



UNIVERSIDADE FEDERAL DO ABC

LABORATÓRIO DE SINAIS E SISTEMAS

**ESTUDO DE ALGORITMOS DE ANÁLISE DE COMPONENTES
INDEPENDENTES PARA SEPARAÇÃO CEGA DE SINAIS DE ÁUDIO**

ALEXANDRE KEITI FUKAMATI

Orientador: Ricardo Suyama

Coorientador: Tito Caco Curimbaba Spadini

Santo André
Setembro de 2021

RESUMO

No problema de *separação cega de fontes* (BSS, *Blind Source Separation*), o objetivo é separar dois ou mais sinais em uma situação em que esses sinais se encontrem misturados. A principal técnica utilizada para a solução desse problema é a chamada *análise de componentes independentes* (ICA, *Independent Component Analysis*), que realiza essa separação partindo do princípio de que esses sinais sejam não gaussianos e independentes entre si. A ICA é uma técnica muito utilizada no tratamento de sinais, tendo diversas aplicações em contextos em que uma mistura de sinais pode afetar a análise de um sinal específico. O presente projeto tem como objetivo realizar a implementação de algoritmos de ICA como forma de obter uma análise comparativa entre eles e, por fim, aplicar esses algoritmos em sinais de áudio.

Sumário

1	INTRODUÇÃO	7
2	FUNDAMENTAÇÃO TEÓRICA	9
2.1	TEORIA DAS PROBABILIDADES	9
2.1.1	Esperança	9
2.1.2	Distribuição Gaussiana	11
2.1.3	Momentos	12
2.1.4	Curtose	14
2.1.5	Descorrelação e Independência	15
2.2	TEORIA DA INFORMAÇÃO	17
2.2.1	Entropia	17
2.2.2	Informação Mútua	18
2.2.3	Negentropia	20
2.3	ANÁLISE DE COMPONENTES PRINCIPAIS	21
2.3.1	Definição	21
2.3.2	PCA pela maximização da variância	22
2.3.3	Branqueamento	24
2.4	ANÁLISE DE COMPONENTES INDEPENDENTES	25
2.4.1	Definição	25
2.4.2	Infomax	26
2.4.3	FastICA	27
3	APLICAÇÃO DE ALGORITMOS	28
3.1	METODOLOGIA EXPERIMENTAL	28
3.2	MÉTRICAS UTILIZADAS	30
3.2.1	Correlação de Pearson	30
3.2.2	Erro quadrático médio	31
3.2.3	Índice de Amari	31
3.2.4	Métricas de avaliação perceptual	32
3.3	EXEMPLO DE APLICAÇÃO	33

4	RESULTADOS E DISCUSSÕES	37
4.1	SINAIS SINTÉTICOS	37
4.1.1	Branqueamento	38
4.1.2	Matriz de inicialização	40
4.1.3	Número de fontes	41
4.2	SINAIS DE ÁUDIO	43
4.2.1	Avaliação por métricas objetivas	43
4.2.2	Avaliação perceptual	44
5	CONCLUSÃO	46

Lista de Figuras

1	Representação esquemática do funcionamento da separação cega de fontes.	7
2	Histogramas com número variado de amostras.	12
3	Formato da curtose para diferentes tipos de distribuição.	15
4	Diagrama de Venn da entropia conjunta entre X e Y , demonstrando de que forma as entropias marginais se relacionam para a obtenção da informação mútua.	20
5	Seta comum apontando para a direita.	21
6	Setas de tamanhos variados.	21
7	Sinais originais S	34
8	Sinais misturados X	34
9	Erros quadráticos entre os sinais de maior correlação. Em cima, para s_1 e y_2 , e embaixo para s_2 e y_1	36
10	Desempenho de cada algoritmo para uma determinada mistura de distribuição de Laplace.	40
11	Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.	41
12	Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.	44
13	Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.	45

Lista de Tabelas

1	Correlações entre os sinais de S e Y	35
2	Métricas de erro quadrático médio entre S e Y	36
3	Métricas de avaliação perceptual e erro quadrático médio entre S e Y .	37
4	Avaliação de algoritmos sem branqueamento para fonte de distribuição de Laplace.	38
5	Avaliação de algoritmos com branqueamento para fonte de distribuição de Laplace.	38
6	Avaliação de algoritmos sem branqueamento para fonte de distribuição uniforme.	39
7	Avaliação de algoritmos com branqueamento para fonte de distribuição uniforme.	39
8	Desempenho dos algoritmos conforme variação no número de fontes para fontes supergaussianas.	42
9	Desempenho dos algoritmos conforme variação no número de fontes para fontes subgaussianas.	42
10	Desempenho para cada algoritmo avaliado pelo índice de Amari, correlação de Pearson e erro quadrático médio.	43
11	Métricas de avaliação perceptual para cada algoritmo.	45

Lista de Siglas e Abreviações

BSS *Blind Source Separation*

EQM Erro Quadrático Médio

ICA *Independent Component Analysis*

PCA *Principal Component Analysis*

SAR *Source-to-Artifacts Ratio*

SDR *Source-to-Distortion Ratio*

SIR *Source-to-Interference Ratio*

SNR *Source-to-Noise Ratio*

1 INTRODUÇÃO

O “problema do coquetel” [1] é um fenômeno muito conhecido na área da psicoacústica. Sua explicação se dá através de uma situação ilustrativa na qual percebe-se que mesmo em uma festa repleta de pessoas e diversos outros elementos distrativos, somos capazes de “selecionar” as informações desejadas e nos comunicarmos naturalmente. Essa situação representa claramente um cenário no qual diversos sinais originalmente independentes se encontram misturados, mas ainda conseguem ser recuperados sem grandes perdas; é nesse princípio que se baseia a *separação cega de fontes* (BSS, *Blind Source Separation*) [2].

O problema de Separação Cega de Fontes se resume à tentativa de recuperar sinais originalmente independentes, a partir de uma situação em que esses sinais estejam misturados. Tal procedimento pode ser aplicado em contextos variados, sendo muito útil na área de telecomunicações e fundamental em situações em que se deseje remover um ruído de uma mistura de sinais através de um processo de separação. Esse problema é representado na Figura 1, que representa um esquemático de como os sinais originais se misturam no ambiente, e podem ser posteriormente separados.

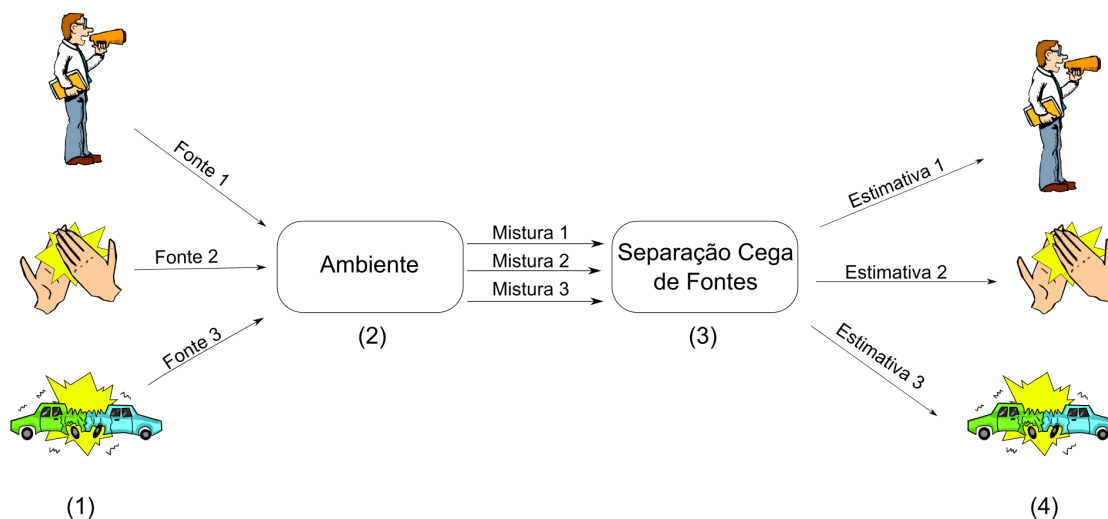


Figura 1: Representação esquemática do funcionamento da separação cega de fontes. Em (1), cada fonte emite seu sinal; em (2), os sinais emitidos pelas fontes são misturados entre si no ambiente; em (3) ocorre o processo de separação das misturas obtidas; e em (4) são obtidas as estimativas dos sinais originais;

Essa separação possibilita o tratamento independente entre esses sinais, de forma que haja a possibilidade de manipulá-los de uma maneira mais precisa e efi-

ciente, como em situações em que um sinal de áudio deva se sobressair em relação aos outros dentro de uma mistura. Esse tipo de problema é recorrente em diversos campos de estudo além da acústica, como por exemplo o tratamento de sinais na área de biomedicina, nos quais frequentemente outros tipos de sinais podem influenciar na obtenção de uma determinada fonte. Assim, surge um cenário muito amplo para o desenvolvimento de métodos para realizar esse processo de separação.

O tratamento de sinais de áudio é um assunto muito estudado no campo de processamento de sinais. Diversas técnicas são desenvolvidas como forma desenvolver esse processo, no entanto, o problema de encontrar sinais misturados é muito recorrente, e a separação pode ser muito útil nesses casos. Um exemplo disto pode ser visto no caso de *softwares* de comunicação, nos quais o sinal da voz captado pelo microfone do usuário pode ocasionalmente ser afetado por ruídos externos, de modo que o sinal principal não seja distribuído de maneira adequada. Diante disto, uma das soluções para este problema é a separação desses ruídos. Desta forma, torna-se evidente a relevância do estudo de métodos para a realização desta separação de sinais, e é nesse contexto que surge a *análise de componentes independentes* (ICA, *Independent Component Analysis*) [3] como uma das possíveis técnicas para a solução do problema de separação cega de fontes.

A análise de componentes independentes é a principal técnica utilizada na área de separação cega de fontes, sendo considerada como uma expansão de uma outra técnica denominada *análise de componentes principais* (PCA, *Principal Component Analysis*) [4]. Desta forma, a compreensão da PCA é fundamental para o entendimento da ICA.

A PCA se trata de uma técnica de análise de dados com o objetivo de extrair informações importantes do conjunto de dados através de transformações ortogonais, obtendo componentes descorrelacionados que comprimem a informação desse conjunto—os chamados *componentes principais*. Ainda que esses componentes não estejam correlacionados, eles precisam ser independentes para que seja possível realizar uma separação entre eles. Com base nisto, é necessário estabelecer critérios para que essas transformações resultem em sinais independentes.

Através de análises estatísticas, chega-se à conclusão de que, para que a separação ocorra sem interferências, os critérios a serem adotados consistem em

duas exigências: 1) que os sinais sejam originalmente independentes; e 2) que os sinais sejam não gaussianos. Assim, partindo desses critérios, é possível aplicar a ICA por diferentes abordagens como forma de realizar essa separação.

A aplicação desta técnica para realizar a separação de sinais de áudio pode solucionar diversos problemas em relação ao tratamento destes sinais, assim, neste estudo serão realizados estudos a respeito dos fundamentos da ICA, e por fim, os algoritmos de separação serão aplicados em sinais de áudio. O principal objetivo é analisar o comportamento e desempenho deste algoritmos quando aplicados a estes tipos de sinais, e entender o funcionamento do processo de separação.

2 FUNDAMENTAÇÃO TEÓRICA

Os princípios e algoritmos da ICA são derivados de diferentes áreas e abordagens. Sendo assim, para realizar um estudo a respeito dessa área, é necessário ter alguma compreensão sobre conceitos básicos de alguns assuntos.

O propósito desta seção é introduzir algumas das principais áreas de estudos utilizadas em ICA, bem como apresentar os principais conceitos destes assuntos.

2.1 TEORIA DAS PROBABILIDADES

Ao lidar com o processamento de sinais, existe uma necessidade constante de estimar valores, criar modelos, realizar análises e executar diversas outras tarefas envolvendo conjuntos de dados. Desta forma, a teoria da probabilidade é uma ferramenta essencial para auxiliar na solução desses problemas.

Desde os princípios da teoria da informação, a teoria das probabilidades foi a base principal para o desenvolvimento dos algoritmos de ICA, possibilitando inúmeras soluções para o problema do BSS. Nesta seção, serão tratados alguns dos principais conceitos desta área.

2.1.1 Esperança

Em teoria da probabilidade, é extremamente comum lidar com variáveis de valor desconhecido. Normalmente, estas variáveis são determinadas por eventos que ocorrem ao acaso, sendo denominadas **variáveis aleatórias**.

Um exemplo de variável aleatória pode ser o lançamento de uma moeda. Sabemos que uma moeda possui os lados *cara* e *coroa*; assim, ao lançar uma moeda, é possível analisar os casos e determinar as probabilidades de qual lado cairá com a face para cima. Entretanto, devido à aleatoriedade deste evento, é impossível determinar assertivamente qual será este lado.

Porém, ainda que seja impossível determinar um valor exato para uma variável aleatória, é possível buscar um valor que a descreva—a **esperança**. A esperança de um evento aleatório é o valor médio deste evento, i.e., o valor que “se espera” dele, consistindo em uma medida que depende apenas do conjunto de dados, facilitando a sua obtenção.

A esperança pode ser calculada de maneiras diferentes, a depender do conjunto de dados em questão. Variáveis aleatórias podem ser divididas em dois principais grupos:

1. Variáveis discretas: Possuem uma determinada quantidade K de valores. Assim, considerando uma variável aleatória X , com valores de x_1 até x_i , considerando que a probabilidade de cada valor x_i é dado por $p(x_i)$, temos que a esperança de X é dada por

$$E\{X\} = \sum_{i=1}^{\infty} x_i p(x_i) , \quad (1)$$

mas vale lembrar que a esperança pode ser considerada como o valor médio de uma variável; sendo assim, caso a probabilidade seja igual para todo x_i , é possível obter a esperança através do cálculo da média aritmética dos valores de X , i.e.,

$$E\{X\} = \frac{1}{K} \sum_{i=1}^K x_i . \quad (2)$$

2. Variáveis contínuas: Os valores são contínuos, então não é possível determinar uma quantidade de valores da variável, logo, não é possível calcular a esperança da mesma forma utilizada para variáveis discretas.

Com isto, considerando que a variável aleatória x é inteiramente contínua com probabilidade $p(x)$, para considerar todos valores de x , sendo x contínuo, podemos utilizar uma integral para calcular a esperança desta variável da seguinte

maneira:

$$E\{X\} = \int_{-\infty}^{\infty} x \cdot p(x) dx \quad (3)$$

2.1.2 Distribuição Gaussiana

Conjuntos de dados podem possuir diferentes tipos de distribuições estatísticas, estes tipos são definidos pela relação que os dados possuem uns com os outros. Cada distribuição possui sua particularidade, impactando na maneira como o conjunto de dados deve ser analisado. Justamente por isto, observar a distribuição de um conjunto de dados é um passo de grande importância em sua análise.

Uma das distribuições estatísticas de maior importância é a *distribuição gaussiana*; para introduzi-la, será realizado um experimento. Para este experimento, considere uma roleta com 12 espaços, cada um com um número inteiro diferente, no intervalo de 1 a 12. A roleta será girada 3 vezes, e cada amostra deste experimento será definida como a soma destes 3 valores. Assim, um número determinado de amostras será coletada e, então, para cada conjunto de amostras, será analisada a distribuição resultante.

No primeiro experimento serão coletadas 5 amostras; no segundo, 50 amostras, seguindo uma progressão geométrica que aumenta em 10 vezes a cada experimento. Ao final, serão realizados 6 experimentos, representados na Figura 2.

A curva observada no último experimento indica que os eventos tendem a se concentrar no valor médio, enquanto as amostras de valores distantes da média acabam sendo bem menos frequentes. Esta curva é conhecida como a curva **gaussiana**, caracterizando a distribuição destes dados como a **distribuição normal** ou **distribuição gaussiana**. Uma variável gaussiana x de esperança μ e desvio padrão σ possui uma *função densidade de probabilidade* definida por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . \quad (4)$$

A esperança e o desvio padrão definem, respectivamente, a média, e a dispersão da distribuição dos dados.

Esta evolução na distribuição amostral do evento para uma distribuição gaussiana é um exemplo do chamado **teorema central do limite**. Este teorema enuncia

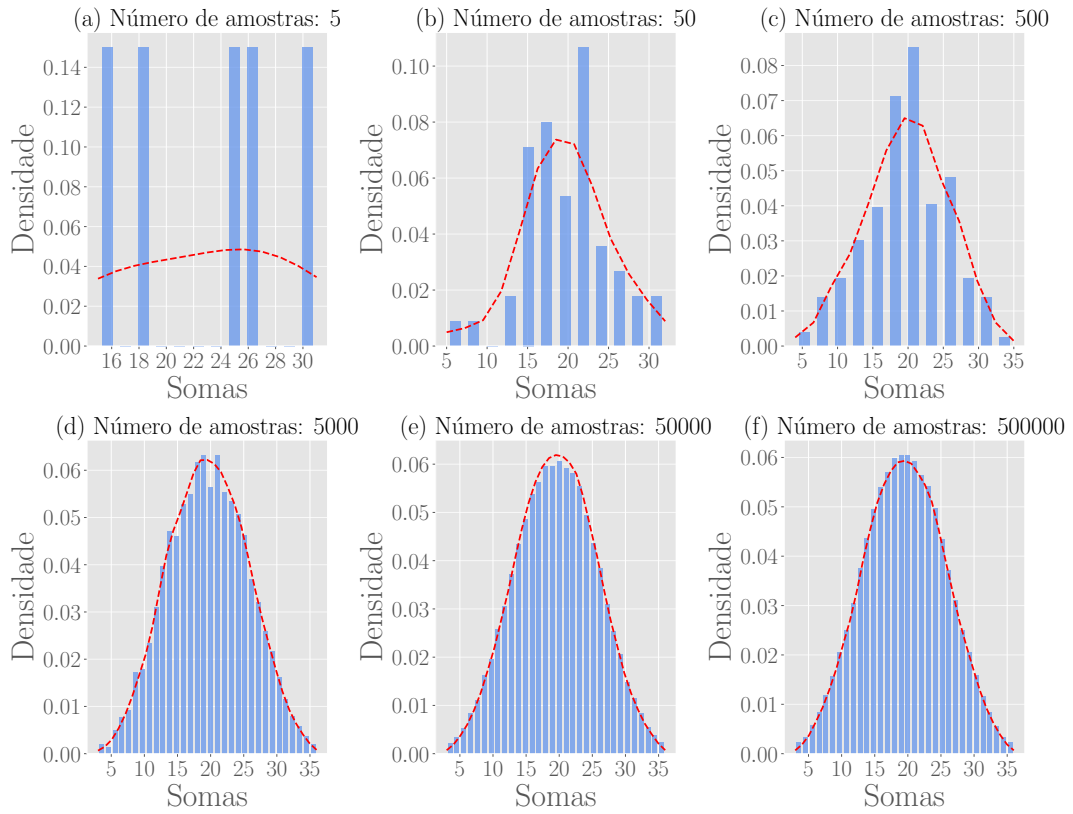


Figura 2: O gráfico (a) representa o experimento realizado com 5 amostras, apresentando uma distribuição muito próxima da uniforme. Em (b), ainda que estejam bem dispersos, é possível observar uma concentração maior no meio do gráfico. Em (c) e (d), isto só se acentua, aumentando significativamente esta concentração no centro do gráfico, que resulta em uma curva de formato sinuoso em (e) e (f).

que, independentemente de sua distribuição, o aumento no número de amostras de uma determinada variável aleatória implica aproximação de sua distribuição amostral a uma distribuição normal. É um teorema fundamental em teoria da probabilidade, devido ao fato de ser válido para qualquer que seja a distribuição da variável em questão, podendo ser útil para manipular diversos tipos de conjuntos de dados.

2.1.3 Momentos

A esperança de uma variável aleatória é uma característica muito importante; no entanto, é possível explorá-la de maneira ainda mais profunda.

Diversos tipos de informação sobre uma variável podem ser obtidas através das

diferentes ordens de sua esperança, estas informações são os chamados **momentos**. Os momentos de uma variável fornecem características de extrema importância. O primeiro momento, por exemplo, é a própria esperança, ou seja, equivale à média. Assim, o n -ésimo momento de uma variável corresponde à esperança da n -ésima potência desta variável. De forma análoga, podemos definir o **momento central**, que corresponde ao cálculo da esperança de uma variável subtraída por sua média.

Desta forma, podemos definir o momento m da ordem k de uma variável x de média \bar{x} como

$$m_k = E\{x^k\} \quad (5)$$

e o momento central μ de ordem k dessa mesma variável como

$$\mu_k = E\{(x - \bar{x})^k\} . \quad (6)$$

É importante ressaltar que o método para o cálculo da esperança pode variar para uma variável contínua ou discreta.

Podemos, então, destacar algumas das principais características que podem ser obtidas a partir de momentos. A variância σ^2 pode ser obtida diretamente a partir do segundo momento central

$$\sigma^2 = E\{(x - \bar{x})^2\} . \quad (7)$$

Assim, pela definição, o desvio padrão σ pode ser obtido a partir da raiz quadrada da variância.

$$\sigma = \sqrt{E\{(x - \bar{x})^2\}} \quad (8)$$

A definição de momento central normalizado é muito útil, pois é utilizada para obter algumas outras características a partir de momentos. A normalização do momento central é feita por uma simples divisão pelo seu desvio padrão σ na mesma ordem do momento em questão

$$\frac{\mu_k}{\sigma^k} = \frac{E\{(x - \bar{x})^k\}}{\sigma^k} . \quad (9)$$

Partindo disto, o terceiro momento central normalizado e o quarto momento central normalizado podem nos fornecer informações de muita relevância. O terceiro momento central normalizado nos fornece a **assimetria**—uma medida relativa ao formato

da distribuição da variável em questão.

$$\frac{\mu_3}{\sigma^3} = \frac{E\{(x - \bar{x})^3\}}{\sigma^3} \quad (10)$$

Uma distribuição com moda, mediana e média iguais, é uma distribuição considerada simétrica, devido ao fato de ter sua curva concentrada em seu centro. Assim, a assimetria mede o quão longe dessas características a distribuição está.

O quarto momento central normalizado determina a **curtose**—uma medida relacionada ao achatamento da curva de probabilidade da distribuição da variável, quantificando o quão achatada ela está. Esta medida é de extrema importância no estudo de ICA, logo, será tratada mais profundamente a seguir.

Momentos também são muito úteis na área de teoria da estimação, na aplicação de uma técnica de estimação de parâmetros chamada *método dos momentos*.

2.1.4 Curtose

Como dito anteriormente, a curtose corresponde ao quarto momento central padronizado, ou seja, se considerarmos uma variável x , podemos obter sua curtose da seguinte maneira:

$$kurt(x) = \frac{\mu_4}{\sigma^4} = \frac{E\{(x - \bar{x})^4\}}{\sigma^4} . \quad (11)$$

No entanto, a variância pode ser representada como na equação (7), assim, obtendo a seguinte outra forma de representar essa relação:

$$kurt(x) = \frac{\mu_4}{\sigma^4} = \frac{E\{(x - \bar{x})^4\}}{(\sigma^2)^2} = \frac{E\{(x - \bar{x})^4\}}{(E\{(x - \bar{x})^2\})^2} . \quad (12)$$

O resultado nos fornece uma medida do achatamento da função densidade de probabilidade da variável, sendo que, quando o resultado é 3, ele corresponde à curva gaussiana. O método anterior é conhecido como **coeficiente de Pearson** para curtose; porém, existe outro modo bastante utilizado para representar este valor—o chamado **coeficiente de Fisher**.

A diferença entre ambos se encontra na medição da curtose para a curva gaussiana, sendo que, pelo método de Pearson, o resultado é 3; pelo método de Fisher, 0. Logo, este segundo coeficiente é obtido a partir de uma simples subtração de 3

unidades em relação ao método de Pearson, resultando em

$$kurt(x) - 3 = \frac{\mu_4}{\sigma^4} - 3 = \frac{E\{(x - \bar{x})^4\}}{(E\{(x - \bar{x})^2\})^2} - 3. \quad (13)$$

Considerando a curtose pelo coeficiente de Fisher, podemos classificar uma distribuição de três diferentes maneiras:

1. $kurt(x) < 0$: A curva é denominada curva **subgaussiana** ou **platicúrtica**.
2. $kurt(x) = 3$: A curva é denominada curva **mesocúrtica**—uma curva **gaussiana**, caracterizando uma distribuição normal
3. $kurt(x) > 3$: A curva é denominada curva **supergaussiana** ou curva **leptocúrtica**.

Na Figura 3, podemos analisar a diferença entre a curtose de cada função densidade de probabilidade.

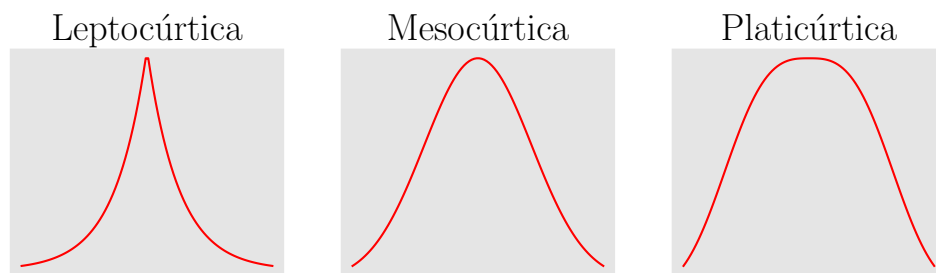


Figura 3: Formato da curtose para diferentes tipos de distribuição.

Medidas de curtose são de muita relevância devido ao fato de ser uma boa forma de se medir a distância de uma determinada distribuição em relação à distribuição normal. Na ICA, um critério de muita importância é a não gaussianidade de um sinal, logo, para trabalhar essa informação, a curtose pode ser uma ferramenta muito interessante.

2.1.5 Descorrelação e Independência

Existem diferentes formas de se medir o impacto que uma variável pode causar em outra, de forma que estas diferentes medidas podem ser de grande importância para realizar análises em relação a estas variáveis. Neste contexto, duas medidas muito utilizadas para realizar a quantificação da relação entre duas variáveis são a **covariância** e a **correlação**.

A covariância tem a função de medir o quanto a variação de uma variável pode impactar na variação de alguma outra variável. Se considerarmos duas variáveis x e y , podemos utilizar momentos para obtê-la da seguinte forma:

$$c_{xy} = E\{(x - \bar{x})(y - \bar{y})\} . \quad (14)$$

Ao considerarmos x e y como vetores, temos

$$C_{xy} = E\{(x - \bar{x})(y - \bar{y})^T\} \quad (15)$$

A correlação é uma medida que, assim como a covariância, mede o quanto a variação de uma variável terá efeito em outra. No entanto, a correlação não considera a média da variável no cálculo da esperança; então, a correlação r_{xy} entre duas variáveis x e y pode ser obtida a partir da esperança conjunta destas duas variáveis, ou seja,

$$r_{xy} = E\{xy\} . \quad (16)$$

Da mesma forma, podemos considerar x e y como vetores, ou seja,

$$R_{xy} = E\{xy^T\} . \quad (17)$$

Assim, podemos considerar que, quando a covariância entre duas variáveis é nula, essas variáveis são consideradas **descorrelacionadas**. No entanto, a descorrelação não garante que variáveis deixem de carregar informações umas sobre as outras. Para determinar essa propriedade, estabelecemos o conceito de **independência**.

Formalmente, variáveis são consideradas independentes quando suas probabilidades marginais se igualam à probabilidade conjunta destas variáveis. Considerando duas variáveis X e Y , a independência estatística ocorre se, e somente se,

$$P(X, Y) = P(X)P(Y) . \quad (18)$$

Variáveis independentes sempre são descorrelacionadas, no entanto, variáveis descorrelacionadas não necessariamente são independentes. Logo, torna-se evidente a necessidade da independência estatística entre variáveis para a ICA, pois é um

critério muito mais forte que a descorrelação. Desta forma, é possível concluir que embora descorrelacionar variáveis não resulte em variáveis independentes, esta técnica pode ser muito útil para facilitar o processo de separação.

2.2 TEORIA DA INFORMAÇÃO

2.2.1 Entropia

Um dos grandes objetivos da teoria da informação é criar formas de transmitir e interpretar a informação. Para isso, um dos principais conceitos utilizados nesta área é a **quantidade de informação**. Considerando uma variável X com variáveis de x_1 a x_n , sendo que cada valor x_i tem probabilidade $p(x_i)$ tal que

$$\sum_{i=1}^n p(x_i) = 1 , \quad (19)$$

a quantidade de informação para cada variável x_i é obtida pelo logaritmo do inverso da probabilidade, ou seja,

$$I(x_i) = \log \left(\frac{1}{p(x_i)} \right) = -\log(p(x_i)) , \quad (20)$$

na qual a base do logaritmo determina a unidade da quantidade de informação, sendo que a base 2 é a mais utilizada, fornecendo a informação em *bits*. Podemos perceber, por exemplo, que, se considerarmos $p(x_i) = 1$, teremos

$$I(x_i) = \log \left(\frac{1}{1} \right) = \log(1) = 0 .$$

Isto se deve ao fato de que, se um evento possui apenas um resultado possível, então nenhuma informação será transmitida por ele, resultando em uma quantidade nula de informação.

No entanto, podemos calcular a quantidade média de informação a partir de uma média ponderada, considerando que cada valor x_i possui peso $p(x_i)$. Levando em conta a equação (19), esta média resulta em

$$H(X) = E\{I(x_i)\} = \frac{\sum_{i=1}^n p(x_i) \log \left(\frac{1}{p(x_i)} \right)}{\sum_{i=1}^n p(x_i)} = -\sum_{i=1}^n p(x_i) \log(p(x_i)) . \quad (21)$$

Em teoria da informação, $H(X)$ é denominada **entropia**. A entropia mede a aleatoriedade de uma variável, representando a incerteza em sua determinação. Isto fica evidente ao analisarmos a sua definição, percebendo que, se $p(x_i)$ for muito próximo de 1, o valor médio será baixo devido ao fato de ser um evento pouco aleatório.

O contrário ocorre quando $p(x_i)$ é dividido em N parcelas iguais, sendo N um número consideravelmente grande. Neste caso, a entropia atingirá um valor numericamente alto, devido ao fato de esse evento possuir um alto grau de aleatoriedade, tendo em vista que pode resultar em um grande número de valores diferentes com probabilidades iguais. Perceba que, se aumentarmos o número de amostras de modo a distribuir a probabilidade entre as amostras, aumentaremos o grau de aleatoriedade do evento; conseqüentemente, a distribuição irá se aproximar da distribuição normal, devido ao teorema central do limite. Com isto, podemos concluir que a entropia máxima resulta na distribuição normal.

Entretanto, a entropia pode estar relacionada tanto com variáveis discretas (**entropia absoluta**), como definido anteriormente, quanto com variáveis contínuas. Para o segundo caso, pode ser calculada a chamada **entropia diferencial**. Considerando que X agora é uma variável contínua de função densidade de probabilidade $f(x)$, podemos obter a entropia diferencial por

$$H(X) = \int_{-\infty}^{\infty} f(x) \log \left(\frac{1}{f(x)} \right) dx = - \int_{-\infty}^{\infty} f(x) \log (f(x)) dx . \quad (22)$$

A entropia diferencial, diferente do caso para valores discretos, pode possuir valores negativos. Desta forma, não é possível comparar o valor da entropia discreta com o valor da entropia absoluta, pois são representados em escalas diferentes.

2.2.2 Informação Mútua

A entropia foi definida anteriormente como a quantidade média de informação que uma variável aleatória pode conter. Porém, duas variáveis aleatórias podem compartilhar uma certa quantidade de informação entre si, essa quantidade é denominada **informação mútua**. Desta forma, quando a informação mútua entre dois sinais é igual a 0, significa que se tratam de sinais independentes; o oposto ocorre quando a informação mútua é máxima, de forma que os sinais compartilhem uma grande quantidade de informação entre si.

Para determinar a informação mútua, é necessário recorrer a alguns conceitos relacionados à entropia. Considerando um sistema de entrada X e saída Y , podemos determinar a entropia desse sistema como um todo pela **entropia conjunta**, dada por

$$H(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log(p(x, y)) \quad (23)$$

Partindo disto, com a entropia de duas variáveis simultâneas, podemos considerar que essa entropia corresponde à união das entropias dessas variáveis. No entanto, também podemos obter a diferença entre essas entropias. Na prática, esta diferença indica a entropia de um valor partindo da entropia de outro; assim, caso nos seja fornecida a entropia da saída Y , podemos obter a entropia de X a partir desse método—a chamada **entropia condicional**. Essa entropia é dada por

$$H(X|Y) = H(X, Y) - H(Y) \quad (24)$$

Assim, da mesma maneira que foram obtidas as entropias anteriores, é possível obter a informação mútua entre a entrada e a saída deste sistema. Foi possível obter valores tanto para a entropia conjunta do sistema quanto para a entropia condicional de uma determinada variável; assim, podemos buscar a entropia compartilhada pela entrada e pela saída do sistema, resultando na informação mútua. Podemos tomar alguma variável como base; neste caso, tomaremos X . Para obter a informação mútua entre estas variáveis, é possível utilizar a variável base e sua entropia condicional, isto se deve ao fato de sua entropia condicional fornecer a entropia específica dessa variável, de forma que, ao subtrairmos as duas, obtemos a informação mútua. Assim, temos:

$$I(X; Y) = H(X) - H(X|Y) . \quad (25)$$

Também é comum ilustrar a ideia de informação mútua aplicando noções de conjuntos, tendo em vista que as entropias estão todas relacionadas. A Figura 4 apresenta um diagrama de Venn, considerando as duas variáveis X e Y explicitando as relações entre suas entropias.

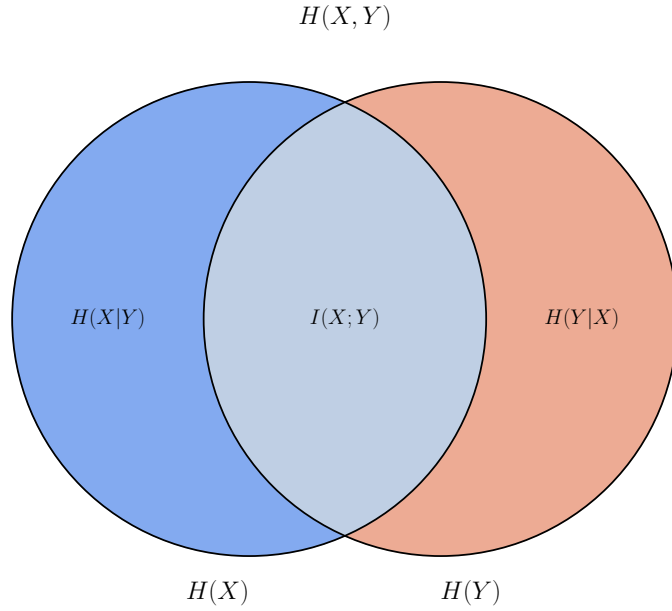


Figura 4: Diagrama de Venn da entropia conjunta entre X e Y , demonstrando de que forma as entropias marginais se relacionam para a obtenção da informação mútua.

2.2.3 Negentropia

Anteriormente, a distribuição normal foi definida como a resultante de uma entropia máxima. Assim, é possível utilizar esta relação para estabelecer uma medida de *não gaussianidade* através da entropia. Sabemos que a entropia será máxima para a distribuição gaussiana, mas não é possível determinar objetivamente este valor; para isto, foi estabelecida a **negentropia**.

A negentropia é uma medida que, ao contrário da entropia, possui valor 0 para a distribuição normal e é sempre positiva, sendo, assim, uma medida baseada na entropia diferencial, mas com não gaussianidade mais fácil de ser obtida devido à escala considerada. Dessa maneira, a negentropia de uma variável x pode ser denotada por $J(x)$, sendo representada por

$$J(x) = H(x_{gauss}) - H(x) \quad (26)$$

Sendo $H(x_{gauss})$ a entropia de um vetor gaussiano de dimensão x de mesma matriz

de covariância Σ de x ,

$$H(x_{gauss}) = \frac{\log(\det|\Sigma|) + n[1 + \log(2\pi)]}{2}. \quad (27)$$

Entretanto, para determinar este valor é necessário obter a função densidade probabilidade, mas isto nem sempre é possível. Nestes casos, são utilizadas outras formas de se obter a negentropia, utilizando diferentes recursos para obter aproximações.

2.3 ANÁLISE DE COMPONENTES PRINCIPAIS

2.3.1 Definição

A análise de componentes principais consiste em uma técnica utilizada para buscar uma representação para um determinado conjunto de dados tal que essa representação possua um menor número de dados, sem perder informação. A Figura 5 carrega em si a informação de uma seta apontando para a direita. Podemos definir uma seta como um objeto composto de um traço e um indicador de direção, de forma que uma seta pode ser representada de inúmeras maneiras desde que esteja de acordo com essa definição, conforme a Figura 6.



Figura 5: Seta comum apontando para a direita.

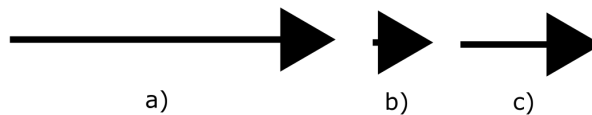


Figura 6: A seta a) mostra uma seta com um traço muito comprido, de forma que a mesma informação possa ser representada de forma mais simples, sem a necessidade de um traço tão longo. Esse excesso na informação transmitida é a chamada **redundância**. No entanto, na seta b), o traço foi reduzido a ponto de não ser muito claro que a figura queira representar uma seta, assim, a temos uma perda nessa transmissão. Já a seta c) é uma representação de uma seta com um comprimento razoável do traço.

Dessa forma, a mesma informação passada pela seta a) foi passada por uma fi-

gura de tamanho menor, pela seta c). Esse é o principal objetivo da PCA; com base em uma informação transmitida, buscar uma representação de menor dimensão. Repare que, caso as setas em questão representassem vetores, essa analogia não poderia ser aplicada, tendo em vista que o comprimento de um vetor contém informações sobre ele. Logo, fica evidente que o resultado do processo está completamente ligado ao tipo da informação em questão.

A obtenção destas componentes principais ocorre através de transformações ortogonais em relação ao conjunto de dados original, de forma que o resultado consiste em um conjunto de dados descorrelacionados. O primeiro passo para este processo é certificar que a variável em questão possui média 0, ou seja, está centralizada, para que a ortogonalização seja realizada. Considerando um vetor x contendo um determinado conjunto de dados, podemos realizar este procedimento a partir de

$$x \leftarrow x - \bar{x} . \quad (28)$$

Este passo é necessário, pois, como os componentes principais estão relacionadas com as maiores variações, quando o vetor não está centralizado, os componentes principais resultantes podem estar sendo influenciados por uma concentração dos dados em um ponto diferente da origem, afetando o seu resultado.

Feita a centralização dos dados, as transformações ortogonais podem ser aplicadas, de forma que o primeiro componente principal represente a maior variância, o segundo componente principal represente a segunda maior variância, e assim por diante. Essas transformações podem ser obtidas de diversas maneiras, dependendo do contexto a ser aplicado e do conhecimento prévio sobre o conjunto de dados.

2.3.2 PCA pela maximização da variância

Foi dito anteriormente que existem diversos métodos para realizar a PCA. No entanto, o método mais básico consiste em obter a maximização da variância através da matriz de covariância como forma encontrar as componentes principais. Além de muito comum, é um método relativamente simples para a PCA.

Inicialmente, é necessário encontrar a matriz de covariância do conjunto de

dados em questão, conforme o procedimento descrito na seção 2.1.5:

$$C_x = E\{(x - \bar{x})(x - \bar{x})^T\} .$$

Partindo disso, podemos maximizar a variância a partir dos autovalores e autovetores da matriz de covariância. Isso pode ser encontrado ao determinarmos os componentes principais pela minimização do erro de reconstrução, de forma que os autovetores e da matriz de covariância C_x podem ser considerados como as direções w dos componentes principais.

$$w_n = e_n . \quad (29)$$

Com a restrição de que w seja um vetor unitário, logo

$$\|w\| = 1 . \quad (30)$$

Autovetores de uma matriz são vetores que não sofrem alterações em sua direção após alguma transformação na matriz. Estes vetores são considerados especiais por possuírem características muito específicas sobre a matriz a que se relacionam, logo são utilizados para caracterizar essas matrizes. Autovetores de uma matriz A podem ser obtidos pela relação

$$Ax = \lambda x , \quad (31)$$

onde x se refere ao autovetor e λ se refere ao **autovalor**, outro valor de grande importância para caracterizar uma matriz. A matriz de covariância pode ser escrita na forma da decomposição pelos autovalores e autovetores:

$$C_x = EDE^T , \quad (32)$$

sendo E uma matriz com os autovetores em suas colunas e D uma matriz com os autovalores em suas diagonais.

Para projetar um determinado vetor x em uma direção dada por um vetor w , podemos utilizar a seguinte relação para obter um vetor resultante y

$$y = w^T \dot{x} . \quad (33)$$

No caso da *PCA*, y são os próprios componentes principais resultantes de x . Dessa forma, para solucionar esse problema devemos definir w corretamente, através dos autovetores, como o mencionado anteriormente. Considerando que os n componentes principais vão se relacionar com os n autovetores, de forma cada direção w_n dos componentes principais é ortogonal as das direções anteriores.

2.3.3 Branqueamento

Um conjunto de dados é considerado branco quando está centralizado, descorrelacionado e possui variância unitária. Isto implica um conjunto de dados centralizado com matriz de covariância no formato de matriz identidade.

O processo de branqueamento é muito útil para a *ICA*, pois como foi citado em 2.1.5, a descorrelação, embora não resulte em independência, pode facilitar consideravelmente o processo. No entanto, por se tratar de um processo de descorrelação, o processo de branqueamento de dados pode ser realizado pela aplicação da *PCA*.

O branqueamento de um conjunto de dados x é feito pelo processo

$$z = V \cdot x , \quad (34)$$

onde z é o conjunto de dados branqueados e V é a chamada **matriz de branqueamento**. Assim, para realizar esse processo, é necessário determinar essa matriz de branqueamento e, como citado anteriormente, é possível aplicar a *PCA* para solucionar esse problema.

A decomposição da matriz de covariância em autovalores e autovetores foi dada na equação (31) e, aplicando a *PCA*, podemos utilizar esses valores para escrever a matriz V

$$V = D^{-\frac{1}{2}} E^T , \quad (35)$$

de forma que

$$E\{zz^T\} = I .$$

Por fim, torna-se possível determinar o processo juntando as equações (34) e (35) resultando na seguinte equação

$$z = D^{-\frac{1}{2}} E^T x . \quad (36)$$

2.4 ANÁLISE DE COMPONENTES INDEPENDENTES

2.4.1 Definição

A ICA é uma das principais técnicas para o problema de separação cega de fontes. Para introduzir este problema, considere um conjunto de sinais S de n componentes $s_1, s_2, s_3, \dots, s_n$, que, quando submetidos a uma matriz de mistura A , resulta em uma mistura de sinais X , também composta de n componentes $x_1, x_2, x_3, \dots, x_n$, de forma que

$$X = AS \quad (37)$$

O problema de separação cega de fontes consiste na tentativa de obter os sinais originais de S , partindo dos sinais misturados de X . Essa estimativa para os sinais pode ser representado por Y , e será obtida manipulando a equação, resultando em

$$Y = XA^{-1} \quad (38)$$

Para simplificar, podemos representar A^{-1} como W , chegando em

$$Y = XW \quad (39)$$

Dessa forma, podemos concluir que se obtivermos acesso apenas aos sinais de mistura em X , uma forma de estimar os sinais originais S é encontrando a matriz de separação W . No entanto, não temos acesso a nenhum tipo de informação em relação aos sinais desejados, assim, serão realizadas suposições a respeito das características destes sinais, de forma que essas características possam ser trabalhadas para a obtenção da matriz de mistura.

O primeiro critério da ICA é: os sinais originais são independentes entre si. A independência estatística foi comentada em [2.1.5](#), e diz respeito a dependência entre duas ou mais variáveis; exigindo que não haja forma alguma de relação entre elas. Tendo em vista que o problema de separação cega de fontes considera que cada sinal em S é um sinal diferente, assumir sua independência é muito válido, e sabendo isso, torna-se possível buscar maneiras de trabalhar a independência em X , como forma de buscar as fontes originais.

O segundo critério da ICA é a não gaussianidade das fontes. Em 2.1.2 foi introduzido o conceito de distribuição gaussiana e sua obtenção pelo teorema central do limite, explicitando como a soma de distribuições de diferentes tipos resulta em uma distribuição gaussiana. Dessa forma, podemos interpretar que uma variável gaussiana pode ser composta de diversos outros tipos de distribuições, de forma que para obter uma variável independente, seja necessário maximizar a não gaussianidade.

Ambos os critérios desencadeiam uma série de abordagens diferentes para a ICA. Neste trabalho, serão abordadas duas dessas abordagens, os algoritmos **Infomax** e **FastICA**.

2.4.2 Infomax

Em 2.2.2, foi apresentado o conceito de informação mútua, que foi definida como quantidade de informação comum entre duas variáveis. A informação mútua é definida com base na entropia, e pode ser utilizada como uma ferramenta para explorar a independência entre sinais.

O princípio de infomax se relaciona com a maximização da informação mútua entre entrada X e saída Y de um determinado neurônio. Isso implica no aumento do fluxo de informação deste neurônio, o que também pode ser interpretado como o aumento na entropia da saída. Temos na teoria da informação a relação entre a entropia conjunta e as entropias marginais

$$H(X_1, \dots, X_n) \leq H(X_1) + \dots + H(X_n). \quad (40)$$

Utilizando essa relação, é possível concluir que com a maximização da entropia de saída, irá resultar em uma igualdade entre os dois lados da relação (40). No entanto, a igualdade entre a entropia conjunta e a soma das entropias marginais implica independência estatística das variáveis em questão. Por fim, com bases nesses conceitos, o algoritmo Infomax busca maximizar a independência das variáveis de saída.

Este processo representa a ideia do algoritmo *Infomax* [5], um dos principais algoritmos de ICA. O algoritmo parte de uma matriz de separação W gerada aleatoriamente e atualiza W de forma a minimizar essa informação mútua, e é determinado por

$$W(n+1) = W(n) + \eta(t) \cdot [I - \varphi(y(n)) \cdot y^T(n)] \cdot W(n), \quad (41)$$

onde η é a taxa de aprendizagem, um número que determina o tamanho do passo dado pelo algoritmo, influenciando diretamente no tempo de convergência e resultado do algoritmo; φ representa uma função não linear a ser determinada com base na curtose do conjunto de dados em questão; e $y(n)$ indica cada amostra do conjunto de dados. Dois valores serão considerados para a função φ :

Para uma curva leptocúrtica, teremos

$$\varphi(y) = \tanh(y) . \quad (42)$$

No caso de uma curva de probabilidade platicúrtica, será considerado

$$\varphi(y) = y - \tanh(y) . \quad (43)$$

2.4.3 FastICA

O *FastICA* [6] é um algoritmo iterativo baseado no método da **iteração de ponto fixo**. Esse método de iteração consiste em atualizar uma determinada função $f(x)$ até o chamado ponto fixo, tal que $f(x_n) = x_n$. Esse algoritmo maximiza a não gaussianidade do sinal como forma de obter os componentes independentes. O procedimento é realizado através da projeção de $w^T x$ em direções que maximizem essa não gaussianidade. No entanto, traremos duas abordagens para a sua execução: a curtose e a negentropia:

1. **FastICA por curtose:** A curtose foi apresentada em 3 como uma característica de extrema importância da distribuição de uma determinada variável aleatória, podendo ser utilizada também como uma medida muito interessante de sua gaussianidade. Podemos, então, nos basear nesse critério para realizar as iterações do algoritmo de forma a maximizar a não gaussianidade e por consequência, obtendo os sinais originais. O algoritmo é dado por

$$w \leftarrow E\{z(w^T z)^3\} - 3w , \quad (44)$$

onde w representa a matriz de mistura a ser atualizada; z , o conjunto de dados

branqueados. Uma das grandes vantagens desse algoritmo é o fato de não ser necessário estabelecer nenhum tipo de parâmetro, de forma que os resultados sejam consideravelmente consistentes.

2. ***FastICA por negentropia:*** A negentropia, introduzida em 2.2.3 é um conceito da teoria da informação diretamente relacionado à entropia. Sua definição surgiu justamente para simplificar a manipulação da entropia, funcionando de maneira oposta, de forma que o valor nulo obtido para a negentropia seja justamente o da máxima entropia. Considerando que a entropia máxima corresponde à distribuição gaussiana, podemos obter a não gaussianidade a partir da maximização da negentropia. Assim, é possível utilizar o *FastICA* para realizar essa tarefa, o algoritmo resulta em

$$w \leftarrow E\{z \cdot g(w^T z)\} - E\{g'(w^T z)\}w . \quad (45)$$

A função não linear g é determinada a partir da derivação do algoritmo e os valores considerados serão

$$g(y) = \tanh(y) . \quad (46)$$

$$g'(y) = 1 - \tanh^2(y) . \quad (47)$$

3 APLICAÇÃO DE ALGORITMOS

Os algoritmos apresentados anteriormente serão aplicados a dois tipos de dados: dados sintéticos e sinais de áudio. Esses testes serão realizados para avaliar o desempenho de cada algoritmo para os diferentes conjuntos de dados, com base em algumas variações em suas aplicações.

3.1 METODOLOGIA EXPERIMENTAL

Os testes serão separados em duas partes: aplicação em sinais sintéticos e aplicação em sinais de áudio. Isso será realizado pois é possível realizar o controle sobre os dados gerados, para podermos analisar algumas características específicas em relação à fonte. Essa separação também é válida pois existem métricas específicas para sinais de áudio, de forma que seja possível avaliar o desempenho dos algoritmos

de forma muito mais precisa. No entanto, o restante do processo utilizado é similar para ambas as partes.

Os dados sintéticos serão gerados com diferentes tipos de distribuição, pois o *Infomax* possui variações em seu algoritmo conforme a curtose da distribuição dos sinais. Assim, será possível verificar a influência da distribuição na separação final, bem como comparar este efeito nos diferentes algoritmos utilizados. Já os sinais de áudios serão separados em diferentes tipos de sinais, contendo músicas e amostras de áudio dos conjuntos **VOICe Dataset** [7] e **MUSDB** [8]. O **MUSDB** é composto de amostras de faixas musicais, enquanto o **VOICe Dataset** é composto de diferentes amostras compostas por diálogos humanos, tanto limpos, como submetidos a ruídos. As amostras de áudio podem ser avaliadas por métricas diferentes, que avaliam de forma mais específica esses tipos de sinais, essas métricas fazem parte do conjunto ***mir_eval***.

Em ambas as partes, os experimentos serão realizados variando o número de fontes. Assim, no caso dos sinais de áudio, serão feitas combinações de amostras do mesmo tipo, acrescentando amostras em determinados conjuntos para realizar essa variação no número de fontes.

Primeiramente, é necessário normalizar os sinais de áudio e certificar de que as taxas de amostragem sejam iguais para todos os casos. Esse pré processamento é essencial para garantir a consistência dos resultados a serem obtidos. Essa etapa também inclui o branqueamento, no entanto, os testes serão realizados tanto com o branqueamento quanto sem o branqueamento, de modo a verificar a influência deste procedimento no resultado em relação a cada algoritmo.

Feito isso, é possível aplicar os algoritmos e efetuar as separações. Além das métricas citadas para os sinais de áudios, serão adotadas métricas gerais, que também serão utilizadas no sinais sintéticos. Essas métricas são a **correlação de Pearson**, o **erro quadrático médio** e o **índice de Amari** e serão comentadas a seguir.

Após coletar os dados a respeito das separações, será possível analisar o impacto de cada alteração feita durante o processo. Resumindo, o passo a passo do procedimento experimental será:

1. Organização dos conjuntos de dados;
2. Determinação do número de fontes;

3. Pré-processamento e branqueamento;
4. Aplicação de algoritmo de separação;
5. Cálculo das métricas;
6. Análise dos resultados.

3.2 MÉTRICAS UTILIZADAS

3.2.1 Correlação de Pearson

Em 2.1.5 foi explicada a diferença entre descorrelação e independência, assim como a definição das medidas de relação entre variáveis, a covariância e a correlação. No entanto, a correlação é medida pela matriz de correlação, logo os resultados dessa medida não são representadas por algum coeficiente. Dessa forma, outra maneira de se medir a correlação pode ser obtido através de um coeficiente, o denominado **coeficiente de correlação de Pearson**.

A correlação de Pearson pode ser obtida a partir da normalização da medida de covariância, sendo possível de ser obtida apenas nos casos em que o número de componentes é definido. Considerando a covariância como o determinado na equação (14) e o desvio padrão como o segundo momento central (equação (8)), a correlação de Pearson r_{xy} entre duas variáveis X e Y com as componentes i entre 1 e n é

$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y} = \frac{E\{(x - \bar{x})(y - \bar{y})\}}{\sqrt{E\{(x - \bar{x})^2\}E\{(y - \bar{y})^2\}}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}. \quad (48)$$

O coeficiente de correlação de Pearson pode assumir valores entre -1 e 1, sendo que -1 é um valor que indica uma correlação inversa entre as variáveis; 1, uma correlação direta entre elas. Assim, a correlação inversa indica que quando uma variável aumenta, a outra diminui; a correlação direta, que ambas estão variando de forma diretamente proporcional.

Este coeficiente é muito útil de diferentes formas. Primeiramente, considerando que o sinal de entrada S possui duas componentes, s_1 e s_2 , não temos a certeza de que a saída y_1 corresponda à entrada s_1 e que y_2 corresponda à s_2 . Desta forma,

podemos verificar a correlação entre as saídas e entradas para associarmos corretamente os pares de entrada com os de saída. Para um caso no qual seja realizada uma separação perfeita e y_1 e y_2 correspondam a s_1 e s_2 , teremos que a correlação entre y_1 e s_1 , assim como entre y_2 e s_2 serão iguais a ± 1 .

Durante o processo de separação, o sinal de saída pode obter uma boa aproximação para o sinal de entrada, no entanto, este sinal pode estar invertido; desta forma, o coeficiente de correlação será negativo, ainda que a separação seja bem sucedida. É por este motivo que podemos considerar tanto valores positivos, quanto negativos para este coeficiente.

Além disso, a correlação pode indicar a precisão da separação, tendo em vista que o coeficiente dos pares de maior correlação de uma separação bem sucedida tendem a ter um módulo maior em relação aos de uma separação mal sucedida.

3.2.2 Erro quadrático médio

O *erro quadrático médio* (EQM) consiste em uma forma de medir a precisão de uma estimativa; resulta da esperança da diferença quadrática de uma valor em relação a seu valor estimado. Esta métrica indica a diferença entre uma variável e seu valor estimado, entretanto, como forma de desconsiderar o efeito de valores negativos, essa diferença é elevada ao quadrado. Assim, considerando um sinal S , sendo estimado pelo sinal Y , temos o EQM como

$$EQM = E\{(Y - S)^2\} = \frac{1}{n} \sum_{i=1}^n (y_i - s_i)^2 \quad (49)$$

É importante ressaltar que, em cenários de mundo real, os sinais em Y não estarão devidamente ordenados em relação aos correspondentes em S . Neste caso, antes de realizar o cálculo do erro quadrático médio, é necessário separar os pares de maior correlação.

3.2.3 Índice de Amari

Em 2.4.1, a ICA foi definida para uma mistura X , sendo resultante de sinais independentes S submetidos a uma matriz de mistura A , na equação (37). De forma análoga, também foi definida a solução Y como aproximação para os sinais S , resul-

tante da matriz W (matriz inversa de A) sendo aplicada a própria mistura X .

Dessa forma, considerando I como uma matriz identidade, temos que, como W é a inversa de A , podemos aplicar a definição de matriz inversa, resultando em

$$W \cdot A = A^{-1} \cdot A = I . \quad (50)$$

Assim, é possível estabelecer um critério de avaliação partindo das matrizes W e A , relacionando essas matrizes com a matriz identidade. Isso pode ser realizado analisando a distância do produto $W \cdot A$ em relação à matriz identidade, de forma que, indiretamente, sejam medidas as distâncias entre as próprias matrizes W e A . Essa medida é conhecida como **índice de Amari** [9], e é dada por

$$E = \sum_{i=1}^n \left(\sum_{j=1}^n \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^n \left(\sum_{i=1}^n \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) . \quad (51)$$

O valor mínimo para esse valor é 0, não possuindo valor máximo. Isso significa que 0 indica uma distância nula entre A e W , implicando igualdade entre ambos.

3.2.4 Métricas de avaliação perceptual

A avaliação perceptual se trata de uma forma de avaliar os sinais separados com base em critérios que dizem respeito aos aspectos audíveis dos sinais. Esse modelo de avaliação para separações, desenvolvido em [10], considera diferentes tipos de distorções que podem influenciar no sinal do áudio e, partindo disso, avalia o desempenho da separação comparando o sinal original com as distorções obtidas do sinal resultante deste processo. Considerando uma fonte S com n componentes, essas métricas surgem com base na decomposição de um determinado sinal s_n de forma que

$$s_n = s_{desejado} + e_{interferencia} + e_{artefatos} + e_{ruído} , \quad (52)$$

onde a letra e representa o erro. Assim, $e_{interferencia}$, $e_{artefatos}$ e $e_{ruído}$ representam os erros de interferência, artefatos externos e ruídos no sinal de áudio, respectivamente. Utilizando esses valores, é possível estabelecer relações que podem ser consideradas como métricas de desempenho para a separação:

1. *Relação fonte-distorção (SDR, Source to Distortion Ratio):*

$$SDR = 10 \log_{10} \frac{\|s_{desejado}\|^2}{\|e_{interferencia} + e_{ruído} + e_{artefatos}\|^2} \quad (53)$$

2. *Relação fonte-interferência (SIR, Source to Interference Ratio):*

$$SIR = 10 \log_{10} \frac{\|s_{desejado}\|^2}{\|e_{interferencia}\|^2} \quad (54)$$

3. *Relação fonte-artefato (SAR, Source to Artifacts Ratio):*

$$SAR = 10 \log_{10} \frac{\|s_{desejado} + e_{interferencia} + e_{ruído}\|^2}{\|e_{artefatos}\|^2} \quad (55)$$

4. *Relação fonte-ruído (SNR, Source to Noise Ratio):*

$$SNR = 10 \log_{10} \frac{\|s_{desejado} + e_{interferencia}\|^2}{\|e_{ruído}\|^2} \quad (56)$$

Estas métricas indicam a presença dos erros em relação ao sinal desejado. Em um cenário ideal, queremos que $e_{distorcao}$ seja o mínimo, o que implica em que cada um dos erros seja o mais próximo de 0 possível. Assim, tanto SDR como SAR, SIR e SNR crescem conforme seus respectivos erros diminuem, de forma que um valor alto de cada um desses critérios, implica um bom resultado da separação em relação a eles.

É importante ressaltar que, para este projeto, o ruído não será considerado. Dessa forma, $e_{ruído} = 0$, logo a SNR não será computada.

Para auxiliar no cálculo dessas métricas, foi utilizado o módulo ***mir_eval*** [11], um módulo em *Python* utilizado para auxiliar na avaliação de sinais de áudio.

3.3 EXEMPLO DE APLICAÇÃO

Realizaremos uma simulação da aplicação do algoritmo *FastICA por negentropia* para ilustrar o procedimento da aplicação de um algoritmo de separação. Para realizar essa simulação, serão utilizadas amostras de áudio extraídas da base ***VOICE dataset***. As amostras em questão se referem a vozes, sem qualquer tipos de interferência, e podem ser observados na Figura 7.

Essas amostras serão normalizadas e submetidas a uma matriz de mistura A , gerando, assim, o conjunto de sinais misturados X . O objetivo é separar esses sinais com base nos princípios da ICA explicados em 2.4.1, obtendo uma aproximação para a matriz de separação W , de forma que W representa uma matriz inversa de A .

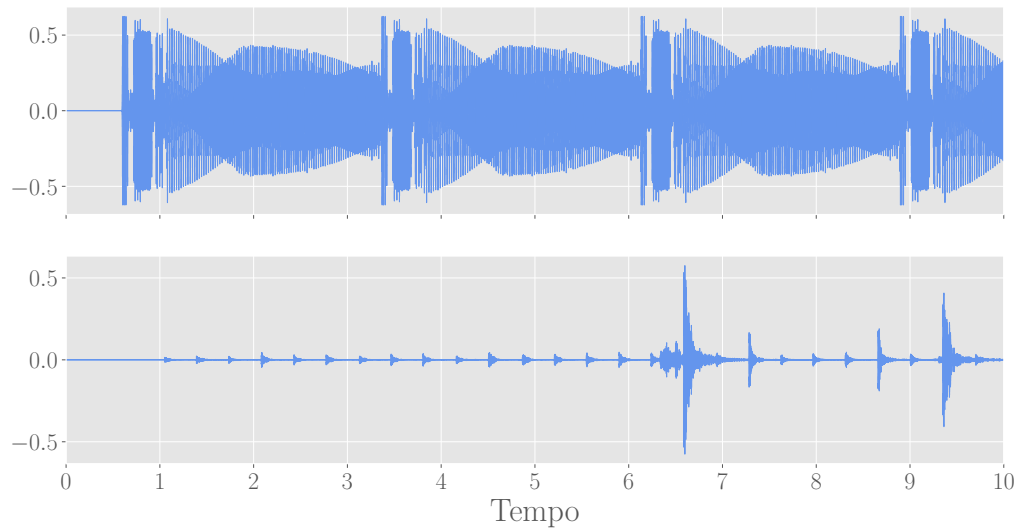


Figura 7: Sinais originais S .

A matriz de mistura A foi gerada aleatoriamente, resultando na mistura X representada pela Figura 8.

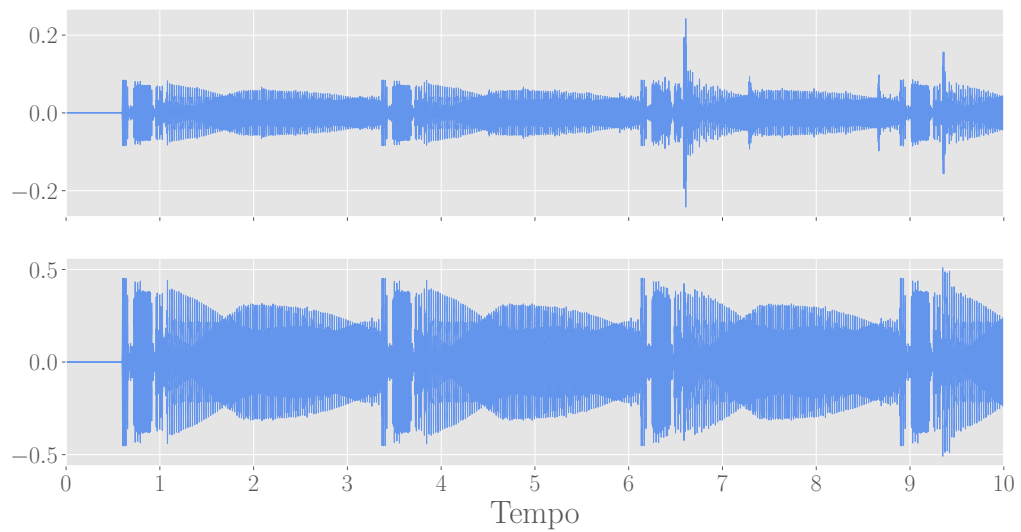


Figura 8: Sinais misturados X .

O primeiro passo para a aplicação do *FastICA* é o branqueamento do con-

Tabela 1: Correlações entre os sinais de S e Y

S	Y	r_{sy}
s_1	y_1	-0.01342
s_1	y_2	0.99991
s_2	y_1	0.99999
s_2	y_2	-0.00091

junto de dados X , utilizando a equação (36). Feito isso, torna-se possível realizar a aplicação do algoritmo, utilizando a iteração com base na equação (45), para obter os sinais separados Y .

Primeiramente, podemos analisar a correlação de Pearson separando em pares os sinais de S e Y . Os valores podem ser observados na Tabela 1.

Observando a Tabela 1, é possível concluir que os pares de maior correlação são s_1 com y_2 e s_2 com y_1 , dessa forma, podemos utilizar essa informação para calcular o EQM. Também é possível analisar que os valores para a correlação foram muito satisfatórios, tendo em vista que o valor máximo possível é de 1 para essa medida.

Como forma de representar o erro quadrático médio desses valores, é possível plotar um gráfico que represente a variação destes valores. Para este caso, os valores estão apresentados na Figura 9.

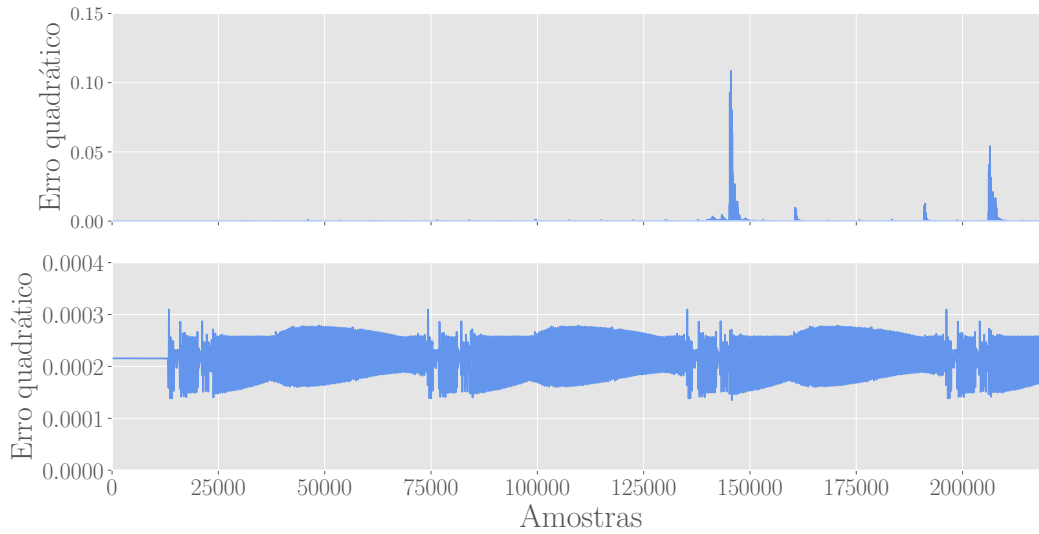


Figura 9: Erros quadráticos entre os sinais de maior correlação. Em cima, para s_1 e y_2 , e embaixo para s_2 e y_1

Assim, considerando os valores médios para cada gráfico, o erro quadrático médio para cada par correspondente pode ser observado na Tabela 2.

Tabela 2: Métricas de erro quadrático médio entre S e Y

S	Y	EQM
s_1	y_2	0.00018
s_2	y_1	0.00021

Como a matriz de mistura A foi gerada, podemos utilizá-la em conjunto com a matriz de separação W obtida para calcular o índice de Amari por meio da equação (51). Realizando os cálculos, obtemos que o índice de Amari para a separação em questão é de

$$E = 0.08874 .$$

Sabendo que o valor mínimo para este índice é 0, podemos considerar que a separação realizada obteve um desempenho consideravelmente bom.

Por fim, também é possível obter a avaliação perceptual das estimações obtidas. Os valores para essas métricas estão representados na Tabela 3.

Tabela 3: Métricas de avaliação perceptual e erro quadrático médio entre S e Y

S	Y	SAR (dB)	SIR (dB)	SDR (dB)
s_1	y_2	37.3050	37.4559	51.9720
s_2	y_1	36.6457	59.5062	36.6682

O principal valor a ser analisado é a SDR, que quanto maior, indicará maior desempenho. Neste caso, podemos comparar o desempenho da separação de um sinal em relação ao outro, chegando a conclusão de que o primeiro par obteve um melhor desempenho em relação aos critérios perceptuais.

4 RESULTADOS E DISCUSSÕES

4.1 SINAIS SINTÉTICOS

Os primeiros testes serão realizados com sinais gerados de forma sintética, para podermos analisar a influência de cada fator em determinado tipo de algoritmo, de forma que possibilite o controle de características das fontes para obter mais precisão nesse tipo de teste. Serão analisadas as influências de algumas variáveis presentes no processo de separação, sendo estas:

1. Branqueamento;
2. Número de fontes;
3. Matriz de inicialização.

Para isso, serão gerados sinais de distribuições subgaussianas e supergaussianas, a distribuição subgaussiana será a uniforme, enquanto a supergaussiana será a distribuição de Laplace. Assim, também será possível avaliar a influência da natureza dos sinais nos testes. Os resultados obtidos na avaliação de cada variável servirão de base para a avaliação das variáveis posteriores, como forma de obter maior consistência em cada teste.

4.1.1 Branqueamento

Para avaliar a influência deste critério, serão aplicados algoritmos de separação com branqueamento, e em seguida sem branqueamento. Este processo será feito tanto para ambos os tipos de fonte. Neste teste, o desempenho para cada algoritmo será testado na situação mais básica da separação de sinais, com apenas duas fontes. O processo utilizará ambos os tipos de sinais, que serão analisados separadamente, com base na média das métricas apresentadas em 3.2. Primeiramente, os algoritmos serão aplicados em sinais de distribuição supergaussiana, seus resultados serão apresentados nas Tabelas 4 e 5.

Tabela 4: Avaliação de algoritmos sem branqueamento para fonte de distribuição de Laplace.

Algoritmo	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	1.310	0.919	0.535
<i>FastICA</i> por negentropia	1.591	0.868	0.650
<i>Infomax</i>	1.518	0.886	0.618

Tabela 5: Avaliação de algoritmos com branqueamento para fonte de distribuição de Laplace.

Algoritmo	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	0.438	0.984	0.563
<i>FastICA</i> por negentropia	0.567	0.977	0.632
<i>Infomax</i>	1.636	0.880	0.678

Observando as Tabelas apresentadas, é possível perceber uma grande diferença no desempenho dos algoritmos *FastICA*, obtendo uma melhora significativa em seus resultados após o processo de branqueamento. Isso demonstra exatamente o que foi apresentado em 2.4.3, a respeito da importância do branqueamento como etapa de pré processamento para o *FastICA*.

No entanto, embora a melhora seja perceptível para o algoritmo *FastICA*, o algoritmo *Infomax* não apresentou grandes diferenças após o processo de branquea-

mento. Analisando os resultados obtidos, também é possível concluir que este algoritmo apresentou um desempenho muito inferior em relação ao *FastICA*.

Em seguida, os resultados serão computados através de fontes de distribuição uniforme, sendo representados nas Tabelas 6 e 7.

Tabela 6: Avaliação de algoritmos sem branqueamento para fonte de distribuição uniforme.

Algoritmo	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	1.806	0.815	0.784
<i>FastICA</i> por negentropia	1.741	0.821	0.828
<i>Infomax</i>	1.846	0.772	0.841

Tabela 7: Avaliação de algoritmos com branqueamento para fonte de distribuição uniforme.

Algoritmo	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	0.358	0.986	0.556
<i>FastICA</i> por negentropia	0.222	0.993	0.529
<i>Infomax</i>	1.7369	0.859	0.692

Mesmo considerando que os resultados obtidos para os sinais da distribuição de Laplace tenham indicado uma melhora impactante nos resultados, este impacto se intensificou consideravelmente para o caso da distribuição uniforme. Isso se deve ao baixo desempenho dos algoritmos no cenário anterior ao branqueamento. Também é possível observar que, diferente do ocorrido ao aplicar a separação nos sinais de distribuição de Laplace, o algoritmo *Infomax* apresentou resultados melhores neste caso, embora, novamente, esteja inferior ao *FastICA*. A razão para essa melhora no algoritmo *Infomax* ocorrida para o caso da distribuição uniforme se deve ao fato de que a distribuição de Laplace havia sido gerada com média 0, enquanto que a distribuição uniforme atende a esse requisito. Desta forma, é possível concluir que, embora o branqueamento não seja um processo essencial ao *Infomax*, a centralização dos dados influencia diretamente em seus resultados.

Outra característica perceptível em relação ao *Infomax* é o seu baixo desem-

penho apresentado. Isso ocorre pois este algoritmo depende de diversos parâmetros, como, por exemplo, sua função de ativação e sua taxa de aprendizagem. Desta forma, sua consistência se torna muito comprometida e pode obter variações significativas dependendo dos sinais a serem separados.

4.1.2 Matriz de inicialização

A matriz de separação W é submetida a uma inicialização aleatória antes de ser atualizada pelos algoritmos. O objetivo deste teste é analisar o quanto a aleatoriedade da geração da matriz W influencia no processo de separação de fontes. Para isso, uma única mistura branqueada de duas fontes será gerada e cada algoritmo realizará 150 tentativas de separação. Neste caso, a única variação em relação a cada tentativa será o valor inicial de W , assim, a variação do desempenho de cada tentativa de separação indicará a influência que a aleatoriedade de W teve neste processo. A Figura 10 e a Figura 11 representam a variação dessas tentativas para cada algoritmo.

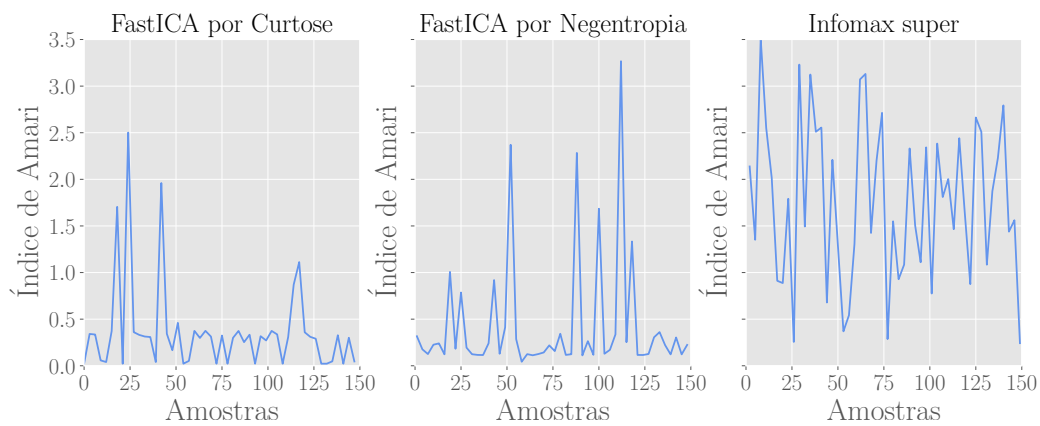


Figura 10: Desempenho de cada algoritmo para uma determinada mistura de distribuição de Laplace.

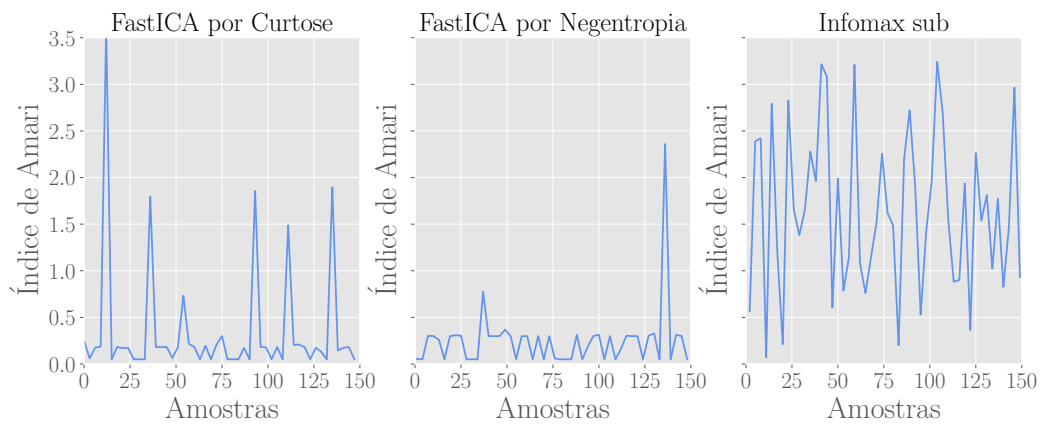


Figura 11: Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.

É possível observar que não existe qualquer diferença explícita em relação ao gráfico de cada distribuição. Assim, em ambos os casos é possível observar que os algoritmos de *FastICA*, embora consigam bons resultados com uma certa consistência, também são afetados pela geração da matriz W , e isso acarreta em alguns resultados bem inferiores.

No entanto, o aspecto que mais chama a atenção está no algoritmo *Infomax*. Este algoritmo obteve uma variação de desempenho significativa para cada tentativa de separação, isso retrata uma inconsistência muito grande deste algoritmo.

4.1.3 Número de fontes

Nos testes anteriores, foram sempre considerados casos simples de separação e sinais, com apenas duas fontes para serem separadas. No entanto, podem ser encontrados diversos outros cenários envolvendo outras quantidades de fontes. Dessa forma, será analisado o comportamento dos algoritmos quando submetidos a variações no número de fontes.

Serão utilizados os mesmos conjuntos de dados branqueados para todos os algoritmos e, após cada teste, será adicionada mais uma fonte à mistura em questão. Assim, os algoritmos serão analisados com base em suas métricas apresentadas nas Tabelas 8 e 9.

Tabela 8: Desempenho dos algoritmos conforme variação no número de fontes para fontes supergaussianas.

Algoritmo	N. de Fontes	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	2	0.528	0.980	0.889
<i>FastICA</i> por curtose	3	2.367	0.934	0.829
<i>FastICA</i> por curtose	4	4.429	0.915	1.174
<i>FastICA</i> por negentropia	2	0.650	0.970	0.828
<i>FastICA</i> por negentropia	3	2.760	0.913	0.703
<i>FastICA</i> por negentropia	4	4.842	0.899	1.195
<i>Infomax</i>	2	1.628	0.907	0.995
<i>Infomax</i>	3	5.208	0.822	0.933
<i>Infomax</i>	4	9.399	0.791	1.389

Tabela 9: Desempenho dos algoritmos conforme variação no número de fontes para fontes subgaussianas.

Algoritmo	N. de Fontes	Índice de Amari	Correlação	EQM
<i>FastICA</i> por curtose	2	0.480	0.979	0.443
<i>FastICA</i> por curtose	3	1.818	0.944	0.983
<i>FastICA</i> por curtose	4	4.317	0.911	0.727
<i>FastICA</i> por negentropia	2	0.369	0.986	0.440
<i>FastICA</i> por negentropia	3	1.621	0.957	1.194
<i>FastICA</i> por negentropia	4	3.625	0.931	0.703
<i>Infomax</i>	2	1.722	0.902	0.628
<i>Infomax</i>	3	4.661	0.850	1.205
<i>Infomax</i>	4	9.353	0.795	0.876

Com base nos dados coletados, é possível observar uma variação similar em ambos os tipos de fontes. Para todos os algoritmos, embora no caso de 3 fontes, os resultados tenham sido razoáveis; ainda assim, é evidente que o aumento no número de fontes causa uma perda notável de desempenho.

4.2 SINAIS DE ÁUDIO

Foram realizadas as aplicações de algoritmos em sinais sintéticos como forma de analisar alguns dos seus aspectos de funcionamento. Com base nestes resultados, serão realizadas aplicações de algoritmo em sinais de áudio como forma de avaliar os processos de separação em sinais reais. Neste caso, torna-se possível realizar a avaliação perceptual das separações de sinais.

Os sinais serão aplicados nos mesmo conjuntos de áudio, considerando novamente a situação mais simples, com apenas duas variáveis. Como foi demonstrada a influência da geração aleatória de W no processo de separação, serão realizadas 50 tentativas de separação para cada mistura de sinais, e os valores para as métricas de cada algoritmo serão dados pela média obtida. No total, serão utilizadas 25 misturas de sinais, com sinais de voz e músicas.

Para este caso, não teremos informações iniciais sobre as fontes, logo, não podemos escolher qual algoritmo de *Infomax* utilizar. Desta forma, este algoritmo será separado em duas partes, sendo uma para cada função de ativação.

4.2.1 Avaliação por métricas objetivas

Para avaliar esse processo de separação, primeiramente, os desempenhos serão dados pelo EQM, pelo índice de Amari e pela correlação de Pearson. Estes resultados estão representados na Tabela 10.

Tabela 10: Desempenho para cada algoritmo avaliado pelo índice de Amari, correlação de Pearson e erro quadrático médio.

Algoritmo	Índice de Amari	Correlação	EQM
FastICA por curtose	0.502	0.978	0.572
FastICA por negentropia	0.606	0.975	0.506
Infomax sub	1.162	0.911	1.058
Infomax super	2.133	0.811	0.465

Os resultados para os algoritmos de *FastICA* novamente se sobressaíram em relação aos de *Infomax*. No entanto, podemos estabelecer uma análise mais es-

pecífica destes casos ao analisar a dispersão dos resultados obtidos. Para isso, podemos analisar o diagrama de caixa representado na Figura 12.

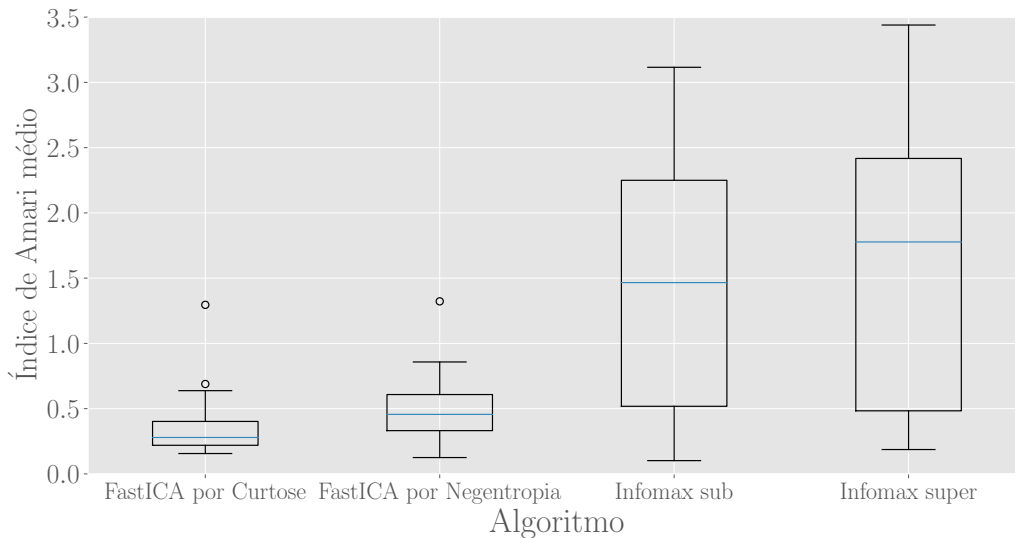


Figura 12: Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.

Uma justificativa para baixo desempenho do *Infomax* pode ser encontrada na figura anterior. Enquanto os algoritmos *FastICA* se mantiveram relativamente consistentes, ocupando limites restritos a valores baixos para o índice de Amari, o *Infomax* apresentou uma oscilação muito expressiva, que pode ser observada pela distância entre os quartis e pelos limites que os valores ocuparam.

O algoritmo *FastICA por curtose* apresentou um melhor resultado, bem como maior consistência entre os algoritmos estudados. A razão pela inconsistência dos algoritmos *Infomax* pode se encontrar principalmente na dependência da definição da função de ativação. No entanto, existem alternativas para a solução deste problema, como, por exemplo, o *Extended Infomax* [12].

4.2.2 Avaliação perceptual

As separações dos sinais de áudio também serão avaliadas conforme as métricas apresentadas em 3.2.4. A principal métrica para avaliação perceptual é a SDR, e a análise de desempenho será analisada principalmente por meio dela. Os resultados para a avaliação perceptual das separações estão apresentados na Figura 11.

Tabela 11: Métricas de avaliação perceptual para cada algoritmo.

Algoritmo	SDR (dB)	SIR (dB)	SAR (dB)
FastICA por curtose	26.164	27.051	50.004
FastICA por negentropia	25.118	26.254	49.931
Infomax sub	17.023	17.023	151.977
Infomax super	9.756	9.756	151.989

Assim como nas métricas observadas anteriormente, a *FastICA por curtose* também apresentou melhor desempenho para as métricas de avaliação perceptual. Neste quesito, o algoritmo *Infomax super* apresentou um desempenho extremamente inferior em relação aos demais. Também podemos analisar a distribuição dos dados obtidos, que está representada pelo diagrama de caixa, na Figura 13.

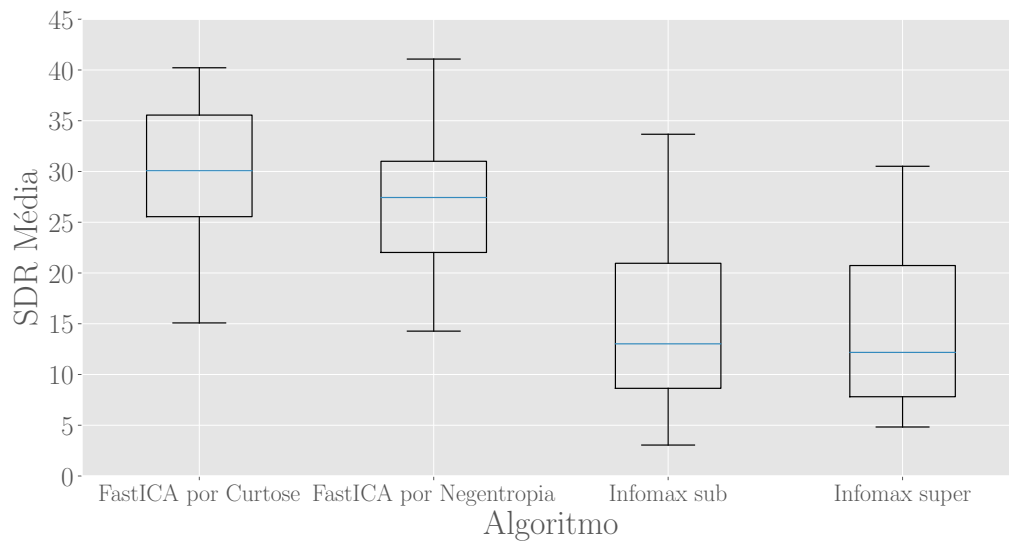


Figura 13: Desempenho de cada algoritmo para uma determinada mistura de distribuição uniforme.

Da mesma forma, os algoritmos *Infomax* não obtiveram uma boa consistência; no entanto, neste caso, ficou evidente que, pela avaliação perceptual, nenhum de seus resultados alcançou o desempenho das boas separações realizadas pelos algoritmos *FastICA*. Assim, os algoritmos *FastICA* possuíram melhor desempenho e maior consistência em relação ao *Infomax*, sendo que o *FastICA por curtose* se saiu ligeiramente melhor em relação ao *FastICA por negentropia*.

5 CONCLUSÃO

Este trabalho se iniciou pela apresentação do problema de BSS, e mostrou que a ICA como uma possível técnica de solução de tal problema. Em seguida, foram apresentadas as principais ferramentas que fundamentam a ICA, como conceitos de teoria da probabilidade e de teoria da informação, bem como a PCA, uma técnica intimamente relacionada à ICA. Partindo disso, foram introduzidos os algoritmos *Infomax*, *FastICA por curtose* e *FastICA por negentropia*.

Após terem sido apresentada a fundamentação teórica dos principais conceitos envolvidos, a fim de solidificar os conceitos desenvolvidos, foram feitas implementações próprias de cada um dos algoritmos abordados no trabalho e, por fim, foram realizados experimentos com diferentes cenários, que envolveram tanto casos com sinais sintéticos de menor complexidade quanto casos mais complexos com sinais sonoros de mundo-real. Os algoritmos foram, então, confrontados diante de tais cenários com base em métricas SDR, SIR e SAR.

A partir das métricas avaliadas, considerando os cenários abordados, é possível concluir que o algoritmo *FastICA por curtose* apresentou um melhor funcionamento devido aos bons resultados e à consistência; no entanto, o *FastICA por negentropia* também se saiu muito bem nos testes, obtendo resultados ligeiramente inferiores. O algoritmo *Infomax*, por sua vez, não apresentou resultados favoráveis, pois depende de informações sobre a distribuição das fontes para ajustar seus parâmetros.

Embora o desempenho dos algoritmos tenha sido corretamente avaliado, ainda é possível projetar possíveis melhorias. Não foi avaliado o custo computacional de cada processo, que é uma informação de extrema importância para uma avaliação completa. Também foi notada a falta de consistência do algoritmo *Infomax*, que poderia ser solucionada pela implementação do algoritmo *Extended Infomax*. De forma geral, alguns algoritmos obtiveram dificuldade em convergir para determinadas misturas, o que também pode ser investigado com mais profundidade. E, por fim, há todo um conjunto de diferentes cenários em que estes testes todos ainda poderiam ser realizados, assim como há diversas outras métricas que poderiam ser incluídas. Estes são aspectos de muita relevância para o estudo, e podem ser tratados em trabalhos futuros.

Referências

- [1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *Journal of the Acoustical Society of America*, vol. 25, no. 5, 1953.
- [2] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, 1991.
- [3] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, no. 3, 1994.
- [4] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [5] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution.," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [6] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, 1999.
- [7] S. Gharib, K. Drossos, E. Fagerlund, and T. Virtanen, "Voice dataset," Jan. 2020.
- [8] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner, "MUSDB18-HQ - an uncompressed version of musdb18," Dec. 2019.
- [9] S. AMARI, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, 1996.
- [10] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, 2006.
- [11] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, D. P. Ellis, and C. C. Raffel, "mir_eval: A transparent implementation of common mir metrics," in *In Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR*, Citeseer, 2014.

- [12] T.-W. Lee, M. Girolami, and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources," *Neural Computation*, vol. 11, pp. 417–441, 02 1999.