

Caminhos da Matemática

História,
Educação e
Aplicações

Bruno Rodrigues de Oliveira
Alan Mario Zuffo
Jorge González Aguilera
(Organizadores)



Pantanal Editora

2019

Bruno Rodrigues de Oliveira
Alan Mario Zuffo
Jorge González Aguilera
(Organizadores)

Caminhos da Matemática

História, Educação e Aplicações



Pantanal Editora

2019

Copyright© Pantanal Editora
Copyright do Texto© 2019 Os Autores
Copyright da Edição© 2019 Pantanal Editora
Editor Chefe: Prof. Dr. Alan Mario Zuffo
Editores Executivos: Prof. Dr. Jorge González Aguilera
Prof. Dr. Bruno Rodrigues de Oliveira

Diagramação: Os editores
Edição de Arte: Os editores e autores
Revisão: Os Autores

Conselho Editorial

- Prof^a. Dr^a. Albys Ferrer Dubois – UO
- Prof. Dr. Caio Cesar Enside de Abreu – UNEMAT
- Prof. Msc. David Chacon Alvarez – UNICENTRO
- Prof. Dr. Fábio Steiner - UEMS
- Prof. Msc. Javier Revilla Armesto – UCG
- Prof. Dr. Julio Cezar Uzinski – UFMT
- Prof. Msc. Lucas Rodrigues Oliveira – Município de Chapadão do Sul
- Prof. Dr. Rafael Felipe Ratke – UFMS
- Prof. Dr. Wéverson Lima Fonseca – UFC
- Prof^a. Dr^a. Yilan Fung Boix – UO

Conselho Técnico Científico

- Esp. Joacir Mário Zuffo Júnior - UNEMAT
- Esp. Maurício Amormino Júnior - UFMG

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
C183	Caminhos da matemática [recurso eletrônico] : história, educação e aplicações / Organizadores Bruno Rodrigues de Oliveira, Alan Mario Zuffo, Jorge González Aguilera. – Nova Xavantina, MT: Pantanal Editora, 2019. 138 p. : il. Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web ISBN 978-65-81460-02-0 1. Matemática – Estudo e ensino. 2. Professores de matemática – Formação. I. Oliveira, Bruno Rodrigues de. II. Zuffo, Alan Mario. III. Aguilera, Jorge González. CDD 510.07
Elaborado por Maurício Amormino Júnior – CRB6/2422	

O conteúdo dos livros e capítulos, seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. O download da obra é permitido e o compartilhamento desde que sejam citadas as referências dos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Pantanal Editora
Rua Abaete, 83, Sala B, Centro. CEP: 78690-000. Nova Xavantina – Mato Grosso - Brasil
Telefone (66)99682-4165 (Whatsapp)
<https://www.editorapantanal.com.br>
contato@editorapantanal.com.br

Apresentação

As principais vertentes no estudo das ciências da matemática são: história da matemática, aspectos relativos ao ensino e aprendizagem da disciplina, o estudo da matemática por si mesma, denominada de matemática pura e também as aplicações das teorias matemáticas.

Neste livro são apresentadas discussões sobre questões relativas a história da matemática, a educação e aplicações.

No primeiro tópico, os autores elaboram um texto que resgata a história do primeiro matemático brasileiro a obter o doutoramento, destacando suas principais contribuições.

No segundo tópico, é apresentada uma pesquisa sobre o ensino/aprendizagem da disciplina de matemática financeira no nível superior, empregando a metodologia de ensino e avaliação *Team Based Learning*.

Mesclando os tópicos de aplicação e educação, temos um trabalho sobre as funções e equações exponenciais e logarítmicas, tratando um pouco sobre seu desenvolvimento e importância históricos e sua utilização até os dias atuais.

Por último, relativo ao tópico de aplicação, duas pesquisas ilustram diferentes teorias matemáticas, concernentes a análise de dados e equações diferenciais, exibindo suas aplicações na análise de sinais de voz e estudo de circuitos elétricos, respectivamente.

Assim sendo, este livro trilha apenas uns pouquíssimos caminhos construídos pela matemática. Pretendemos que esta obra seja ampliada para que esta disciplina tão essencial para o desenvolvimento da ciência possa ser melhor compreendida em suas mais diferentes abordagens.

Bruno Rodrigues de Oliveira

Alan Mario Zuffo

Jorge González Aguilera

SUMÁRIO

Capítulo 1	5
Joaquim Gomes De Souza (1829-1864): A Construção de uma Imagem de Souzinha	
Capítulo 2	37
Aplicações de funções e equações exponenciais e logarítmicas	
Capítulo 3	51
Ensino de matemática financeira no curso de licenciatura em matemática e bacharelado em ciências contábeis: reflexões dos impactos no processo de ensino/aprendizagem e/ou no mercado de trabalho	
Capítulo 4	61
Extração de Características em Sinais de Voz por meio da Análise de Componentes Independentes	
Capítulo 5	76
Equações Diferenciais Ordinárias na Aplicação de Circuitos Elétricos	

Extração de Características em Sinais de Voz por meio da Análise de Componentes Independentes

Bruno Rodrigues de Oliveira^{1*}

INTRODUÇÃO

Dentre os algoritmos de aprendizagem não supervisionados utilizados nas pesquisas de processamento de sinais, ICA (*Independent Component Analysis*) tem despertado bastante interesse na comunidade científica. Seu sucesso se deve em parte, a sua extensa aplicabilidade nos mais variados problemas, dos quais: 1. Separação cega de sinais de voz, de imagens, de instrumentos musicais, e também aplicações médicas como análise de EEG (Eletroencefalograma), ECG (Eletrocardiograma), ECG fetal, onde os sinais são obtidos misturados em cada eletrodo; 2. No estudo de séries temporais de dados financeiros com o objetivo de detectar anomalias, fornecendo o real impacto do retorno anormal (Franco, 2008); 3. Extração de características dos sinais, que permite, por exemplo, reconhecimento de locutor, de fonemas e caracteres (Kwon e Lee, 2004), redução de ruído (Lee *et al.*, 2000) e compressão. 4. Na área de processamento de texto combinado a linguística, na extração de conceitos, onde um conceito é visto como uma combinação linear de verbos e predicados, e a análise de componentes independentes extrai informações sobre as relações dos verbos e substantivos (Chagnaa *et al.*, 2007).

Outra peculiaridade do método ICA e que também contribuiu para sua popularização, é que ele utiliza estatísticas de ordem superior a dois, como a curtose. Tais estatísticas possuem melhores informações sobre as características dos sinais, pois estatísticas de ordem elevada capturam melhor a variabilidade dos dados (Ozawa e Kotami, 2014). Vários trabalhos em reconhecimento de padrões tem utilizado essa propriedade e obtidos bons resultados, dentre eles: 1. Lee *et al.* (2000), apresenta taxa de erro no percentual de 2.0%, utilizando 20 vetores bases, na tarefa de reconhecimento de locutor, enquanto que o método que utiliza análise mel-cepstral obteve 3.8% de erro; e 2. Kwon e Lee (2004) no estudo de reconhecimento de fonemas,

¹ Rua Londrina, 2038, Sibipiruna, Chapadão do Sul-MS.

* Autor de Correspondência: bruno@cerradosites.com.

que utiliza em conjunto o método PCA (*Principal Component Analysis*) para redução da dimensão dos dados, obtiveram taxas de acerto de aproximadamente 50%.

A última característica que contribuiu para o sucesso do ICA é o algoritmo de ponto fixo FastICA, que possibilitou a utilização em larga escala devido sua eficiência computacional e ainda a facilidade de implementação.

Segundo Hyvärinen *et al.* (2001) em meados de 1980 J. Hérou, C. Jutten e B. Ans propuseram um método que utilizava uma simples rede neural e um modelo de decorrelação não linear para resolver o problema da codificação de movimento em contração muscular, onde sensores, as saídas x_1, x_2 , mediam a contração muscular e as entradas s_1, s_2 , a posição angular e velocidade do movimento. Embora ainda nessa época o nome ICA não tenha sido cogitado, o modelo $x(t) = As(t)$ de mistura de fontes, foi utilizado.

No início década de 90, A. Cichocki e R. Unbehauen, propuseram novos métodos para o problema da separação de fontes, introduzindo a técnica conhecida como PCA não-linear. No entanto as pesquisas nesta época estavam restritas aos grupos franceses, não tendo alcançado interesse internacional.

Somente em meados dos anos 90, após avanços no campo da otimização estatística, cresceu o interesse pela técnica ICA após a publicação do artigo de A. J. Bell e T. J. Sejnowski abordando o princípio INFOMAX para a resolução do problema da separação de fontes, com a contribuição de S. I. Amari que introduziu a utilização do gradiente natural. Alguns anos mais tarde os mesmos autores trabalharam para desenvolvimento do algoritmo de ponto fixo FastICA.

Hoje grande parte das pesquisas em ICA se concentra em universidades da Finlândia, Coreia e Japão. Na universidade de Helsinki, Finlândia, vários softwares tem sido desenvolvidos para a implementação do ICA, inclusive o pacote para MATLAB® denominado FastICA (Hyvärinen *et al.*, 2001).

ASPECTOS TEÓRICOS

ANÁLISE DE COMPONENTES INDEPENDENTES

A Análise de Componentes Independentes (ICA) é um método estatístico multivariado inicialmente proposto para resolver o problema da separação de fontes num contexto neurofisiológico (Hyvärinen *et al.*, 2001).

Neste problema consideramos um sistema MIMO (*Multiple-Input-Multiple-Output*) onde p sinais de entrada (fontes) são combinados linearmente em n sinais de saída (sensores) (Cichocki *et al.*, 2009):

$$\mathbf{x}_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{ip}s_p \quad \forall i = 1, \dots, n \quad (1)$$

sendo s_i ($i = 1, \dots, p$) variáveis latentes mutuamente estatisticamente independentes; a_{ij} ($i, j = 1, \dots, m$) coeficientes reais que ponderam as fontes para os sinais de mistura \mathbf{x}_i , sendo estes os únicos valores diretamente observáveis. Em notação matricial/vetorial se tem:

$$\begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \dots & \ddots & \dots \\ a_{n1} & \dots & a_{np} \end{bmatrix} \begin{bmatrix} s_1 \\ \dots \\ s_p \end{bmatrix} \Rightarrow \mathbf{x} = A\mathbf{s} \quad (2)$$

onde a matriz $A = (a_{ij})$ é chamada de matriz de mistura. Como seus valores são desconhecidos, assim como os das fontes s_i (dai o nome “separação cega”), resolver o problema da Separação Cega de Fontes, *Blind Source Separation* (BSS) é estimar $W = A^+$ encontrando assim os sinais de entrada – as fontes estimadas $\hat{\mathbf{s}}$, onde A^+ é a pseudo-inversa da matriz A . Ou seja:

$$\hat{\mathbf{s}} = W\mathbf{x} \quad (3)$$

A independência estatística acima mencionada implica que, se $\rho_i(s_i)$ é a função densidade de probabilidade marginal de cada fonte s_i e $\rho(s_1, s_2, \dots, s_p)$ a função densidade de probabilidade conjunta das s_i , então as fontes s_i são mutuamente estatisticamente independentes se, e somente se:

$$\rho(s_1, s_2, \dots, s_i) = \rho_1(s_1)\rho_2(s_2)\dots\rho_p(s_p) \quad (4)$$

Em outras palavras, dizer que as fontes são estatisticamente independentes significa dizer que a informação de uma determinada fonte s_p não diz nada sobre outra fonte s_q , ou seja, as fontes não estão relacionadas.

Em situações reais, principalmente quando se trata de sinais de voz, essa exigência da independência das fontes não restringe a aplicabilidade do método ICA, mas pelo contrário, facilita sua implementação visto que sinais naturais geralmente são independentes.

Outra restrição que se deve considerar diz respeito a distribuição das fontes: que essas devem ter distribuição não-gaussiana, ou somente uma das fontes terem distribuição gaussiana. Isso se deve ao fato do método ICA utilizar estatísticas de alta ordem, como a curtose, e esta, por exemplo, tem valor nulo para variáveis aleatórias com distribuição gaussiana (Hyvärinen *et al.*, 2001).

Para estimar as fontes \mathbf{s} na equação (2), partindo do modelo $\mathbf{s} = A^+\mathbf{x} = W\mathbf{x}$, as componentes independentes (ICs) são estimadas considerando uma combinação linear dos \mathbf{x}_j ,

denotada por $\mathbf{y} = \mathbf{b}^T \mathbf{x}$, sendo \mathbf{b} um vetor a ser determinado. Isso implica que $\mathbf{y} = \mathbf{b}^T \mathbf{A} \mathbf{s}$, logo, \mathbf{y} também é uma combinação linear dos s_j . Denotando $\mathbf{b}^T \mathbf{A}$ por \mathbf{q}^T , tem-se $\mathbf{y} = \mathbf{b}^T \mathbf{A} \mathbf{s} = \mathbf{q}^T \mathbf{s}$.

Então se \mathbf{b} é uma das colunas de \mathbf{W} , $\mathbf{b}^T \mathbf{x} = [b_1 \ b_2 \ \dots \ b_n][x_1 \ x_2 \ x_3 \ \dots \ x_n]^T = s_j$ é uma das ICs, portanto, $s_j = \mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s} \Rightarrow \mathbf{q} = [0 \ \dots \ 1_j \ \dots \ 0]^T$, ou seja, apenas um elemento de \mathbf{q} é 1 e o restante igual a 0. Como \mathbf{A} é desconhecida, não se pode determinar \mathbf{b} exatamente, deve-se então estimá-lo.

Do Teorema Central do Limite, sabe-se que a soma de duas ou mais variáveis aleatórias é mais gaussiana que as variáveis originais. Então $\mathbf{y} = \mathbf{q}^T \mathbf{s} = q_1 s_1 + q_2 s_2 + \dots + q_p s_p$ é mais gaussiana que qualquer s_j , e torna-se menos gaussiana quando de fato é igual a um dos s_j , o que ocorre quando $\mathbf{q} = [0 \ \dots \ 1_j \ \dots \ 0]^T$, o que por sua vez implica que todos os s_j tem distribuição idêntica.

A ideia então é variar os coeficientes de \mathbf{q} e observar como a distribuição de $\mathbf{y} = \mathbf{q}^T \mathbf{s}$ muda. Mas como não se conhece \mathbf{q} faz-se isso com $\mathbf{b}^T \mathbf{x}$, variando os coeficientes de \mathbf{b} . Se variando os coeficientes de \mathbf{b} , encontra-se um $\mathbf{q} = [0 \ \dots \ 1_j \ \dots \ 0]^T$, então se encontra uma das ICs s_j .

Em outras palavras, deve-se encontrar um \mathbf{q} que tenha a maior quantidade de elementos iguais a zero, exceto um, ou seja, caminha-se no sentido de encontrar um vetor \mathbf{q} que torne a distribuição de \mathbf{y} menos gaussiana (mais não-gaussiana) quanto possível, já que quanto mais valores de \mathbf{q} forem não zero mais próximo a distribuição de \mathbf{y} estará da gaussiana, logo, nesse sentido se está maximizando a não-gaussianidade. Realizando manipulações algébricas concluímos que \mathbf{b} deve ser tal que:

$$\mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s} \Rightarrow \mathbf{b}^T \mathbf{A} \mathbf{s} = \mathbf{q}^T \mathbf{s} \Rightarrow \mathbf{b}^T \mathbf{A} = \mathbf{q}^T \Rightarrow (\mathbf{b}^T \mathbf{A})^T = (\mathbf{q}^T)^T \Rightarrow \mathbf{A}^T \mathbf{b} = \mathbf{q} \quad (5)$$

NEGENTROPIA

Uma medida de não-gaussianidade que pode ser utilizada para estimação da matriz \mathbf{W} é a negentropia, que é baseada numa quantidade da entropia diferencial. A entropia de um vetor aleatório \mathbf{y} com função densidade $\rho_{\mathbf{y}}(\mathbf{y})$ é definida como $H(\mathbf{y}) = - \int \rho_{\mathbf{y}}(\boldsymbol{\eta}) \log \rho_{\mathbf{y}}(\boldsymbol{\eta}) d\boldsymbol{\eta}$ então a negentropia é

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (6)$$

onde \mathbf{y}_{gauss} é um vetor aleatório gaussiano com mesma correlação e covariância de \mathbf{y} . $J(\mathbf{y})$ goza das seguintes propriedades (Pereira, 2003): i) É sempre não negativa; ii) É zero se, e

somente se, \mathbf{y} tem distribuição gaussiana; iii) É invariante para transformações lineares invertíveis e para mudanças de escala.

Dado o custo computacional para se calcular $J(\mathbf{y})$ utiliza-se uma aproximação $J(\mathbf{y}) \approx \frac{1}{12} [\mathcal{E}\{\mathbf{y}^3\}]^2 + \frac{1}{48} kurt(\mathbf{y})^2$ (Zuben e Attux, 2010), onde $\mathcal{E}\{.\}$ é a esperança matemática. Porém dada a não robustez da curtose, para algumas distribuições de probabilidade, outra aproximação é sugerida. Fazendo-se $\kappa_1 = \frac{1}{12}, \kappa_2 = \frac{1}{48}, f(\mathbf{y}) = \mathbf{y}^3, g(\mathbf{y}) = \mathbf{y}^4$, resultando em $J(\mathbf{y}) \approx \kappa_1 [\mathcal{E}\{f(\mathbf{y})\}]^2 + \kappa_2 [\mathcal{E}\{g(\mathbf{y}) - g(\mathbf{v})\}]^2$, onde \mathbf{v} é uma variável gaussiana de média $\mathbf{0}$ e variância 1. Portanto, a aproximação para a negentropia torna-se (Hyvärinen *et. al.*, 2001):

$$J(\mathbf{y}) \propto [\mathcal{E}\{G(\mathbf{y})\} - \mathcal{E}\{G(\mathbf{v})\}]^2 \quad (7)$$

Algumas escolhas adequadas para a função não-quadrática $G(\mathbf{z})$, para alguma variável aleatória \mathbf{z} são (Hyvärinen *et al.*, 2001): $G_1(\mathbf{z}) = \vartheta^{-1} \log \cosh(\vartheta \mathbf{z})$ e $G_2(\mathbf{z}) = -\exp(-\mathbf{z}^2/2)$, onde $1 \leq \vartheta \leq 2$.

GRADIENTE DESCENDENTE ESTOCÁSTICO

Para maximizar a não-gaussianidade utiliza-se o método do Gradiente Descendente Estocástico, cujo objetivo é minimizar uma função custo, iterativamente, tipicamente dada por

$$C(\mathbf{w}) = \mathcal{E}\{G(\mathbf{w}^T \mathbf{x})\} \quad (8)$$

sendo \mathbf{x} um vetor de observação aleatório, e \mathbf{w} uma das colunas da matriz W .

Para isso toma-se um ponto inicial $\mathbf{w}(0)$, computa o gradiente de $C(\mathbf{w})$ nesse ponto, e move-o na direção do gradiente negativo, até que $\|\mathbf{w}(t) - \mathbf{w}(t - 1)\|$ atinja um valor pequeno, ou seja, caminha-se na direção mais íngreme da curva da função. Tem-se a seguinte regra, para $t = 1, 2, \dots$, e taxa de aprendizagem α :

$$\mathbf{w}(t) = \mathbf{w}(t - 1) - \alpha(t) \frac{\partial C(\mathbf{w})}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}(t-1)} \quad (9)$$

Denotando por $\Delta \mathbf{w} = \mathbf{w}(t) - \mathbf{w}(t - 1)$ a diferença entre o valor novo e o valor anterior de \mathbf{w} , a regra em na equação (9) torna-se:

$$\Delta \mathbf{w} = -\alpha \frac{\partial C(\mathbf{w})}{\partial \mathbf{w}} \Rightarrow \Delta \mathbf{w} \propto - \frac{\partial C(\mathbf{w})}{\partial \mathbf{w}} \quad (10)$$

Isso significa que o vetor $\Delta \mathbf{w}$ tem a mesma direção do vetor gradiente. No entanto, existe um escalar α que pode ter seu tamanho ajustado, e este determina o tamanho do passo na direção do gradiente negativo.

A regra de aprendizagem na equação (9) é conhecida como aprendizagem em lote. Como a função custo na equação (8) depende do cálculo das médias das amostras, o que pode ser

dispendioso computacionalmente, adota-se o paradigma da aprendizagem *on-line*, onde ao invés de considerar as amostras inteiras utiliza-se apenas a última observação do vetor \mathbf{x} , portanto, a esperança matemática na equação (8) pode ser descartada e a função custo assume a forma $C(\mathbf{w}) = G(\mathbf{w}^T \mathbf{x})$ (Hyvärinen *et al.*, 2001).

Uma desvantagem desse método é que para funções não quadráticas como as funções G_1, G_2 acima descritas, que podem possuir vários mínimos locais, as interações no algoritmo ocasionalmente conduzem para algum destes extremos locais ao invés dos globais. Neste caso, o método não consegue escapar deste extremo e fornece então uma solução que não é a desejada. Para evitar este problema escolhas adequadas do valor inicial de $\mathbf{w}(0)$ devem ser feitas. O algoritmo FastICA abaixo anunciado, utiliza um vetor aleatório com norma unitária para $\mathbf{w}(0)$.

GRADIENTE DESCENDENTE ESTOCÁSTICO UTILIZANDO NEGENTROPIA

A ideia principal do algoritmo é iniciar \mathbf{w} aleatoriamente, com norma unitária. Obter a direção na qual o valor da negentropia de $\mathbf{y} = \mathbf{w}^T \mathbf{x}$ cresça e mover \mathbf{w} nessa direção, até que convirja, sendo \mathbf{x} um vetor aleatório que foi branqueado e portanto $\mathcal{E}\{(\mathbf{w}^T \mathbf{x})^2\} = \|\mathbf{w}\|^2 = 1$, sendo $\|\cdot\|$ a norma euclidiana.

Assim, tomando como função custo a negentropia de \mathbf{w} , ou seja, $C(\mathbf{w}) = [\mathcal{E}\{G(\mathbf{w}^T \mathbf{x})\} - \mathcal{E}\{G(\mathbf{v})\}]^2$, e normalizando \mathbf{w} , após cada interação, para que a variância de $\mathbf{w}^T \mathbf{x}$ permaneça constante, pode-se resumir o algoritmo (Hyvärinen *et al.*, 2001) para estimar \mathbf{w} :

$$\Delta \mathbf{w} \propto \mathcal{E}\{G(\mathbf{w}^T \mathbf{x})\} - \mathcal{E}\{G(\mathbf{v})\} \mathcal{E}\{\mathbf{x} g(\mathbf{w}^T \mathbf{x})\} \quad (11)$$

Para garantir que cada \mathbf{w} obtido seja diferente é necessário alguma restrição. Toma-se portanto uma ortogonalização entre as direções de projeção (Zuben e Attux, 2010), utilizando a ortogonalização de Gram-Schmidt com um método deflacionário, onde os vetores \mathbf{w}_p são estimados um por um, sendo que após cada interação subtrai-se de \mathbf{w}_{p+1} as projeções, ou seja, $\mathbf{w}_{p+1} = \mathbf{w}_{p+1} - \sum_{j=1}^{p-1} (\mathbf{w}_{p+1}^T \mathbf{w}_j) \mathbf{w}_j$.

FASTICA

Utilizando a abordagem de ponto-fixo, pode-se melhorar significativamente o algoritmo acima proposto em relação a velocidade de convergência e também a facilidade de implementação. Uma interação de ponto-fixo sugerida para o gradiente na equação (11) é eliminar a parcela $\mathcal{E}\{G(\mathbf{w}^T \mathbf{x})\} - \mathcal{E}\{G(\mathbf{v})\}$ da equação, pois esta de qualquer jeito seria cancelada pela normalização de \mathbf{w} que segue na interação, e multiplicar ambos os lados por $\alpha \mathbf{w}$,

o que melhora as propriedades de convergência (Hyvärinen *et al.*, 2001). Deste modo, obtém-se a seguinte equação iterativa:

$$\begin{aligned} (1 + \alpha)\mathbf{w} &\leftarrow \mathcal{E}\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} + \alpha\mathbf{w} \\ \mathbf{w} &\leftarrow \mathbf{w}/\|\mathbf{w}\| \end{aligned} \quad (12)$$

Uma escolha adequada de α torna FastICA uma aproximação do Método de Newton, resolvendo a equação $\mathcal{E}\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} + \beta\mathbf{w} = \mathbf{0}$. O resultado fornece enfim um método de interação de ponto-fixo (Hyvärinen *et al.*, 2001):

$$\mathbf{w} \leftarrow \mathcal{E}\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - \mathcal{E}\{\mathbf{x}g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w} \quad (13)$$

Abaixo resumimos o algoritmo FastICA, utilizado para obter as colunas \mathbf{w}_p de W , supondo que o vetor \mathbf{x} foi, antes, branqueado e centralizado (Hyvärinen *et al.*, 2001).

1. Defina a quantidade N de componentes a estimar, e faça $p \leftarrow 1$.
2. Escolha um valor inicial aleatoriamente, com norma unitária para \mathbf{w}_p .
3. Faça $\mathbf{w}_p \leftarrow \mathcal{E}\{\mathbf{x}g(\mathbf{w}_p^T\mathbf{x})\} - \mathcal{E}\{\mathbf{x}g'(\mathbf{w}_p^T\mathbf{x})\}\mathbf{w}_p$.
4. Ortogonalize \mathbf{w}_p fazendo $\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1}(\mathbf{w}_p^T\mathbf{w}_j)\mathbf{w}_j$.
5. Normalize \mathbf{w}_p fazendo $\mathbf{w}_p \leftarrow \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|}$. Se \mathbf{w}_p não convergir¹ volte para o passo 3.
6. Faça $p \leftarrow p + 1$. Enquanto $p \leq N$ volte ao passo 2.

ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (PCA) é uma transformação linear capaz de decorrelacionar variáveis correlacionadas projetando-as em um espaço de menor dimensão, onde cada dimensão é denominada de Componente Principal, sendo que a primeira componente principal tem a maior variabilidade dos dados, a segunda componente é perpendicular a primeira e tem a segunda maior variabilidade e assim sucessivamente (Loesch e Hoeltgebaum, 2012).

Dado um vetor aleatório $\mathbf{x} = [x_1 \quad x_2 \quad \dots \quad x_p]^T$, com média zero e variância unitária, afim de obter as componentes principais (PCs), primeiro determinam-se os autovalores λ_i na equação característica:

$$|C_x - \lambda I| = 0 \quad (14)$$

sendo C_x a matriz de covariância de \mathbf{x} , assim definida para p variáveis:

¹ \mathbf{w}_p convergirá quanto o valor novo de \mathbf{w}_p e o anterior tiverem a mesma direção, isto é, o valor do produto escalar é aproximadamente igual a 1.

$$C_x = \begin{bmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix} = \mathcal{E}\{(\mathbf{x} - m_x)(\mathbf{x} - m_x)^T\} \quad (15)$$

onde $c_{ij} = \mathcal{E}\{(x_i - m_i)(x_j - m_j)\}$ é a covariância de x_i, x_j , m_k a média de x_k , e $\mathcal{E}\{x_i x_j\} = \int_{-\infty}^{\infty} x_i x_j \rho_x(\mathbf{x}) d\mathbf{x}$ a esperança de x_i, x_j .

Os autovalores obtidos satisfazem a desigualdade $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Associado a cada autovalor λ_i tem-se um autovetor $\mathbf{v}_i = [v_{i1} \ v_{i2} \ \dots \ v_{ip}]^T$. Logo, a i -ésima componente principal será $\mathbf{y}_i = \mathbf{v}_i \mathbf{x}^T$, o que mostra que os eixos do novo sistema de coordenadas no qual a PCA está projetando os dados, estão na direção dos autovetores \mathbf{v}_i .

Se $\text{var}(\mathbf{x})$ denota a variância da variável \mathbf{x} , então teremos $\text{var}(y_1) > \text{var}(y_2) > \dots > \text{var}(y_p)$. Como as PCs são descorrelacionadas $\text{corr}(y_i, y_j) = 0$, sendo $\text{corr}(\cdot)$ a correlação entre as variáveis, então, y_1 é a primeira componente, y_2 a segunda e assim por diante.

A técnica da PCA pode ser utilizada como pré-processamento na implementação do método ICA: 1. Projetando os dados em um subespaço de menor dimensão, o que reduz algumas componentes menos significativas (Green *et al.*, 2002); e 2. Realizando o branqueamento dos dados (Zeman, 2000).

O branqueamento é uma técnica de pré-processamento implementada na maioria dos algoritmos ICA. Ela é útil, pois após a transformação de branqueamento a matriz de separação W é ortogonal. Assim, os algoritmos ICA restringem sua busca ao espaço das matrizes ortogonais, o que por sua vez reduz o número de parâmetros livres do sistema. Além disso, tais matrizes induzem ao surgimento de propriedades numéricas importantes como maior velocidade de convergência e estabilidade se comparado ao espaço das matrizes em geral (Hyvärinen, 2014).

Uma variável aleatória \mathbf{x} é dita branca, por definição, se os elementos dessa variável tem variância unitária e são descorrelacionados, ou seja, $\mathcal{E}\{\mathbf{x}\mathbf{x}^T\} = I$. Então, dado um vetor aleatório $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$ com média zero, uma transformação B de branqueamento é tal que $\mathbf{z} = B\mathbf{x}$ é branco, onde

$$B = D^{-1/2}V^T \quad (16)$$

sendo $V = [v_1 \ v_2 \ \dots \ v_n]$ uma matriz de autovetores tendo norma unitária, e $D = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_n)$ uma matriz diagonal com os autovalores da matriz de covariância C_x .

EXTRAÇÃO DE CARACTERÍSTICAS

Em geral métodos de extração de características procuram uma metodologia para representar certos dados de alguma forma mais conveniente para determinada finalidade. O objetivo principal na extração de características baseada em ICA é encontrar uma transformação linear tal que os coeficientes resultantes, sejam tão estatisticamente independentes quanto possível (Kwon e Lee, 2004).

Pelos artigos de Lee *et al.* (2000) e Hyvärinen *et al.* (1999), para extrair características de um sinal \mathbf{x} consideramos este sinal segmentado $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$, e aplicamos o algoritmo ICA para obter componentes independentes \mathbf{u} de \mathbf{x} , pela transformação linear:

$$\mathbf{u} = W\mathbf{x} \quad (17)$$

O algoritmo ICA assume que o sinal observado \mathbf{x} é uma combinação linear dos componentes \mathbf{u} , ou seja:

$$\mathbf{x} = W^+\mathbf{u} \quad (18)$$

O objetivo é então estimar W obtendo W^+ , tal que, $W^+W = \Lambda P$, sendo P uma matriz de permutação e Λ uma matriz diagonal de escala. Portando, quando as saídas \mathbf{u} são as fontes \mathbf{s} , ou seja, $\mathbf{u} = \Lambda P\mathbf{s}$, alteradas sua ordem e multiplicadas por alguma constante, as funções base (vetores característicos), são as colunas da matriz W (Bell e Sejnowski, 1996), onde \mathbf{s} são as fontes originais.

As Figuras 1 e 2 exibem a forma de onda de um sinal de voz e seus vetores característicos, respectivamente (Kwon & Lee, 2004). Tais vetores foram obtidos atualizando as matrizes de separação a cada 1000 segmentos, e estão ordenados pela norma l_2 , assim definida

$$\|A\|_2 = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2}.$$

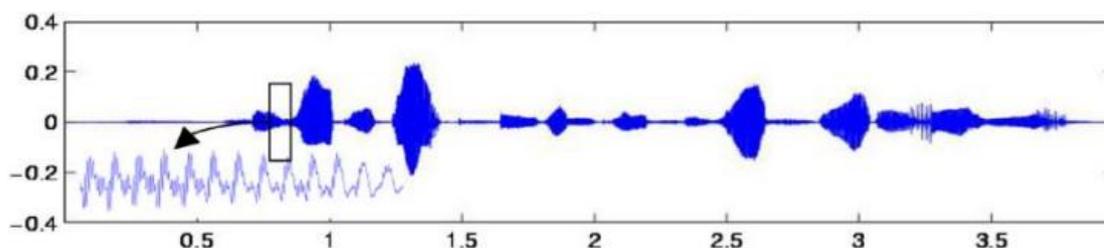


Figura 1. Forma de onda de um sinal de voz. Fonte: Kwon e Lee (2004).

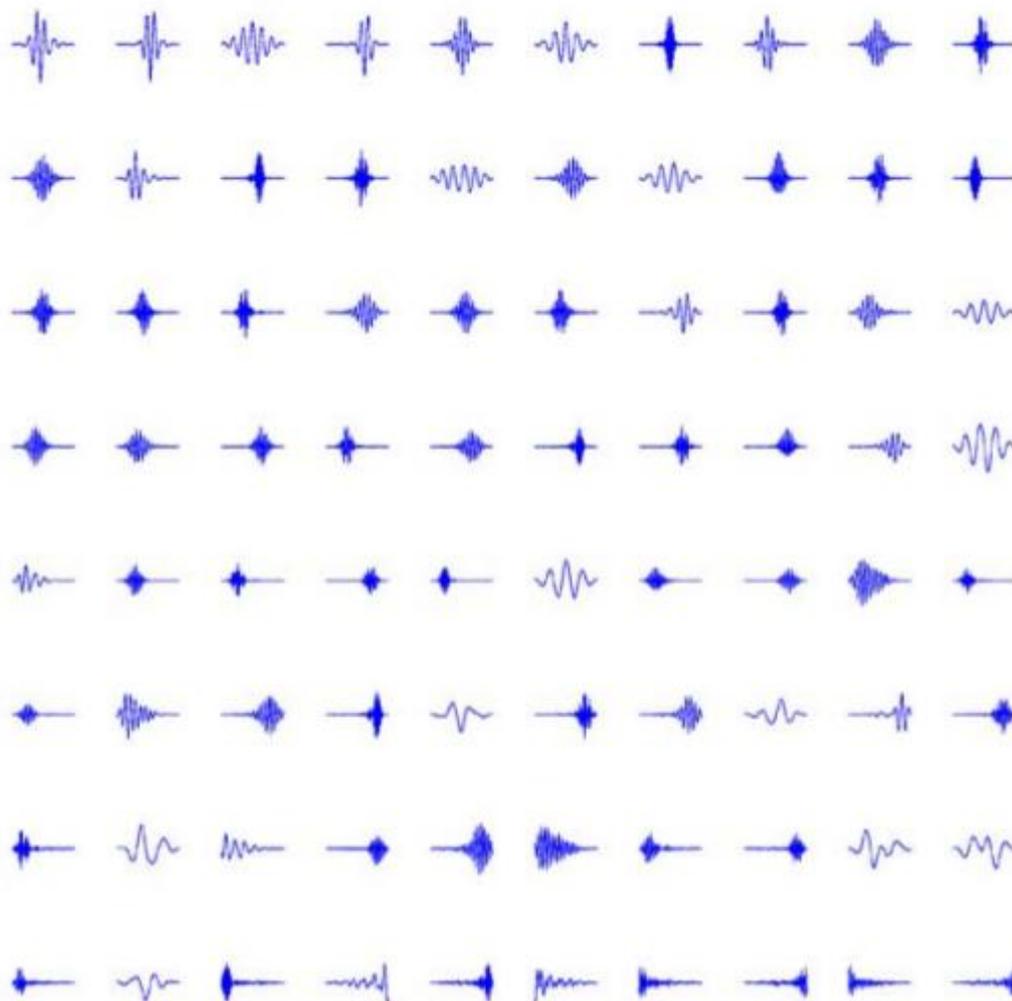


Figura 2. Vetores Característicos do sinal de voz exibido na figura anterior. Fonte: Kwon e Lee, (2004).

REDUÇÃO DE RUÍDO BASEADO NA ICA

CODIFICAÇÃO ESPARSA

Nos métodos de codificação esparsa, uma variável aleatória com média zero é dita esparsa quando sua função densidade de probabilidade (f.d.p) tem um pico em zero e caudas pesadas, ou seja, uma distribuição supergaussiana (Hyvärinen *et al.*, 2001). Como pode ser observado na Figura 3 que exibe algumas formas de ondas de sinais de voz e seus respectivos histogramas, sinais de voz tem esse tipo distribuição.

O método ICA se assemelha a codificação esparsa no sentido que seu objetivo é encontrar as direções em que as componentes decorrelacionadas são tão não-gaussianas quanto possível, enquanto que na codificação esparsa o objetivo é encontrar as componentes tão

esparças quanto possível. Quando as componentes são esparças, que é o caso de sinais de voz, então o ICA busca componentes decorrelacionadas tão esparças quanto possível. Na Figura 3 estão ilustrados três sinais de voz e as respectivas densidades, onde pode-se observar um pico em zero, já que a média foi removida.

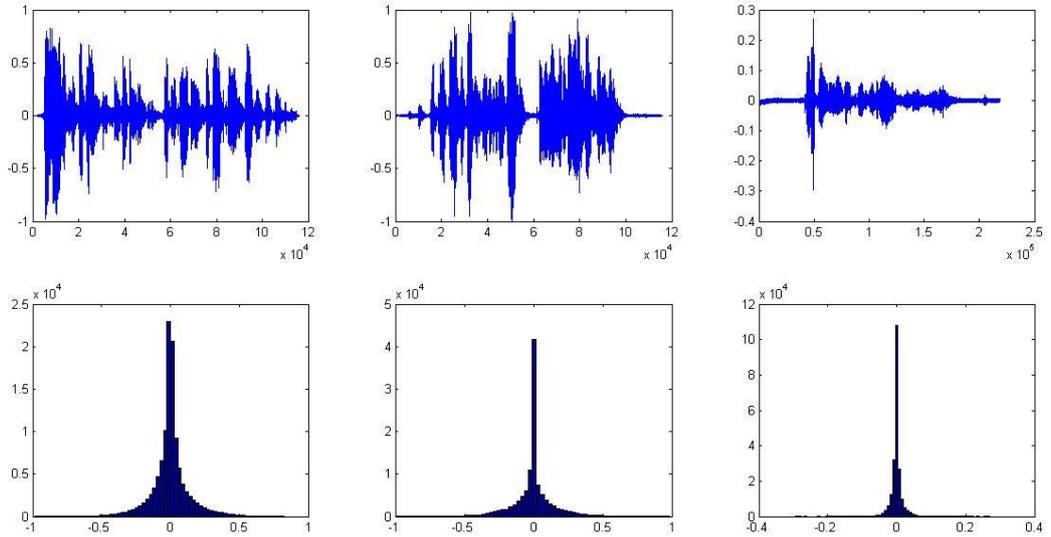


Figura 3. Formas de onda de sinais de voz e seus respectivos histogramas. Fonte: Autor do trabalho.

Em um sinal contaminado por ruído aditivo gaussiano, as componentes esparças têm valores absolutos próximos de zero. Então, estas podem ser consideradas puramente ruído. Para a redução de ruído tenta-se, utilizando uma abordagem de limiar suave, encolher essas componentes esparças, semelhante ao método *Wavelet Shrinkage* (Hyvärinen *et al.*, 1999).

ESTIMADOR PARA REDUÇÃO DE RUÍDO

Para a tarefa de redução de ruído gaussiano de variáveis aleatórias não-gaussianas, que é o caso considerado aqui, deve-se utilizar o estimador MAP (*Maximum a posteriori estimator*), $\hat{\theta}_{MAP}$ que é definido como o vetor de parâmetro θ que maximiza a densidade $\rho_{\theta|x}(\theta|x_T) = \frac{\rho_{x|\theta}(x_T|\theta)\rho_{\theta}(\theta)}{\rho_x(x_T)}$ de θ dada as medições de x_T , sendo $\hat{\theta}_{MAP}$ o valor de ocorrência mais provável do vetor de parâmetros θ dos dados de x_T (Hyvärinen *et al.*, 2001).

Considera-se o sinal ruidoso \tilde{x} formado pelo sinal limpo x e ruído gaussiano v com média zero e variância σ^2 :

$$\tilde{x} = x + v \quad (19)$$

O estimador MAP pode então ser aplicado para fornecer uma versão com ruído reduzido $\hat{\mathbf{x}}$, do sinal ruidoso $\tilde{\mathbf{x}}$, sendo que este estimador fornece a seguinte estimação

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[\frac{1}{2\sigma^2} (\tilde{\mathbf{x}} - \mathbf{x})^2 + f(\mathbf{x}) \right] \quad (20)$$

onde $f(\mathbf{x}) = -\log p(\mathbf{x})$ é o logaritmo negativo da função densidade de probabilidade de \mathbf{x} .

Supondo f estritamente convexa e diferenciável, a minimização na equação (20) é equivalente a resolver a equação $\frac{1}{\sigma^2} (\hat{\mathbf{x}} - \tilde{\mathbf{x}}) + f'(\hat{\mathbf{x}}) = \mathbf{0}$ (Hyvärinen *et al.*, 1999) que fornece a função de encolhimento $h(\tilde{\mathbf{x}})$

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} - \sigma^2 f'(\hat{\mathbf{x}}) = h(\tilde{\mathbf{x}}) \quad (21)$$

que tem a seguinte inversa

$$h(\mathbf{x})^{-1} = \mathbf{x} + \sigma^2 f'(\mathbf{x}) \quad (22)$$

Em casos onde a inversa na equação acima não possa ser calculada, uma aproximação é sugerida por Hyvärinen *et al.* (1999):

$$\hat{\mathbf{x}}^o = \operatorname{sign}(\tilde{\mathbf{x}}) \max(0, |\tilde{\mathbf{x}}| - \sigma^2 |f'(\tilde{\mathbf{x}})|) \quad (23)$$

sendo $\operatorname{sign}(\tilde{\mathbf{x}})$ o sinal de $\tilde{\mathbf{x}}$, e $\max(\cdot)$ o valor máximo do vetor.

Como pode ser observado na equação (22) o cálculo da inversa da função de encolhimento depende do sinal \mathbf{x} , no entanto, não temos esse sinal disponível, mas somente o sinal ruidoso $\tilde{\mathbf{x}}$. Hyvärinen *et al.* (1999) sugere neste caso a utilização de algum sinal livre de ruído com as mesmas características estatísticas que \mathbf{x} para obter $h(\mathbf{x})^{-1}$.

ALGORITMO PARA REDUÇÃO DE RUÍDO

Resumindo o algoritmo para redução de ruído temos os seguintes passos (Lee *et al.*, 2000):

1. Forneça como entrada para o algoritmo FastICA (ou algum outro algoritmo ICA) o sinal ruidoso $\tilde{\mathbf{x}}$ com segmentos de tamanho n , para obter uma matriz $W_{n \times n}$.
2. Estime as funções de encolhimento h_i para cada vetor base (colunas da matriz de separação W), ou utilize a aproximação dada na equação (23).
3. Calcule os coeficientes ruidosos \mathbf{y} do sinal ruidoso $\tilde{\mathbf{x}}$, ou seja, $\mathbf{y} = W\tilde{\mathbf{x}}$.
4. Obtenha os coeficientes sem ruído $\hat{\mathbf{s}} = h(\mathbf{y})$.
5. Recupera-se o sinal com ruído reduzido $\hat{\mathbf{x}} = W^{-1}\hat{\mathbf{s}}$.

Os resultados obtidos por Lee *et al.* (2000) estão exibidos na Figura 4.

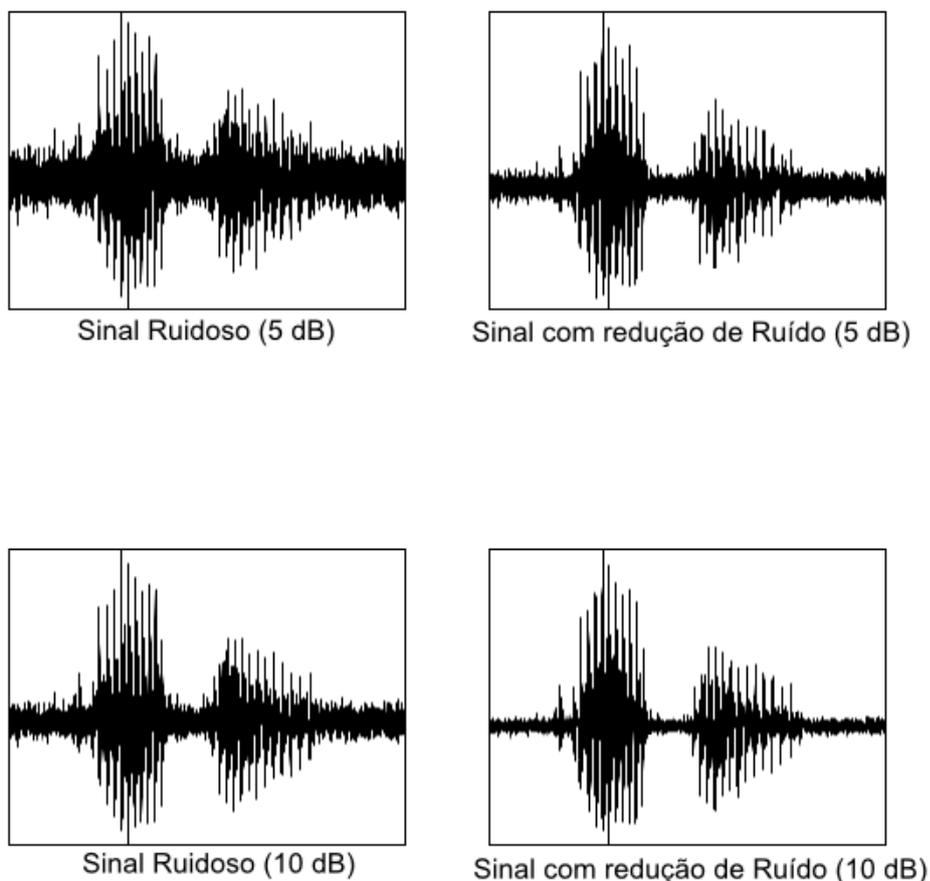


Figura 4. Resultados da redução de ruído implementada por Lee et. al. (2000). Fonte: Adaptado de Lee *et al.* (2000) com legendas traduzidas pelo autor.

CONCLUSÃO

A Análise de Componentes Independentes permite a extração de vetores de características de sinais de voz, que por sua vez podem ser utilizados para diversas finalidades, como na remoção de ruído, e permitem uma representação eficiente destes sinais.

O fato de utilizar estatísticas de ordem superior faz com que este método seja vantajoso, pois algumas informações relevantes dos sinais não são possíveis de serem analisadas utilizando estatísticas de ordem menor que três. A estrutura de fase, que contém as informações espaciais e temporais, as características do sinal, por exemplo, são perceptíveis somente pelas estatísticas de alta ordem. Na tarefa de reconhecimento de locutor, os métodos baseados em ICA tem mostrado vantagem em relação aos métodos baseados na análise cepstral.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bell AJ, Sejnowski TJ (1996). Learning the higher-order structure of a natural sound. *Network: Computation in Neural Systems*, 7: 261-266.
- Cavalcante CC (2004). Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário. Tese de Doutorado, UNICAMP, São Paulo, 297 p.
- Cichocki A, Zdunek R, Phan AH, Amari S (2009). Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and Blind Source Separations. New Jersey: Wiley. 477 p.
- Chagnaa A, Ock C, Lee C, Jaimai P (2007). Feature Extraction of Concepts by Independent Component Analysis. *International Journal of Information Processing Systems*, 3(1):33-37.
- Franco AL (2008). Aplicação da Análise de Componentes Independentes em Estudo de Eventos em Finanças. Tese de Doutorado, UFRS, Porto Alegre, 155 p.
- Green CG, Nandy RR, Cordes D (2002). PCA-Preprocessing of fMRI Data Adversely Affects the Results of ICA. Disponível em <<http://cds.ismrm.org/ismrm-2002/PDF5/1445.PDF>>. Acesso em: 15/04/2014.
- Hyvärinen A (2014). Independent Component analysis: recent advances. Disponível em <<http://rsta.royalsocietypublishing.org/content/371/1984/20110534.full>>. Acesso em: 12/05/2014.
- Hyvärinen A, Karuhunen J, Oja E (2001). Independent Component analysis. New Jersey: Wiley, 479p.
- Hyvärinen A, Karuhunen J, Oja E (1999). Image Denoising by Sparse Code Shrinkage. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.9957>>. Acesso em: 07/04/2014.
- Hyvärinen A (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE – Transactions on Neural Networks*, 10(3):626-634.
- Kwon O, Lee T (2004). Phoneme recognition using ICA-based feature extraction and transformation. *Signal Processing*, 84(6):1005-1019.
- Lee J, Jung H, Lee T, Lee S (2000). Speech Coding and Noise Reduction Using ICA-Based Speech Features. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.27.3670>>. Acesso em: 08/04/2014.
- Lee J, Jung H, Lee T, Lee S (2000). Speech Feature Extraction Using Independent Component Analysis. *IEEE International Conference*, 3:1631-1634.

- Loesch C, Hoeltgebaum M (2012). Métodos Estatísticos Multivariados. São Paulo: Editora Saraiva, 288p.
- Ozawa S, Kotanu M (2014). A Study of Feature Extraction and Seleccion Using Independent Component Analysis. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.2914>>. Acesso em: 06/05/2014.
- Pereira JMCC (2003). Análise de Componentes Independentes na Separação Cega de Fontes Coincidentes no Espaço. Dissertação de Mestrado, Universidade do Porto, Porto – Portugal.
- Zeman T (2000). BSS - Preprocessing Steps for Separation Improvement. Disponível em: <<http://noel.feld.cvut.cz/vyu/prj-czs-asi/zeman/pca.pdf>>. Acesso em: 15/04/2014.
- Zuben FJV, ATTUX RRF (2010). Análise de Componentes Independentes (ICA). Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004_1s10/notas_de_aula/>. Acesso em: 03/04/2014.



Alan Mario Zuffo

Graduado em Agronomia pela UNEMAT. Mestre em Agronomia - Fitotecnia (Produção Vegetal) UFPI. Doutor em Agronomia - Fitotecnia (Produção Vegetal) UFLA. Pós-Doutorado em Agronomia na UEMS. Prof. UFMS em Chapadão do Sul.



Jorge González Aguilera

Graduado em Agronomia pelo ISCA-B (Cuba). Especialista em Biotecnologia pela Universidad de Oriente (Cuba). Mestrado em Fitotecnia e Doutorado em Genética e Melhoramento pela UFV e Post Doutorado na Embrapa Trigo. Prof. UFMS em Chapadão do Sul.



Bruno Rodrigues de Oliveira

Graduado em Matemática pela UEMS. Mestrado em Engenharia Elétrica UNESP-Ilha Solteira e Doutorado em andamento na mesma instituição. Pesquisador independente.

Pantanal Editora

Rua Abaete, 83, Sala B, Centro. CEP: 78690-000
Nova Xavantina – Mato Grosso - Brasil
Telefone (66)99682-4165 (Whatsapp)
<https://www.editorapantanal.com.br>
contato@editorapantanal.com.br

ISBN 978-658146002-0

