb{R}^{I_1 \times I_2}$ é uma matriz de Hankel⁴ de *rank* L_r associada a uma fonte, $\mathbf{c}_n \in \mathbb{R}^{I_3 \times 1}$ é um vetor associado à matriz de composição \mathbf{A} e N é o número de fontes, facilmente calculado como o *rank* de \mathbf{Y} . A construção do tensor \mathcal{T} se é feita da seguinte forma: inicialmente calcula-se uma matriz de Hankel para cada linha de \mathbf{Y} e, em seguida, concatena-se todas as matrizes de Hankel em um tensor de terceira ordem \mathcal{T} .

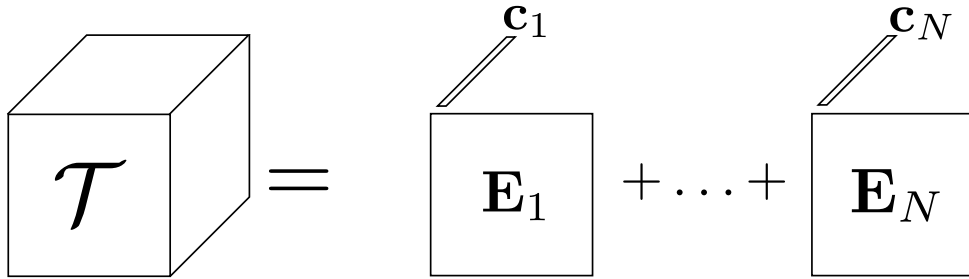


Figura 3.4: Block Term Decomposition de um tensor \mathcal{T} .

Decompondo o tensor \mathcal{T} de acordo com o modelo (3.44) fornece N matrizes \mathbf{E}_n e vetores \mathbf{c}_n . Estima-se as T amostras de uma linha de \mathbf{S} calculando a média das T anti-diagonais de uma matriz \mathbf{E}_n . Fazendo isso para todas as matrizes, obtém-se todas as fontes. Assim como todos os métodos de separação de fontes, o BSS-BTD (Algoritmo 6)

⁴Uma matriz de Hankel $\mathbf{M} = [m_{i,j}]$ é aquela que, dada uma sequência $\{s_1, s_2, \dots\}$, satisfaz a seguinte relação: $m_{i,j} = s_{i+j-1}$. Dessa forma, todas as suas anti-diagonais são constantes.

também está sujeito a indeterminações de escala e permutação.

Entrada: Matriz de observação \mathbf{X} .

Saída: Matriz de composição \mathbf{A} e matriz de fontes \mathbf{S} .

1. Inicialização

- Definir o número N de fontes através das SVD de \mathbf{X} .
- Definir L_n para $n = 1, 2, \dots, N$.

2. Para cada linha de \mathbf{X} :

- Calcular a sua matriz de Hankel.

3. Concatenar todas as matrizes de Hankel, gerando o tensor de terceira ordem \mathcal{T} .

4. Calcular a sua BTD.

5. A partir dessa decomposição...

- Estimar \mathbf{S} através das matrizes de *rank* L_n ; ou
- Estimar a matriz de decomposição $\hat{\mathbf{A}}$ e calcular $\mathbf{S} \leftarrow \hat{\mathbf{A}}^{-1}\mathbf{X}$.

Algoritmo 6: Algoritmo BSS-BTD.

Apesar do BSS-BTD apresentar performance superior que alguns métodos de SCF em (LATHAUWER, 2011) para fontes oscilatórias, é assumido que L_n é conhecido para cada n -ésima fonte. Entretanto ainda não há um método automático para a seleção desses parâmetros, por isso, em situações práticas, é preciso buscar manualmente os parâmetros que forneçam a separação ótima.

3.5.3 ICA via matriz de “carrelação”

Momentos e cumulantes são estatísticas obtidas a partir da expansão de Taylor das duas primeiras funções características, como visto na Seção 2.3. Cumulantes de ordem superior apresentam uma propriedade interessante: eles são nulos quando as fontes são mutuamente independentes. Essa propriedade permite o desenvolvimento de algoritmos que exploram essa ideia a fim de estimar as fontes. Pode-se representar naturalmente essas estatísticas por meio de tensores, porém o uso de ferramentas multilineares acarreta no aumento da complexidade computacional e notacional. Técnicas baseadas na diagonalização de tensores de cumulantes precisam de uma quantidade razoável de amostras para garantir a qualidade da estimação dos cumulantes. Em (SLAPAK; YEREDOR, 2012b; SLAPAK; YEREDOR, 2012c) os autores propuseram uma nova ferramenta matemática que tenta contornar os problemas citados por meio de uma generalização da matriz de covariância que consegue capturar informações de ordem superior sem utilizar a aborda-

gem multilinear. Essa nova ferramenta é fundamentada em duas estatísticas chamadas **carmédia**⁵ e **carrelação**, definidas a seguir.

Seja $\mathbf{x} \in \mathbb{R}^{N \times 1}$ um vetor aleatório e $\mathbf{g}(\cdot) : \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}^{L \times 1}$. A carmédia de $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^{L \times 1}$ em relação a \mathbf{x} em um **ponto de processamento** $\tau \in \mathbb{R}^{N \times 1}$ arbitrário é definida como

$$\eta_{\mathbf{x}}[\mathbf{g}(\mathbf{x}); \tau] = \frac{\mathbb{E}[\mathbf{g}(\mathbf{x})e^{\mathbf{x}^T \tau}]}{\mathbb{E}[e^{\mathbf{x}^T \tau}]} \in \mathbb{R}^L. \quad (3.45)$$

Em (SLAPAK; YEREDOR, 2012b) mostra-se que a carmédia possui as mesmas propriedades do operador esperança, como linearidade e separabilidade no caso de independência estatística. De fato, a carmédia se degenera na esperança quando $\tau = \mathbf{0}$.

Agora considera-se as funções $\mathbf{g}_1(\cdot) : \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}^{L_1}$ e $\mathbf{g}_2(\cdot) : \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}^{L_2 \times 1}$. A matriz de carrelação cruzada entre $\mathbf{g}_1(\mathbf{x})$ e $\mathbf{g}_2(\mathbf{x})$ em relação a \mathbf{x} em um ponto de processamento $\tau \in \mathbb{R}^{N \times 1}$ arbitrário é definida como

$$\Psi_{\mathbf{x}}[\mathbf{g}_1(\mathbf{x}), \mathbf{g}_2(\mathbf{x}); \tau] = \eta_{\mathbf{x}}[\mathbf{g}_1(\mathbf{x})\mathbf{g}_2^T(\mathbf{x}); \tau] - \eta_{\mathbf{x}}[\mathbf{g}_1(\mathbf{x}); \tau]\eta_{\mathbf{x}}^T[\mathbf{g}_2(\mathbf{x}); \tau]. \quad (3.46)$$

A matriz de carrelação de $\mathbf{g}(\mathbf{x})$ é definida como a carrelação cruzada entre $\mathbf{g}(\mathbf{x})$ e ele mesmo:

$$\Psi_{\mathbf{x}}[\mathbf{g}(\mathbf{x}); \tau] = \Psi_{\mathbf{x}}[\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{x}); \tau]. \quad (3.47)$$

Seja uma transformação linear $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{c}$, no qual $\mathbf{C} \in \mathbb{R}^{L \times N}$ e $\mathbf{c} \in \mathbb{R}^L$ são uma matriz e um vetor constantes. As matrizes de carrelação de \mathbf{x} e \mathbf{y} relacionam-se da seguinte forma:

$$\Psi_{\mathbf{x}}[\mathbf{y}; \tau] = \mathbf{C}\Psi_{\mathbf{x}}[\mathbf{x}; \tau]\mathbf{C}^T \quad (3.48)$$

$$\Psi_{\mathbf{y}}[\mathbf{y}; \tau] = \mathbf{C}\Psi_{\mathbf{x}}[\mathbf{x}; \mathbf{C}^T \tau]\mathbf{C}^T. \quad (3.49)$$

Mostra-se que se um vetor \mathbf{x} pode ser particionado em dois vetores mutuamente independentes $\mathbf{x}_1 \in \mathbb{R}^{N_1}$ e $\mathbf{x}_2 \in \mathbb{R}^{N_2}$ com $N = N_1 + N_2$ então $\Psi_{\mathbf{x}}[\mathbf{x}, \tau]$ é diagonal em blocos para todo τ .

Um método de separação de fontes com abordagem ICA foi proposto em (SLAPAK; YEREDOR, 2012c) utilizando essas duas propriedades. Seja o modelo linear instantâneo (1.2). Busca-se uma matriz $\mathbf{B} \approx \mathbf{A}^{-1}$. A matriz de carrelação da equação (1.2) pode ser expressa por meio da equação (3.49):

$$\Psi_{\mathbf{x}}(\tau) = \mathbf{A}\Psi_{\mathbf{s}}(\mathbf{A}^T \tau)\mathbf{A}^T. \quad (3.50)$$

⁵O prefixo “car-” reflete a relação com a função característica.

Multiplicando a equação (3.50) por \mathbf{B} pela esquerda e por \mathbf{B}^T pela direita, obtém-se:

$$\mathbf{B}\Psi_{\mathbf{x}}(\tau)\mathbf{B}^T = \Psi_{\mathbf{s}}(\mathbf{A}^T\tau). \quad (3.51)$$

Como o vetor \mathbf{s} contém componentes mutuamente independentes, então as matrizes $\Psi_{\mathbf{s}}(\mathbf{A}^T\tau)$ e $\mathbf{B}\Psi_{\mathbf{x}}(\tau)\mathbf{B}^T$ são diagonais. Portanto, busca-se uma matriz $\hat{\mathbf{B}}$ que busque diagonalizar a matriz de correlação $\Psi_{\mathbf{x}}(\tau)$. Essa ideia foi formalizada em um problema de otimização não-linear *Weighted Least-Squares* (WLS):

$$\min_{\hat{\mathbf{B}}} \left\{ \mathbf{h}^T \left(\hat{\Psi}_{\mathbf{x}}(\tau_1), \dots, \hat{\Psi}_{\mathbf{x}}(\tau_M); \hat{\mathbf{B}} \right) \mathbf{W} \mathbf{h} \left(\hat{\Psi}_{\mathbf{x}}(\tau_1), \dots, \hat{\Psi}_{\mathbf{x}}(\tau_M); \hat{\mathbf{B}} \right) \right\}, \quad (3.52)$$

no qual

$$\mathbf{h} \left(\hat{\Psi}_{\mathbf{x}}(\tau_1), \dots, \hat{\Psi}_{\mathbf{x}}(\tau_M); \hat{\mathbf{B}} \right) = \Theta \left[\hat{\mathbf{B}}\Psi_{\mathbf{x}}(\tau_1)\hat{\mathbf{B}}^T, \dots, \hat{\mathbf{B}}\Psi_{\mathbf{x}}(\tau_M)\hat{\mathbf{B}}^T \right],$$

é uma vetorização das matrizes fornecidas como argumento do operador $\Theta(\cdot)$ feita de tal forma que apenas os elementos superiores à diagonal principal dessas matrizes sejam considerados e \mathbf{W} é uma matriz de pesos.

Slapak e Yeredor propuseram um algoritmo do tipo Gauss-Newton para resolver o problema não-linear (3.52), resultando em um algoritmo computacionalmente complexo mas eficiente, apresentando melhores resultados que algoritmos como FastICA e JADE. Nota-se que esse algoritmo assume que os pontos pontos de processamento τ_1, \dots, τ_M são conhecidos. Na prática isso não acontece. Porém em (SLAPAK; YEREDOR, 2012c) é proposto uma forma de se estimar um conjunto de pontos de processamento que minimizem o traço da matriz de covariância de $\hat{\mathbf{B}}$, oferecendo um possível conjunto de parâmetros que forneçam uma boa separação de fontes.

Recentemente alguns trabalhos foram publicados utilizando essas ferramentas. Em (SLAPAK; YEREDOR, 2012a) propôs-se um método de identificação cega de um processo média-móvel não-Gaussiano utilizando carmédia. Já em (GU *et al.*, 2013; GU *et al.*, 2014) foram propostos novos métodos para identificação de misturas sub-determinadas. Observa-se que métodos baseados em matrizes de correlação têm atraído a atenção da comunidade científica, motivando a pesquisa de novos algoritmos de separação fontes e aperfeiçoar os métodos existentes.

3.6 Conclusão

Neste Capítulo foram apresentada os métodos e algoritmos clássicos de separação de fontes. Foi definido o conceito de contraste, útil para definir critérios de otimização que resultem na separação cega de fontes. Foram apresentados alguns métodos algébricos de separação de fontes seguindo uma abordagem de Teoria da Informação. Devido ao custo

prático dessa abordagem, contrastes baseados em cumulantes de ordem superior foram propostos por meio da aproximação de medidas como negentropia. Métodos algébricos podem atingir um alto custo computacional e não rastrear mudanças rápidas no sistema de composição. Motivado por essas desvantagens, métodos iterativos como FastICA e RobustICA foram propostos, apresentando maior eficiência computacional e simplicidade na implementação. Finalmente, foi feito um rápido estudo bibliográfico de algumas novas técnicas de SCF, tais como BTM, NMF e métodos baseados em matrizes de correlação.

4. RESULTADOS E DISCUSSÃO

4.1 Introdução

Neste Capítulo serão apresentados resultados de simulações e experimentos realizados através dos algoritmos apresentados no Capítulo 3. A primeira parte deste Capítulo consiste em simulações com sinais sintéticos, realizadas a fim de analisar o desempenho dos algoritmos de SCF em diferentes cenários de ruído Guassiano, quantidade de amostras e quantidade de fontes. Já a segunda parte visa apresentar uma aplicação prática de separação de fontes, especificamente o a separação de eletrocardiograma fetal do materno através de observações bioelétricas feitas por eletrodos no abdômem da mãe.

Nos experimentos e simulações feitos, os seguintes algoritmos foram utilizados:

- JADE (Subseção 3.3.2). Implementação MATLAB disponível em <http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m>.
- FastICA (Subseção 3.4.1) no modo deflação com não-linearidade $g(u) = u^3$ com 1.000 iterações por fonte. Implementação MATLAB disponível em <http://research.ics.aalto.fi/ica/fastica/>.
- RobustICA (Subseção 3.4.2) no modo deflação com 1.000 iterações por fonte. Implementação MATLAB disponível em <http://www.i3s.unice.fr/~zarzoso/robustica.html>.

4.2 Experimentos com sinais sintéticos

Considera-se os seguintes sinais:

$$s_1(k) = \text{sen}(k/2) \quad (4.1)$$

$$s_2(k) = \left[\frac{\text{mod}(k, 23) - 11}{9} \right]^5 \quad (4.2)$$

$$s_3(k) = \frac{\text{mod}(k, 27) - 13}{9} \quad (4.3)$$

$$s_4(k) = x \ln y, \quad k = 0, 1, \dots, K - 1, \quad (4.4)$$

no qual K é o número de amostras disponíveis, $\text{mod}(a, b)$ representa o resto da divisão entre a e b , x é uma variável aleatória uniforme discreta com valores em $\{-1, 1\}$ e y é uma variável aleatória uniformemente distribuída entre $]0; 1[$. Armazena-se K amostras

desses sinais na matriz

$$\mathbf{S} = \begin{bmatrix} s_1(0) & s_1(1) & \dots & s_1(K-1) \\ s_2(0) & s_2(1) & \dots & s_2(K-1) \\ s_3(0) & s_3(1) & \dots & s_3(K-1) \\ s_4(0) & s_4(1) & \dots & s_4(K-1) \end{bmatrix}.$$

Cada sinal é mostrado na Figura 4.2(a) para $K = 500$. A fim de se observar algumas propriedades estatísticas das fontes, foi calculado o histograma de cada sinal (Figura 4.1) para a quantidade de $K = 500.000$ amostras e 100 intervalos de valores. Ao se observar os histogramas 4.1(a) e 4.1(c), conclui-se que $s_1(k)$ e $s_3(k)$ são sinais sub-Gaussianos, enquanto os histogramas 4.1(b) e 4.1(d) indicam que $s_2(k)$ e $s_4(k)$ são sinais super-Gaussianos. Essa análise é confirmada pelos valores estimados da curtose.

Considera-se um modelo instantâneo linear de mistura com $M = 4$ observações, resultando em uma matriz \mathbf{A} gerada aleatoriamente de dimensão (4×4) . Os sinais são misturados, gerando uma matriz $\mathbf{X} = \mathbf{AS}$ com 4 sinais de observação, mostrados na Figura 4.2(b). Em seguida, estimou-se as fontes através do método JADE (Figura 4.2(c)). Para isso, foi calculado o filtro separador \mathbf{W} tal que a estimação das fontes é dada por $\mathbf{Y} = \mathbf{WX}$. Nota-se que as fontes estimadas estão permutadas e multiplicadas por um certo fator de escala. A fim de desfazer essas ambiguidades, calculou-se a correlação de cada fonte estimada com todas as fontes geradas. Dessa forma, é possível encontrar a ordem correta das fontes. Para se encontrar o fator de escala correto de cada sinal estimado, busca-se o fator α que minimiza a distância entre n -ésimo sinal estimado \mathbf{y}_n (com permutação corrigida) e o n -ésimo sinal gerado \mathbf{s}_n :

$$\min_{\alpha} \|\mathbf{y}_n - \alpha \mathbf{s}_n\|^2.$$

Resolvendo esse problema, obtém-se:

$$\alpha^* = \frac{\mathbf{s}_n^T \mathbf{y}_n}{\|\mathbf{y}_n\|^2}.$$

Após correção das ambiguidades, as fontes estimadas pelo algoritmo JADE são mostradas na Figura 4.2(d). Observa-se que os sinais calculados são bastante próximos às fontes originais. Os outros algoritmos forneceram estimações com a mesma qualidade, portanto conclui-se que os três algoritmos conseguem facilmente separar os sinais sintéticos gerados. Todavia, este resultado não permite a comparar diretamente os três algoritmos. Portanto, outros cenários serão desenvolvidos a seguir.

A comparação quantitativa de algoritmos dá-se por meio da definição de critérios de desempenho e da comparação deles entre os algoritmos. Um critério utilizado na literatura

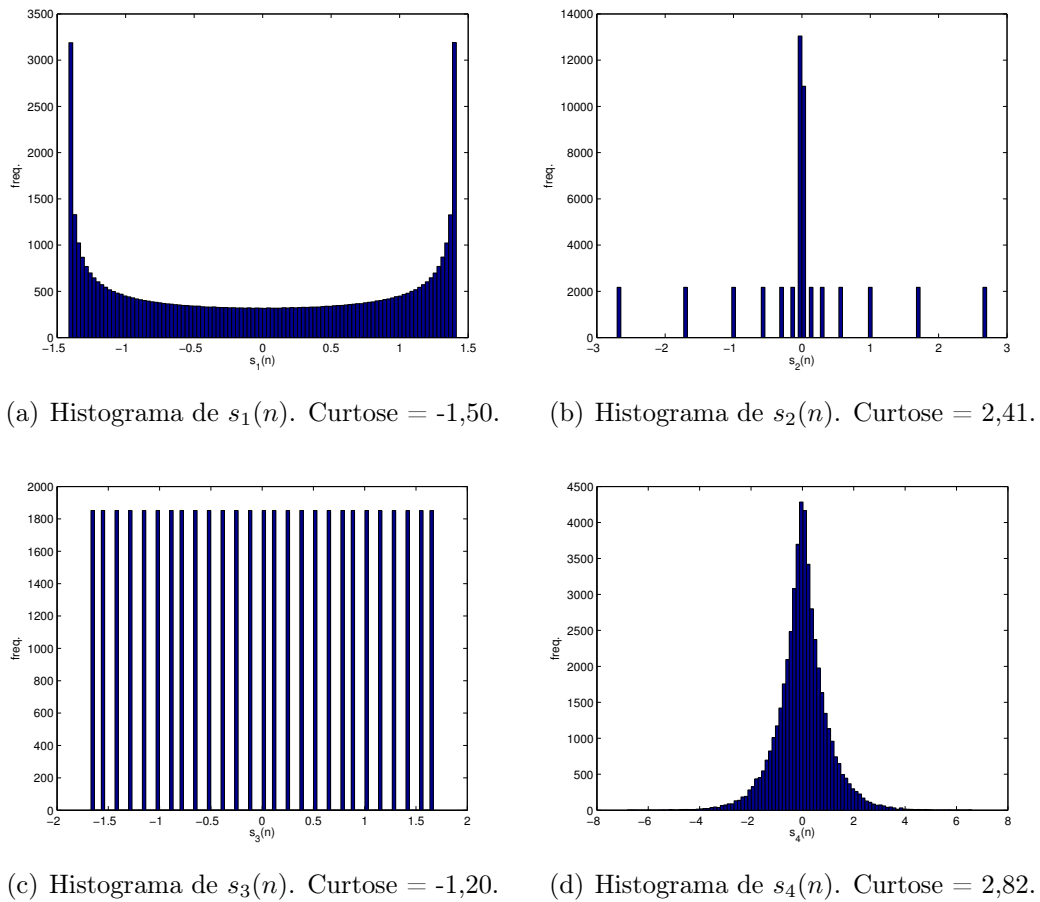


Figura 4.1: Histograma das fontes sintéticas.

(LATHAUWER, 2011) é o *Interference-to-Signal Ratio* (ISR) é definido como

$$\text{ISR} = 10 \log_{10} \left[\frac{\sum_{n=1}^N p_{nn}^2}{\sum_{m \neq n} p_{mn}^2} \right] \text{ [dB]}, \quad (4.5)$$

no qual $\mathbf{P} = \mathbf{W}\mathbf{A}$. Um outro critério utilizado neste trabalho é o *mean squared error* entre as estimações e as fontes:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^N \left[\frac{1}{K} \sum_{k=0}^{K-1} (s_n(k) - y_n(k))^2 \right]. \quad (4.6)$$

O MSE mede o desvio médio de todas as fontes estimadas em relação às suas fontes originais.

O desempenho dos algoritmos foi analisado em três cenários distintos. Primeiramente considerou a separação cega de fontes com observações corrompidas da seguinte forma:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \sigma\mathbf{N},$$

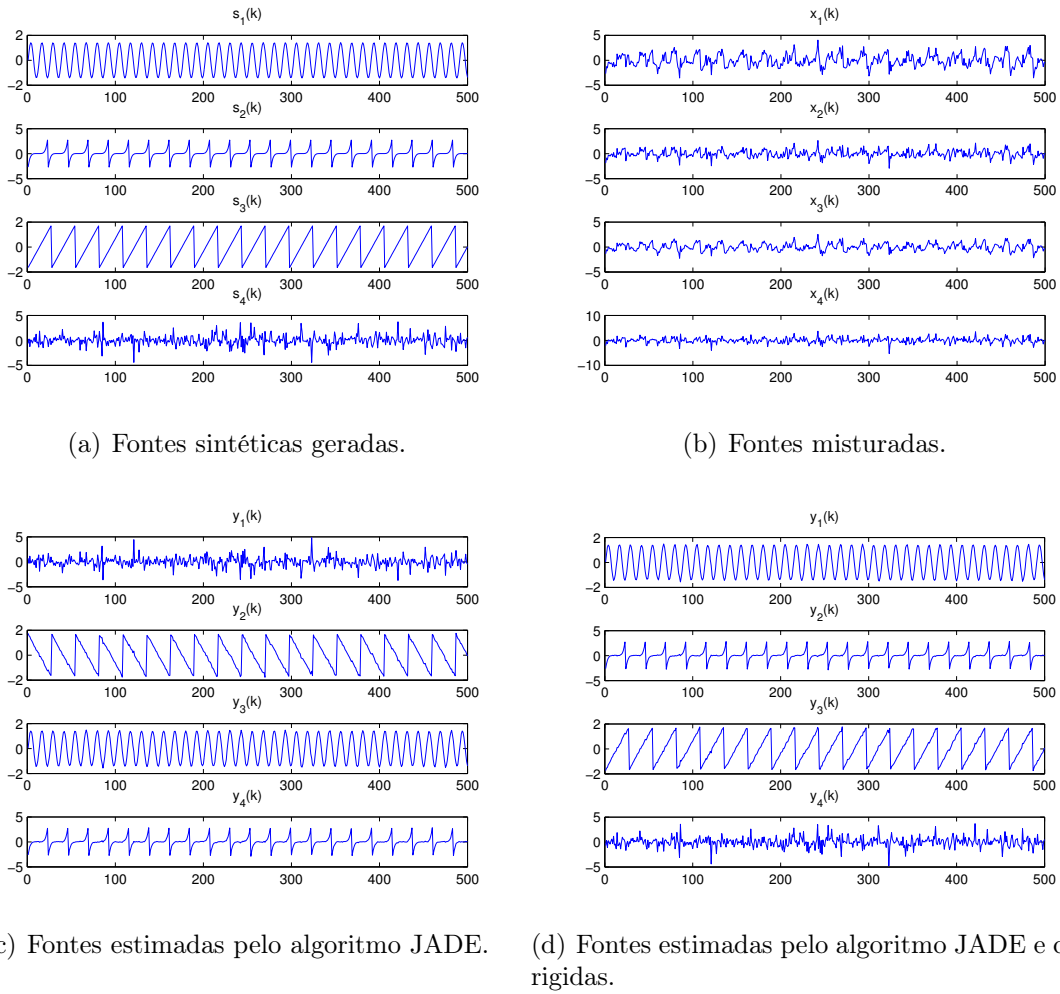


Figura 4.2: Experimento com sinais sintéticos utilizando o algoritmo JADE.

no qual $\mathbf{N} \in \mathbb{R}^{M \times M}$ é uma matriz de ruído branco e Gaussiano e σ é um fator que controla a potência do ruído. A relação sinal-ruído (SNR) é definida como:

$$\text{SNR}(\sigma) = 20 \log_{10} \frac{\|\mathbf{S}\|_F}{\sigma \|\mathbf{N}\|_F}. \quad (4.7)$$

Para se analisar o desempenho dos algoritmos considerando sinais de observação corrompidos por ruído branco aditivo e Gaussiano, fixou-se o número de amostras para 2.000, utilizaram-se as quatro fontes (4.1)-(4.4) e calcularam-se os critérios de desempenho *ISR* e *MSE* por meio de uma simulação de Monte Carlo com 1.000 *runs*. A Figura 4.3 mostra a relação entre SNR e o critério *ISR*, enquanto a Figura 4.4 descreve o *MSE* em função da SNR. Observa-se que para valores de SNR baixa, os algoritmos apresentam o mesmo desempenho, enquanto que para valores elevados, o RobustICA apresenta a melhor performance sob ruído.

Analisou-se o efeito da quantidade de amostras nos algoritmos através de uma simulação de Monte Carlo. As quatro fontes sintéticas foram misturadas sem considerar ruído.

Os critérios de desempenho foram calculados através de 1.000 *runs* de Monte Carlo. As Figuras 4.5 e 4.6 descrevem os critérios *ISR* e *MSE* em função da *SNR* respectivamente. Como as curvas do RobustICA estão abaixo das curvas dos outros algoritmos, conclui-se que o RobustICA é menos sensível à quantidade de amostras.

Para se investigar o efeito da quantidade de fontes a serem estimadas sob o desempenho dos algoritmos, utilizou-se fontes senóidais com frequências diferentes, pois, assim, pode-se gerar facilmente um determinado número de fontes sintéticas. A n -ésima fonte é definida como

$$s_n(k) = \text{sen}\left(\frac{k}{2n+1}\right), \quad k = 0, \dots, K-1.$$

Os critérios de desempenho foram calculados através de simulações de Monte Carlo de 1.000 *runs*. Nessas simulações, fixou-se o número de amostras em $K = 2.000$ e os sinais foram misturados sem a adição de ruído. As Figuras 4.7 e 4.8 descrevem os critérios de desempenho em função da quantidade de fontes. Como esperado, a performance dos algoritmos piora à medida que mais fontes são utilizadas. Observa-se que os algoritmos apresentam, em média, o mesmo desempenho.

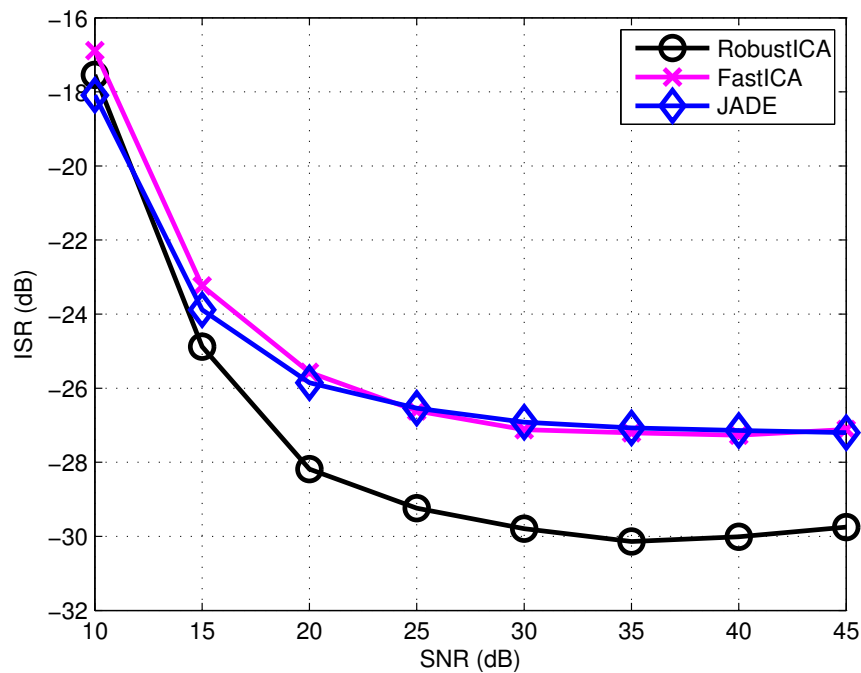


Figura 4.3: Curvas de SNR vs. ISR.

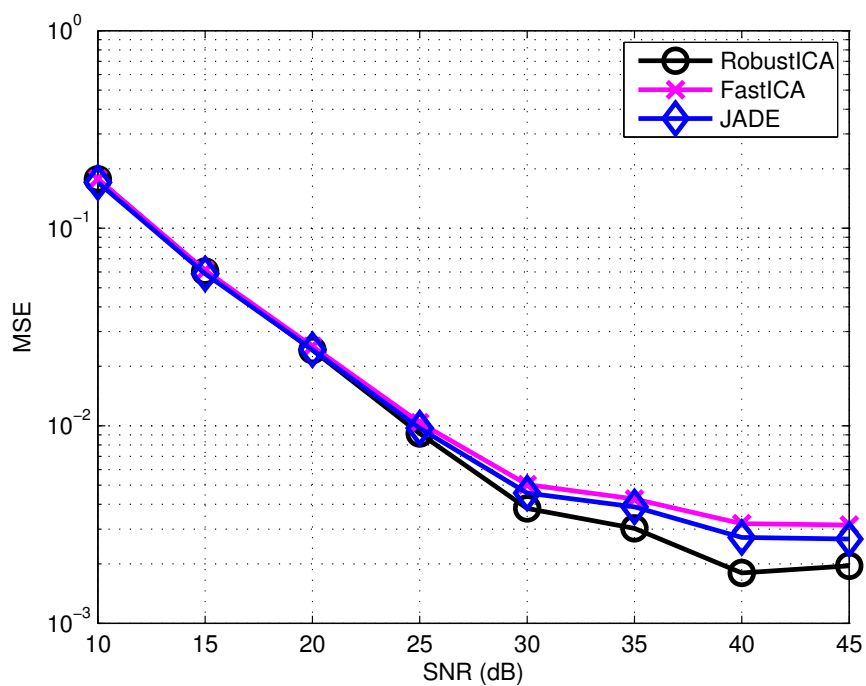


Figura 4.4: Curvas SNR vs MSE.

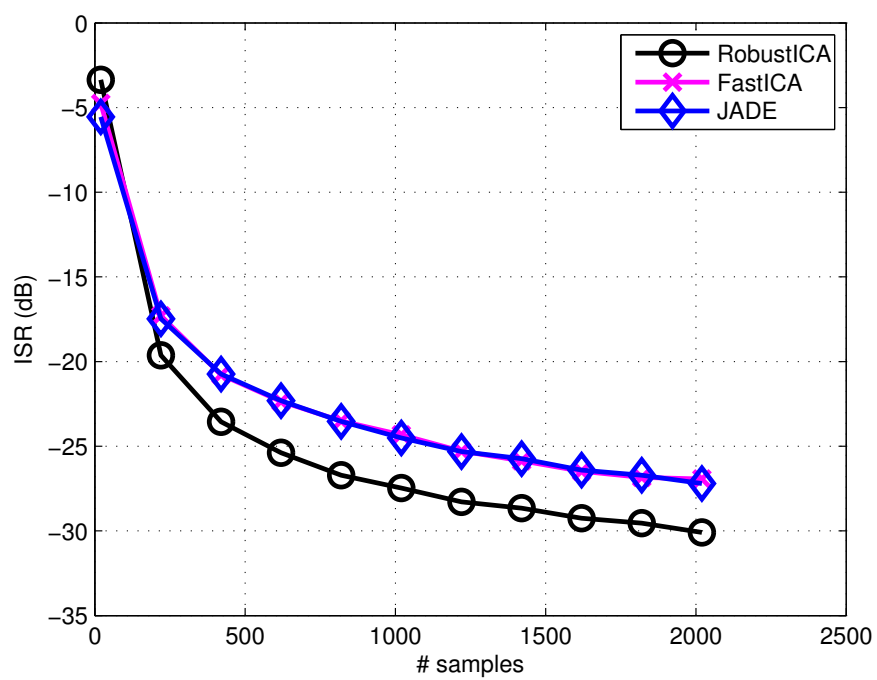


Figura 4.5: Curvas de número de amostras vs. ISR.

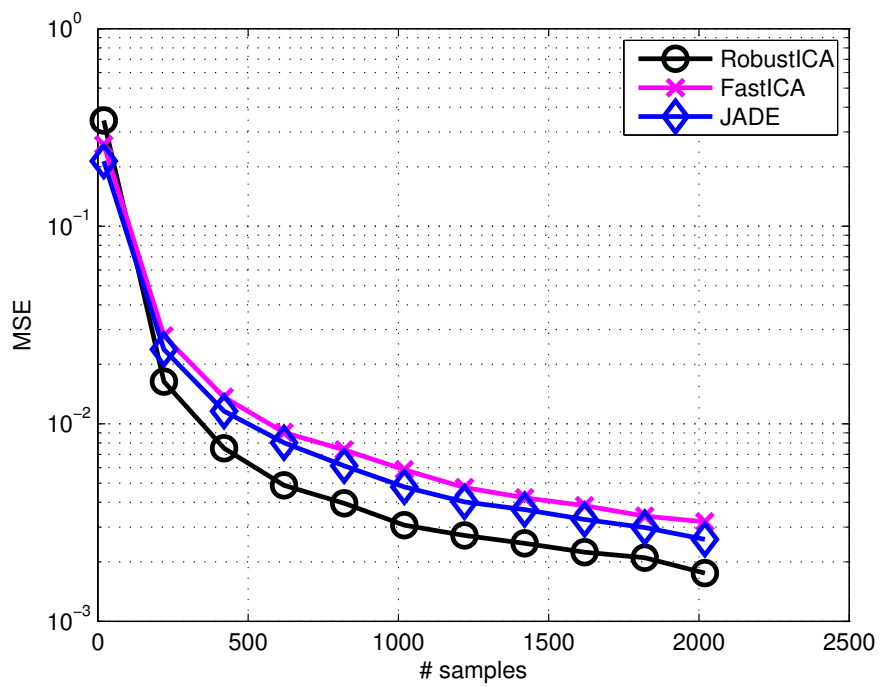


Figura 4.6: Curvas de número de amostras vs. MSE.

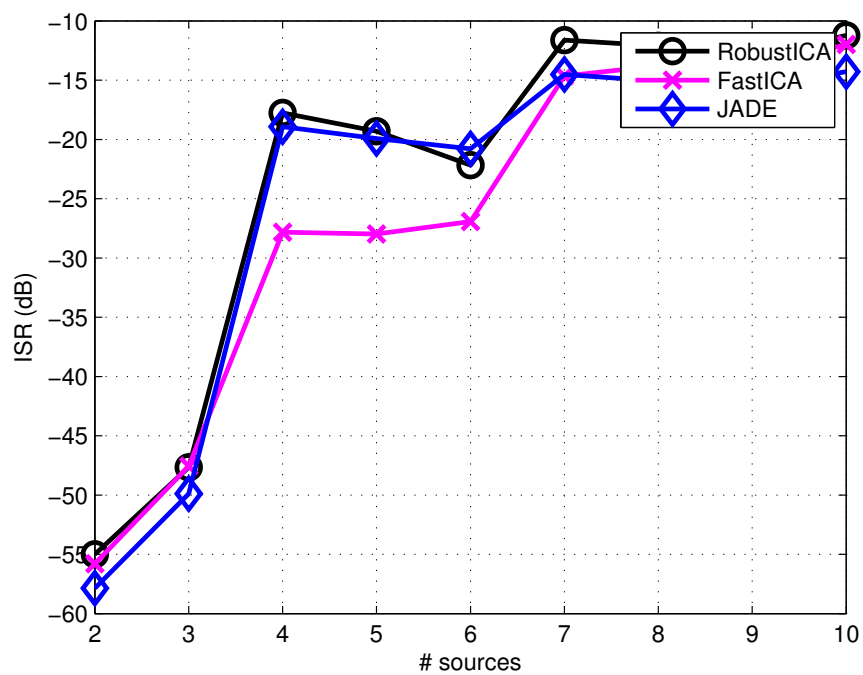


Figura 4.7: Curvas de número de fontes vs. ISR.

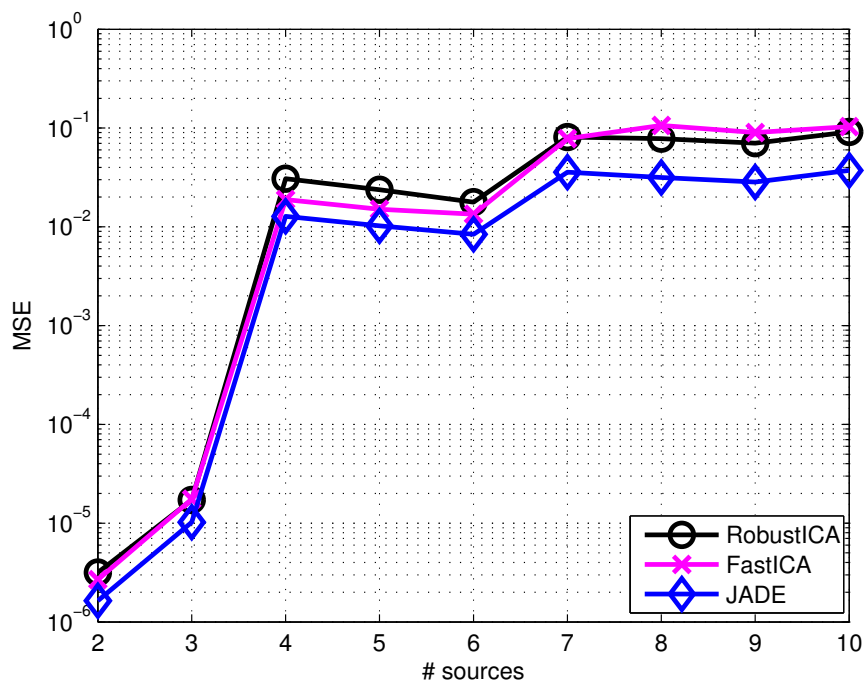


Figura 4.8: Curvas de número de fontes vs. MSE.

4.3 Experimentos práticos

Graças ao advento da eletrônica e do processamento digital de sinais, é possível analisar órgãos internos do corpo humano de forma não-invasiva de tal forma que informações importantes sobre a saúde desses órgãos sejam obtidas de forma confiável. Tem-se muito interesse na análise de sinais bioelétricos provenientes do coração, captados através de eletrodos dispostos na superfície do corpo do paciente. Dá-se o nome de **eletrocardiograma** (ECG) para os sinais capturados através desses sensores. Um ECG é geralmente um sinal quase periódico, e apresenta um *layout* bem definido, composto pelas ondas P e T e pelo complexo QRS (SÖRNMO; LAGUNA, 2005), como mostrado na Figura 4.9. Pode-se estimar a taxa cardíaca do paciente detectando-se corretamente os picos R em um ECG.

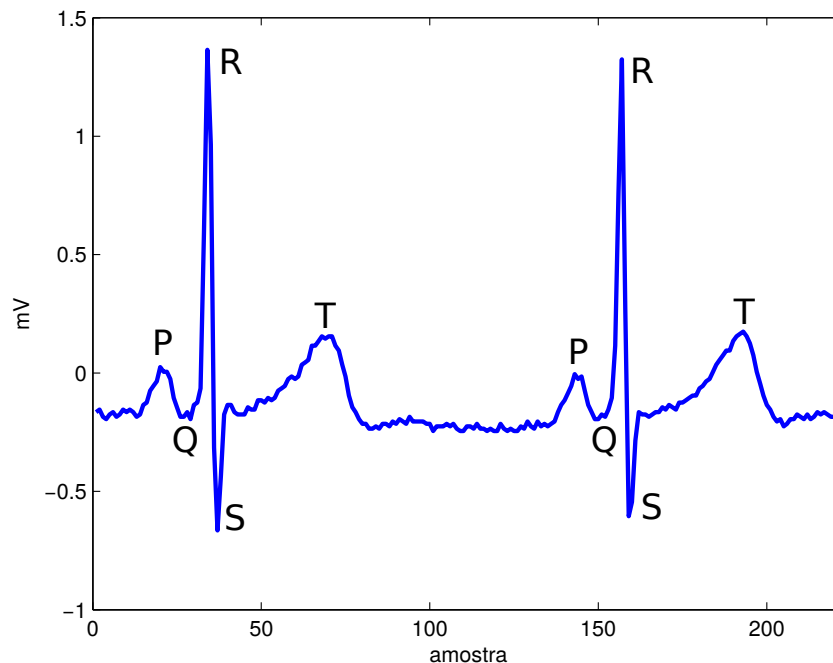


Figura 4.9: Eletrocardiograma ilustrando as onda P e T juntamente com complexo QRS. Sinal ECG obtido do MIT-BIH *normal sinus rhythm database*, paciente 16272.

Na Medicina, uma forma simples de monitorar a saúde de um feto se dá pelo monitoramento da sua taxa cardíaca. A forma mais conhecida de calculá-lo é por meio do ultrassom. Apesar de ser uma técnica precisa e simples, o ultrassom necessita ser operado por um médico ou enfermeira bem treinado para o manuseio do transdutor. Além disso, não é recomendável utilizá-lo por períodos prolongados. Em contraste, o uso do ECG abdominal apresenta uma maior praticidade, pois não há necessidade de manuseio de seus elementos para a extração de sinais e não há restrições em relação à duração de uso. O sinal obtido pelos eletrodos são processados, removendo as interferências geradas pela

mãe (ECG materno, artefatos de movimento, etc). Soluções como filtragem em frequência ou modelos paramétricos não são bem adaptados a esse problema (ZARZOSO; NANDI; BACHARAKIS, 1997), pois o sinal de ECG possui uma grande variabilidade e o ECG fetal possui baixa potência, impedindo o uso de algumas técnicas. Uma solução clássica para esse problema utiliza filtros adaptativos (ZARZOSO; NANDI, 2001). Existem fortes indícios de que os ECG da mãe e do filho sejam independentes, motivando o uso de técnicas de SCF.

Nesta seção, será feito um experimento para a extração do ECG fetal a partir de observações ECG abdominal da mãe utilizando o algoritmo RobustICA, devido ao seu bom desempenho observado na seção anterior. Utilizou-se o conjunto de observações ECG disponível no banco de dados DaISy (MOOR, 2010). Esse conjunto consiste em um 10s de observações divididos em 8 canais (5 abdominais e 3 torácicos) amostrados na taxa de 250 Hz.

Será utilizado apenas os 5 ECGs abdominais, pois são eles os sinais que possuem maior contribuição do ECG fetal. A fim de eliminar a componente de flutuação e atenuar a interferência da linha de alimentação em 60 Hz dos sinais obtidos, aplicou-se um filtro IIR Tchebychev tipo II com banda de passagem entre 0.5 Hz e 55 Hz (SÖRNMO; LAGUNA, 2005), resultando nos ECGs mostrados na Figura 4.10. É possível observar o ECG fetal nos canais 1, 2 e 3.

Em seguida, aplicou-se o algoritmo RobustICA nesse ECG, resultando nos sinais mostrados na Figura 4.11. Ao identificar as ondas R nos sinais calculados, conclui-se que as fontes 1 e 2 correspondem ao ECG materno e as fontes 3 e 4 são relacionadas ao ECG fetal. Apesar da fonte 4 estar bastante ruidosa, ainda é possível observar os pontos R em sincronia com os mesmos pontos da fonte 3. Através do RobustICA foi possível extrair um sinal ECG fetal que possui claramente um complexo QRS, permitindo o cálculo da taxa cardíaca do feto. Em contraste, nota-se que não é possível identificar claramente as outras componentes do ECG como onda P e T. Alternativamente, foram propostos algoritmos de SCF otimizados para o ECG fetal (REAZ; WEI, 2004; NIKNAZAR *et al.*, 2013).

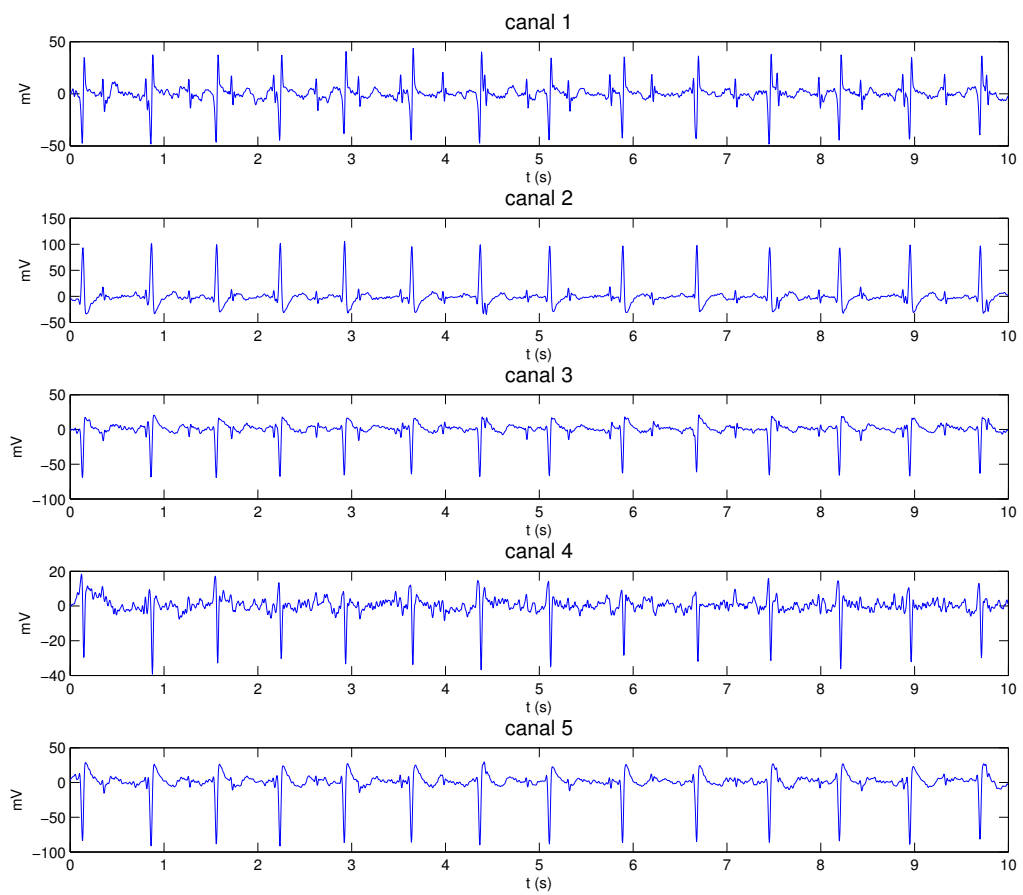


Figura 4.10: Eletrocardiogramas abdominais da mãe.

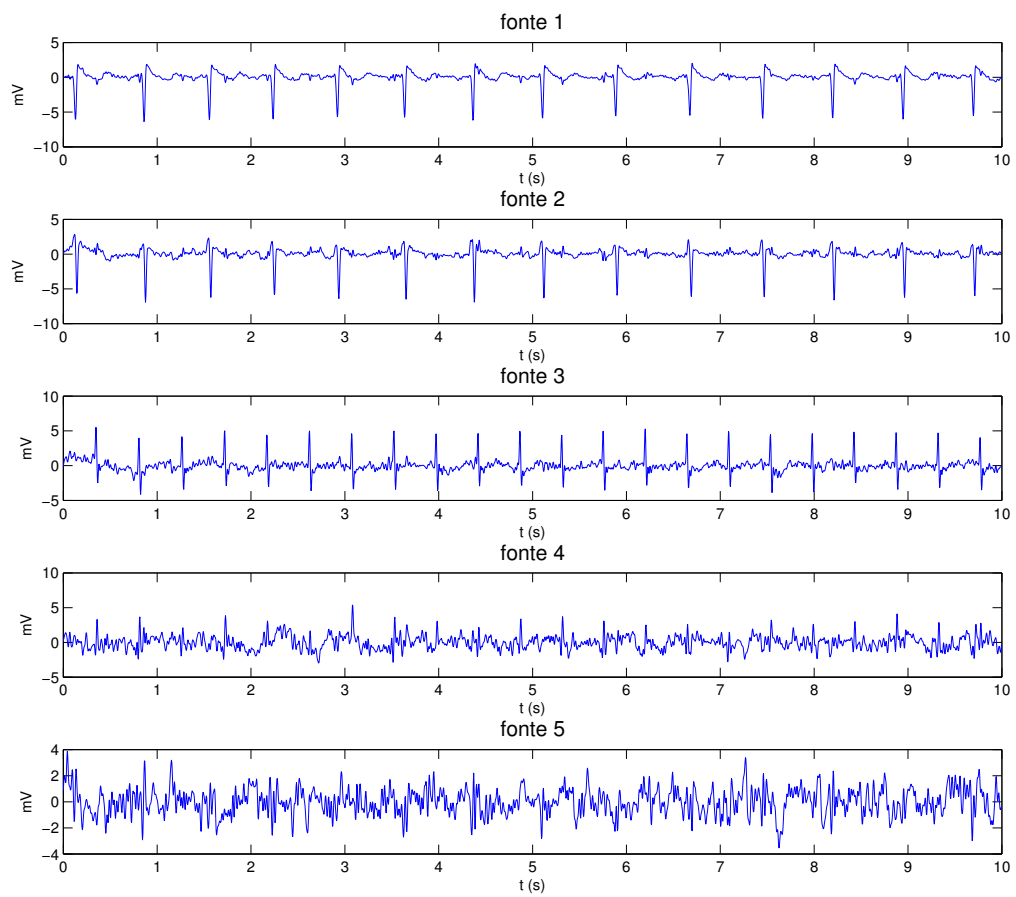


Figura 4.11: Componentes independentes dos ECGs maternos, fornecendo o ECG fetal.

5. CONCLUSÃO

Inicialmente foi feita uma introdução ao problema de separação cega de fontes e descreveu-se alguns modelos matemáticos que caracterizam esse problema. Foram apresentadas algumas aplicações desse problema na área de telecomunicações, engenharia biomédica e acústica. Em seguida, foi feita uma revisão sobre algumas ferramentas matemáticas necessárias para a compreensão dos métodos de SCF. Posteriormente, soluções algébricas e iterativos foram discutidas e tendências de pesquisa nessa área foram comentadas.

Os algoritmos RobustICA, FastICA e JADE foram comparados através de simulações computacionais com sinais sintéticos. Através dessas simulações, pode-se concluir que o algoritmo RobustICA apresenta um desempenho melhor em cenários com ruído e também é menos sensível à quantidade de amostras disponíveis. Em seguida, investigou-se o desempenho do algoritmo RobustICA no problema prático de separação de ECG fetal e materno. Observou-se que essa técnica permite a correta estimação da taxa cardíaca do feto de forma não-invasiva.

Através deste trabalho, pode-se concluir que o campo de pesquisa em SCF tem recebido muita atenção da comunidade acadêmica devido ao seu vasto campo de aplicações em várias áreas do conhecimento, como Medicina, Biologia, Engenharia, Astronomia, Economia, etc. Engenheiros e cientistas têm se esforçado para melhorar o desempenho dos métodos de SCF para que eles sejam mais eficazes na prática. Para isso, é preciso novos resultados teóricos para sistemas de composição não-linear e problemas sub-determinados.

Essa discussão abre espaço para trabalhos futuros. Com a ajuda das ferramentas tensoriais, é possível implementar novas técnicas de SCF em problemas sub-determinados. Como mostrado neste trabalho, métodos baseados em correlação mostram-se promissores. Portanto, propõe-se modelar essa nova modalidade de ICA utilizando ferramentas tensoriais. Uma outra possibilidade de trabalho futuro seria melhorar e estender o algoritmo RobustICA. É possível estendê-lo para o caso convolutivo e MIMO, assim como é possível utilizar estimadores robustos de cumulantes, para melhorar a sua capacidade de lidar com *outliers*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRARD, F.; DEVILLE, Y. Blind separation of dependent sources using the "time-frequency ratio of mixtures" approach. In: IEEE. *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*. [S.l.], 2003. v. 2, p. 81–84.
- AGEE, B. *et al.* Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays. *Proceedings of the IEEE*, IEEE, v. 78, n. 4, p. 753–767, 1990.
- BELL, A. J.; SEJNOWSKI, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, MIT Press, v. 7, n. 6, p. 1129–1159, 1995.
- BELOUCHRANI, A. *et al.* A blind source separation technique using second-order statistics. *Signal Processing, IEEE Transactions on*, IEEE, v. 45, n. 2, p. 434–444, 1997.
- BINGHAM, E.; HYVÄRINEN, A. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International journal of neural systems*, World Scientific, v. 10, n. 01, p. 1–8, 2000.
- BOFILL, P.; ZIBULEVSKY, M. Underdetermined blind source separation using sparse representations. *Signal processing*, Elsevier, v. 81, n. 11, p. 2353–2362, 2001.
- CARDOSO, J.-F. Blind signal separation: statistical principles. *Proceedings of the IEEE*, IEEE, v. 86, n. 10, p. 2009–2025, 1998.
- CARDOSO, J.-F. High-order contrasts for independent component analysis. *Neural computation*, MIT Press, v. 11, n. 1, p. 157–192, 1999.
- CARDOSO, J.-F.; SOULOUMIAC, A. Blind beamforming for non-gaussian signals. In: IET. *IEE Proceedings F (Radar and Signal Processing)*. [S.l.], 1993. v. 140, n. 6, p. 362–370.
- CHABRIEL, G. *et al.* Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications. *Signal Processing Magazine, IEEE*, IEEE, v. 31, n. 3, p. 34–43, 2014.
- CHERRY, E. C. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, Acoustical Society of America, v. 25, n. 5, p. 975–979, 1953.
- CICHOCKI, A. *et al.* Tensor decompositions for signal processing applications. *From Two-way to Multiway Component Analysis, ESAT-STADIUS Internal Report*, p. 13–235, 2014.
- CICHOCKI, A. *et al.* *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. [S.l.]: John Wiley & Sons, 2009.

- COMON, P. Independent component analysis, a new concept? *Signal processing*, Elsevier, v. 36, n. 3, p. 287–314, 1994.
- COMON, P. Tensors: a brief introduction. In: *IEEE Sig. Proc. Magazine*. [S.l.: s.n.], 2014. v. 31, n. 3.
- COMON, P. *et al.* Performance of contrast-based blind source separation. In: IEEE. *Signal Processing Advances in Wireless Communications, First IEEE Signal Processing Workshop on*. [S.l.], 1997. p. 345–348.
- COMON, P.; JUTTEN, C. *Handbook of Blind Source Separation: Independent component analysis and applications*. [S.l.]: Academic press, 2010.
- COMON, P.; MOREAU, E. Improved contrast dedicated to blind separation in communications. In: IEEE. *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*. [S.l.], 1997. v. 5, p. 3453–3456.
- DOGAN, M. C.; MENDEL, J. M. Cumulant-based blind optimum beamforming. *IEEE Transactions on Aerospace and Electronic Systems*, IEEE, v. 30, n. 3, p. 722–741, 1994.
- GERVEN, S. V.; COMPERNOLLE, D. V. Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness. *IEEE Transactions on Signal Processing*, IEEE, v. 43, n. 7, p. 1602–1612, 1995.
- GOLUB, G. H.; LOAN, C. F. V. *Matrix computations*. [S.l.]: JHU Press, 2012.
- GOOCH, R.; LUNDELL, J. The cm array: An adaptive beamformer for constant modulus signals. In: IEEE. *IEEE International Conference on ICASSP'86 Acoustics, Speech, and Signal Processing*. [S.l.], 1986. v. 11, p. 2523–2526.
- GOROKHOV, A.; LOUBATON, P. Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, IEEE, v. 44, n. 9, p. 813–820, 1997.
- GU, F. *et al.* Blind identification of underdetermined mixtures with complex sources using the generalized generating function. *Circuits, Systems, and Signal Processing*, Springer, p. 1–13, 2014.
- GU, F. *et al.* Generalized generating function with tucker decomposition and alternating least squares for underdetermined blind identification. *EURASIP Journal on Advances in Signal Processing*, Springer, v. 2013, n. 1, p. 1–9, 2013.
- HÉRAULT, J. *et al.* Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In: GRETSI, GROUPE D'ETUDES DU TRAITEMENT DU SIGNAL ET DES IMAGES. *10 Colloque sur le traitement du signal et des images, FRA, 1985*. [S.l.], 1985.
- HUBER, P. J. Projection pursuit. *The annals of Statistics*, JSTOR, p. 435–475, 1985.
- HUNYADI, B. *et al.* Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, Springer International Publishing, v. 2014, n. 1, 2014. Disponível em: <<http://dx.doi.org/10.1186/1687-6180-2014-139>>.

- HYVARINEN, A. One-unit contrast functions for independent component analysis: A statistical analysis. In: IEEE. *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*. [S.l.], 1997. p. 388–397.
- HYVÄRINEN, A. New approximations of differential entropy for independent component analysis and projection pursuit. In: *Neural Information Processing Systems*. [S.l.: s.n.], 1998. v. 10, p. 273–279.
- HYVARINEN, A. Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, IEEE, v. 10, n. 3, p. 626–634, 1999.
- HYVÄRINEN, A. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 371, n. 1984, p. 20110534, 2013.
- HYVÄRINEN, A. *et al.* *Independent component analysis*. [S.l.]: John Wiley & Sons, 2004.
- KIM, T. *et al.* Independent vector analysis: definition and algorithms. In: IEEE. *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*. [S.l.], 2006. p. 1393–1396.
- KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. *SIAM review*, SIAM, v. 51, n. 3, p. 455–500, 2009.
- LACOUME, J.-L. *et al.* Statistiques d'ordre supérieur pour le traitement du signal. 1997.
- LATHAUWER, L. D. Decompositions of a higher-order tensor in block terms-part ii: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, SIAM, v. 30, n. 3, p. 1033–1066, 2008.
- LATHAUWER, L. D. Blind separation of exponential polynomials and the decomposition of a tensor in rank- $(L_r, L_r, 1)$ terms. *SIAM Journal on Matrix Analysis and Applications*, SIAM, v. 32, n. 4, p. 1451–1474, 2011.
- LATHAUWER, L. D. *et al.* An introduction to independent component analysis. *Journal of chemometrics*, v. 14, n. 3, p. 123–149, 2000.
- LEE, D. D.; SEUNG, H. S. Algorithms for non-negative matrix factorization. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2001. p. 556–562.
- LEE, H. *et al.* Nonnegative tensor factorization for continuous eeg classification. *International journal of neural systems*, World Scientific, v. 17, n. 04, p. 305–317, 2007.
- MCCULLAGH, P. *Tensor methods in statistics*. [S.l.]: Chapman and Hall London, 1987.
- MOOR, B. D. *Database for the identification of systems (DaISy)*. [S.l.]: EE Dept., ESAT/SCD, KU Leuven, Belgium, <http://www.esat.kuleuven.ac.be/sista/daisy>, 2010.
- NIKIAS, C. L.; MENDEL, J. M. Signal processing with higher-order spectra. *IEEE signal processing magazine*, v. 10, n. 3, p. 10–37, 1993.

- NIKNAZAR, M. *et al.* Robust 3-way tensor decomposition and extended state kalman filtering to extract fetal ecg. In: IEEE. *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*. [S.l.], 2013. p. 1–5.
- NOCEDAL, J.; WRIGHT, S. J. Numerical optimization, second edition. *Numerical optimization*, Springer New York, p. 497–528, 2006.
- O'GRADY, P. D. *et al.* Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*, Wiley Online Library, v. 15, n. 1, p. 18–33, 2005.
- PAATERO, P. Least squares formulation of robust non-negative factor analysis. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 37, n. 1, p. 23–35, 1997.
- PAATERO, P.; TAPPER, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, Wiley Online Library, v. 5, n. 2, p. 111–126, 1994.
- PEDERSEN, M. S. *et al.* A survey of convolutive blind source separation methods. *Multichannel Speech Processing Handbook*, Springer, p. 1065–1084, 2007.
- PUIGT, M. *et al.* Validity of the independence assumption for the separation of instantaneous and convolutive mixtures of speech and music sources. In: *Independent Component Analysis and Signal Separation*. [S.l.]: Springer Berlin Heidelberg, 2009. p. 613–620.
- REAZ, M. B. I.; WEI, L. S. Adaptive linear neural network filter for fetal ecg extraction. In: IEEE. *Intelligent Sensing and Information Processing, 2004. Proceedings of International Conference on*. [S.l.], 2004. p. 321–324.
- ROLL, J. Contribution de la proprioception musculaire à la perception et au contrôle du mouvement chez l'homme (thèse de doctorat ès sciences). *Université Aix-Marseille I*, 1981.
- SHAMSUNDER, S.; GIANNAKIS, G. B. Multichannel blind signal separation and reconstruction. *IEEE Transactions on Speech and Audio Processing*, IEEE, v. 5, n. 6, p. 515–528, 1997.
- SLAPAK, A.; YEREDOR, A. Charm-based estimator for non-gaussian moving-average process. In: IEEE. *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*. [S.l.], 2012. p. 1–5.
- SLAPAK, A.; YEREDOR, A. Charrelation and charm: Generic statistics incorporating higher-order information. *Signal Processing, IEEE Transactions on*, IEEE, v. 60, n. 10, p. 5089–5106, 2012.
- SLAPAK, A.; YEREDOR, A. Charrelation matrix based ica. In: *Latent Variable Analysis and Signal Separation*. [S.l.]: Springer, 2012. p. 107–114.
- SMITH, D. *et al.* An analysis of the limitations of blind signal separation application with speech. *Signal Processing*, Elsevier, v. 86, n. 2, p. 353–359, 2006.

- SÖRNMO, L.; LAGUNA, P. *Bioelectrical signal processing in cardiac and neurological applications*. [S.l.]: Academic Press, 2005.
- STARCK, J.-L.; FADILI, M. J. An overview of inverse problem regularization using sparsity. In: *Image Processing (ICIP), 2009 16th IEEE International Conference on*. [S.l.: s.n.], 2009. p. 1453–1456. ISSN 1522-4880.
- TALEB, A.; JUTTEN, C. Source separation in post-nonlinear mixtures. *Signal Processing, IEEE Transactions on*, IEEE, v. 47, n. 10, p. 2807–2820, 1999.
- TALWAR, S. *et al.* Blind estimation of multiple co-channel digital signals arriving at an antenna array. In: IEEE. *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993*. [S.l.], 1993. p. 349–353.
- TOMASI, G.; BRO, R. A comparison of algorithms for fitting the parafac model. *Computational Statistics & Data Analysis*, Elsevier, v. 50, n. 7, p. 1700–1734, 2006.
- WANG, B.; PLUMBLEY, M. D. Musical audio stream separation by non-negative matrix factorization. In: *Proc. DMRN summer conf.* [S.l.: s.n.], 2005. p. 23–24.
- WANG, Z. *et al.* Single-trial decoding of bistable perception based on sparse nonnegative tensor decomposition. *Computational intelligence and neuroscience*, Hindawi Publishing Corporation, v. 2008, 2008.
- WEINSTEIN, E. *et al.* Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech and Audio Processing*, IEEE, v. 1, n. 4, p. 405–413, 1993.
- ZARZOSO, V.; COMON, P. Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *Neural Networks, IEEE Transactions on*, IEEE, v. 21, n. 2, p. 248–261, 2010.
- ZARZOSO, V. *et al.* Maternal and foetal ecg separation using blind source separation methods. *Mathematical Medicine and Biology*, IMA, v. 14, n. 3, p. 207–225, 1997.
- ZARZOSO, V.; NANDI, A. K. Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. *Biomedical Engineering, IEEE Transactions on*, IEEE, v. 48, n. 1, p. 12–18, 2001.
- ZARZOSO, V.; NANDI, A. K. Blind mimo equalization with optimum delay using independent component analysis. *International Journal of Adaptive Control and Signal Processing*, Wiley Online Library, v. 18, n. 3, p. 245–263, 2004.