



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA E DE COMPUTAÇÃO



Correntropia Complexa: Definição, Propriedades e Aplicações

João Paulo Ferreira Guimarães

Orientador: Prof. Allan De Medeiros Martins

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da UFRN (área de concentração: Engenharia de Computação) como parte dos requisitos para obtenção do título de Doutor em Ciências.

Número de ordem PPgEEC: D258
Natal, RN, Setembro de 2019

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Guimarães, João Paulo Ferreira.

Correntropia Complexa: definição, propriedades e aplicações/ João Paulo Ferreira Guimarães - 2019

117 f.: il.

Tese (Doutorado)-Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós Graduação em Engenharia Elétrica e de Computação, Natal, 2019

Orientador: Dr. Allan de Medeiros Martins.

1. Correntropia - Tese. 2. Dados Complexos - Tese. 3. Similaridade - Tese.
I. Martins, Allan de Medeiros. II. Título.

RN/UF/BCZM

CDU 621.3:004

*Não existe um caminho para a
felicidade. A felicidade é o caminho.
Thich Nhat Hanh*

Agradecimentos

Ao meu amigo e colega Aluísio. Sem ele esse trabalho nem teria começado. Aluísio me apresentou a área, me motivou, me deu as primeiras aulas e, muito mais do que isso, mostrou que era possível encontrar alegria no caminho, não só no destino dessa jornada rumo ao doutorado.

Ao meu orientador, Professor Allan, um exemplo de Professor e orientador, em todos os aspectos. Serei eternamente grato pelas aulas e lições que vão muito além da academia e que levo para minha vida pessoal e profissional.

Ao Professor e amigo Joilson, pelo incentivo, por todas as ideias, cafezinhos, risadas e, principalmente, pelo contagiante entusiasmo por essa tal de matemática.

Ao Professor von Borries e Felipe, pelo acolhimento, amizade e lições em terras estrangeiras.

Aos meus muitos amigos e amigas, pelos vários momentos de alegria e companheirismo.

Aos meus pais e irmão, pelo amor e suporte incontestáveis durante toda a minha vida.

À minha esposa Giovanna, pelo companheirismo, paciência, amor e carinho.

Ao meu filho Rafael, por me impulsionar a ser uma pessoa melhor.

À Deus, pelo seu imenso amor gratuito e incondicional.

Ao IFRN, por investir no desenvolvimento acadêmico do seu corpo docente e por me mostrar que é possível fazer pesquisa de ponta aliada a um ensino público, gratuito e de qualidade.

Resumo

Estudos recentes demonstraram que a correntropia é uma ferramenta eficiente para analisar momentos estatísticos de ordem superior em ambientes de ruído não gaussianos. Embora a correntropia tenha sido utilizada com dados complexos, nenhum estudo teórico foi realizado para elucidar suas propriedades ou seu uso em problemas de otimização. Sendo assim, baseado em uma interpretação probabilística, este trabalho apresenta uma nova medida de similaridade entre duas variáveis aleatórias complexas, definida como correntropia complexa. As suas propriedades são estudadas bem como seu uso enquanto função custo no Critério de Máxima Correntropia Complexa (*Maximum Complex Correntropy Criterion* – MCCC). O MCCC é aplicado a um algoritmo de gradiente ascendente e numa nova solução recursiva baseada em ponto fixo. Para testar a robustez dessa nova medida em ambientes com ruídos não gaussianos, aplica-se a correntropia complexa como função custo em um problema de identificação de sistemas, num problema de equalização de canal e de amostragem compressiva. Também é mostrado como a correntropia complexa pode ser aplicada ao problema de similaridade entre ângulos. As simulações demonstram vantagens proeminentes quando os métodos propostos são comparados a métodos clássicos da literatura.

Palavras-chave: Correntropia, dados complexos, medida de similaridade

Abstract

Recent studies have demonstrated that correntropy is an efficient tool for analyzing higher-order statistical moments in non-Gaussian noise environments. Although correntropy has been used with complex-valued data, no theoretical study was pursued to elucidate its properties, nor how to best use it for optimization. By using a probabilistic interpretation, this work presents a novel similarity measure between two complex-valued random variables, which is defined as complex correntropy. Its properties are studied as well as a new recursive solution for the Maximum Complex Correntropy Criterion (MCCC) and two algorithms are derived, one based on the ascendent gradient and a second one on a fixed-point solution. Simulations were made in order to evaluate how robust this new measure is to impulsive noise in different problems: liner system identification, channel equalization and in a compressive sensing problem. It is also shown the application of complex correntropy as a tool to analyse the similarity between angles. The results demonstrate prominent advantages of the proposed method when compared with the classical algorithms in the literature.

Keywords: Correntropy, complex-valued data, Similarity measure

Sumário

Sumário	i
Lista de Figuras	iv
Lista de Tabelas	vi
Lista de Algoritmos	vii
Lista de Símbolos e Abreviaturas	viii
1 Introdução	2
1.1 Motivação	2
1.2 Objetivos	3
1.3 Contribuições	3
1.4 Organização do trabalho	3
2 Fundamentação Teórica	5
2.1 Números Complexos	5
2.1.1 Histórico	5
2.1.2 Introdução	5
2.1.3 Complexo Conjugado	6
2.1.4 Módulo ou Valor Absoluto de um número complexo	6
2.1.5 Geometria dos números complexos	6
2.1.6 Vetores e Matrizes Complexas	7
2.1.7 Operador Hermitiano	8
2.1.8 Funções Complexas	8
2.1.9 Cálculo de Wirtinger	11
2.2 Espaço de Hilbert	13
2.2.1 Espaço vetorial	13
2.2.2 Produto interno	13
2.2.3 Espaços vetoriais dotados de produto interno	14
2.2.4 Norma	14
2.2.5 Espaço vetorial normado	14
2.2.6 Métrica	14
2.2.7 Completude	15
2.2.8 Espaço de Banach	15

2.2.9	Espaço de Hilbert	15
2.2.10	Kernel Positivo Definido	15
2.2.11	Reproducing Kernel Hilbert Spaces	16
2.3	Correntropia	17
2.3.1	Definição	17
2.3.2	Estimação da correntropia com o <i>kernel</i> gaussiano	18
2.3.3	Propriedades	18
2.3.4	Interpretação probabilística da correntropia	20
2.3.5	Critério de Máxima Correntropia	21
3	Correntropia Complexa	24
3.1	Estado da Arte	24
3.2	Interpretação Probabilística	25
3.3	Definição	26
3.4	Propriedades	26
3.5	Critério de Máxima Correntropia Complexa	35
3.5.1	Correntropia complexa como função custo	35
3.5.2	Solução usando o Gradiente Ascendente	36
3.5.3	Solução de Ponto fixo	38
3.5.4	Análise dos algoritmos	40
3.6	Estratégias de <i>kernel</i> adaptativo	41
3.6.1	MCCC com Largura de Kernel Adaptativa	41
3.6.2	MCCC com Chaveamento de Largura de Kernel	42
3.6.3	MCCC com Largura de Kernel Variável	42
4	Aplicações	44
4.1	Simulação de Ambientes com Ruído Não-Gaussiano	44
4.1.1	Ruído Bimodal	44
4.1.2	Ruído Alfa-Estável	45
4.2	Métricas de avaliação	46
4.3	Identificação de Sistemas Lineares	47
4.3.1	Resultados usando <i>kernel</i> adaptativo	50
4.4	Equalização de Canal	53
4.5	Amostragem Compressiva	58
4.5.1	CCIM como aproximação da norma ℓ_0	60
4.5.2	ℓ_0 -MCCC	62
4.6	Similaridade entre Ângulos	68
5	Conclusões	73
5.1	Lista de publicações	74
5.1.1	Publicações relacionadas a esta tese	74
5.1.2	Publicações provenientes de colaborações	74
	Referências bibliográficas	75

A	Correntropia como integral de f_{XY} na reta $x = y$	81
B	Análise da Expansão em Série de Taylor da Correntropia com <i>kernel</i> Gaussiano	84
C	Correntropia complexa como integral de f_{XYZW} no plano $x = y$ e $z = w$	88
D	Análise de convergência	91
E	Algoritmo do MCCC com Ponto Fixo sem usar o Cálculo de Wirtinger	94
F	Correntropia Complexa com resposta complexa	99
	F.1 Prova que $\exp(z)$, com $z \in \mathbb{C}$, é diferenciável:	100
	F.2 Prova que $k'_\sigma = \exp(xy^*)$ é positivo-definido	101

Lista de Figuras

2.1	Plano complexo formado pelo eixo real e imaginário. Número complexo $z = a + jb$, formado pelo par ordenado $(\Re(z), \Im(z))$. Módulo ($ z $) e fase (θ) também podem ser usados para representar z . O seu complexo conjugado $z^* = a - jb$ também é mostrado.	7
2.2	Diagrama mostrando a relação entre os tipos de espaços vetoriais.	16
4.1	Histograma com amostra do ruído bimodal cuja parte real é $0.80\mathcal{N}(0, 1) + 0.20\mathcal{N}(13, 2)$, e a parte imaginária $0.90\mathcal{N}(-15, 5) + 0.10\mathcal{N}(8, 1)$	45
4.2	Histograma com amostra do ruído com distribuição alfa estável com GSNR = 20dB e $\alpha = 1.5$, tanto para parte real quanto imaginária.	46
4.3	Típico problema de identificação de sistemas.	47
4.4	Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS. Ambiente com ruído bimodal.	48
4.5	Desvio padrão referente as simulações mostradas na Figura 4.4.	49
4.6	Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS num ambiente com ruído modelado pela distribuição alfa estável, com GSNR = 15 dB e $\alpha = 1.5$	50
4.7	Desvio padrão referente as simulações mostradas na Figura 4.6.	51
4.8	Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS num ambiente com ruído modelado pela distribuição alfa estável, com GSNR = 15 dB e $\alpha = 1.5$	52
4.9	Desvio padrão referente as simulações mostradas na Figura 4.8.	53
4.10	Diagrama de blocos contendo o problema de Equalização de Canal.	53
4.11	Eventos da equalização de canal	54
4.12	Média de 10^5 curvas BER para os algoritmos CLMS, CRLS, LAD e MCCC com t	55
4.13	Desvio padrão referente aos resultados mostrados na Figura 4.12.	56
4.14	Curvas BER para o algoritmo do MCCC com o tamanho de <i>kernel</i> fixo em $\sigma = 1$ para diferentes valores de GSNR e de α	56
4.15	Influência de diferentes tamanhos de <i>kernel</i> no comportamento das curvas BER para o MCCC com diferentes valores de GSNR e índice de estabilidade ($\alpha = 1.5$) fixo.	57

4.16	Típico problema de amostragem compressiva. Deseja-se reconstruir um vetor esparso \mathbf{x} com tamanho $L \times 1$ a partir de um vetor menor \mathbf{y} com tamanho $M \times 1$, onde $M < L$	58
4.17	Comparativo entre o percentual de reconstruções com sucesso usando a aproximação da norma ℓ_0 com a CCIM com o NESTA, que utiliza a norma ℓ_1 . O vetor vetor complexo esparso que é recuperado tem tamanho $L = 512$ e S elementos não zero em função do número de medidas (M) realizadas.	61
4.18	Ambiente com ruído impulsivo	67
4.19	SER médio de 10^3 simulações mostrando o desempenho dos algoritmos na reconstrução de um vetor esparso complexo, sobre diferentes valores de GSNR e um $\alpha = 1.5$ fixo. S é o número de elementos não nulos no vetor esparso. Note que, quanto maior for o GSNR, menor a potência do ruído.	68
4.20	SER médio de 10^3 simulações mostrando o desempenho dos algoritmos testados sob múltiplos números de medidas realizadas M e parâmetros do ruído fixo em GSNR=25 dB e $\alpha = 1.5$	69
4.21	SER médio de 10^2 simulações do ℓ_0 -MCCC em função do GSNR e do índice de estabilidade, α , ambos parâmetros de controle da distribuição alfa estável que modela o ruído impulsivo.	69
4.22	SER médio de 10^2 simulações do ℓ_0 -MCCC com $\alpha = 1.5$ e em função da potência do ruído GSNR e do tamanho do <i>kernel</i> σ	70
4.23	Desempenho do ℓ_0 -MCCC variando os parâmetros λ , que pondera a importância da aproximação do gradiente de ℓ_0 , e de β , que controla a zona de atuação do atrator. A Figura é uma média de 10^2 simulações de Monte Carlo.	70
4.24	Número sobre círculo unitário	71
4.25	Ângulos dentro do círculo unitário	72

Lista de Tabelas

4.1	Média de 10^3 SER computado para cada algoritmo em 3 diferentes cenários: Sem ruído, com ruído gaussiano e com ruído impulsivo.	65
4.2	Correntropia entre os vetores de ângulos que seguem distribuição gaussiana com média é z_i , com $i = [1, 5]$ e variância 1.	71
4.3	Correntropia complexa entre os vetores de números complexos equivalentes aos ângulos da Tabela 4.2.	72

Lista de Algoritmos

1	Gradiente com MCCC	38
2	Ponto fixo com MCCC	41
3	ℓ_0 -MCCC	64

Lista de Símbolos e Abreviaturas

\mathbb{C}	Conjunto dos números complexos
\mathbb{R}	Conjunto dos números reais
f_V	Função densidade de probabilidade da variável aleatória V
$j = \sqrt{-1}$	Unidade imaginária
AMCCC	<i>Adaptive Kernel Width MCCC</i> MCCC com Largura de Kernel Adaptativa
BER	<i>Bit Error Rate</i> (Taxa de erros de bit)
CCIM	<i>Complex Correntropy Induced Metric</i> (Métrica Induzida da Correntropia Complexa)
CIM	<i>Correntropy Induced Metric</i> (Métrica Induzida da Correntropia)
CLMS	<i>Complex Least Mean Square</i> (Método dos Mínimos Quadrados Complexos)
CRLS	<i>Complex Recursive Least Square</i> (Método dos Mínimos Quadrados Complexos Recursivo)
CS	<i>Compressive Sensing</i> (Amostragem Compressiva)
ED	<i>Euclidean Distance</i> (Distância euclidiana)
GSNR	<i>Generalized Signal-to-Noise Ratio</i> (Relação Sinal-Ruído Generalizada)
LAD	<i>Least Absolute Deviation</i> (Método do Mínimo Desvio Absoluto)
MCC	<i>Maximum Correntropy Criterion</i> (Critério de Máxima Correntropia)

- MCCC *Maximum Complex Correntropy Criterion*
(Critério de Máxima Correntropia Complexa)
- NESTA *Nesterov's Algorithm*
(Algoritmo de Nesterov)
- PDF *Probability Density Function*
(Função densidade de probabilidade)
- QAM *Quadrature Amplitude Modulation*
(Modulação de Amplitude em Quadratura)
- RIP *Restricted Isometry Property*
(Propriedade de Isometria Restrita)
- RKHS *Reproducing Kernel Hilbert Space*
(Espaço de Hilbert com Kernel Reprodutor)
- SER *Signal-to-Error Ratio*
(Relação Sinal Erro)
- SMCCC *Adaptive Kernel Width MCCC*
(MCCC com chaveamento de largura de *kernel*)
- VKW-MCCC *Variable Kernel Width MCCC*
(MCCC com Largura de *kernel* Variável)

Notação

Esta tese utiliza o seguinte padrão de notação:

- Escalares são escritas por letras minúsculas em itálico: *a*;
- Constantes e variáveis aleatórias são representadas por letras maiúsculas: *B*;
- Vetores são representados por letras minúsculas em negrito: **c**;
- Utilizam-se sempre vetores coluna;
- Matrizes são escritas usando letras maiúsculas em negrito: **D**;
- O superescrito $(\cdot)^T$ é usado para indicar transposição de matrizes e vetores;

Capítulo 1

Introdução

Neste capítulo são apresentados em linhas gerais as motivações, objetivos, contribuições e organização desta tese.

1.1 Motivação

Sinais pertencentes ao domínio dos números complexos são encontrados em muitos problemas de engenharia tais como comunicações, radar, biomedicina, geofísica entre outras áreas de ciências aplicadas [Schreier e Scharf 2010]. Isso se dá em virtude da modelagem do problema ou pela própria natureza da aplicação [Mandic e Goh 2009]. Dessa forma, é conveniente utilizar algoritmos que possam processar esses dados nesse domínio, mantendo suas características originais [Adali e Haykin 2010].

A correntropia é uma medida de similaridade que foi definida em 2006 pelo grupo de pesquisa do laboratório de computação em neuro-engenharia (CNEL) liderado pelo professor José Carlos Príncipe [Santamaria et al. 2006]. Ela é capaz de extrair informações estatísticas de ordens superiores, generalizando a correlação [Liu et al. 2007]. Isso faz com que essa medida tenha um desempenho superior quando comparada a métodos que utilizam estatística de segunda ordem, principalmente em ambientes contaminados por ruídos não gaussianos [Príncipe 2010].

Apesar do relativo pouco tempo desde sua primeira publicação, o estudo dessa medida tem crescido na comunidade científica e seu uso tem se massificado nas mais diferentes áreas da engenharia e em aplicações como reconhecimento facial [He, Zheng e Hu 2011], filtragem adaptativa [Chen et al. 2014], classificação [Cao et al. 2018], reconhecimento de padrões [He, Zheng, Hu e Kong 2011], análise de componentes principais [He, Hu, Zheng e Kong 2011], filtros de Kalman [Dang et al. 2019], *clustering* [Hao et al. 2015], reconhecimento de patologias vocais [Fontes et al. 2014], entre muitos outros.

Entretanto, a correntropia foi definida somente no domínio real [Santamaria et al. 2006, Liu et al. 2007, Príncipe 2010], o que dificulta a aplicação da correntropia a toda uma classe de problemas que envolvem dados de natureza complexa.

1.2 Objetivos

O objetivo desta tese é fornecer o embasamento teórico necessário para que a correntropia possa ser aplicada a problemas que envolvam dados complexos.

1.3 Contribuições

Destaca-se as seguintes contribuições desta tese:

- Este trabalho define uma nova medida de similaridade entre variáveis aleatórias complexas denominada correntropia complexa;
- Propriedades dessa nova medida são estudadas para mostrar que a correntropia complexa herda características importantes da correntropia convencional;
- É definido o critério de máxima correntropia complexa (Maximum Complex Correntropy Criterion - MCCC), que estuda o uso da correntropia complexa enquanto função custo;
- Mostra-se duas soluções a partir do MCCC, uma baseado no gradiente ascendente e uma segunda em ponto-fixa;
- O MCCC é aplicado ao problema de identificação de sistemas lineares, equalização de canal, amostragem compressiva e similaridade entre ângulos;
- São propostas 3 diferentes estratégias de *kernel* adaptativo para o MCCC que são estudadas no problema de identificação de sistemas;
- É definida a métrica induzida da correntropia complexa (Complex Correntropy Induced Metric - CCIM);
- O MCCC e a CCIM são aplicados a um problema de amostragem compressiva para reconstruir sinais complexos esparsos.

1.4 Organização do trabalho

Esta tese está organizada em mais 5 capítulos além desta introdução. A fundamentação teórica para entendimento desta tese é apresentada no capítulo 2. São abordados brevemente conceitos relacionados aos números complexos, espaço de Hilbert e da medida de similaridade correntropia. O Capítulo 3 define a correntropia complexa, estuda suas propriedades e seu uso enquanto função custo no MCCC. É apresentada uma solução baseada no gradiente ascendente e uma segunda em ponto fixo. Estratégias de *kernel* adaptativo também são estudadas nesse capítulo. A correntropia complexa é aplicada a problemas de identificação de sistemas lineares, equalização de canal, amostragem compressiva e similaridade entre ângulos no Capítulo 4. As conclusões e perspectivas de trabalhos futuros são discutidas no Capítulo 5.

Além disso, 6 anexos complementam o texto. O anexo A mostra a prova da correntropia como integral de f_{xy} na reta $x = y$. Já o anexo B estuda a expansão em série de Taylor da correntropia com o *kernel* gaussiano. A correntropia complexa como integral de f_{xyzs} no plano $x = y$ e $z = s$ é visto no anexo C. A análise da convergência do algoritmo

de gradiente ascendente com a correntropia complexa como função custo é mostrado no anexo D. Por sua vez, o anexo E mostra a dedução do algoritmo de ponto fixo sem usar o cálculo de Wirtinger. Por fim, o anexo F mostra uma proposta de *kernel* com retorno complexo e prova que o mesmo é positivo definido.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta brevemente a fundamentação teórica necessária ao entendimento desta tese. Primeiramente, um resumo sobre o domínio dos números complexos é apresentado. Em seguida, elementos necessários ao entendimento do espaço de Hilbert são apresentados. Por fim, explora-se a medida de similaridade denominada correntropia, suas propriedades e seu uso enquanto função custo.

2.1 Números Complexos

2.1.1 Histórico

Os primeiros relatos referentes a raízes de números negativos remetem a meados do século I, nos trabalhos de Heron de Alexandria acerca de volumes de corpos geométricos [Smith 1925]. Os registros seguintes são de aproximadamente 250 anos depois, quando Diphantus buscava encontrar os lados de um triângulo de diâmetro 12 unidades e área 7 unidades quadradas, resultando numa equação com raízes complexas. Mandic e Goh (2009) indicam a origem dos números complexos como uma consequência do desenvolvimento humano. Assim como os números naturais estão relacionados a contagem ou frações estão relacionadas a compartilhamento, os números complexos surgiram da necessidade de solucionar equações do tipo $x^2 = -1$. Hoje, os números complexos são estudados e usados em diversas áreas do conhecimento como física, processamento de sinais e matemática aplicada.

2.1.2 Introdução

Seja o número complexo $z \in \mathbb{C}$, tal que $z = a + jb$, onde $a, b \in \mathbb{R}$, a é chamado parte real do número complexo z , e pode ser obtido a partir da operação $\Re(z) = a$, enquanto que b é chamada parte imaginária do número complexo z , obtida pela operação $\Im(z) = b$. $j = \sqrt{-1}$ é a unidade imaginária. Um número complexo z cuja parte real vale $a = 0$, é denominado puramente imaginário. Se o número complexo z tiver a parte imaginária $b = 0$, trata-se de um número real, $z \in \mathbb{R}$. Zero é o único número que é, ao mesmo tempo, real e puramente imaginário. Dois números complexos $z_1, z_2 \in \mathbb{C}$ são iguais se, e somente se, suas

partes reais forem iguais ($\Re(z_1) = \Re(z_2)$) assim também como suas partes imaginárias ($\Im(z_1) = \Im(z_2)$) [Ahlfors 1979].

2.1.3 Complexo Conjugado

O superescrito $(\cdot)^*$ é usado nesse trabalho para indicar o conjugado do respectivo número complexo. Assim, se $z \in \mathbb{C}$ e $z = a + jb$, então $z^* = a - jb$, onde a, b são números reais. Vale salientar as seguintes propriedades. Sejam $z, w \in \mathbb{C}$:

$$(z \pm w)^* = z^* \pm w^*, \quad (2.1)$$

$$(zw)^* = z^* w^*, \quad (2.2)$$

$$\left(\frac{z}{w}\right)^* = \frac{z^*}{w^*}, \quad (2.3)$$

$$(z^*)^* = z. \quad (2.4)$$

É possível obter as partes reais e imaginárias de um número complexo manipulando ele e seu conjugado:

$$\Re(z) = \frac{z + z^*}{2}, \quad (2.5)$$

$$\Im(z) = \frac{z - z^*}{2j}. \quad (2.6)$$

2.1.4 Módulo ou Valor Absoluto de um número complexo

O produto de um número complexo z pelo seu conjugado z^* é sempre positivo ou zero. A raiz quadrada desse produto é denominado módulo ou valor absoluto, ou seja:

$$zz^* = |z|^2 = a^2 + b^2. \quad (2.7)$$

Seja $z, w \in \mathbb{C}$, vale ressaltar as seguintes propriedades:

$$|z| = |z^*| \quad (2.8)$$

$$|zw| = |z||w| \quad (2.9)$$

$$|z + w|^2 = |z|^2 + |w|^2 + 2\Re(zw^*) \quad (2.10)$$

$$|z + w| \leq |z| + |w| \quad (2.11)$$

A Inequação (2.11) é chamada desigualdade triangular.

2.1.5 Geometria dos números complexos

Uma representação comum do número complexo z se dá a partir do uso de um par ordenado $(\Re(z), \Im(z))$ num plano formado pelo eixo dos números reais e pelo eixos dos números imaginários, que é denominado plano complexo. Neste caso, pode-se referir ao

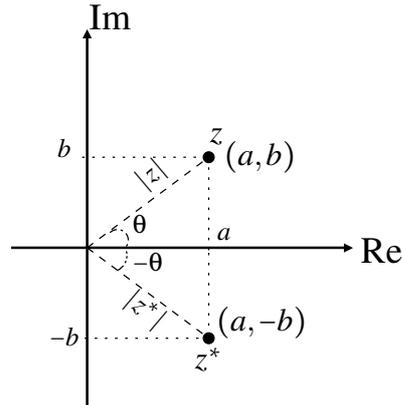


Figura 2.1: Plano complexo formado pelo eixo real e imaginário. Número complexo $z = a + jb$, formado pelo par ordenado $(\Re(z), \Im(z))$. Módulo ($|z|$) e fase (θ) também podem ser usados para representar z . O seu complexo conjugado $z^* = a - jb$ também é mostrado.

número complexo z como ponto z [Ahlfors 1979]. Uma outra alternativa de representação do ponto z usando o plano complexo se dá através de suas coordenadas polares, ou seja, usando o seu módulo, também chamado de valor absoluto de z (distância entre o ponto z e a origem do plano complexo) e sua fase, também chamado de argumento (ângulo entre o módulo e o eixo real positivo). Além disso, a partir da fórmula de Euler $\exp(jx) = \cos(x) + j\sin(x)$, é possível escrever um número complexo z como

$$z = |z|(\cos(\theta) + j\sin(\theta)) = |z|\exp(j\theta). \quad (2.12)$$

Essa representação também é chamada de forma trigonométrica. A figura 2.1 sintetiza os conceitos descritos acima.

2.1.6 Vetores e Matrizes Complexas

Por padrão, esta tese trabalha com vetores coluna. Utiliza-se a notação $\mathbf{z} \in \mathbb{C}^{M \times 1}$ ou simplesmente $\mathbf{z} \in \mathbb{C}^M$ para descrever a dimensão de \mathbf{z} , assim:

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_M \end{bmatrix}, \quad (2.13)$$

onde $z_k \in \mathbb{C}$, e $k = 1, \dots, M$.

Seja \mathbf{Z} uma matriz complexa de dimensão $M \times N$, ou seja, M linhas e N colunas, escreve-se $\mathbf{Z} \in \mathbb{C}^{M \times N}$:

$$\mathbf{Z} = \begin{bmatrix} z_{11} & \cdots & z_{1N} \\ z_{21} & \cdots & z_{2N} \\ \vdots & \ddots & \vdots \\ z_{M1} & \cdots & z_{Mn} \end{bmatrix}, \quad (2.14)$$

onde $z_{kl} \in \mathbb{C}$, $k = 1, \dots, M$ e $l = 1, \dots, N$.

2.1.7 Operador Hermitiano

O operador hermitiano $(\cdot)^H$ é usado na literatura relacionada ao processamento de sinais complexos. Ele equivale ao conjugado do transposto do vetor ou matriz em questão. Assim, por exemplo, o hermitiano do vetor complexo \mathbf{Z} da Equação (2.13) é

$$\mathbf{z}^H = [z_1^* \quad z_2^* \quad \cdots \quad z_m^*], \quad (2.15)$$

enquanto que o hermitiano da matriz complexa \mathbf{Z} , expressa na Equação (2.14) é:

$$\mathbf{Z}^H = (\mathbf{Z}^*)^T = (\mathbf{Z}^T)^* = \begin{bmatrix} z_{11}^* & \cdots & z_{M1}^* \\ z_{12}^* & \cdots & z_{M2}^* \\ \vdots & \ddots & \vdots \\ z_{1N}^* & \cdots & z_{MN}^* \end{bmatrix}. \quad (2.16)$$

2.1.8 Funções Complexas

Ahlfors (1979) ressalta que o estudo de funções de variáveis complexas busca estender o cálculo ao domínio complexo. Seja f uma função complexa $f: \mathbb{D} \rightarrow \mathbb{C}$, é possível escrever f da seguinte forma [Mandic e Goh 2009]:

$$f(z) = u(z) + jv(z), \quad (2.17)$$

ou ainda

$$f(z) = f(a, b) = u(a, b) + jv(a, b). \quad (2.18)$$

Como no domínio real, o limite de uma função $f(z)$ quando z tende a z_0 é ζ :

$$\lim_{z \rightarrow z_0} f(z) = \zeta \quad (2.19)$$

só é válido se, e somente se, para qualquer $\varepsilon > 0$ existir um número $\delta > 0$ tal que $|f(z) - \zeta| < \varepsilon$ e $0 < |z - z_0| < \delta$. A função f será contínua se

$$\lim_{z \rightarrow z_0} f(z) = f(z_0), \quad (2.20)$$

para todo $z_0 \in \mathbb{D}$.

Uma vez estabelecido o conceito de limite é possível dizer que, para que uma função complexa $f(z) = u(z) + jv(z)$ seja diferenciável em \mathbb{C} , o limite (2.21) deve convergir para um único valor complexo independente de como $\Delta z \rightarrow 0$. Ou seja,

$$\frac{df}{dz} = f'(z) = \lim_{\Delta z \rightarrow 0} \frac{[u(z + \Delta z) + jv(z + \Delta z)] - [u(z) + jv(z)]}{\Delta z} \quad (2.21)$$

Entretanto, como $z = a + jb$, deve-se considerar tanto $\Delta a \rightarrow 0$ quanto $\Delta b \rightarrow 0$. Assim, é possível reescrever a Equação (2.21) como

$$f'(a, b) = \lim_{\Delta a \rightarrow 0, \Delta b \rightarrow 0} \frac{[u(a + \Delta a, b + \Delta b) + jv(a + \Delta a, b + \Delta b)] - [u(a, b) + jv(a, b)]}{\Delta a + j \Delta b} \quad (2.22)$$

Vale ressaltar que não importa como Δz se aproxime de 0, seja pela parte real $\Delta a \rightarrow 0$, ou pela parte imaginária $\Delta b \rightarrow 0$, o limite da Equação (2.22) deve convergir para o mesmo valor. Mandic e Goh (2009) afirma que é conveniente considerar nesse estudo os seguintes casos [Needham 1997, Mathews e Howell 2012]:

Caso 1: $\Delta b = 0$ e $\Delta a \rightarrow 0$

Neste caso, temos:

$$f'(a, b) = \lim_{\Delta a \rightarrow 0} \frac{[u(a + \Delta a, b) + jv(a + \Delta a, b)] - [u(a, b) + jv(a, b)]}{\Delta a} \quad (2.23)$$

Rearranjando a Equação (2.23) para agrupar parte real e parte imaginária, temos:

$$\begin{aligned} f'(a, b) &= \lim_{\Delta a \rightarrow 0} \frac{u(a + \Delta a, b) - u(a, b)}{\Delta a} + j \frac{v(a + \Delta a, b) - v(a, b)}{\Delta a} \\ &= \frac{\partial u(a, b)}{\partial a} + j \frac{\partial v(a, b)}{\partial a} \end{aligned} \quad (2.24)$$

Caso 2: $\Delta a = 0$ e $\Delta b \rightarrow 0$

$$f'(a, b) = \lim_{\Delta b \rightarrow 0} \frac{[u(a, b + \Delta b) + jv(a, b + \Delta b)] - [u(a, b) + jv(a, b)]}{j \Delta b} \quad (2.25)$$

A Equação (2.25) pode ser rearranjada para agrupar a parte real e parte imaginária, observando a unidade imaginária j no denominador:

$$\begin{aligned}
f'(a,b) &= \lim_{\Delta b \rightarrow 0} \frac{u(a,b+\Delta b) - u(a,b)}{j \Delta b} + j \frac{v(a,b+\Delta b) - v(a,b)}{j \Delta b} \\
&= \lim_{\Delta b \rightarrow 0} \frac{u(a,b+\Delta b) - u(a,b)}{j \Delta b} + \frac{v(a,b+\Delta b) - v(a,b)}{\Delta b} \\
&= \lim_{\Delta b \rightarrow 0} \frac{v(a,b+\Delta b) - v(a,b)}{\Delta b} - j \frac{u(a,b+\Delta b) - u(a,b)}{\Delta b} \\
&= \frac{\partial v(a,b)}{\partial b} - j \frac{\partial u(a,b)}{\partial b}
\end{aligned} \tag{2.26}$$

Dessa forma, como as equações (2.24) e (2.26) devem ser iguais. Igualando as partes real e imaginária das equações (2.24) e (2.26), temos:

$$\frac{\partial u(a,b)}{\partial a} = \frac{\partial v(a,b)}{\partial b}, \quad \frac{\partial v(a,b)}{\partial a} = -\frac{\partial u(a,b)}{\partial b}, \tag{2.27}$$

que são as chamadas equações de Cauchy-Riemann [Ahlfors 1979, Mandic e Goh 2009]. Assim, para que uma função $f(z) : \mathbb{C} \rightarrow \mathbb{C}$ seja diferenciável, não apenas as derivadas parciais de $\partial u(a,b)/\partial a$ e $\partial v(a,b)/\partial b$ devem existir mas, também devem satisfazer as equações (2.27). Essa classe de funções é denominada analítica ou holomórfica.

Funções holomórficas

Neste sentido, na definição clássica de função, apenas as funções holomórficas (analíticas) são diferenciáveis no domínio complexo [Ahlfors 1979]. Para se testar se uma função $f(z)$ é holomórfica, basta aplicar o teste das equações (2.27). Por exemplo, a função $f(z) = z^2$. Primeiro devemos reescreve-la da forma $f(a,b)$:

$$\begin{aligned}
f(z) &= f(a,b) = z^2 \\
&= (a + jb)^2 = a^2 + 2jab + j^2b^2 \\
&= (a^2 - b^2) + j(2ab)
\end{aligned} \tag{2.28}$$

Assim, temos que $u(a,b) = (a^2 - b^2)$ e que $v(a,b) = 2ab$. Calculando as derivadas parciais, temos:

$$\frac{\partial u(a,b)}{\partial a} = 2a \tag{2.29}$$

$$\frac{\partial u(a,b)}{\partial b} = -2b \tag{2.30}$$

$$\frac{\partial v(a,b)}{\partial a} = 2b \tag{2.31}$$

$$\frac{\partial v(a,b)}{\partial b} = 2a \tag{2.32}$$

Podemos então verificar que as derivadas parciais acima atendem as equações de

Cauchy-Riemann (2.27). Dessa forma, $f(z) = z^2$ é uma função analítica. As regras de diferenciação se aplicam normalmente, portanto, $f'(z) = 2z$.

Por último, vamos analisar um exemplo pertinente, $f(z) = zz^*$. Como no primeiro exemplo, vamos reescrevê-la na forma $f(a, b)$:

$$\begin{aligned} f(z) &= f(a, b) = zz^* \\ &= (a + jb)(a - jb) \\ &= a^2 - j^2b^2 \\ &= a^2 + b^2 \end{aligned} \tag{2.33}$$

Dessa vez, $u(a, b) = a^2 + b^2$ enquanto que $v(a, b) = 0$. Calculando as derivadas parciais, temos:

$$\frac{\partial u(a, b)}{\partial a} = 2a \tag{2.34}$$

$$\frac{\partial u(a, b)}{\partial b} = 2b \tag{2.35}$$

$$\frac{\partial v(a, b)}{\partial a} = 0 \tag{2.36}$$

$$\frac{\partial v(a, b)}{\partial b} = 0 \tag{2.37}$$

Como se pode ver, por não atender as equações de Cauchy-Riemann, $f(z) = zz^*$ não é analítica em \mathbb{C} , ou seja, não pode ser diferenciada ou integrada. Neste caso, $f(z) = zz^*$ é uma função $f: \mathbb{C} \rightarrow \mathbb{R}$. Esse tipo de função é muito comum na área de processamento de sinais onde, em geral, funções custo com resposta real e argumento complexo são usadas em filtros adaptativos [Mandic e Goh 2009]. Para resolver esse tipo de problema, um feramental matemático denominado Cálculo de Wirtinger é apresentado a seguir. Maiores detalhes sobre diferenciação e integração de variáveis complexas podem ser encontradas em Rudin (1987).

2.1.9 Cálculo de Wirtinger

O cálculo de Wirtinger, também chamado de $\mathbb{C}\mathbb{R}$ -Cálculo [Mandic e Goh 2009], foi introduzido pela primeira vez em [Wirtinger 1927] e é baseado na dualidade entre os espaços \mathbb{C} e \mathbb{R}^2 . Trabalhar em \mathbb{R}^2 pode tornar os cálculos muito mais longos, deseja-se aproveitar a estrutura dos números complexos para simplificar o processo.

Como já fizemos outras vezes nesse capítulo, seja a função complexa $f: \mathbb{C} \rightarrow \mathbb{C}$, podemos trabalhar com a sua versão no \mathbb{R}^2 . Assim, a função $f(z)$ pode ser escrita também como

$$f(z) = f(a + jb) = f(a, b) = u(a, b) + jv(a, b) \tag{2.38}$$

Escrevendo a e b usando a Equação (2.5), temos:

$$a = \frac{z+z^*}{2} \text{ e } b = \frac{z-z^*}{2j}. \quad (2.39)$$

Usando a regra da cadeia, é possível escrever:

$$\begin{aligned} \frac{\partial f}{\partial z} &= \frac{\partial f}{\partial a} \frac{\partial a}{\partial z} + \frac{\partial f}{\partial b} \frac{\partial b}{\partial z} \\ &= \frac{\partial f}{\partial a} \frac{1}{2} + \frac{\partial f}{\partial b} \frac{1}{2j} \\ &= \frac{\partial f}{\partial a} \frac{1}{2} - \frac{\partial f}{\partial b} \frac{j}{2} \end{aligned} \quad (2.40)$$

e

$$\begin{aligned} \frac{\partial f}{\partial z^*} &= \frac{\partial f}{\partial a} \frac{\partial a}{\partial z^*} + \frac{\partial f}{\partial b} \frac{\partial b}{\partial z^*} \\ &= \frac{\partial f}{\partial a} \frac{1}{2} - \frac{\partial f}{\partial b} \frac{1}{2j} \\ &= \frac{\partial f}{\partial a} \frac{1}{2} + \frac{\partial f}{\partial b} \frac{j}{2} \end{aligned} \quad (2.41)$$

O ponto chave é considerar as variáveis z e z^* como independentes entre si [Adali e Haykin 2010]. Assim, colocando $1/2$ em evidência na Equação (2.40), a *derivada de Wirtinger*, também conhecida como *W-Derivada* de f é definida da seguinte forma [Bouboulis e Theodoridis 2011]:

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial a} - j \frac{\partial f}{\partial b} \right) \quad (2.42)$$

Da mesma forma, colocando $1/2$ em evidência na Equação (2.41), a *derivada conjugada de Wirtinger*, também chamada de *CW-Derivada* de f pode ser definida como:

$$\frac{\partial f}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial a} + j \frac{\partial f}{\partial b} \right) \quad (2.43)$$

O trabalho de [Bouboulis e Theodoridis 2011] apresenta uma revisão completa sobre as propriedades associadas ao Cálculo de Wirtinger. Em síntese, para computar a W-Derivada de uma dada função f , que é expressa em termos de z e z^* , basta aplicar as regras de diferenciação sempre considerando z^* como uma constante. De forma semelhante, seguindo o mesmo raciocínio, para computar a CW-Derivada de uma função f , também expressa em termos de z e z^* , deve-se aplicar as regras de diferenciação padrão considerando z como uma constante [Bouboulis e Theodoridis 2011]. Assim, considerando f como $f(z) = zz^*$, como exemplo, pode-se concluir que:

$$\frac{\partial f}{\partial z} = z^* \quad \text{e} \quad \frac{\partial f}{\partial z^*} = z \quad (2.44)$$

Uma outra grande vantagem do Cálculo de Wirtinger é que quando aplicado a funções analíticas em \mathbb{C} , obtêm-se a mesma resposta que utilizando a diferenciação padrão.

2.2 Espaço de Hilbert

Antes de apresentar a definição de espaço de Hilbert (*Hilbert Space*) é necessário mostrar, mesmo que brevemente, alguns conceitos importantes para o entendimento do mesmo.

2.2.1 Espaço vetorial

Um espaço vetorial é uma conjunto de objetos denominados vetores que podem ser somados e multiplicados por números (escalares), permanecendo no mesmo conjunto [Vetterli et al. 2014]. Para qualquer vetor $\mathbf{x}, \mathbf{y}, \mathbf{z}$ em \mathcal{V} e $a, b \in \mathbb{C}$, os seguintes axiomas devem ser seguidos:

- *Comutatividade*: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$;
- *Associatividade*: $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ e $(ab)\mathbf{x} = a(b\mathbf{x})$
- *Distributividade*: $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ e $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$
- *Elemento neutro*: Existe o elemento $\mathbf{0}$ em \mathcal{V} tal que $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ para todo \mathbf{x} em \mathcal{V} ;
- *Elemento oposto*: Para cada \mathbf{x} em \mathcal{V} existe um elemento único $-\mathbf{x}$ em \mathcal{V} tal que $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$;
- *Identidade multiplicativa*: Para cada \mathbf{x} em \mathcal{V} , $1 \cdot \mathbf{x} = \mathbf{x}$.

Um exemplo de espaço vetorial é \mathbb{C}^N que foi apresentado na seção anterior. Trata-se do espaço vetorial formado de vetores complexos de dimensão finita:

$$\mathbb{C}^N = \{ \mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T \mid x_n \in \mathbb{C}, n \in \{1, 2, \dots, N\} \} \quad (2.45)$$

Outro exemplo interessante é o espaço vetorial de funções complexas sobre \mathbb{R} , ou seja:

$$\mathbb{C}^{\mathbb{R}} = \{ x \mid x(t) \in \mathbb{C}, t \in \mathbb{R} \} \quad (2.46)$$

2.2.2 Produto interno

Sejam $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$ e $(a \in \mathbb{C})$, designa-se como o produto interno em um espaço vetorial \mathcal{V} sobre \mathbb{C} a função complexa $\langle \cdot \rangle$ definida em $\mathcal{V} \times \mathcal{V}$ com as seguintes propriedades:

- *Distributividade*: $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;
- *Linearidade no Primeiro argumento*: $\langle a\mathbf{x}, \mathbf{y} \rangle = a\langle \mathbf{x}, \mathbf{y} \rangle$;
- *Simetria Hermitiana*: $\langle \mathbf{x}, \mathbf{y} \rangle^* = \langle \mathbf{y}, \mathbf{x} \rangle$;
- *Positivo definido*: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, e $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ se, e somente se, $\mathbf{x} = \mathbf{0}$.

É interessante notar que, por causa da segunda e terceira propriedade, temos que: $\langle \mathbf{x}, a\mathbf{y} \rangle = a^* \langle \mathbf{x}, \mathbf{y} \rangle$.

O produto interno padrão em \mathbb{C}^N é

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^N x_n y_n^* = \mathbf{y}^H \mathbf{x} \quad (2.47)$$

2.2.3 Espaços vetoriais dotados de produto interno

Um espaço vetorial dotado de uma operação de produto interno conforme definido na seção 2.2.2 é chamado de espaço de produto interno. Uma outra denominação para esses tipos de espaços vetoriais é de espaço pré-hilbertiano.

2.2.4 Norma

A norma é uma função que impõe a um vetor a ideia de comprimento ou tamanho. É o equivalente vetorial a magnitude de um escalar.

Seja $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$ e $(a \in \mathbb{C})$, a norma em um espaço vetorial \mathcal{V} em \mathbb{C} é uma função real $\|\cdot\|$ definida em \mathcal{V} , com as seguintes propriedades:

- *Positivo definido:* $\|\mathbf{x}\| \geq 0$, e $\|\mathbf{x}\| = 0$ se, e somente se, $\mathbf{x} = \mathbf{0}$.
- *Escalabilidade positiva:* $\|a\mathbf{x}\| = |a| \|\mathbf{x}\|$
- *Desigualdade Triangular:* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Um produto interno pode ser usado para induzir uma norma. Neste caso se diz que a norma é induzida pelo produto interno. Por exemplo, a norma padrão em \mathbb{C}^N , induzida pelo produto interno é:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \quad (2.48)$$

As normas induzidas por produtos internos tem algumas propriedades específicas [Vetterli et al. 2014]. Destaca-se aqui a desigualdade de Cauchy-Schwarz:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|, \quad (2.49)$$

com a igualdade sendo atingida se, e somente se, $\mathbf{x} = a\mathbf{y}$, para algum escalar a .

2.2.5 Espaço vetorial normado

Um espaço vetorial dotado de uma operação de norma é chamado espaço vetorial normado.

2.2.6 Métrica

A norma de um vetor \mathbf{x} , ou seja, seu comprimento, pode ser vista como a distância desse vetor a origem. Dessa forma, em um espaço vetorial normado, define-se uma métrica (ou distância) entre os vetores \mathbf{x} e \mathbf{y} como a norma da diferença entre eles, ou seja:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|. \quad (2.50)$$

2.2.7 Completude

Um espaço vetorial normado \mathcal{V} é dito completo quando toda sequência de Cauchy em \mathcal{V} converge para um vetor também em \mathcal{V} . Recomenda-se aqui a leitura do capítulo 2 do livro [Vetterli et al. 2014] para o estudo dos conceitos formais de completude. De forma intuitiva, um espaço vetorial é completo quando é garantido que qualquer sequência deve convergir e permanecer no mesmo espaço vetorial.

O espaço vetorial formado por sequências de números racionais \mathbb{Q} não é um espaço vetorial completo, já que, por exemplo:

$$x_k = \sum_{n=0}^k \frac{1}{n!}, \quad (2.51)$$

é racional para qualquer inteiro não negativo k mas, fazendo o limite dessa sequência para $k \rightarrow \infty$, a sequência converge para e , que é irracional.

2.2.8 Espaço de Banach

Um espaço vetorial normado completo é chamado de Espaço de Banach [Rudin 1987].

2.2.9 Espaço de Hilbert

Como foi dito anteriormente, um espaço vetorial dotado de uma operação de produto interno é chamado de espaço pré-hilbertiano. O espaço de Hilbert é um espaço vetorial que além de ser dotado de produto interno também é completo. A Figura 2.2 resume os conceitos descritos na seção mostrando a relação entre os tipos de espaços vetoriais.

2.2.10 Kernel Positivo Definido

Mercer e Forsyth (1909) mostraram que, considerando um conjunto χ , seja o *kernel* $\kappa: \chi \times \chi \rightarrow \mathbb{C}$, κ é dito como um *kernel* positivo-definido se, para todo $x_1, \dots, x_n \in \chi$ e $c_1, \dots, c_n \in \mathbb{C}$

$$\sum_{i=1}^n \sum_{k=1}^n \kappa(x_i, x_k) c_i c_k^* \geq 0. \quad (2.52)$$

Proposição 1: Seja $\kappa_i: \chi \times \chi \rightarrow \mathbb{C}$ ($i = 1, 2, \dots$) *kernels* positivos definidos, as relações a seguir também formam *kernels* positivos definidos [Aronszajn 1950]:

- Combinação positiva: $a\kappa_1 + b\kappa_2$, $a, b \geq 0$
- Produto: $\kappa_1 \kappa_2$, $(\kappa_1(x, y) \kappa_2(x, y))$
- Limite: $\lim_{i \rightarrow \infty} \kappa_i(y)$, assumindo que os limites existem.

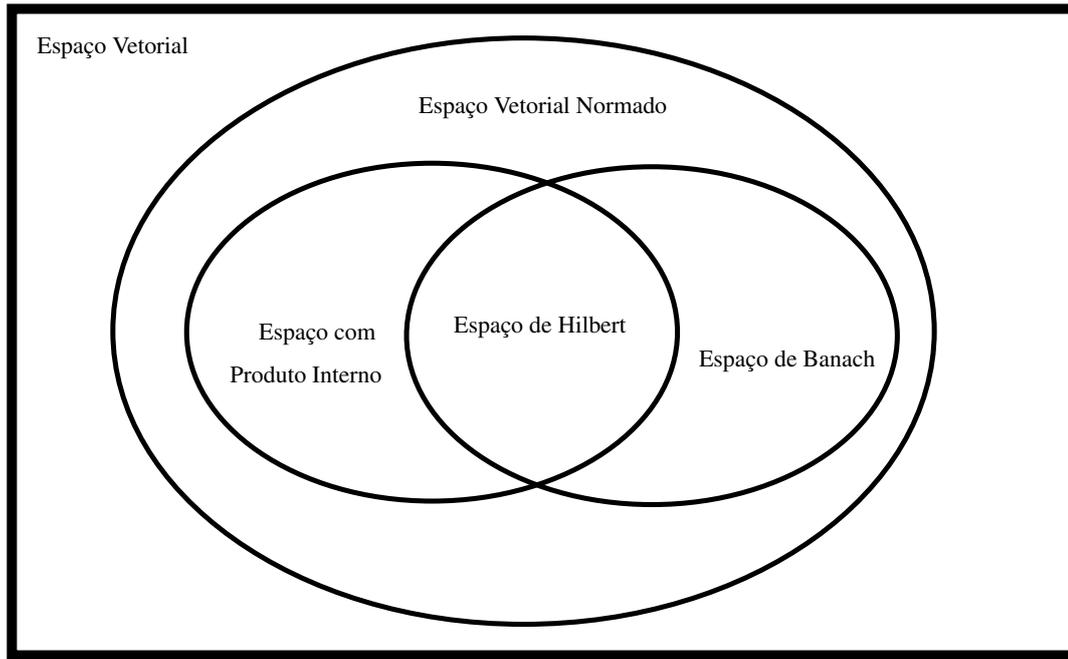


Figura 2.2: Diagrama mostrando a relação entre os tipos de espaços vetoriais.

Exemplo:

$$\kappa'(x, y) = e^{\kappa(x, y)} = 1 + \kappa + \frac{1}{2}\kappa^2 + \frac{1}{3!}\kappa^3 + \dots \quad (2.53)$$

é positivo definido.

Proposição 2: Seja $\kappa : \chi \times \chi \rightarrow \mathbb{C}$ um *kernel* positivo definido e $f : \chi \rightarrow \mathbb{C}$, uma função arbitrária, então:

$$\tilde{\kappa}(x, y) = f(x) \kappa(x, y) \overline{f(y)}, \quad (2.54)$$

onde $\overline{(\cdot)}$ é o conjugado da função argumento.

$\tilde{\kappa}(x, y)$ é positivo definido assim como $f(x)\overline{f(y)}$ [Aronszajn 1950].

2.2.11 Reproducing Kernel Hilbert Spaces

De acordo com o teorema de Mercer [Mercer e Forsyth 1909], todo *kernel* positivo-definido pode ser escrito como:

$$\kappa(x, y) = \sum_{n=0}^{\infty} \lambda_n \phi_n(x) \phi_n(y) \quad (2.55)$$

onde $\{\phi_n(\cdot), n = 1, 2, \dots\}$ e $\{\lambda_n, n = 1, 2, \dots\}$ são a sequência de autofunções e seus respectivos autovalores correspondentes ao *kernel* κ .

Aronszajn (1950) mostrou que todo *kernel* positivo definido mapeia os dados originais x (espaço de entrada) num espaço de Hilbert funcional único dotado de uma propriedade

de reprodução do *kernel* (*reproducing property*):

$$f(x) = \langle \kappa(\cdot, x), f \rangle, \forall f \in \mathcal{F}. \quad (2.56)$$

Esse espaço é chamado de *Reproducing Kernel Hilbert Space* (Espaço de Hilbert com Kernel Reprodutor - RKHS). Nele, uma transformação não linear

$$\kappa(\cdot, x) = \Phi(x) \mapsto \sqrt{\lambda_n} \phi_n, n = 1, 2, \dots \quad (2.57)$$

leva os dados do espaço de entrada (x) para um espaço de características (*feature space*) onde pode-se computar o produto interno

$$\langle \Phi(x), \Phi(y) \rangle = \langle \kappa(\cdot, x), \kappa(\cdot, y) \rangle = \kappa(x, y) \quad (2.58)$$

A Equação (2.58) é conhecida como truque do *kernel*. Dependendo do kernel utilizado, a dimensão do espaço de características pode ser arbitrariamente grande e até infinito. Dessa forma, os dados x de um espaço de entrada podem ser levados ao um espaço de alta dimensionalidade. A partir dessa ideia, é possível obter versões não-lineares de qualquer algoritmo linear expresso em termos de um produto interno sem precisar conhecer ao certo o mapeamento Φ [Principe 2010].

2.3 Correntropia

Esta seção faz uma pequena revisão da medida de similaridade denominada correntropia. Sua definição é apresentada junto com algumas de suas propriedades. Em especial, a propriedade que indica o significado probabilístico da correntropia é aprofundado já que o mesmo foi usado como base na dedução da correntropia complexa, principal contribuição dessa tese. Ao fim dessa seção, o uso da correntropia enquanto função custo é apresentado.

2.3.1 Definição

Sejam X e Y variáveis aleatórias reais arbitrárias, define-se como correntropia a medida de similaridade generalizada [Liu et al. 2007]:

$$V_\sigma(X, Y) = E_{XY}[k_\sigma(X, Y)] \quad (2.59)$$

em que $E[\cdot]$ é o operador de valor esperado e $k_\sigma(\cdot, \cdot)$ é um *kernel* positivo-definido contínuo. σ é um parâmetro livre chamado largura ou tamanho do *kernel* $k(\cdot, \cdot)$. Como o valor esperado de uma variável aleatória X é dado por

$$E_X[X] = \int_{-\infty}^{\infty} x f_x(x) dx, \quad (2.60)$$

onde $f_x(x)$ é a função densidade de probabilidade de X , e que

$$E_X[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx, \quad (2.61)$$

onde $g(X)$ é uma função da variável X [Papoulis 1984], temos que a Equação 2.59 também pode ser escrita como

$$V_{\sigma}(X, Y) = \iint k_{\sigma}(X, Y)f_{XY}(x, y)dxdy, \quad (2.62)$$

onde $f_{XY}(x, y)$ é a função densidade de probabilidade conjunta de X e Y .

2.3.2 Estimação da correntropia com o *kernel* gaussiano

Como será mostrado em seguida, o *kernel* gaussiano

$$G_{\sigma}(x-y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \quad (2.63)$$

apresenta propriedades interessantes, sendo bastante utilizado na literatura [Liu et al. 2007]. De modo geral, a função densidade de probabilidade conjunta de X e Y é desconhecida e apenas um número finito de amostras $\{(x_i, y_i)\}_{i=1}^N$ estão disponíveis. Dessa forma, é possível estimar a correntropia como [Liu et al. 2007]

$$\hat{V}_{N, \sigma}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i, y_i). \quad (2.64)$$

O símbolo $(\hat{\cdot})$ é sempre utilizado nessa tese para indicar estimações, nesse caso, da correntropia entre X e Y .

Vale salientar que, ao estimar a correntropia, deve-se selecionar o valor do tamanho do *kernel*, σ . Esse valor pode assumir qualquer valor finito [Principe 2010], sendo assim, um parâmetro livre.

2.3.3 Propriedades

As propriedades da correntropia foram exploradas inicialmente no trabalho [Santamaria et al. 2006]. Já [Liu et al. 2007] reorganizou-as e foi complementado em [Principe 2010]. São apresentadas aqui algumas propriedades pertinentes ao entendimento desta tese.

Propriedade 1: Para *kernels* simétricos, a correntropia também é simétrica:

$$V_{\sigma}(X, Y) = E_{XY}[k_{\sigma}(X, Y)] = E_{YX}[k_{\sigma}(Y, X)] = V_{\sigma}(Y, X) \quad (2.65)$$

Propriedade 2: Para o *kernel* Gaussiano, a correntropia é positiva e limitada. A sua estimacão, \hat{V}_{σ} , é sempre limitada entre zero, menor valor possível, e $1/\sqrt{2\pi\sigma}$, maior valor que só é atingido quando $X = Y$.

$$0 \leq \hat{V}_\sigma(X, Y) \leq \frac{1}{\sqrt{2\pi\sigma}} \quad (2.66)$$

Propriedade 3: Para o *kernel* gaussiano, a correntropia é a soma ponderada de todos os momentos de ordem par da variável aleatória $X - Y$. A Equação (2.67) mostra a expansão em série de Taylor da correntropia usando o *kernel* gaussiano.

$$V_\sigma(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E_{XY}[(X - Y)^{2n}]. \quad (2.67)$$

É importante dizer que, quando o tamanho do *kernel* for suficientemente grande, os momentos estatísticos de segunda ordem tendem a se sobressair em relação aos demais. O Anexo B prova a Equação (2.67) e mostra com maiores detalhes essa característica.

Propriedade 4: Usando o *kernel* gaussiano com tamanho σ , fazendo com que o tamanho de *kernel* σ decresça até zero implica que a correntropia se aproxime do valor da densidade de probabilidade associada ao evento $X = Y$.

$$\lim_{\sigma \rightarrow 0} V_\sigma(X, Y) = \int f_{XY}(x, x) dx \quad (2.68)$$

Devido a importância dessa propriedade para entendimento desta tese, ela é estudada com mais detalhes na seção a seguir.

Propriedade 4.1: Assumindo que dados independentes e identicamente distribuídos $\{(x_i, y_i)_{i=1}^N\}$ são amostrados da distribuição conjunta f_{xy} , enquanto $\hat{f}_{\sigma, xy}$ é a estimação de Parzen com tamanho de *kernel* σ , a correntropia estimada com o tamanho de *kernel* $\sigma' = \sigma\sqrt{2}$ corresponde a integral de $\hat{f}_{\sigma, xy}$ sobre a linha $x = y$.

$$\hat{V}_{\sigma'}(X, Y) = \int_{-\infty}^{\infty} \hat{f}_{\sigma, XY}(x, y) du \Big|_{x=y=u} \quad (2.69)$$

A dedução completa dessa propriedade é apresentada no Anexo A.

Propriedade 5: Seja $E = X - Y$ com PDF f_E , sob as condições $N \rightarrow \infty$, $\hat{V}_{N, \sigma}(X, Y)$ é um estimador consistente em erro médio quadrado de $V_\sigma(X, Y)$. Além disso, sob as condições $N\sigma \rightarrow \infty$ e $\sigma \rightarrow 0$, $\hat{V}_{N, \sigma}(X, Y)$ é um estimador assintoticamente não enviesado de $f_E(0)$ e consistente em erro médio quadrado.

Propriedade 6: A correntropia é uma estatística de segunda ordem dos dados projetados $\Phi(X)$ e $\Phi(Y)$ em um *kernel* reproduzidor de um espaço de Hilbert (*Reproducing Kernel Hilbert Space - RKHS*). O que implica em

$$V_\sigma(X, Y) = E[k_\sigma(X, Y)] = E[\langle \Phi(X), \Phi(Y) \rangle_H] = E[\Phi(X)^T \Phi(Y)] \quad (2.70)$$

onde $\langle \cdot \rangle_H$ é o operador de produto interno em um espaço de Hilbert.

Propriedade 7: A estimação da correntropia como o *kernel* gaussiano induz uma métrica no espaço de eventos. Sejam X e Y duas variáveis aleatórias reais, a função

$$\Psi_{\sigma}(X, Y) = \sqrt{G_{\sigma}(0) - V_{\sigma}(X, Y)} \quad (2.71)$$

define uma métrica chamada métrica induzida de correntropia (*Correntropy Induced Metric* - CIM).

Propriedade 7.1: A CIM é capaz de aproximar a norma ℓ_0 quando o tamanho de *kernel* tende a zero, $\sigma \rightarrow 0$ [Seth e Principe 2008].

2.3.4 Interpretação probabilística da correntropia

Nesta seção deseja-se mostrar que calcular a correntropia, $V(X, Y)$, entre duas variáveis aleatórias reais X e Y significa estimar a densidade de probabilidade do evento $X = Y$. Essa interpretação foi primeiramente apresentada em [Liu et al. 2006] e depois transformada em propriedade em [Liu et al. 2007]. Nesta tese, tal característica foi apresentada na propriedade 4. Como essa concepção é expandida no capítulo seguinte para tratar similaridades entre números complexos, apresenta-se aqui em detalhes para melhorar o entendimento.

O conceito de correntropia está diretamente relacionado a estimação de quão similares duas variáveis são quando o estimador de Parzen [Parzen 1962] é aplicado à estimação da PDF conjunta f_{XY} [Liu et al. 2006]. A densidade de probabilidade relacionada ao evento $X = Y$ pode ser obtida por:

$$P(X = Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \delta(x - y) dx dy \quad (2.72)$$

Na maioria dos casos, a distribuição $f_{XY}(x, y)$ é desconhecida e apenas um número finito de amostras $(x_n, y_n), n = 1, 2, \dots, N$ está disponível. Entretanto, quando utiliza-se o estimador l -dimensional de Parzen com o *kernel* gaussiano, a probabilidade conjunta é dada por [Silverman 1986]:

$$\hat{f}_{X^1, X^2, \dots, X^L}(x^1, x^2, \dots, x^L) = \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^L G_{\sigma}(x^l - x_n^l). \quad (2.73)$$

$G_{\sigma}(x)$ foi definido na Equação (2.63). A notação x_n^l representa a n -ésima amostra do l -ésimo componente L -dimensional da variável aleatória. Vale ressaltar que σ é o tamanho do *kernel*. Para definir a correntropia sobre os números reais, o trabalho de [Liu et al. 2006] considera $L=2$. Assume-se também que $X^1 = X$ e $X^2 = Y$ na Equação (2.73), resultando em:

$$\hat{f}_{XY}(x, y) = \frac{1}{N} \sum_{n=1}^N G_{\sigma}(x - x_n) G_{\sigma}(y - y_n) \quad (2.74)$$

Substituindo (2.74) em (2.72), obtêm-se:

$$\hat{P}(X = Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}(x - x_n) G_{\sigma}(y - y_n) \delta(x - y) dx dy \quad (2.75)$$

Uma vez que apenas valores diferentes de zero ocorrem ao longo da bissetriz do espaço conjunto (por causa da função delta), $x = y$, pode-se reescrever a Equação (2.75) da seguinte forma:

$$\hat{P}(X = Y) = \frac{1}{N} \sum_{n=1}^N \int_{-\infty}^{\infty} G_{\sigma}(u - x_n) G_{\sigma}(u - y_n) du \quad (2.76)$$

onde u é o valor assumido por x e y na linha $x = y$.

Como o produto de gaussianas resulta também numa gaussiana com o tamanho de *kernel* igual a raiz quadrada do original (ver Anexo A para demonstração), a Equação (2.76) pode ser escrita como:

$$\hat{P}(X = Y) = \frac{1}{N} \sum_{n=1}^N G_{\sqrt{2}\sigma}(x_n - y_n) \quad (2.77)$$

Utilizando o *kernel* gaussiano em (2.59), obtêm-se:

$$V_{\sigma}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) G_{\sigma}(x - y) dx dy \quad (2.78)$$

Comparando as Equações (2.72) e (2.78) é possível assumir que, de fato, a correntropia pode ser interpretada como a densidade do evento $X = Y$. Por enquanto, isso se diz respeito a duas variáveis aleatórias analisadas com um tamanho de *kernel* pequeno e para PDF's suaves. Além disso, a equação (2.77) provê a mesma solução para estimar a correntropia proposta por [Santamaria et al. 2006] que foi mostrada em (2.62), podendo ser interpretada como a estimação não paramétrica da densidade de probabilidade correspondente ao evento $X = Y$. O valor da correntropia tenderá ao valor da densidade de probabilidade se duas condições impostas pelo estimador de Parzen [Parzen 1962] forem satisfeitas: $N \rightarrow \infty$ enquanto $\sigma \rightarrow 0$.

2.3.5 Critério de Máxima Correntropia

Trata-se do uso da correntropia como função custo em problemas de otimização. Nessa seção considera-se a notação em negrito \mathbf{p} ou \mathbb{P} como um vetor/matriz contendo amostras da variável aleatória P , por exemplo. Considerando um modelo linear e definindo o vetor erro $\mathbf{e} = \mathbf{d} - \mathbf{y}$, como a diferença entre um sinal desejado \mathbf{d} e uma saída estimada $\mathbf{y} = \mathbf{w}^T \mathbf{X}$, onde $\mathbf{e}, \mathbf{d}, \mathbf{y} \in \mathbb{R}^{1 \times N}$, $\mathbf{w} \in \mathbb{R}^{1 \times M}$, onde M é o número de pesos, e $\mathbf{X} \in \mathbb{R}^{M \times N}$. Dessa forma, considera-se o critério de máxima correntropia (*Maximum Correntropy Criterion* -MCC) como

$$J_{MCC} = V(D, Y) = E[k_{\sigma}(D, Y)] \quad (2.79)$$

Deseja-se encontrar qual \mathbf{w} maximiza a correntropia entre D e Y . Em outras palavras, deseja-se maximizar a probabilidade de D ser igual a Y ou que o erro $\mathbf{e} = \mathbf{d} - \mathbf{y}$ seja zero através da busca de um \mathbf{w} ótimo.

$$\max_{\mathbf{w}} J_{MCC} \quad (2.80)$$

Essa função custo é utilizada em diversos trabalhos [Li et al. 2007, Bessa et al. 2009] associadas geralmente a um gradiente ascendente. Após um ajuste do tamanho do *kernel*, os resultados em ambientes dotados de ruído não gaussiano são geralmente superiores a métodos de segunda ordem.

$$J_{MCC} = E[G_{\sigma}(\mathbf{d} - \mathbf{y})] \quad (2.81)$$

Computando-se a estimativa de J_{MCC} , têm-se:

$$J_{MCC} = \frac{1}{\sqrt{2\pi\sigma}} \cdot \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(d_i - y_i)^2}{2\sigma^2}\right) \quad (2.82)$$

$$J_{MCC} = \frac{1}{\sqrt{2\pi\sigma}} \cdot \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(d_i - \mathbf{w}^T x_i)^2}{2\sigma^2}\right), \quad (2.83)$$

onde x_i é a i -ésima coluna da matriz X , $x_i = X(\cdot, i)$.

Solução usando o gradiente ascendente

Para implementar uma solução usando o gradiente ascendente, temos:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \nabla J_n \quad (2.84)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{N\sqrt{2\pi\sigma^3}} \sum_{i=1}^N \exp\left(-\frac{e_i^2}{2\sigma^2}\right) e_i x_i \quad (2.85)$$

onde $e_i = d_i - \mathbf{w}^T(n)x_i$. Usando o gradiente estocástico:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{N\sqrt{2\pi\sigma^3}} \exp\left(-\frac{e(n)^2}{2\sigma^2}\right) e(n)\mathbf{x}(n), \quad (2.86)$$

onde $e(n) = d(n) - \mathbf{w}^T(n)x(n)$, e $x(n)$ é a n -ésima coluna da matriz X , $x_n = X(\cdot, n)$.

Solução usando ponto fixo

O trabalho de [Singh e Principe 2010] introduziu um algoritmo de ponto fixo associado ao critério de máxima correntropia. Esse algoritmo produz uma convergência muito mais rápida em troca de um custo computacional maior decorrente de uma inversão matricial. Para conseguir a equação de atualização de \mathbf{w} é necessário derivar a função custo em relação aos pesos ótimos \mathbf{w} e iguala-la a zero:

$$\frac{\partial J_{MCC}}{\partial \mathbf{w}} = 0 \quad (2.87)$$

$$\frac{\partial E[G_{\sigma}(\mathbf{d} - \mathbf{y})]}{\partial \mathbf{w}} = 0 \quad (2.88)$$

$$E[G_{\sigma}(\mathbf{d} - \mathbf{y})(\mathbf{d} - \mathbf{y})\mathbf{X}] = 0 \quad (2.89)$$

$$E[G_{\sigma}(\mathbf{d} - \mathbf{y})(\mathbf{d} - \mathbf{X}^T \mathbf{w})\mathbf{X}] = 0 \quad (2.90)$$

$$E[G_{\sigma}(\mathbf{e})\mathbf{dX}] = \mathbf{w}E[G_{\sigma}(\mathbf{e})\mathbf{XX}^T] \quad (2.91)$$

$$\mathbf{w} = [E[G_{\sigma}(\mathbf{e})\mathbf{XX}^T]]^{-1}[E[G_{\sigma}(\mathbf{e})\mathbf{dX}]] \quad (2.92)$$

Portanto, obtêm-se:

$$\mathbf{w}(n+1) = \left[\sum_{i=1}^N G(\mathbf{e}_i) \mathbf{X}_i \mathbf{X}_i^T \right]^{-1} \left[\sum_{i=1}^N G(\mathbf{e}_i) \mathbf{d}_i \mathbf{X}_i \right] \quad (2.93)$$

Lembrando que $\mathbf{e}_i = \mathbf{d}_i - \mathbf{w}^T(n) \mathbf{x}_i$, implicando na recursividade característica do ponto fixo. Detalhes relacionados a implementação da equação (2.93) podem ser encontrados em [Singh e Principe 2010].

Capítulo 3

Correntropia Complexa

Neste capítulo define-se uma nova medida de similaridade entre variáveis aleatórias complexas chamada correntropia complexa. Suas propriedades são estudadas para mostrar a semelhança dessa nova medida com a correntropia convencional. É definido o Critério de Máxima Correntropia Complexa (MCCC), que visa usar a correntropia complexa como função custo. Dois algoritmos são apresentados, um baseado no gradiente ascendente e um segundo em ponto fixo. Além disso, são apresentadas estratégias de kernel adaptativo para o MCCC.

3.1 Estado da Arte

Como foi mencionado no capítulo anterior, a correntropia foi definida somente para variáveis aleatórias reais [Liu et al. 2007]. Por essa limitação, não são encontrados muitos trabalhos que utilizam a correntropia quando se tem dados complexos. Destacam-se a seguir alguns trabalhos que utilizaram simplificações ou adaptações para contornar essa limitação.

Para poder usar a correntropia com dados complexos, Kim (2010) por exemplo, utiliza um passo de gradiente com parte real e imaginária derivados separadamente para atuar numa equalização de canal cega envolvendo modulações complexas. Já o trabalho de Azam Khalili (2015) usa a exponencial do módulo do erro para conseguir tratar o dado complexo e derivar um algoritmo adaptável de estimativa de deslocamento de frequência de portadora. O trabalho de Ogunfunmi e Paul (2015) usa uma abordagem semelhante para propor um algoritmo que usa o critério de máxima correntropia com quatérnions. Além desses trabalhos, pode-se encontrar métodos de *kernel* complexos na literatura [Bouboulis e Theodoridis 2011, Boloix-Tortosa et al. 2015]. Entretanto, a aplicação da correntropia a dados complexos não foi propriamente formalizada e nenhum estudo teórico sobre suas propriedades ou seu uso em problemas de otimização foi realizado.

Rêgo (2014) generalizou em seu trabalho a correntropia sob diversas circunstâncias. Particularmente, o que Rêgo (2014) denominou como extensão 2 trata da aplicação da correntropia a variáveis aleatórias de 2×2 , o que pode ser aplicado no caso da igualdade entre números complexos, onde as partes real e imaginárias devem ser iguais ao mesmo tempo. Partindo dessa ideia, a partir da interpretação probabilística da correntro-

pia definida para os reais, deriva-se nesse capítulo uma nova medida de similaridade para variáveis aleatórias complexas que visa manter as características originais da correntropia convencional.

3.2 Interpretação Probabilística

Nesta seção, basicamente se aplicará a mesma metodologia usada para obter a expressão da correntropia convencional mostrada no capítulo anterior que consiste em usar o estimador de Parzen para medir a similaridade entre duas variáveis complexas. Sendo assim, assumindo duas variáveis complexas $C_1 = X + jZ$ e $C_2 = Y + jS$, onde $C_1, C_2 \in \mathbb{C}$, e X, Y, Z, S são variáveis aleatórias reais, é possível estimar a densidade de probabilidade do evento $C_1 = C_2$ como a densidade de probabilidade de $X = Y$ e $Z = S$. Isso se deve porque dois números complexos são iguais quando tanto a parte real quanto a parte imaginária são iguais, simultaneamente. Seja f_{XYZS} a função densidade de probabilidade conjunta, a estimação da probabilidade associada ao evento $x = y$ e $z = s$ que implica em $C_1 = C_2$ pode ser escrita da seguinte forma:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) \delta(x - y) \delta(z - s) dx dy dz ds \quad (3.1)$$

Quando $x = y$ e $z = s$, a Equação (3.1) pode ser reescrita por:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, x, z, z) dx dz$$

É possível substituir \hat{f}_{XYZS} pelo estimador de Parzen definido na Equação (2.73), assumindo $L = 4$ obtêm-se:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}(x - x_n) G_{\sigma}(x - y_n) G_{\sigma}(z - z_n) G_{\sigma}(z - s_n) dx dz \quad (3.2)$$

Resolvendo a integral dupla em (3.2) obtêm-se:

$$\hat{P}(C_1 = C_2) = \frac{1}{N} \sum_{n=1}^N G_{\sigma\sqrt{2}}(x_n - y_n) G_{\sigma\sqrt{2}}(z_n - s_n),$$

o que corresponde a estimar a correntropia para duas variáveis aleatórias complexas C_1 e C_2 . Mais detalhes dessa resolução podem ser encontrados no Anexo C. Logo, define-se como correntropia para duas variáveis complexas, ou simplesmente correntropia complexa, (usando o *kernel* Gaussiano):

$$V_{\sigma}^c(C_1, C_2) = E_{C_1 C_2} [G_{\sigma\sqrt{2}}^c(C_1 - C_2)] \quad (3.3)$$

em que $G_{\sigma}^c(C_1 - C_2)$ vale:

$$G_{\sigma}^c(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (3.4)$$

Assim, o estimador não paramétrico da correntropia complexa com o estimador de Parzen pode ser escrito como:

$$\hat{V}_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (3.5)$$

Ou, da forma alternativa:

$$\hat{V}_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \quad (3.6)$$

3.3 Definição

Embora a interpretação probabilística tenha resultado na equação (3.3) a partir do uso do *kernel* Gaussiano, não há razão pela qual não definir a correntropia complexa para qualquer *kernel* positivo definido, tal como a correntropia convencional. Sendo assim, expandimos aqui a definição de correntropia complexa entre duas variáveis aleatórias complexas, C_1 e $C_2 \in \mathbb{C}$ da seguinte forma:

$$V_{\sigma}^c(C_1, C_2) = E_{C_1 C_2} [k_{\sigma}(C_1, C_2)]. \quad (3.7)$$

As propriedades estudadas a seguir mostram a estreita relação entre as duas medidas.

3.4 Propriedades

Esta seção estuda diversas propriedades da correntropia complexa bem como suas respectivas provas matemáticas. Como essa medida foi desenvolvida de forma a herdar características da correntropia, algumas propriedades como simetria, informação de alta ordem estatística, interpretação probabilística, estimador consistente e não enviesado e perspectiva do ponto de vista do espaço de Hilbert (propriedades 1, 3, 4, 5 e 8, respectivamente) são semelhantes com as propriedades da correntropia convencional. Outras propriedades tiveram que ser adaptadas, como a propriedade 2, que define os limites para o *kernel* gaussiano e propriedade 9 que define uma métrica induzida a partir da correntropia complexa. Além disso, três novas propriedades foram estudadas. Uma delas conecta numericamente a correntropia complexa com a correntropia convencional (propriedade 6), enquanto a segunda mostra uma forma alternativa de representação polar para a medida, derivada da natureza própria dos números complexos (propriedade 7).

Propriedade 1

Simetria: Para *kernels* simétricos, a correntropia complexa também é simétrica.

$$V_{\sigma}^c(C_1, C_2) = E_{C_1 C_2}[k_{\sigma}(C_1, C_2)] = E_{C_1 C_2}[k_{\sigma}(C_2, C_1)] = V_{\sigma}^c(C_2, C_1) \quad (3.8)$$

Prova: Essa propriedade segue os conceitos de *kernels* definidos positivo e simétricos. ■

Propriedade 2

Resposta limitada e positiva: Para o *kernel* gaussiano, a correntropia complexa é sempre positiva e limitada. A estimação sempre vale entre zero (mínimo) é $1/2\pi\sigma^2$, valor máximo encontrado apenas quando $C_1 = C_2$.

$$0 \leq V_{\sigma}^c(C_1, C_2) \leq \frac{1}{2\pi\sigma^2} \quad (3.9)$$

Prova: Aplicando o *kernel* gaussiano, G_{σ}^c , definido em 3.10, na definição 3.7, obtém-se 3.11:

$$G_{\sigma}^c(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (3.10)$$

$$V_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \quad (3.11)$$

Para $C_1 = C_2$, $x = y$ e $z = s$, $\forall n$, obtemos $V^c(C_1, C_2) = 1/2\pi\sigma^2$. Caso contrário, quando C_1 é muito diferente de C_2 , o termo negativo da exponencial em (3.11) faz com que o valor de $V^c(C_1, C_2)$ tenda a zero, o que valida essa propriedade. ■

Propriedade 3

A correntropia complexa fornece informações estatísticas de alta ordem: Usando o *kernel* gaussiano, a correntropia complexa é a soma ponderada de todos os infinitos momentos de ordem par da variável aleatória $(C_1 - C_2)$. Além disso, aumentar o tamanho do *kernel* faz com que a correntropia se aproxime do valor da correlação entre C_1 e C_2 .

Prova: Lembrando que $C_1 = X + jZ$ e $C_2 = Y + jS$, pode-se analisar a Equação (3.11) de acordo com a expansão em série de Taylor:

$$\begin{aligned}
 V_{\sigma}^c(C_1, C_2) = & \frac{1}{2\pi\sigma^2} E \left[1 - \frac{(X-Y)^2}{2\sigma^2} + \frac{(X-Y)^4}{8\sigma^4} - \frac{(X-Y)^6}{48\sigma^6} + \frac{(X-Y)^8}{384\sigma^8} + \right. \\
 & - \frac{(Z-S)^2}{2\sigma^2} + \frac{(Z-S)^4}{8\sigma^4} - \frac{(Z-S)^6}{48\sigma^6} + \frac{(Z-S)^8}{384\sigma^8} + \frac{(X-Y)^2(Z-S)^2}{4\sigma^4} + \\
 & - \frac{(X-Y)^2(Z-S)^4}{16\sigma^6} - \frac{(X-Y)^4(Z-S)^2}{16\sigma^6} + \frac{(X-Y)^2(Z-S)^6}{96\sigma^8} \\
 & \left. + \frac{(X-Y)^4(Z-S)^4}{64\sigma^8} + \frac{(X-Y)^6(Z-S)^2}{96\sigma^8} + \dots \right] \quad (3.12)
 \end{aligned}$$

Agrupando os termos que contém σ^2 no denominador e definindo h_{σ} como uma variável contendo todos os termos de alta ordem do somatório, é possível escrever:

$$V_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} - \frac{1}{2\pi\sigma^2} E \left[\frac{(X-Y)^2 + (Z-S)^2}{2\sigma^2} \right] + h_{\sigma} \quad (3.13)$$

$$V_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} - \frac{1}{4\pi\sigma^4} E[(C_1 - C_2)(C_1 - C_2)^*] + h_{\sigma} \quad (3.14)$$

$$V_{\sigma}^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} - \frac{1}{4\pi\sigma^4} R[C_1, C_2] + h'_{\sigma} \quad (3.15)$$

onde $R[C_1, C_2] = E[C_1 C_2^*]$ é a correlação entre C_1 e C_2 .

Nota-se na Equação (F.19) que h'_{σ} , variável que envolve os termos de alta ordem, tende a zero mais rápido que o segundo termo da equação com o aumento do valor de σ , o que corresponde exatamente a correlação envolvendo duas variáveis complexas C_1 e C_2 . Sendo assim, a medida que o tamanho do *kernel* vai crescendo, a correntropia complexa tende a se aproximar da correlação analogamente a correntropia convencional para o caso com dados reais. ■

Propriedade 4:

Interpretação probabilística: Seja σ o tamanho do *kernel* gaussiano, fazendo σ tender a zero faz com que a correntropia complexa se aproxime do valor da densidade de probabilidade associado ao evento $(C_1 = C_2)$.

$$\lim_{\sigma \rightarrow 0} V^c(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 = P(C_1 = C_2) \quad (3.16)$$

onde $f_{XYZS}(u_1, u_1, u_2, u_2)$ é a função densidade de probabilidade (*Probability Density Function* - PDF) conjunta entre X, Y, Z, S .

Prova:

Deve-se usar a definição apresentada na Equação (3.7): $V^c(C_1, C_2) = E_{C_1 C_2}[K(C_1, C_2)]$. Dois número complexo são iguais quando tanto a parte real quanto a parte imaginária também são. Sendo assim, usando o *kernel* gaussiano e expandindo os termos associados a C_1 e C_2 , pode-se obter:

$$V^c(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(x, y, z, s) G_{\sigma}(x-y) G_{\sigma}(z-s) dx dy dz ds \quad (3.17)$$

Da teoria de distribuições, a seguinte igualdade pode ser demonstrada:

$$\lim_{\sigma \rightarrow 0} G_{\sigma}(x) \equiv \delta(x) \quad (3.18)$$

O que implica em:

$$\lim_{\sigma \rightarrow 0} V^c(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(x, y, z, s) \delta(x-y) \delta(z-s) dx dy dz ds \quad (3.19)$$

Fazendo $x = y = u_1$ e $z = s = u_2$, obtém-se:

$$\lim_{\sigma \rightarrow 0} V^c(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 = P(C_1 = C_2) \quad (3.20)$$

■

A correntropia complexa foi definida para herdar o sentido probabilístico da correntropia definida para os reais. A Equação (3.20) representa uma transformação de variável da PDF conjunta f_{XYZS} , sendo avaliada com os argumentos u_1 e u_2 . Usando o método de Parzen [Parzen 1962] para estimar f_{XYZS} , quando σ tender a zero e o produto $N\sigma$ tender a infinito, a estimativa da PDF conjunta, $\hat{f}_{XYZS}(x, y, z, s)$, tende a $f_{XYZS}(x, y, z, s)$ assintoticamente no sentido do erro médio quadrado. Sendo assim, o *kernel* gaussiano quando utilizado com um tamanho de *kernel* pequeno no estimador de Parzen de f_{xyzs} faz com que a Equação (3.20) forneça um escalar que é justamente o valor da probabilidade do evento $(C_1 = C_2)$.

Propriedade 4.1

Correntropia complexa como valor da integral no plano $x = y$ e $z = s$: Assumindo que dados independentes e identicamente distribuídos $\{(x_i, y_i, z_i, s_i)_{i=1}^N\}$ são amostrados da distribuição conjunta f_{xyzs} , enquanto $\hat{f}_{\sigma_{xyzs}}$ é a estimação de Parzen com tamanho de *kernel* σ , a correntropia complexa estimada com o tamanho de *kernel* $\sigma' = \sigma\sqrt{2}$ corresponde a integral de $\hat{f}_{\sigma_{xyzs}}$ sobre o plano formado pela condição $x = y$ e $z = s$.

$$\hat{V}_{\sigma'}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{\sigma_{XYZS}}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} \quad (3.21)$$

Prova: Devido ao tamanho da prova, ela foi movida para o Apêndice C.

Propriedade 5

Estimador consistente e não enviesado Sob as condições $N \rightarrow \infty$ e $\hat{V}_{N,\sigma}(X, Y)$, a estimação da correntropia $v_\sigma(X, Y)$ é consistente em erro médio quadrado. Além disso, sob as condições $N\sigma \rightarrow \infty$ e $\sigma \rightarrow 0$, $\hat{V}_{N,\sigma}(X, Y)$ o estimador $\hat{V}_\sigma(X, Y)$ é assintoticamente não enviesado para P_E e consistente em erro médio quadrático.

Prova: Para provar essa propriedade é preciso seguir as condições propostas pelo estimador de Parzen expostas em [Parzen 1962].

$$E[\hat{V}_{N,\sigma}^c(X, Y)] = V_\sigma^c(X, Y) \quad (3.22)$$

Usando o *kernel* gaussiano e fazendo $N \rightarrow \infty$ e $\sigma \rightarrow 0$:

$$\lim_{N \rightarrow \infty, \sigma \rightarrow 0} E[\hat{V}_{N,\sigma}^c(X, Y)] = p_E(0) \quad (3.23)$$

$$\text{var}[\hat{V}_{N,\sigma}^c(X, Y)] = N^{-1} \text{var}[G_\sigma(E)] \quad (3.24)$$

$$\lim_{N \rightarrow \infty, \sigma \rightarrow 0} E[\hat{V}_{N,\sigma}^c(X, Y)] = p_E(0) \int G_1^2(u) du \quad (3.25)$$

onde $G_1(u)$ é o *kernel* Gaussiano com tamanho de *kernel* $\sigma = 1$ e $\text{var}[\cdot]$ é o operador de variância.

■

Propriedade 6

Relação entre a correntropia convencional e a correntropia complexa: Para duas variáveis aleatórias R_1 e $R_2 \in \mathbb{R}$, usando o *kernel* gaussiano com tamanho de *kernel* σ para ambas as funções, para obter o mesmo valor computado pela correntropia convencional, é necessário multiplicar a correntropia complexa pela constante $(\sqrt{2\pi\sigma})$.

$$V_\sigma^c(R_1, R_2) \sqrt{2\pi\sigma} = V_\sigma(R_1, R_2) \quad (3.26)$$

Prova: Expandindo o lado esquerdo da Equação (3.26), tem-se:

$$V_\sigma^c(R_1, R_2) \sqrt{2\pi\sigma} = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \sqrt{2\pi\sigma} \quad (3.27)$$

Se R_1 e R_2 não possuem parte imaginária, pode-se escrever a Equação (3.27) como:

$$V_{\sigma}^c(R_1, R_2) \sqrt{2\pi\sigma} = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2}{2\sigma^2}\right) \quad (3.28)$$

Que é exatamente a expressão para se calcular $V_{\sigma}(C_1, C_2)$.

■

É importante mencionar que, como consequência da propriedade 6, é possível obter $V_{\sigma}^c(R_1, R_2) = V_{\sigma}(R_1, R_2)$ apenas com o tamanho de *kernel* $\sigma = \frac{1}{\sqrt{2\pi}}$.

Propriedade 7

Expressando a correntropia complexa em coordenadas polares: Usando o *kernel* gaussiano, a correntropia complexa por ser expressa em coordenadas polares da seguinte forma:

$$V^c(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{|C_1|^2 + |C_2|^2}{2\sigma^2} + \frac{|C_1||C_2|\cos(\theta - \phi)}{\sigma^2}\right) \quad (3.29)$$

Prova: Já que C_1 e C_2 podem ser escritos na forma polar como $C_1 = |C_1|\cos(\phi) + j|C_1|\sin(\phi)$, e $C_2 = |C_2|\cos(\theta) + j|C_2|\sin(\theta)$, a Equação (3.11) pode ser reescrita como:

$$\begin{aligned} (x - y)^2 &= (|C_1|\cos(\theta) - |C_2|\cos(\phi))^2 = \\ &= |C_1|^2\cos^2(\theta) - 2|C_1||C_2|\cos(\theta)\cos(\phi) + |C_2|^2\cos^2(\phi)^2 \end{aligned} \quad (3.30)$$

Expandindo o termo $(z - s)^2$, tem-se:

$$(z - s)^2 = (|C_1|\sin(\theta) - |C_2|\sin(\phi))^2 = |C_1|^2\sin^2(\theta) - 2|C_1||C_2|\sin(\theta)\sin(\phi) + |C_2|^2\sin^2(\phi)^2 \quad (3.31)$$

Fazendo $(x - y)^2 + (z - s)^2$ a partir da soma das equações (3.30) e (3.31), tem-se:

$$|C_1|^2 + |C_2|^2 - 2|C_1||C_2|(\cos(\theta)\cos(\phi) + \sin(\theta)\sin(\phi)) = |C_1|^2 + |C_2|^2 - 2|C_1||C_2|\cos(\theta - \phi) \quad (3.32)$$

■

Propriedade 8

Correntropia Complexa na Perspectiva do Espaço de Hilbert Da forma como foi definida na Equação (3.7), a correntropia complexa, como valor esperado de um *kernel*

positivo-definido, sugere uma interpretação na perspectiva do espaço de Hilbert. Isso porque, como Aronszajn (1950) mostrou em seu teorema, todo *kernel* positivo definido mapeia os dados originais (espaço de entrada) num RKHS de alta dimensionalidade. Entretanto, não existe limitação para o uso de variáveis complexas no espaço de entrada [Vetterli et al. 2014]. Dessa forma, sejam Q e W variáveis aleatórias complexas:

$$V_{\sigma}^c(Q, W) = E[k_{\sigma}(Q, W)] = E[\langle \Phi(Q), \Phi(W) \rangle] = E[\Phi(W)^H \Phi(Q)]. \quad (3.33)$$

onde $\Phi(\cdot)$ é uma transformação não linear que mapeia os dados do espaço de entrada para um espaço de características.

Dessa forma, a correntropia complexa pode ser interpretada, assim como no caso real, como um produto interno num RKHS. ■

Propriedade 9

Métrica Induzida da Correntropia Complexa: Assim como a correntropia induz uma métrica denominada CIM [Seth e Principe 2008], pode-se definir a métrica induzida da correntropia complexa (*Complex Correntropy Induced Metric - CCIM*), ψ^c , como:

$$\begin{aligned} \psi_{\sigma}^c(Q, W) &= \sqrt{V_{\sigma}^c(0, 0) - V_{\sigma}^c(Q, W)} \\ &= \sqrt{\frac{G_{\sigma}^c(0)}{L} \sum_{i=1}^L \left\{ 1 - \exp\left(-\frac{(q_i - w_i)(q_i - w_i)^*}{2\sigma^2}\right) \right\}}. \end{aligned} \quad (3.34)$$

onde Q e W são duas variáveis aleatórias complexas e $\mathbf{q} = (q_1, q_2, \dots, q_L)$ e $\mathbf{w} = (w_1, w_2, \dots, w_L)$ são vetores contendo L amostras dessas variáveis aleatórias. Algumas propriedades devem ser atendidas para que uma métrica seja considerada válida. São elas a não negatividade, identidade, simetria e desigualdade triangular. A prova de cada uma dessas propriedades é mostrada a seguir:

Simetria e não negatividade

A prova dessas propriedades acompanham a definição da correntropia complexa que especifica o uso de um *kernel* simétrico e positivo definido. Assim, sendo $V_{\sigma}^c(Q, W) = V_{\sigma}^c(W, Q)$ assegurado pela propriedade 1 então $\psi_{\sigma}^c(Q, W) = \psi_{\sigma}^c(W, Q)$. Além disso, pro caso do uso do *kernel* Gaussiano, pode-se afirmar que $0 \leq V_{\sigma}^c(Q, W) \leq 1/(2\pi\sigma^2)$. Portanto, a CCIM é simétrica e não negativa.

Identidade

Usando o *kernel* Gaussiano, se $Q = W$, então $V_{\sigma}^c(Q, W) = G_{\sigma}^c(0) = 1/(2\pi\sigma^2)$, que implica em $\psi_{\sigma}^c(Q, W) = 0$ para $Q = W$.

Desigualdade triangular

Estende-se aqui a prova do caso real presente em [Principe 2010]. Deseja-se mostrar que

$$\Psi_{\sigma}^c(Q, T) \leq \Psi_{\sigma}^c(Q, W) + \Psi_{\sigma}^c(W, T). \quad (3.35)$$

Pode-se escrever $\tilde{Q} = [\Phi(q_1), \Phi(q_2), \dots, \Phi(q_L)]$ e $\tilde{W} = [\Phi(w_1), \Phi(w_2), \dots, \Phi(w_L)]$ no espaço de Hilbert onde $\Phi(\cdot)$ é um mapeamento não linear do espaço de entrada para um RKHS [Principe 2010]. Dessa forma, a distância Euclidiana $ED(\tilde{Q}, \tilde{W})$ vale

$$\begin{aligned} ED(\tilde{Q}, \tilde{W}) &= \sqrt{\langle \tilde{Q} - \tilde{W}, \tilde{Q} - \tilde{W} \rangle} \\ &= \sqrt{\langle \tilde{Q}, \tilde{Q} \rangle - 2\langle \tilde{Q}, \tilde{W} \rangle + \langle \tilde{W}, \tilde{W} \rangle} \end{aligned} \quad (3.36)$$

ou

$$ED(\tilde{Q}, \tilde{W}) = \sqrt{2L(G_{\sigma}^c(0) - V_{\sigma}^c(Q, W))} = \sqrt{2L}\Psi^c(Q, W). \quad (3.37)$$

Então,

$$\Psi_{\sigma}^c(Q, T) = \frac{ED(\tilde{Q}, \tilde{T})}{\sqrt{2L}} \leq \frac{ED(\tilde{Q}, \tilde{W})}{\sqrt{2L}} + \frac{ED(\tilde{W}, \tilde{T})}{\sqrt{2L}} \quad (3.38)$$

ou

$$\Psi_{\sigma}^c(Q, T) \leq \Psi_{\sigma}^c(Q, W) + \Psi_{\sigma}^c(W, T), \quad (3.39)$$

que completa a prova. ■

Propriedade 9.1

O quadrado da CCIM aproxima a norma ℓ_0 quando $\sigma \rightarrow 0$ num problema de minimização: Seja \mathbf{z} um vetor complexo de tamanho L , $\mathbf{z} \in \mathbb{C}^L$

$$\min \|\mathbf{z}\|_0 \approx \min [\Psi_{\sigma \rightarrow 0}^c(\mathbf{z}, 0)]^2 = \frac{G_{\sigma}^c(0)}{L} \sum_{i=1}^L \left\{ 1 - \exp\left(-\frac{z_i z_i^*}{2\sigma^2}\right) \right\}. \quad (3.40)$$

Prova: Considerando que \mathbf{t} é o vetor obtido pela minimização da norma ℓ_0 , e \mathbf{r} quando o quadrado da CCIM é minimizado, pode-se afirmar que

$$\Psi_{\sigma}^c(\mathbf{R}, 0) \leq \Psi_{\sigma}^c(\mathbf{T}, 0) \quad (3.41)$$

e usando (3.34), pode-se escrever (3.41) como

$$\sum_{i=1}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \geq \sum_{i=1}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \quad (3.42)$$

O somatório em (3.42) pode ser reescrito como

$$\sum_{i=1}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1, \\ r_i r_i^* = 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) + \sum_{\substack{i=1, \\ r_i r_i^* \neq 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \quad (3.43)$$

e

$$\sum_{i=1}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1, \\ t_i t_i^* = 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) + \sum_{\substack{i=1, \\ t_i t_i^* \neq 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \quad (3.44)$$

Entretanto, os somatórios das equações (3.43) e (3.44) relacionados a $r_i r_i^* = 0$ e $t_i t_i^* = 0$, podem ser reescritos como

$$\sum_{\substack{i=1, \\ r_i r_i^* = 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1, \\ r_i r_i^* = 0}}^L \exp(0) = L - \|\mathbf{r}\|_0 \quad (3.45)$$

e

$$\sum_{\substack{i=1, \\ t_i t_i^* = 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = \sum_{\substack{i=1, \\ t_i t_i^* = 0}}^L \exp(0) = L - \|\mathbf{t}\|_0. \quad (3.46)$$

Substituindo (3.45) e (3.46) em (3.43) e (3.44), respectivamente, pode-se escrever

$$\sum_{i=1}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) = L - \|\mathbf{r}\|_0 + \sum_{\substack{i=1, \\ r_i r_i^* \neq 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \quad (3.47)$$

e

$$\sum_{i=1}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right) = L - \|\mathbf{t}\|_0 + \sum_{\substack{i=1, \\ t_i t_i^* \neq 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right). \quad (3.48)$$

Usando (3.47) e (3.48), é possível dizer que (3.42)

$$\|\mathbf{r}\|_0 - \sum_{\substack{i=1, \\ r_i r_i^* \neq 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) \leq \|\mathbf{t}\|_0 - \sum_{\substack{i=1, \\ t_i t_i^* \neq 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right)$$

ou

$$\|\mathbf{r}\|_0 - \|\mathbf{t}\|_0 \leq \sum_{\substack{i=1, \\ r_i r_i^* \neq 0}}^L \exp\left(-\frac{r_i r_i^*}{2\sigma^2}\right) - \sum_{\substack{i=1, \\ t_i t_i^* \neq 0}}^L \exp\left(-\frac{t_i t_i^*}{2\sigma^2}\right), \quad (3.49)$$

e usando do fato de que $\|\mathbf{r}\|_0 \geq \|\mathbf{t}\|_0$, pode-se escolher um valor para o tamanho de *kernel*

σ pequeno o bastante a fim de tornar o lado direito de (3.49) pequeno o bastante para se aproximar arbitrariamente de 0^+ . Assim,

$$\|\mathbf{r}\|_0 \approx \|\mathbf{t}\|_0,$$

o que completa a prova. ■

As propriedades estudadas até então são quase todas referentes ao *kernel* gaussiano que, como a propriedade 2 ressaltou, resultam num valor real para a correntropia complexa. Entretanto, a correntropia complexa foi definida para qualquer *kernel* positivo definido e, para algumas aplicações, uma função que tenha uma resposta complexa pode ser necessária. O anexo F mostra um exemplo de *kernel* positivo definido que provê uma resposta complexa a correntropia complexa.

3.5 Critério de Máxima Correntropia Complexa

Visto que a correntropia foi definida para o domínio dos complexos é necessário atualizar o seu uso enquanto função custo para viabilizar a sua aplicabilidade em problemas de otimização. Esse capítulo apresenta o critério de máxima correntropia complexa (*Maximum Complex Correntropy Criterion* - MCCC). O cálculo de Wirtinger, que foi apresentado na seção anterior, é utilizado para derivar dois algoritmos sendo um deles baseado em um gradiente estocástico e um segundo em ponto-fixa.

3.5.1 Correntropia complexa como função custo

Seja \mathbf{d} e \mathbf{y} vetores contendo N amostras das variáveis aleatórias D e Y , respectivamente, considerando um modelo linear e definindo o erro, $\mathbf{e} = \mathbf{d} - \mathbf{y}$, como a diferença entre o sinal desejado, \mathbf{d} , e a saída do filtro $\mathbf{y} = \mathbf{w}^H \mathbf{X}$, e lembrando que $\mathbf{e}, \mathbf{d}, \mathbf{y} \in \mathbb{C}^{1 \times N}$, $\mathbf{w} \in \mathbb{C}^{1 \times M}$, onde M é o número de pesos, e $\mathbf{X} \in \mathbb{C}^{M \times N}$, podemos definir uma nova função custo J_{MCCC} como a máxima correntropia complexa entre D e Y :

$$J_{MCCC} = V_{\sigma}^C(D, Y). \quad (3.50)$$

Devido as propriedades apresentadas na seção anterior, escolheu-se o *kernel* gaussiano G_{σ}^C , assim:

$$\begin{aligned} J_{MCCC} &= E[G_{\sigma}(\mathbf{d} - \mathbf{w}^H \mathbf{X})] = E[G_{\sigma}^C(\mathbf{e})] \\ &= \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(d_i - y_i)(d_i - y_i)^*}{2\sigma^2}\right) \\ &= \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right), \end{aligned} \quad (3.51)$$

onde \mathbf{w} é o parâmetro que controla o erro entre o sinal estimado e o desejado.

A saída natural para usar o MCCC em problemas de otimização seria, como no caso real, derivar a equação 3.50 em relação a \mathbf{w} . A equação (3.50) depende de parâmetros complexos, (\mathbf{d}, \mathbf{y}) , mas, a partir do uso do *kernel* gaussiano, terá sempre uma resposta real, como ressalta a propriedade 2 apresentada na seção anterior. Sendo assim, tal como no exemplo 2 do capítulo de fundamentação teórica, funções com resposta real e argumento complexo violam as condições impostas pelas equações de Cauchy-Riemman. Assim, a equação (3.50) não é analítica no domínio dos complexos e portanto, a diferenciação padrão não pode ser aplicada. Uma possível saída a esse problema é considerar a função custo definida no domínio euclidiano com dimensionalidade dobrada (\mathbb{R}^2), entretanto, essa abordagem traz consigo onerosos cálculos pela quantidade de variáveis envolvidas, o que dificulta as deduções [Bouboulis e Theodoridis 2011]. O cálculo de Wirtinger, apresentado brevemente no capítulo anterior, provê de uma forma elegante a obtenção de gradientes de funções com resultado real mas que são definidas no domínio complexo [Mandic e Goh 2009, Bouboulis e Theodoridis 2011].

3.5.2 Solução usando o Gradiente Ascendente

Uma solução usando o gradiente ascendente pode ser escrita como

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \nabla J_{MCCC} \quad (3.52)$$

Primeiramente, vamos expandindo o argumento da exponencial, $e_i e_i^*$, da Equação (3.51):

$$\begin{aligned} e_i e_i^* &= (d_i - y_i)(d_i - y_i)^* \\ e_i e_i^* &= (d_i - y_i)(d_i^* - y_i^*) \\ e_i e_i^* &= (d_i d_i^* - d_i y_i^* - y_i d_i^* + y_i y_i^*) \end{aligned} \quad (3.53)$$

Sabendo que $y_i = \mathbf{w}^H \mathbf{x}_i$, onde \mathbf{x}_i é o i -ésima coluna da matriz de entrada \mathbf{X} , temos que

$$\begin{aligned} y_i &= \mathbf{w}^H \mathbf{x}_i \\ y_i^* &= (\mathbf{w}^H \mathbf{x}_i)^* \\ y_i^* &= (\mathbf{w}^H)^* (\mathbf{x}_i)^* \end{aligned} \quad (3.54)$$

Mas, $\mathbf{w}^H = (\mathbf{w}^*)^T = (\mathbf{w}^T)^*$, então:

$$\begin{aligned} y_i^* &= (\mathbf{w}^H)^* (\mathbf{x}_i)^* \\ y_i^* &= ((\mathbf{w}^T)^*)^* (\mathbf{x}_i)^* \\ y_i^* &= \mathbf{w}^T \mathbf{x}_i^* \end{aligned} \quad (3.55)$$

Assim, reescrevendo 3.53, temos:

$$\begin{aligned}
 e_i e_i^* &= (d_i d_i^* - d_i y_i^* - y_i d_i^* + y_i y_i^*) \\
 e_i e_i^* &= (d_i d_i^*) - (d_i \mathbf{w}^T \mathbf{x}_i^*) - (\mathbf{w}^H \mathbf{x}_i d_i^*) + (\mathbf{w}^H \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i^*)
 \end{aligned} \tag{3.56}$$

Usando o cálculo de Wirtinger para obter o gradiente de J_{MCCC} , temos:

$$\nabla J_{MCCC} = \frac{\partial J_{MCCC}}{\partial \mathbf{w}^*} = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(d_i - y_i)(d_i - y_i)^*}{2\sigma^2}\right) \frac{(-1)}{2\sigma^2} \frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} \tag{3.57}$$

Fazendo agora a derivada do argumento da exponencial, temos:

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = \frac{\partial(d_i d_i^* - d_i y_i^* - y_i d_i^* + y_i y_i^*)}{\partial \mathbf{w}^*} \tag{3.58}$$

Que pode ser reescrito como:

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = \frac{\partial(d_i d_i^*)}{\partial \mathbf{w}^*} - \frac{\partial(d_i \mathbf{w}^T \mathbf{x}_i^*)}{\partial \mathbf{w}^*} - \frac{\partial(\mathbf{w}^H \mathbf{x}_i d_i^*)}{\partial \mathbf{w}^*} + \frac{\partial(\mathbf{w}^H \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i^*)}{\partial \mathbf{w}^*} \tag{3.59}$$

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = \frac{\partial(d_i d_i^*)}{\partial \mathbf{w}^*} - \frac{\partial(d_i \mathbf{w}^T \mathbf{x}_i^*)}{\partial \mathbf{w}^*} - \frac{\partial((\mathbf{w}^*)^T \mathbf{x}_i d_i^*)}{\partial \mathbf{w}^*} + \frac{\partial((\mathbf{w}^*)^T \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i^*)}{\partial \mathbf{w}^*} \tag{3.60}$$

Que resulta em

$$\begin{aligned}
 \frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} &= 0 - 0 - \mathbf{x}_i d_i^* + \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i^* \\
 &= -\mathbf{x}_i d_i^* + \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i^* \\
 &= -(\mathbf{d}_i^* - \mathbf{w}^T \mathbf{x}_i^*) \mathbf{x}_i \\
 &= -(\mathbf{d}_i^* - (\mathbf{w}^T \mathbf{x}_i^*)^*) \mathbf{x}_i \\
 &= -(\mathbf{d}_i^* - (\mathbf{w}^H \mathbf{x}_i)^*) \mathbf{x}_i \\
 &= -(\mathbf{d}_i - \mathbf{w}^H \mathbf{x}_i)^* \mathbf{x}_i \\
 \frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} &= -(\mathbf{d}_i - \mathbf{w}^H \mathbf{x}_i)^* \mathbf{x}_i = -(\mathbf{d}_i - y_i)^* \mathbf{x}_i = -e_i^* \mathbf{x}_i
 \end{aligned} \tag{3.61}$$

Sempre usando a notação de denominador na derivada de um escalar por um vetor.

Assim, podemos escrever 3.57 como:

$$\nabla J_n = \frac{\partial J_n}{\partial \mathbf{w}^*} = +\frac{1}{4\pi\sigma^4} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) (e_i^* \mathbf{x}_i) \tag{3.62}$$

O que resulta numa regra de atualização:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{N4\pi\sigma^3} \sum_{i=1}^N \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i^* \mathbf{x}_i \quad (3.63)$$

onde $e_i = d_i - \mathbf{w}^H(n)\mathbf{x}_i$. Usando o gradiente estocástico, tem-se:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{N4\pi\sigma^3} \exp\left(-\frac{\mathbf{e}(n)\mathbf{e}(n)^*}{2\sigma^2}\right) \mathbf{e}(n)^* \mathbf{x}(n) \quad (3.64)$$

Análise de Convergência

Devido ao seu tamanho, a análise de convergência foi movida para o anexo D.

Algoritmo

O algoritmo 1 sintetiza o procedimento para utilizar o gradiente estocástico com MCCC.

Algorithm 1 Gradiente com MCCC

Inicialização:

Escolher passo do gradiente η , largura de *kernel* σ , vetor de pesos inicial \mathbf{w} , sinal desejado $\mathbf{d}^{N \times 1}$ e entrada \mathbf{X}

Computação:

```

1: procedure GRADIENTE ESTOCÁSTICO–MCC:
2:   % N é a dimensão do vetor desejado d
3:   % Inicializando o contador de iterações
4:   i=1
5:   for (i ≤ N) do
6:     % Extraíndo a i-ésima coluna da Matriz X
7:      $\mathbf{x}(i) = \mathbf{X}(\cdot, i)$ 
8:     % Computando o erro na iteração i
9:      $\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{w}^H \mathbf{x}(i)$ 
10:     $\mathbf{w} = \mathbf{w} + \frac{\mu}{N4\pi\sigma^3} \exp\left(-\frac{\mathbf{e}(i)\mathbf{e}(i)^*}{2\sigma^2}\right) \mathbf{e}(i)^* \mathbf{x}(i)$ 
11:
12:   end for
13: end procedure

```

3.5.3 Solução de Ponto fixo

Para encontrar uma solução de ponto fixo para os pesos ótimos \mathbf{w} , basta derivar a função custo J_{MCCC} em relação a \mathbf{w}^* e igualar a zero usando o cálculo de Wirtinger.

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}^*} = 0 \quad (3.65)$$

Usa-se aqui um pequeno truque de derivar em relação a \mathbf{w}^* para obter a expressão de \mathbf{w} de forma direta. Um outro caminho seria derivar em função de \mathbf{w} , como se faz no caso real, por exemplo. O problema dessa abordagem é que depois de derivar obtêm-se apenas \mathbf{w}^* . O \mathbf{w} acaba sumindo da expressão, fazendo com que seja necessário aplicar o operador $*$ ao final do processo de derivação para, finalmente, encontrar o \mathbf{w} otimizado:

$$\begin{aligned}\frac{\partial J_{MCCC}}{\partial \mathbf{w}^*} &= \frac{\partial E[G_{\sigma}^c(\mathbf{e})]}{\partial \mathbf{w}^*} = E\left[G_{\sigma}^c(\mathbf{e})\frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*}\right] = \mathbf{0} \\ \frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*} &= \frac{\partial[(\mathbf{d} - \mathbf{w}^H \mathbf{X})(\mathbf{d} - \mathbf{w}^H \mathbf{X})^*]}{\partial \mathbf{w}^*} \\ \frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*} &= \frac{\partial[(\mathbf{d} - \mathbf{w}^H \mathbf{X})(\mathbf{d}^* - \mathbf{w}^T \mathbf{X}^*)]}{\partial \mathbf{w}^*}\end{aligned}$$

Usa-se então o cálculo de Wirtinger para obter a derivada de $\mathbf{e}\mathbf{e}^*$ em relação a \mathbf{w}^* :

$$\begin{aligned}\frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*} &= \frac{\partial(\mathbf{d}\mathbf{d}^* - \mathbf{d}\mathbf{w}^T \mathbf{X}^* - \mathbf{w}^H \mathbf{X}\mathbf{d}^* + \mathbf{w}^H \mathbf{X}\mathbf{w}^T \mathbf{X}^*)}{\partial \mathbf{w}^*} \\ \frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*} &= \frac{\partial(\mathbf{d}\mathbf{d}^* - \mathbf{d}\mathbf{X}^H \mathbf{w} - \mathbf{w}^H \mathbf{X}\mathbf{d}^* + \mathbf{w}^H \mathbf{X}\mathbf{X}^H \mathbf{w})}{\partial \mathbf{w}^*} \\ \frac{\partial \mathbf{d}\mathbf{d}^*}{\partial \mathbf{w}^*} &= 0 \\ \frac{\partial \mathbf{d}\mathbf{X}^H \mathbf{w}}{\partial \mathbf{w}^*} &= 0 \\ \frac{\partial \mathbf{w}^H \mathbf{X}\mathbf{d}^*}{\partial \mathbf{w}^*} &= \frac{\partial \mathbf{X}^T \mathbf{w}^* \mathbf{d}^*}{\partial \mathbf{w}^*} = \mathbf{X}\mathbf{d}^* \\ \frac{\partial \mathbf{w}^H \mathbf{X}\mathbf{X}^H \mathbf{w}}{\partial \mathbf{w}^*} &= \frac{\partial \mathbf{w}^T \mathbf{X}^* \mathbf{X}^T \mathbf{w}^*}{\partial \mathbf{w}^*} = \mathbf{X}\mathbf{X}^H \mathbf{w}\end{aligned}$$

Sempre considerando o resultado como um vetor coluna. Assim, têm-se:

$$\frac{\partial(\mathbf{e}\mathbf{e}^*)}{\partial \mathbf{w}^*} = (\mathbf{X}\mathbf{d}^* - \mathbf{X}\mathbf{X}^H \mathbf{w})$$

$$E[G_{\sigma}^c(\mathbf{e})(\mathbf{X}\mathbf{d}^* - \mathbf{X}\mathbf{X}^H \mathbf{w})] = \mathbf{0}$$

$$E[G_{\sigma}^c(\mathbf{e})\mathbf{X}\mathbf{d}^*] = E[G_{\sigma}^c(\mathbf{e})\mathbf{X}\mathbf{X}^H] \mathbf{w} \quad (3.66)$$

Note que, embora a equação (3.66) tenha a mesma forma funcional da solução de Wiener, ela não é analítica devido a inclusão no erro do parâmetro \mathbf{w} na exponencial gaussiana em cada um dos lados da equação. Entretanto, uma solução de ponto fixo pode ser apresentada:

$$\mathbf{w} = \left[\sum_{n=1}^N G_{\sigma}^c(\mathbf{e}_n) \mathbf{X}_n \mathbf{X}_n^H \right]^{-1} \left[\sum_{n=1}^N G_{\sigma}^c(\mathbf{e}_n) \mathbf{d}_n^* \mathbf{X}_n \right] \quad (3.67)$$

A equação (3.67) é uma solução iterativa para obtenção do valor ótimo de \mathbf{w} . Mesmo que a convergência seja obtida após poucas iterações, existe o inconveniente de ter que computar todo o somatório apresentado até atingir o valor final. Isso torna essa forma apresentada um tanto quanto inadequada em problemas de aprendizagem em tempo real. Sendo assim, uma solução de ponto fixo estocástico pode ser derivada baseada no algoritmo apresentado a correntropia limitado aos números reais [Singh e Principe 2010] mas baseado na equação (3.67). Primeiramente, define-se uma matriz de autocorrelação ponderada do sinal de entrada complexo. Depois, um vetor de correlação cruzada ponderado entre o conjugado desejado e o vetor de entrada:

$$R_n = \sum_{n=1}^N G_{\sigma}^c(\mathbf{e}_n) \mathbf{X}_n \mathbf{X}_n^H \quad \text{e} \quad P_n = \sum_{n=1}^N G_{\sigma}(\mathbf{e}_n) \mathbf{d}_n^* \mathbf{X}_n \quad (3.68)$$

Como mostrado em [Singh e Principe 2010], para o caso real as equações (3.68) se assemelham à solução de Wiener mas, ao invés de usar a simples média, têm-se a ponderação da exponencial da Gaussiana do erro. Sendo assim, a matriz de autocorrelação e o vetor de correlação cruzada podem ser atualizados de forma recursiva, como nos clássicos algoritmos LMS ou RLS [Haykin 2002]. Assim:

$$R_n = R_{n-1} + G_{\sigma}^c(\mathbf{e}_n) \mathbf{X}_n \mathbf{X}_n^H \quad \text{e} \quad P_n = P_{n-1} + G_{\sigma}(\mathbf{e}_n) \mathbf{d}_n^* \mathbf{X}_n \quad (3.69)$$

Deve-se ressaltar que para implementar as expressões recursivas mostradas acima, é necessário definir os valores iniciais para \mathbf{w} , bem como para R_0 e P_0 . Outro ponto que merece destaque é a semelhança da solução de ponto-fixo com o clássico algoritmo RLS para números complexos, CRLS. Basicamente, a diferença entre eles é a ponderação apresentada pela exponencial da gaussiana. Como poderá ser visto na seção de resultados, para grandes tamanhos de *kernel*, a solução de ponto fixo para o MCCC se aproxima numericamente do CRLS.

O algoritmo 2 resume o procedimento de implementação dessa estratégia. O anexo E mostra a dedução desse mesmo algoritmo sem a utilização do cálculo de Wirtinger, chegando a mesma regra de atualização.

3.5.4 Análise dos algoritmos

É importante notar que os algoritmos 1 e 2 generalizam para o caso complexo as soluções baseadas no gradiente ascendente e o ponto fixo com MCC introduzidos em [Singh e Principe 2010]. Basta comparar as regras de atualização dos algoritmos definidos nessa seção com os algoritmos apresentados na seção 2.3.5. Vale salientar também que, fazendo σ grande o suficiente, conforme a Propriedade 3 ressalta, a correntropia complexa se aproxima da correlação entre \mathbf{d} e \mathbf{y} . Para essa condição, os algoritmos aqui apresentados, assim como no caso real, generalizam também os algoritmos *Complex Least Mean Square* (CLMS) [Mandic e Goh 2009], e o *Complex Recursive Least Squares* (CRLS)

Algorithm 2 Ponto fixo com MCCC*Inicialização:*largura de *kernel* σ , vetor de pesos inicial \mathbf{w} , sinal desejado $\mathbf{d}^{N \times 1}$ e entrada \mathbf{X} *Computação:*

```

1: procedure GRADIENTE ESTOCÁSTICO–MCC:
2:   % N é a dimensão do vetor desejado  $\mathbf{d}$ 
3:   % Inicializando matrizes que auxiliares
4:   % Inicializando o contador de iterações
5:   k=1
6:   for (i<N) do
7:     % Extraindo a  $i$ -ésima coluna da Matriz  $X$ 
8:      $\mathbf{x}(i) = \mathbf{X}(\cdot, i)$ 
9:     % Computando o erro na iteração  $i$ 
10:     $\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{w}^H \mathbf{x}(i)$ 
11:    % Variável auxiliar com cálculo da exponencial do erro
12:     $\beta = \exp\left(-\frac{\mathbf{e}(i)\mathbf{e}(i)^*}{2\sigma^2}\right)$ 
13:     $P = P + \beta d^* \mathbf{x}_i$ 
14:     $R = R + \beta \mathbf{x}_i \mathbf{x}_i^H$ 
15:     $\mathbf{w} = R^{-1} P$ 
16:   end for
17: end procedure

```

[Diniz 2008]. Isso é mostrado por meio de simulações no próximo capítulo.

3.6 Estratégias de *kernel* adaptativo

Como foi mostrado na seção das propriedades, assim como no caso real, a correntropia complexa tem um parâmetro livre denominado tamanho ou largura de *kernel*, σ . O seu valor deve ser escolhido pelo usuário e varia de acordo com a natureza dos dados. Dessa forma, escolher um valor ótimo para o tamanho de *kernel* não é uma tarefa trivial [Principe 2010].

Por causa disso, vários trabalhos propuseram métodos de determinar um valor ótimo para essa variável [Wang et al. 2015, Huang et al. 2017, Zhao et al. 2012, Wang et al. 2017]. Entretanto, todos os estudos relacionados são referentes a correntropia clássica aplica a dados reais. Sendo assim, essa seção visa atualizar alguns desses algoritmos para que possam ser utilizados também pela correntropia complexa.

3.6.1 MCCC com Largura de Kernel Adaptativa

A estratégia definida MCC com largura de *kernel* adaptativo (*Adaptive Kernel Width MCC*, AMCC) [Wang et al. 2017] consiste em selecionar um valor de *kernel* baseado na combinação da regra de Silverman [Silverman 1986] e o quadrado do erro previsto. Wang

et al. (2017) afirmam que isso implica numa convergência mais rápida, principalmente quando o vetor de pesos que se deseja encontrar está bastante longe do ideal. Esse método tem uma implementação simples que não aumenta o custo computacional e nem introduz parâmetros livres.

Para adaptar essa estratégia para dados de natureza complexa, podemos definir uma nova regra de atualização

$$\sigma(n)^2 = e(n)e(n)^* + \sigma^2, \quad (3.70)$$

onde σ é um tamanho de *kernel* predefinido (a regra de Silverman pode continuar a ser usada), e $e(n)$ é o erro na iteração n . Essa adaptação no A-MCC é denominada MCCC com Largura de Kernel Adaptativa (*Adaptive Kernel Width MCCC - AMCCC*).

3.6.2 MCCC com Chaveamento de Largura de Kernel

Com o objetivo de aumentar a velocidade de convergência do método, Wang et al. (2015) propôs o algoritmo de Chaveamento de Largura de Kernel do MCC (*Switch Kernel Width MCC - S-MCC*) onde ou um tamanho de *kernel* fixo ou a metade do quadrado do erro é usada como tamanho de *kernel*. O *kernel* escolhido é o maior entre esses dois valores. Dessa forma, para adaptar essa técnica ao MCC basta fazer:

$$\sigma^2(n) = \text{máximo} \left(\frac{e(n)e(n)^*}{2}, \sigma^2 \right) \quad (3.71)$$

Essa estratégia é denominada de MCCC com chaveamento de largura de *kernel* (*Switch Kernel Width MCCC - S-MCCC*). Esse é mais um exemplo de estratégia que não adiciona parâmetros livres ao algoritmo.

3.6.3 MCCC com Largura de Kernel Variável

O MCC com o tamanho de *kernel* variável (*Variable Kernel Width MCC - VKW*) foi proposto em [Huang et al. 2017] como uma estratégia de cálculo da largura de *kernel* a cada iteração através da maximização da exponencial negativa do erro ao quadrado $\exp(-e^2/2\sigma^2)$. Além disso, uma modificação na função custo é aplicada de modo a diminuir a influência da largura de *kernel* na parte externa da exponencial. Ao invés de maximizar $J_n = E[G_{\sigma(e)}]$, a proposta de Huang et al. (2017) é maximizar $J_k = E[\sigma^2 G_{\sigma}(e)]$. É possível usar da mesma estratégia para o caso complexo:

$$J_n = E[\sigma^4 G_{\sigma}(e)] = E[\sigma^4 G_{\sigma}(D - Y)] \quad (3.72)$$

O que resulta numa equação de atualização do gradiente estocástico:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{4\pi N} \exp\left(-\frac{\mathbf{e}(n)\mathbf{e}(n)^*}{2\sigma^2}\right) \mathbf{e}(n)^* \mathbf{x}(n). \quad (3.73)$$

A cada iteração, depois de calcular o erro $\mathbf{e}(n)$, o tamanho de *kernel* é atualizado a partir de uma nova otimização para fazê-lo caminhar na direção da minimização do erro. Assim o tamanho de *kernel* é escrito como uma função do erro:

$$\max_{\sigma_n} J(e_n) = \exp\left(-\frac{e_n e_n^*}{2\sigma_n(e)^2}\right). \quad (3.74)$$

Diferenciando a Equação (3.74) com relação a e_n usando o cálculo de Wirtinger, temos:

$$\frac{\partial J_n}{\partial e^*} = -\exp\left(-\frac{e_n e_n^*}{2\sigma_n(e)^2}\right) \frac{e_n^* \sigma(e)_n [\sigma_n(e) - 2e_n \sigma'_n(e)]}{2\sigma_n^4(e)} \quad (3.75)$$

Depois, igualando a Equação (3.75) a zero, temos:

$$\sigma_n = 2e_n \sigma'_n = 2k_\sigma |e_n| \quad (3.76)$$

Essa estratégia é denominada VKW-MCCC.

Capítulo 4

Aplicações

Este capítulo tem como objetivo exemplificar o uso da correntropia complexa em algumas aplicações. Inicialmente, é apresentada também uma breve explanação sobre as diferentes estratégias de simulação de ambientes de ruído não-gaussiano que são usados nas diferentes aplicações do capítulo além de métricas usadas para avaliação das respostas dos algoritmos. São apresentados resultados na área de identificação de sistemas lineares, equalização de canal e amostragem compressiva. Além disso, são estudados efeitos de estratégias de *kernel* adaptativo bem como a aplicação da correntropia complexa na solução de problemas envolvendo dados circulares.

4.1 Simulação de Ambientes com Ruído Não-Gaussiano

Com o objetivo de avaliar o desempenho dos algoritmos propostos em ambientes com ruído não-gaussiano são usadas duas estratégias nas simulações: Através de uma função densidade de probabilidade com duas modas (bimodal) e uma distribuição alfa estável. Ambas as estratégias são apresentadas a seguir.

Nesta tese são usados ruídos complexos. Sendo assim, seja $\eta \in \mathbb{C}$ a variável responsável por modelar o ruído alfa estável ou bimodal, gera-se a parte real ($\eta^{Re} \in \mathbb{R}$) e imaginária ($\eta^{Im} \in \mathbb{R}$) de forma independente, cada uma seguindo parâmetros próprios. Elas são combinadas *a posteriori*, ou seja, $\eta = \eta^{Re} + j\eta^{Im}$.

4.1.1 Ruído Bimodal

Uma das estratégias utilizadas para simular ruído impulsivo é somar duas distribuições normais. A notação $\mathcal{N}(\mu, \sigma)$ é utilizada para descrever uma distribuição normal com média μ e variância σ^2 . Para descrever o percentual de dados concentrado em cada moda escreve-se $\Xi\mathcal{N}(\mu, \sigma)$, onde Ξ é o percentual referente a moda descrita pela gaussiana em seguida. A Figura 4.1 mostra um típico histograma de um ruído bimodal com as seguintes características: Para a parte real usa-se $0.8\mathcal{N}(0, 1) + 0.2\mathcal{N}(13, 2)$, ou seja, trata-se de um ruído bimodal onde 80% dos dados estão concentrados na moda cuja média é zero e a variância é 1. Os outros 20% dos dados estão concentrados numa segunda moda que tem média 13 e desvio padrão 2. Seguindo a mesma lógica, a parte imaginária é caracterizada como $0.90\mathcal{N}(-15, 5) + 0.10\mathcal{N}(8, 1)$.

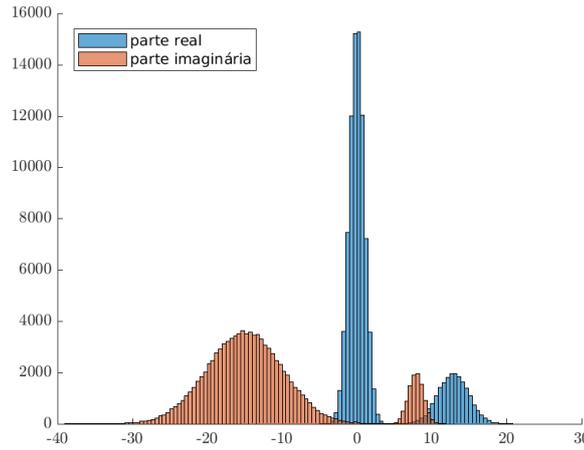


Figura 4.1: Histograma com amostra do ruído bimodal cuja parte real é $0.80\mathcal{N}(0, 1) + 0.20\mathcal{N}(13, 2)$, e a parte imaginária $0.90\mathcal{N}(-15, 5) + 0.10\mathcal{N}(8, 1)$.

4.1.2 Ruído Alfa-Estável

Esta tese utiliza a distribuição de alfa estável de Lévy, também conhecida como distribuição estável, ou alfa estável para modelar um ambiente de ruído impulsivo e gaussiano [Samorodnitsky e Taqqu 1994].

A distribuição estável é caracterizada por quatro parâmetros: índice de estabilidade ($0 < \alpha \leq 2$), parâmetro de assimetria ($-1 \leq \beta \leq 1$), parâmetro de escala ($\sigma \geq 0$) e um parâmetro de deslocamento ($\mu \in \mathbb{R}$) [Samorodnitsky e Taqqu 1994]. A forma mais comum de introduzir variáveis aleatórias com essa distribuição é através de sua função característica:

$$E[\exp(j\theta X)] = \exp \left\{ -\sigma^\alpha |\theta|^\alpha \left(1 - j\beta \operatorname{sign}(\theta) \tan \frac{\pi\alpha}{2} \right) + j\mu\theta \right\}, \quad (4.1)$$

se $\alpha \neq 1$, e

$$E[\exp(j\theta X)] = \exp \left\{ -\sigma|\theta| \left(1 - \frac{2j\beta}{\pi} \operatorname{sign}(\theta) \ln|\theta| \right) + j\mu\theta \right\}, \quad (4.2)$$

se $\alpha = 1$. A função sign é definida como

$$\operatorname{sign}(u) = \begin{cases} -1 & u < 0, \\ 0 & u = 0, \\ 1 & u > 0. \end{cases} \quad (4.3)$$

O índice de estabilidade α é responsável por controlar o quão impulsivo será o ruído. O quão menor for o valor de α , maior vai ser a cauda da distribuição. É importante ressaltar que esta tese trabalha com distribuições estáveis simétricas, isto é, β e μ são sempre selecionados como 0 em todas as simulações que usam esse tipo de ruído. Isso porque, quando o valor de $\beta, \mu = 0$, e $\alpha = 2$, a distribuição estável é equivalente a uma distribuição gaussiana. [Samorodnitsky e Taqqu 1994].

A relação sinal-ruído generalizada (*Generalized Signal-to-Noise Ratio - GSNR*) [Nikias e Shao 1995] em dB, definida como

$$\text{GSNR} = 10 \log \left(\frac{P_S}{\sigma^\alpha} \right), \quad (4.4)$$

é usada para determinar o parâmetro de escala σ . P_S é a potência do sinal sem ruído, e α é o índice de estabilidade. Dessa forma, para um dado valor de GSNR e α , o parâmetro de escala σ é calculado usando a Equação (4.4).

Nesta tese, todos os ruídos gerados a partir da distribuição estável tem os mesmos parâmetros para parte real e imaginária. Um exemplo de histograma de uma distribuição alfa estável pode ser visto na Figura 4.2.

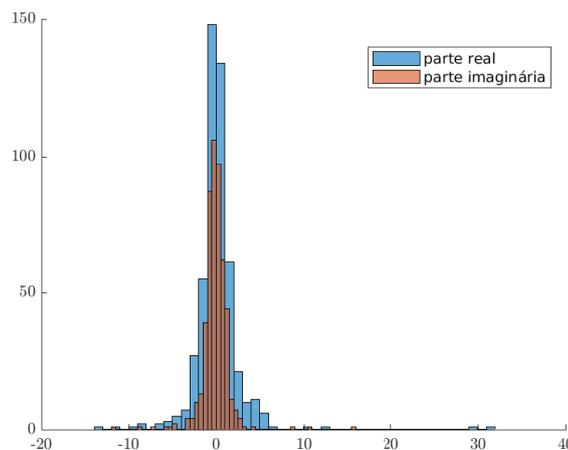


Figura 4.2: Histograma com amostra do ruído com distribuição alfa estável com GSNR = 20dB e $\alpha = 1.5$, tanto para parte real quanto imaginária.

4.2 Métricas de avaliação

A métrica mais usada para avaliar os algoritmos apresentados nesse capítulo é a razão entre o sinal original \mathbf{s} e o erro \mathbf{e} (*Signal to Error Ratio - SER*), e é definida em dB como:

$$\text{SER}_{dB} = 10 \log \left(\frac{\|\mathbf{s}\|^2}{\|\mathbf{e}\|^2} \right), \quad (4.5)$$

onde $\mathbf{e} = \mathbf{s} - \mathbf{w}$, e \mathbf{w} é o vetor estimado pelo algoritmo. Vale lembrar que o operador $\|\cdot\|$ de norma 2 aplicado aos complexos $\mathbf{s}, \mathbf{e} \in \mathbb{C}^L$ resulta em:

$$\text{SER}_{dB} = 10 \log \left(\frac{\mathbf{s}^H \mathbf{s}}{(\mathbf{s} - \mathbf{e})^H (\mathbf{s} - \mathbf{w})} \right). \quad (4.6)$$

Outra métrica utilizada na avaliação dos desempenhos dos algoritmos num cenário de equalização de canal é a taxa de erro por bit (*Bit Error Rate - BER*).

4.3 Identificação de Sistemas Lineares

Para testar a robustez da correntropia complexa como função custo a ambientes de ruído impulsivo, utiliza-se do MCCC numa típica tarefa de identificação de sistemas. Um diagrama esquemático do problema pode ser visto na Figura 4.3. Deseja-se identificar o vetor complexo $\bar{\mathbf{w}} = [(+1 - 2j), (-3 + 4j)]^T$, cujos valores foram escolhidos arbitrariamente. Um sinal complexo de entrada \mathbf{x} , cuja a parte real é caracterizada pela PDF $\mathcal{N}(1.0, 0.5)$ enquanto que a parte imaginária é caracterizada pela PDF $\mathcal{N}(2.0, 1.5)$ é multiplicado por $\bar{\mathbf{w}}$. Depois, esse vetor resultante é somado com um ruído η , gerando o vetor desejado \mathbf{d} . Através das medidas ruidosas do vetor \mathbf{d} e da entrada \mathbf{x} , deseja-se obter o valor de $\bar{\mathbf{w}}$.

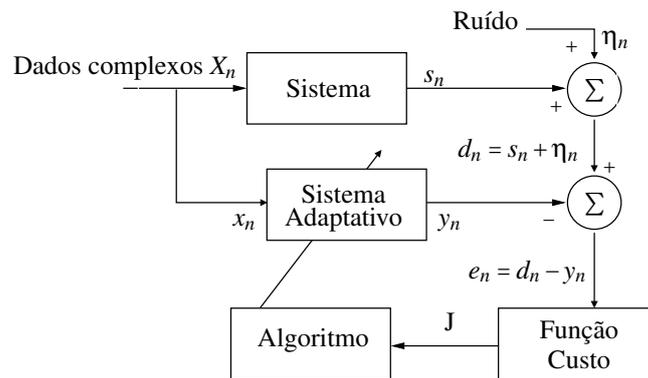


Figura 4.3: Típico problema de identificação de sistemas.

Para isso utiliza-se o MCCC aplicado ao método do gradiente e ao ponto fixo, conforme descrito no capítulo anterior. Além disso, 3 tamanhos de *kernel* foram utilizados. Esses algoritmos são confrontados com 3 outros algoritmos da literatura: CLMS, [Mandic e Goh 2009], CRLS [Diniz 2008] e o *Least Absolute Deviation* (LAD) [Alliney e Ruzinsky 1994]. O SER é usado como métrica para avaliação dos resultados.

Num primeiro momento, o ruído impulsivo foi modelado através de uma distribuição bimodal $0.90\mathcal{N}(0.0, 0.05) + 0.10\mathcal{N}(10.0, 5.0)$ tanto para a parte real quanto para a parte imaginária. O resultado da identificação do sistema pelos algoritmos pode ser visto na Figura 4.4, que mostra a média de 10^4 simulações de Monte Carlo. \mathbf{w} sempre é inicializado com zeros e as simulações são finalizadas sempre após 300 iterações de cada algoritmo. Já a Figura 4.5 mostra o desvio padrão referente a essa mesma simulação.

Conforme esperado, uma rápida convergência é obtida através dos métodos baseado em ponto-fixo, caso do CRLS e dos ponto fixo que usam o MCCC. Os algoritmos baseados no gradiente ascendente tem uma convergência mais lenta (Gradiente com MCCC, CLMS e CLAD). Todos eles usaram um passo fixo de $\mu = 0.02$. É possível observar também que o ruído impulsivo deteriora o desempenho dos métodos de segunda ordem como o CRLS e o CLMS. Conforme esperado, os métodos que usam o MCCC e o LAD como função custo sofrem menos com ruído impulsivo. A figura 4.4 mostra a queda do SER dos métodos baseados em segunda ordem.

O tamanho de *kernel* é um parâmetro livre do MCCC. Conforme foi mostrado nas

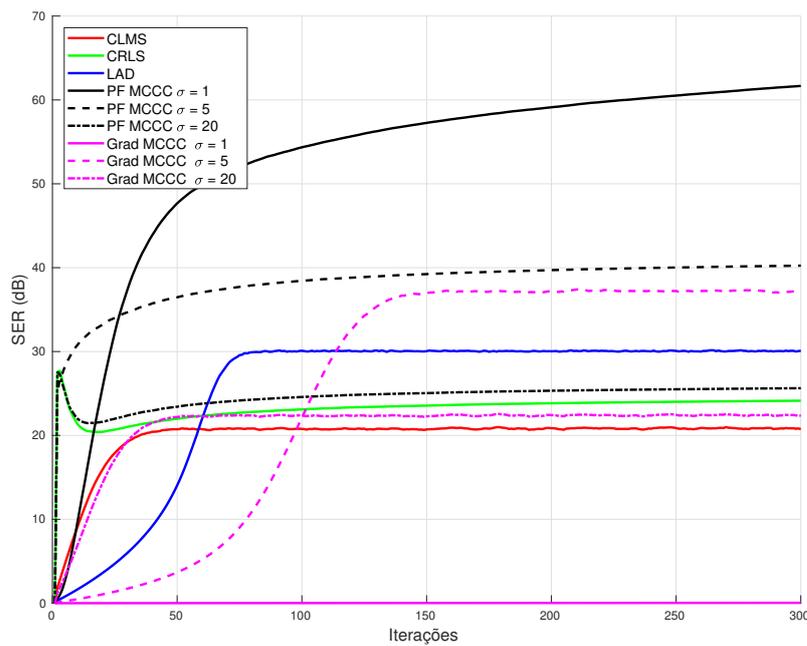


Figura 4.4: Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS. Ambiente com ruído bimodal.

propriedades da correntropia complexa, ele pondera os momentos estatísticos que estão sendo usados pela medida. Os três valores de *kernel* escolhidos ajudam a mostrar o efeito desse parâmetro no algoritmo. Por exemplo, conforme o tamanho do *kernel* cresce, os momentos de alta ordem decrescem rapidamente, fazendo com que os métodos baseados no MCCC se aproximem do desempenho dos métodos que usam estatística de segunda ordem, caso do CRLS, e do CLMS.

O MCCC com o ponto fixo foi o algoritmo que em média atingiu os maiores valores de SER. Um valor pequeno do *kernel* ajuda a ignorar os *outliers* cateterísticos do ruído bimodal. Isso pode fazer também com que a convergência do método seja prejudicada, tanto que, analisando a Figura 4.5, nota-se que o ponto fixo com MCCC e tamanho de *kernel* 1 tem o maior desvio padrão da simulação, indicando que algumas simulações podem não terem convergido para $\bar{\mathbf{w}}$. A velocidade de convergência também é prejudicada, para confirmar isso basta confrontar o ponto fixo com MCCC e tamanho de *kernel* 1 com o ponto fixo e MCCC mas com tamanho de *kernel* 5.

É importante também fazer um comparativo em relação ao MCCC usando o gradiente e o ponto fixo. A diminuição do tamanho do *kernel* faz com que a superfície de desempenho se torne mais rugosa, o que dificulta bastante o gradiente ascendente. Por isso, é possível ver que o MCCC com gradiente e tamanho de *kernel* 1 não convergiu. Também vale ressaltar que o algoritmo baseado em ponto fixo necessita realizar inversão matricial para calcular os pesos ótimos, tendo assim um custo computacional maior.

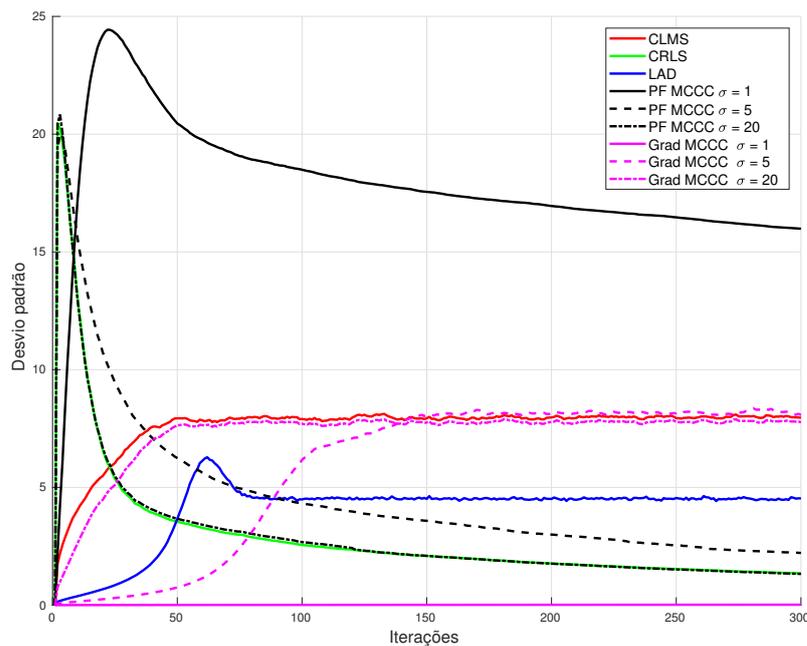


Figura 4.5: Desvio padrão referente as simulações mostradas na Figura 4.4.

O mesmo problema foi testado quando a distribuição alfa estável é usada para modelar o ruído impulsivo, com parâmetros GSNR = 10dB e $\alpha = 1.5$. Os resultados da média do SER podem ser vistos na Figura 4.6. Os desvios padrão correspondentes são vistos na Figura 4.7.

Apesar dos desempenho entre os métodos serem parecidos ao exemplo anterior, é interessante ver que o tamanho do *kernel* precisou ser ajustado para que o MCCC pudesse ter resultados satisfatórios. Mudando a natureza dos dados, seja do vetor de entrada ou do ruído, o tamanho do *kernel* precisa ser ajustado. A variância infinita que caracteriza a distribuição de probabilidade do ruído alfa estável torna o processo de identificação mais difícil, resultando em níveis de SER menores para todos os algoritmos. Assim, o tamanho de *kernel* $\sigma = 1$ que obteve os melhores resultados no ambiente de ruído bimodal já não convergiu quando o ruído segue a distribuição alfa estável, sendo considerado muito pequeno. O gradiente com tamanho de *kernel* 1 também não convergiu.

De forma geral, como as simulações nos dois ambientes puderam mostrar, a escolha apropriada do tamanho do *kernel* faz com que o MCCC com ponto fixo seja um método que consegue aliar a robustez ao ruído impulsivo do algoritmo LAD como a rápida convergência do CRLS. Caso o custo computacional seja considerável ao problema, pode-se recorrer ao gradiente com MCCC.

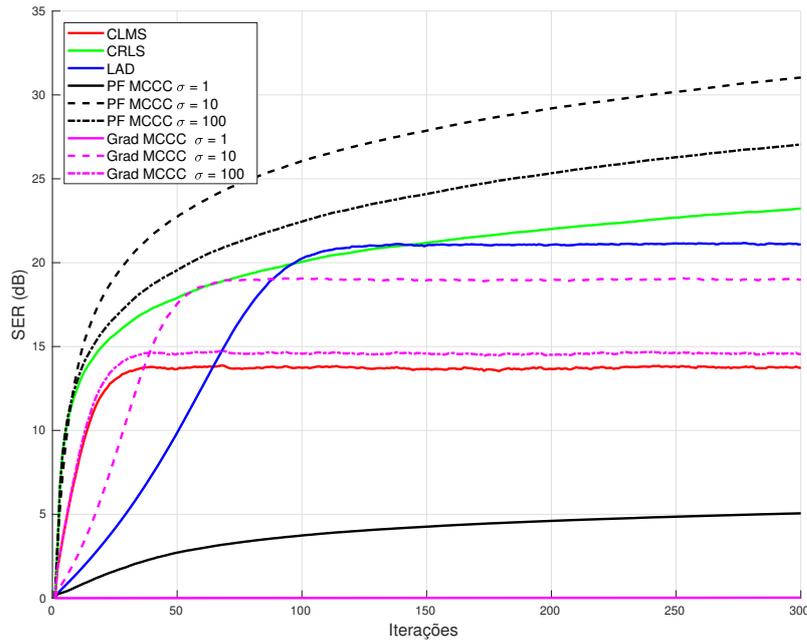


Figura 4.6: Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS num ambiente com ruído modelado pela distribuição alfa estável, com GSNR = 15 dB e $\alpha = 1.5$.

4.3.1 Resultados usando *kernel* adaptativo

Esta seção compara os resultados obtidos na seção anterior com as estratégias de *kernel* adaptativos apresentadas nesta tese. É considerado o mesmo cenário de identificação de sistema com os mesmos parâmetros do vetor de entrada. A distribuição alfa estável modela o ambiente de ruído impulsivo, com GSNR = 15dB e $\alpha = 1.5$. A estratégia VKW-MCCC deixa a convergência do algoritmo mais lenta. Dessa forma, ao invés de 300 iterações são analisados 1500 passos de cada algoritmo. Uma outra mudança foi a exclusão dos algoritmos baseados em ponto fixo do comparativo, já que todas as estratégias de *kernel* adaptativo são desenvolvidas para o gradiente ascendente e não se desejava ter um gráfico com excesso de algoritmos, dificultando o comparativo. Como o ambiente e os parâmetros são os mesmos, pode-se comparar as curvas de resultados desta seção com os resultados da seção anterior. Uma outra mudança foi a exclusão do gradiente com MCCC e tamanho de *kernel* fixo $\sigma = 1$ já que nas simulações da seção anterior esse valor de *kernel* impediu a conversão do método. Optou-se então por escolher um valor de $\sigma = 5$ como o menor *kernel* fixo testado no gradiente ascendente com MCCC.

As três estratégias implementadas, AMCCC, SMCCC e VKW-MCCC calculam um sigma ótimo baseado no erro de estimação do sistema. Apesar de serem estratégias que visam tornar o ajuste do *kernel* automático, todas elas precisam de um valor de *kernel* inicial, não eliminando o parâmetro livre totalmente. Maiores detalhes podem ser encon-

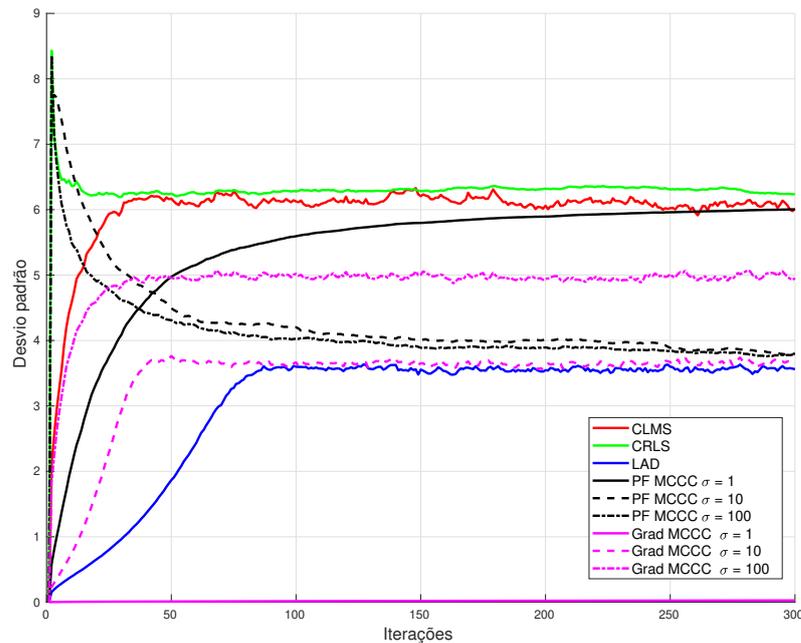


Figura 4.7: Desvio padrão referente as simulações mostradas na Figura 4.6.

trados na seção 3.6 do capítulo anterior. Todas as simulações desta seção utilizaram um tamanho de *kernel* inicial para os algoritmos de $\sigma_i = 0.2$. Nos testes realizados, dependendo do nível de ruído, tamanho do vetor entrada, etc, os algoritmos tem o desempenho mais ou menos sensível a escolha desse parâmetro inicial, com destaque para o VKW, que se mostrou a estratégia mais sensível a variação desse parâmetro.

A Figura 4.8 mostra o SER médio de 10^4 simulações de Monte Carlo. Nela, é possível verificar que o VKW-MCCC foi a estratégia que atingiu o maior valor de SER médio, em torno de 23 dB após 1500 iterações. A título de comparação, no mesmo ambiente, o melhor resultado até então foi o ponto fixo com MCCC usando um tamanho de *kernel* $\sigma = 10$, que atingiu em média cerca de 30dB depois de 300 iterações. Apesar do bom resultado, o VKW-MCCC foi o algoritmo com a convergência mais lenta dentre todos os testados, ultrapassando o SER médio do MCCC com gradiente ascendente e tamanho de *kernel* fixo $\sigma = 5$ somente após cerca de 550 iterações.

Já a estratégia AMCCC mostrou-se ter um SER médio menor do que o LAD e o MCCC com gradiente ascendente e tamanho de *kernel* fixo $\sigma = 10$, e maior do que a estratégia adaptativa SMCCC.

Tanto o AMCCC quanto o SMCCC superaram em média o desempenho do gradiente com MCCC e um tamanho de *kernel* $\sigma = 100$ (maior testado) ou do próprio CLMS.

O desvio padrão das simulações da Figura 4.8 é mostrada na Figura 4.9.

O uso da correntropia complexa como função custo, assim como no caso real, traz como desvantagem o fato da introdução do tamanho do *kernel* como parâmetro livre no método. Como foi mostrado nas simulações, ele é responsável por controlar a velocidade

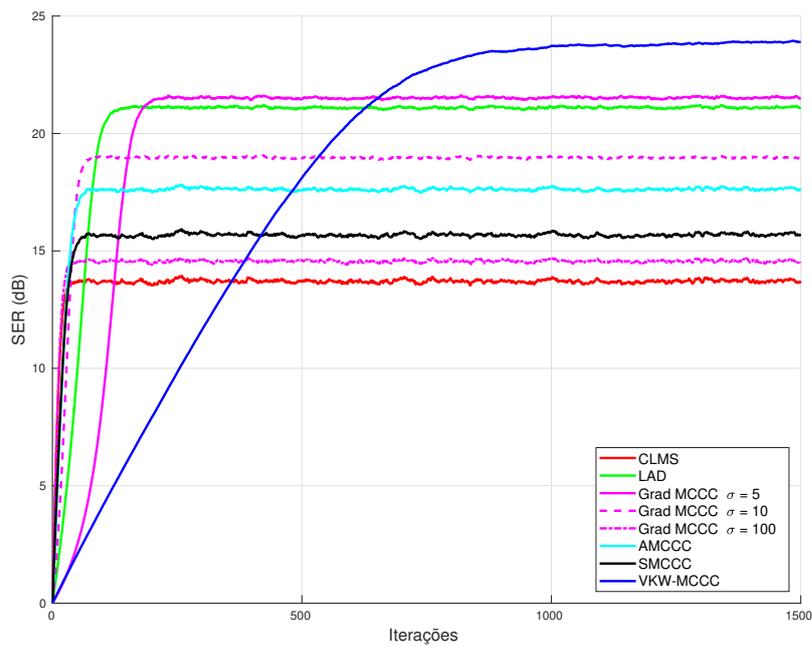


Figura 4.8: Média de 10^4 SER comparando os algoritmos do gradiente com MCCC e o ponto fixo com MCCC com os algoritmos clássicos da literatura CLMS, LAD, CRLS num ambiente com ruído modelado pela distribuição alfa estável, com GSNR = 15 dB e $\alpha = 1.5$.

de convergência e desempenho em regime do algoritmo já que seleciona quais momentos estatísticos estão sendo utilizados. As estratégias de usar um *kernel* adaptativo mostraram-se mais eficazes do que escolher um valor não ótimo aleatório para o tamanho do *kernel* do algoritmo. Por outro lado, pode-se notar que o algoritmo ainda é sensível ao *kernel* fixo que ainda é necessário como parâmetro para iniciar os algoritmos.

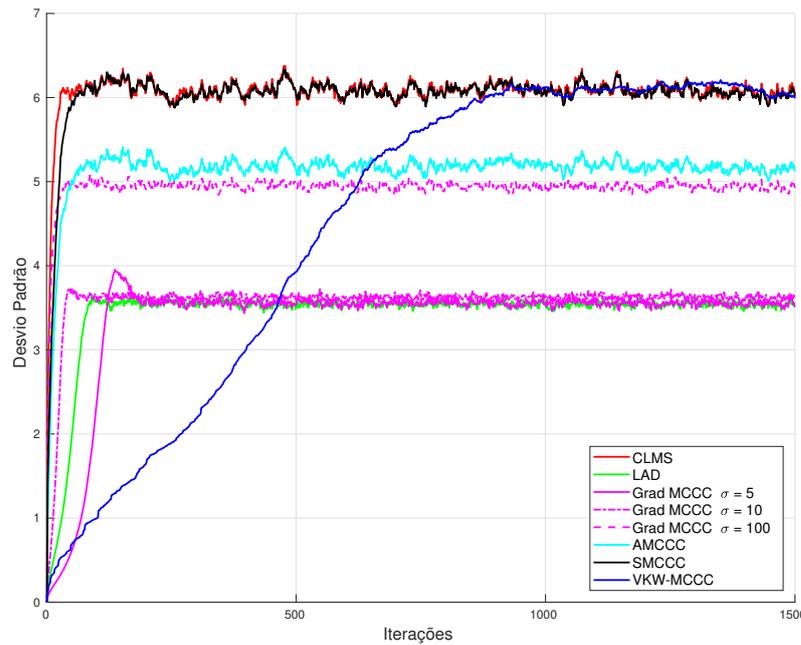


Figura 4.9: Desvio padrão referente as simulações mostradas na Figura 4.8.

4.4 Equalização de Canal

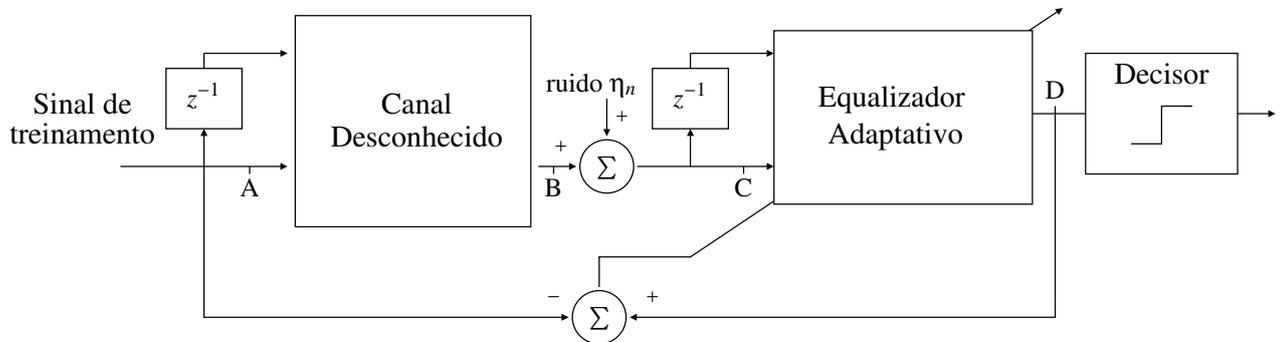


Figura 4.10: Diagrama de blocos contendo o problema de Equalização de Canal.

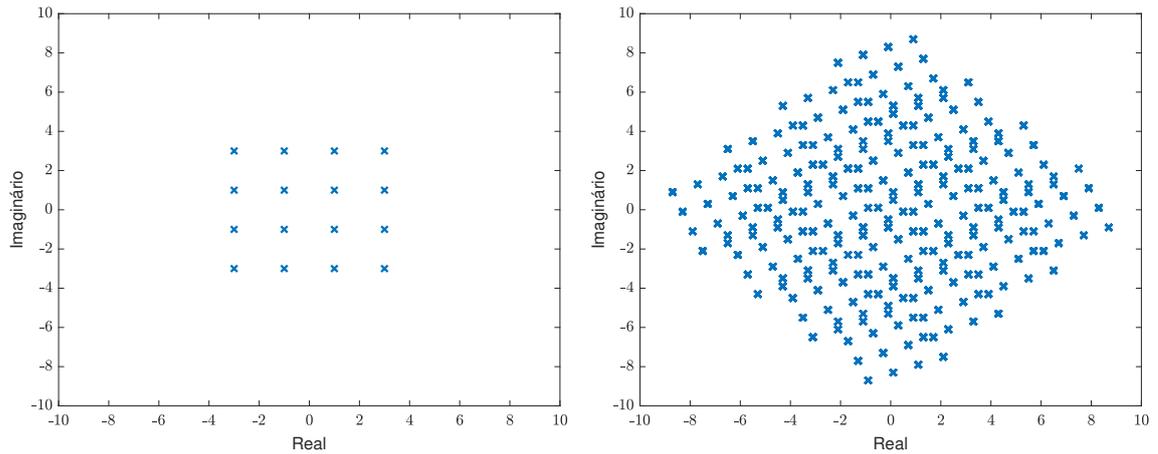
Um problema de equalização de canal foi selecionado como um segundo cenário para simulações utilizando os mesmos algoritmos para contrastar o MCCC. A Figura 4.10 mostra um diagrama de blocos de um sistema de comunicação associado a um equalizador de canal adaptativo.

Um sinal de treinamento conhecido formado pela sequência de 5000 símbolos complexos espaçados associados a uma modulação de amplitude em quadratura 16-QAM (Quadrature Amplitude Modulation) cujos valores podem assumir: $(\pm 1 \pm j1, \pm 2 \pm j2, \dots, \pm 3 \pm j3)$, como pode ser visto na Figura 4.11-a. O componente de fase $A_{k;l}$ e de quadratura

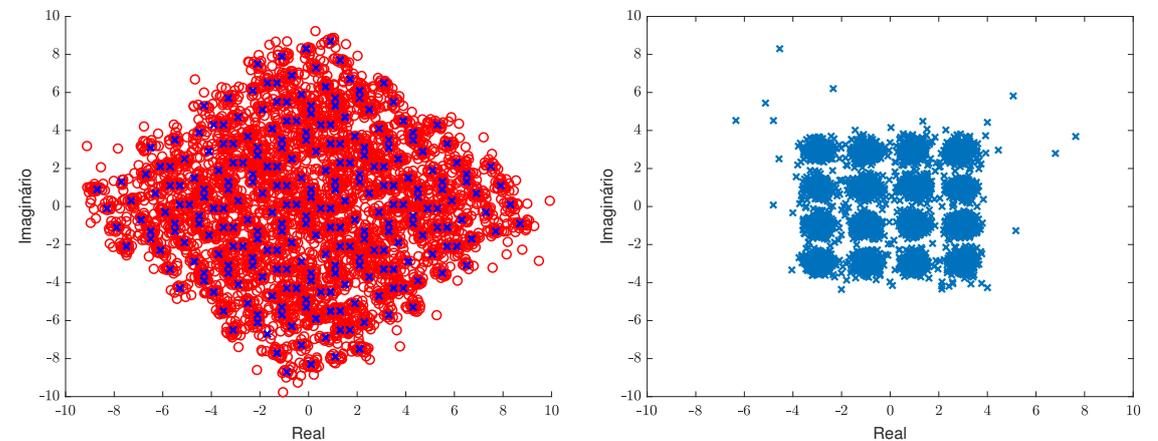
$A_{k;Q}$ no A -ésimo instante e seu respectivo atraso são submetidos a um canal desconhecido \mathbf{w} , modelado como $\mathbf{w} = [w_1, w_2]^T$, onde w_1 e $w_2 \in \mathbb{C}$. O resultado disso é um sinal distorcido denominado B_k :

$$B_K = [A_K \ A_{K-1}] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad (4.7)$$

onde w_1 e w_2 foram arbitrariamente escolhidos como $(1.1 - j1.1)$ e $(0.5 - j0.2)$ respectivamente. O sinal distorcido B_K pode ser observado na Figura 4.11-b.



(a) Sinal de entrada limpo. Ponto A da Figura 4.10. (b) Sinal de entrada distorcido pelo canal. Ponto B da Figura 4.10.



(c) Sinal de entrada distorcido pelo canal e corrompido com ruído com distribuição alfa estável e GSNR = 20 dB. Ponto C da Figura 4.10. (d) Sinal corrigido pela equalização de canal utilizando o MCCC e $\text{kernel size } \sigma = 1$. Ponto D da Figura 4.10.

Figura 4.11: Fluxo dos dados no processo de equalização de canal em cada ponto da Figura 4.10.

Como foi observado que nas simulações de identificação de sistemas, o ruído construído a partir da distribuição alfa estável é bem mais severo do que o bimodal, como mostram as curvas SER da seção anterior. Assim, para testar a resposta dos algoritmos a

um ambiente de ruído impulsivo, as simulações conduzidas na equalização de canal utilizaram somente um ruído aditivo η_k ao sinal B_K com distribuição de probabilidade alfa estável, gerando o sinal $C_K = B_K + \eta_K$, mostrado na Figura 4.11-c, que foi gerada a partir de um $\alpha = 1.5$.

O sinal C_K e seu respectivo atraso são usado enquanto entradas do equalizador adaptativo assim como o erro entre o sinal de treinamento e a saída do equalizador D_K . Nas simulações, w sempre é inicializado com zeros. Enquanto o MCCC e o CRLS são métodos de ponto fixo, o CLMS e o LAD utilizam o gradiente ascendente com coeficiente de aprendizado $\mu = 0.01$. Após finalizado o processo de equalização, o parâmetro w encontrado ao fim das iterações é usado para gerar uma curva de taxa de erro (*Bit Error Rate* - BER) para cada um dos algoritmos e para diferentes valores de GSNR. O resultado desse comparativo pode ser visto na Figura 4.12.

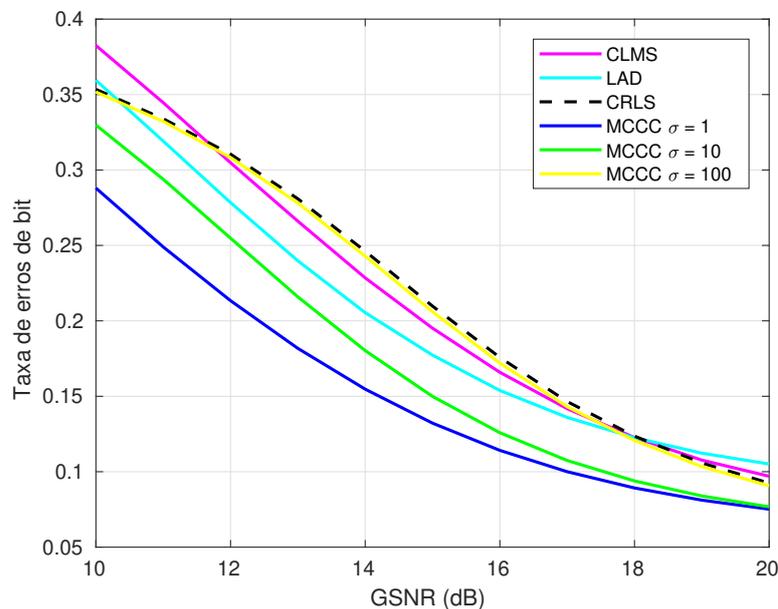


Figura 4.12: Média de 10^5 curvas BER para os algoritmos CLMS, CRLS, LAD e MCCC com t .

Além dos métodos clássicos já mencionados, três valores de *kernel* foram escolhidos para a solução de ponto fixo com MCCC: 1, 10, e 100. Além disso, vale salientar que o desempenho dos algoritmos são avaliados na média de 10^5 simulações de Monte Carlo. A Figura 4.13 mostra os desvios padrão correspondentes a cada um dos algoritmos nas 10^5 simulações de Monte Carlo.

Como pode ser visto na Figura 4.12, todos os algoritmos conseguiram resultados satisfatórios em um cenário com alto valor de GSNR. Entretanto, a medida em que esse parâmetro decai, graças a natureza do ruído utilizado, a probabilidade do aparecimento de *outliers* aumenta. Esse tipo de ambiente degrada o desempenho de métodos de segunda ordem, como o CRLS e o CLMS. A figura 4.14 mostra o comportamento do MCCC para valores de α que variam de 1 a 2 enquanto que o tamanho do *kernel* é fixo em $\sigma = 1$. De

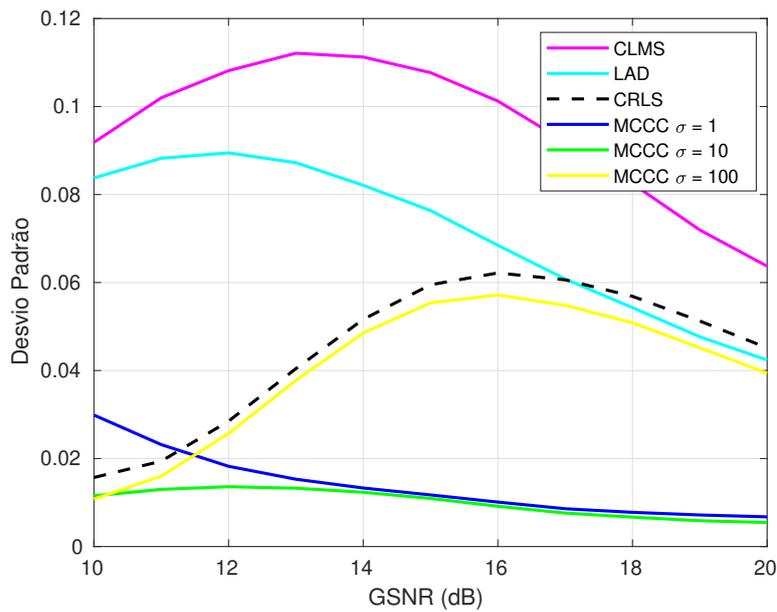


Figura 4.13: Desvio padrão referente aos resultados mostrados na Figura 4.12.

acordo com a propriedade 3, usando o *kernel* gaussiano, a correntropia complexa contém informação de alta ordem estatística, fazendo com que o MCCC seja robusto a ruído impulsivo como o LAD.

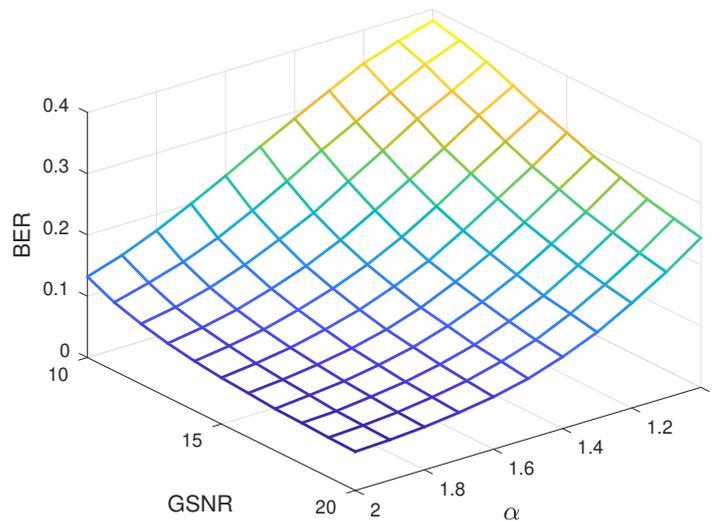


Figura 4.14: Curvas BER para o algoritmo do MCCC com o tamanho de *kernel* fixo em $\sigma = 1$ para diferentes valores de GSNR e de α .

Além disso, o desvio padrão das simulações de Monte Carlo que são vistas na Figura 4.13 mostra que, para toda a faixa de tamanhos de *kernel* testada e de valores de GSNR,

o MCCC teve o menor desvio padrão entre todos os algoritmos testados. Como pode ser atestado pela Figura 4.12, o desempenho do MCCC depende diretamente da escolha correta do tamanho de *kernel*, que é um parâmetro livre do método. Os melhores resultados foram obtidos com $\sigma = 1$. A figura 4.11-d mostra o sinal corrigido, D_K , produzido pelo método MCCC usando $\sigma = 1$ e $GSNR = 20dB$.

Pode-se notar que, para alguns valores de GSNR, o algoritmo LAD consegue resultados superiores ao MCCC com tamanho de *kernel* $\sigma = 100$. Como foi ressaltado pela propriedade 4, com o crescimento do tamanho do *kernel*, a correntropia complexa tende a se comportar como a correlação da mesma maneira que a correntropia o faz com dados reais. Sendo assim, conforme esperado, tanto em média (Figura 4.12-a) quanto no desvio padrão (Figura 4.12-b), o desempenho do MCCC se assemelha ao do CRLS com o maior tamanho de *kernel* testado, $\sigma = 100$. Para enfatizar a importância do ajuste do tamanho do *kernel*, foi produzida a Figura 4.15, que mostra a superfície formada por cada uma das curvas BER criada com tamanhos de *kernel* variando de 5 a 50.

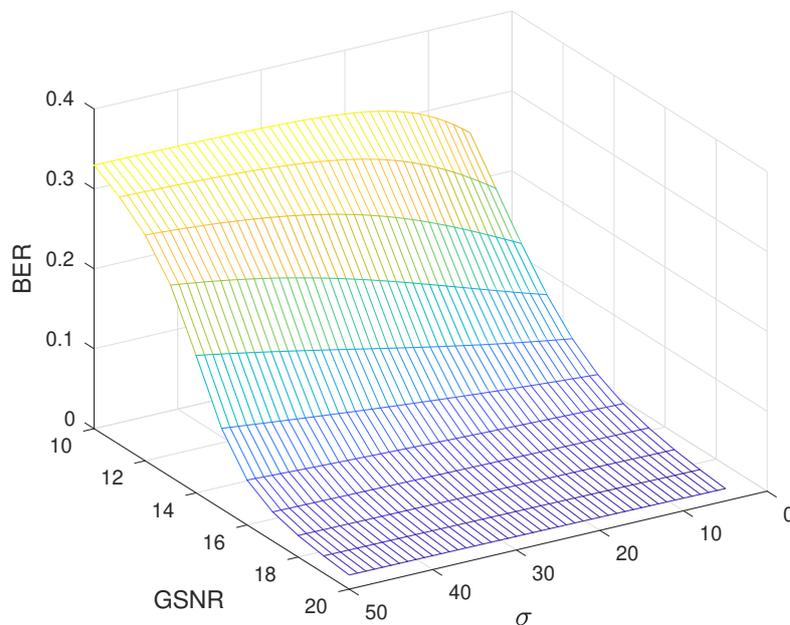


Figura 4.15: Influência de diferentes tamanhos de *kernel* no comportamento das curvas BER para o MCCC com diferentes valores de GSNR e índice de estabilidade ($\alpha = 1.5$) fixo.

Com o tamanho do *kernel* ajustado, o MCCC foi capaz de obter menores valores de erros que os algoritmos tradicionais para todos os valores de GSNR testados, mostrando ser uma ferramenta válida em ambientes de ruído impulsivo.

4.5 Amostragem Compressiva

Recentemente, um grande interesse surgiu na comunidade científica em torno de uma área chamada amostragem compressiva (*Compressive sensing* - CS). Trata-se de tirar vantagem de sinais esparsos para reconstruí-los usando menos medidas do que a teoria da amostragem de Nyquist-Shannon indicariam como adequada [Candes e Wakin 2008]. Por isso, a amostragem compressiva foi aplicada com êxito em diversas áreas como prospecção sísmica [Sun et al. 2018] ou até monitoramento de sinais biométricos [Meng et al. 2018, Zhu et al. 2019, Meng e Zhu 2019]. As técnicas de amostragem compressiva têm se mostrado de certa forma robusta na tarefa de reconstrução de sinais esparsos sob influência de ruído gaussiano [Akcakaya e Tarokh 2010, Zhu e Baron 2013, Candès 2008, Lin 2017], entretanto, ruído impulsivo tende a deteriorar o desempenho de algoritmos tradicionais da área que são baseados em métodos de otimização que usam informação estatística de segunda ordem [Candes e Wakin 2008, Becker et al. 2011, Zhang et al. 2018].

Seja um sinal \mathbf{x} de tamanho L , com esparsidade S , ou seja, das L amostras, apenas S são diferentes de zero, é possível reconstruir \mathbf{x} a partir de um sinal menor \mathbf{y} com tamanho M , ou seja, $M < L$, através do mapeamento de uma matriz \mathbf{A} de dimensão $M \times L$. Essa matriz é chamada de matriz de medição. Matematicamente, isso pode ser expresso como um sistema possível indeterminado (com múltiplas soluções) $\mathbf{y} = \mathbf{A}\mathbf{x}$. A Figura 4.16, adaptada do trabalho [Baraniuk 2007], mostra essa relação.

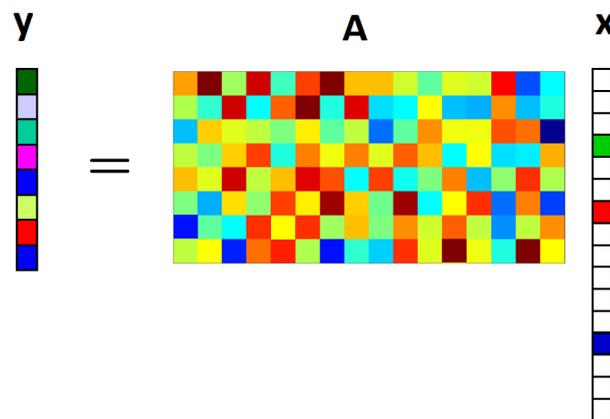


Figura 4.16: Típico problema de amostragem compressiva. Deseja-se reconstruir um vetor esparsa \mathbf{x} com tamanho $L \times 1$ a partir de um vetor menor \mathbf{y} com tamanho $M \times 1$, onde $M < L$.

A "norma" ℓ_0 , indicado pelo operador $\|\cdot\|_0$, corresponde a quantidade de elementos diferentes de zero de um vetor, ou seja

$$\|\mathbf{x}\|_0 = \#\{i | x_i \neq 0\}, \quad (4.8)$$

onde $\#(\cdot)$ é operador de cardinalidade.

Esta tese faz uso de um abuso de notação chamando ℓ_0 de norma, assim como grande

parte dos trabalhos de amostragem compressiva. Formalmente, a "norma" ℓ_0 não pode ser qualificada como norma porque não é uma operação homogênea.

A amostragem compressiva requer uma esparsidade $S = \|\mathbf{x}\|_0$ muito menor do que o comprimento do sinal \mathbf{x} , $S \ll L$. Além disso, a matriz de medição \mathbf{A} deve ter uma isometria restrita (*Restricted Isometry Property* - RIP) para que o sinal consiga ser reconstruído com alta probabilidade [Candès 2008]. De forma prática, submatrizes da transformada discreta de Fourier, transformada discreta do cosseno e até matrizes aleatórias atendem aos critérios de isometria restrita podendo então serem utilizadas em problemas de amostragem compressiva [Candes e Wakin 2008].

Levando em consideração o ruído, pode-se escrever o problema de amostragem compressiva como $\mathbf{y} = \mathbf{Ax} + \gamma$, onde γ é um ruído aditivo. Ainda em relação a modelagem do problema, quando um sinal \mathbf{x} não é esparso, pode-se usar uma transformação bijetiva T de forma que $T\mathbf{x} = \hat{\mathbf{x}}$ seja esparso. Ou seja, $\mathbf{y} = \mathbf{AT}^{-1}\hat{\mathbf{x}} + \gamma$. Sem perda de generalidade, é possível escrever $\mathbf{y} = \mathbf{A}\hat{\mathbf{x}} + \gamma$, redefinindo a matriz de medição \mathbf{A} .

Vários algoritmos foram propostos para resolver o problema do sistema $\mathbf{y} = \mathbf{Ax} + \gamma$, tais como programação linear e otimização convexa. A grande maioria deles implementa uma solução aproximada da minimização de $\|\mathbf{x}\|_0$ sujeito a restrições:

$$\begin{aligned} & \underset{x}{\text{minimizar}} \quad \|\mathbf{x}\|_0 \\ & \text{sujeito a} \quad \|\mathbf{Ax} - \mathbf{b}\|_2^2 \leq \varepsilon, \end{aligned} \tag{4.9}$$

onde \mathbf{b} é o vetor de medidas que é poluído pelo ruído aditivo e ε é um fator de tolerância.

Entretanto, minimizar (4.9) é um problema NP-difícil [Candes e Wakin 2008], fazendo com que sua solução seja muito demorada [Seth e Principe 2008]. Por esse motivo, o que grande parte da literatura propõe é minimizar a norma ℓ_1 [Seth e Principe 2008, Becker et al. 2011] ao invés da norma ℓ_0 :

$$\begin{aligned} & \underset{x}{\text{minimizar}} \quad \|\mathbf{x}\|_1 \\ & \text{sujeito a} \quad \|\mathbf{Ax} - \mathbf{b}\|_2^2 \leq \varepsilon, \end{aligned} \tag{4.10}$$

O trabalho de Seth e Principe (2008) mostrou que a métrica induzida da correntropia poderia aproximar a norma ℓ_0 , reduzindo o número de medidas necessário a reconstrução de sinais reais quando comparado a norma ℓ_1 .

Já He et al. (2019) incluiu o critério de máxima correntropia na restrição em conjunto com uma estratégia de atrator de zeros para aproximar a norma ℓ_0 e conseguir resultados animadores em ambientes de ruído impulsivo mas sempre envolvendo dados reais.

Vários são as aplicações de amostragem compressiva que necessitam de dados complexos [Zhu e Bamler 2012, Sung et al. 2017, Bettens et al. 2017, Tivive e Bouzerdoum 2013, Zhu e Bamler 2011]. Sendo assim, esta tese aplicou a CCIM para estender a vetores complexos o trabalho de Seth e Principe (2008). Além disso, desenvolveu um novo algoritmo chamado ℓ_0 -MCCC que usa o MCCC para reconstruir vetores complexos a partir de medidas contaminadas com ruído impulsivo. Essas duas contribuições a área de amostragem compressiva são descritas nas seções a seguir. Em ambos os casos, os algoritmos

propostos são confrontados com soluções da literatura que são capazes de lidar com dados de natureza complexa.

4.5.1 CCIM como aproximação da norma ℓ_0

Como foi mostrado na propriedade 9.1 da correntropia complexa (seção 3.4), fazendo o tamanho de *kernel* $\sigma \rightarrow 0$, a CCIM (ψ^c) é capaz de aproximar a norma ℓ_0 num problema de otimização. Dessa forma, uma possível solução a (4.10) é fazer:

$$\mathbf{z} = \min \psi_{\sigma}^c(\mathbf{z}, \mathbf{0}) \text{ sujeito a } \mathbf{y} = \mathbf{A}\mathbf{z}, \quad (4.11)$$

onde $\mathbf{y} \in \mathbb{C}^M$ é o vetor de medidas complexo, enquanto $\mathbf{z} \in \mathbb{C}^L$ é o vetor esparso que deseja se reconstruir. Foi usado uma matriz de medição complexa $\mathbf{A} \in \mathbb{C}^{M \times L}$ aleatória que segue a PDF $\mathcal{N}(0, 1)$ tanto para parte real quanto para a parte imaginária.

Como no caso real desenvolvido em [Seth e Principe 2008], utiliza-se nesta tese o método da projeção do gradiente [Horie e Aiyoshi 1999] para realizar a minimização. Nesse método, o vetor gradiente \mathbf{g} é computado e projetado em \mathbf{A}^T . Como nas aplicações acima, a Equação (3.40) não é analítica no domínio dos complexos, veja:

$$\min \|\mathbf{z}\|_0 \approx \min [\psi_{\sigma \rightarrow 0}^c(\mathbf{z}, 0)]^2 = \frac{G_{\sigma}^c(0)}{L} \sum_{i=1}^L \left\{ 1 - \exp\left(-\frac{z_i z_i^*}{2\sigma^2}\right) \right\},$$

portanto o cálculo de Wirtinger deve ser aplicado mais uma vez para obter:

$$[\mathbf{g}]_{L \times 1} = \left[\frac{\partial \psi_{\sigma}^c(\mathbf{z}, 0)}{\partial z_1^*}, \dots, \frac{\partial \psi_{\sigma}^c(\mathbf{z}, 0)}{\partial z_L^*} \right], \quad (4.12)$$

onde

$$\frac{\partial \psi_{\sigma}^c(\mathbf{z}, 0)}{\partial z_i^*} = \frac{G_{\sigma}^c(0) z_i}{2L\sigma^2} \exp\left(-\frac{z_i z_i^*}{2\sigma^2}\right). \quad (4.13)$$

Então, a projeção pode ser obtida como

$$\tilde{\mathbf{g}} = [\mathbf{I} - \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{A}] \mathbf{g}, \quad (4.14)$$

então, para atualizar \mathbf{z} , temos:

$$\mathbf{z}(n+1) = \mathbf{z}(\mathbf{n}) - \eta \frac{\tilde{\mathbf{g}}}{\|\tilde{\mathbf{g}}\|_2}. \quad (4.15)$$

Como foi mostrado, o tamanho de *kernel* próximo de 0 é necessário para que a CCIM aproxime a norma ℓ_0 . Entretanto, deixar o valor do *kernel* muito pequeno dificulta a descida do gradiente, produzindo muitos mínimos locais. Para superar esse problema, foi adotado a mesma solução proposta ao caso real, uma técnica de anelamento de *kernel* (*kernel annealing*) proposto por Erdogmus (2002). Os resultados presentes nessa tese usam os mesmos parâmetros do trabalho de [Seth e Principe 2008] para o caso real com exceção do passo do gradiente que foi mudado para $\eta = 0.05$. Basicamente, trata-se da

descida exponencial do valor do tamanho *kernel* com o passar das iterações.

Nos resultados aqui apresentados, foi reconstruído um sinal complexo \mathbf{z} esparsos de tamanho $L = 512$. Quatro valores de esparsidade foram testadas, $S = 16, 32, 64, 128$. Os valores não nulos são selecionados aleatoriamente e uniformemente distribuídos dentro do vetor \mathbf{z} . O valor propriamente dito foi gerado usando uma distribuição normal com média 0 e variância 1, tanto para a parte real quanto para a parte imaginária. Vários números de medida foram testados, de $M = 40$ até $M = 300$ medidas, com um passo de 4.

A Figura 4.17 mostra a média de 10^3 iterações de Monte Carlo para cada valor de M testado. Nessa simulação nenhum ruído é acrescentado ao sinal. Uma reconstrução do vetor esparsos é considerada realizada com sucesso quando $\|\mathbf{z} - \mathbf{z}_r\|_2 < 10^{-3}$. O gradiente descendente é inicializado com a solução de mínima energia $\mathbf{z}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$. O método aqui proposto é comparado ao NESTA (Nesterov's Algorithm)[Becker et al. 2011], algoritmo que minimiza a norma ℓ_1 ao invés da CCIM na aproximação da norma ℓ_0 e que é capaz de lidar com dados complexos.

As simulações numéricas, mostradas pela Figura 4.17, apontam que não só a CCIM é capaz de aproximar a norma ℓ_0 como menos medidas são necessárias para reconstruir o vetor esparsos complexo, isso quando comparado ao NESTA, que usa a norma ℓ_1 . Como esperado no problema de amostragem compressiva, a medida em que se aumenta o número de elementos não nulos (S) no vetor esparsos, mais medidas são necessárias para se reconstruir com sucesso o vetor original complexo.

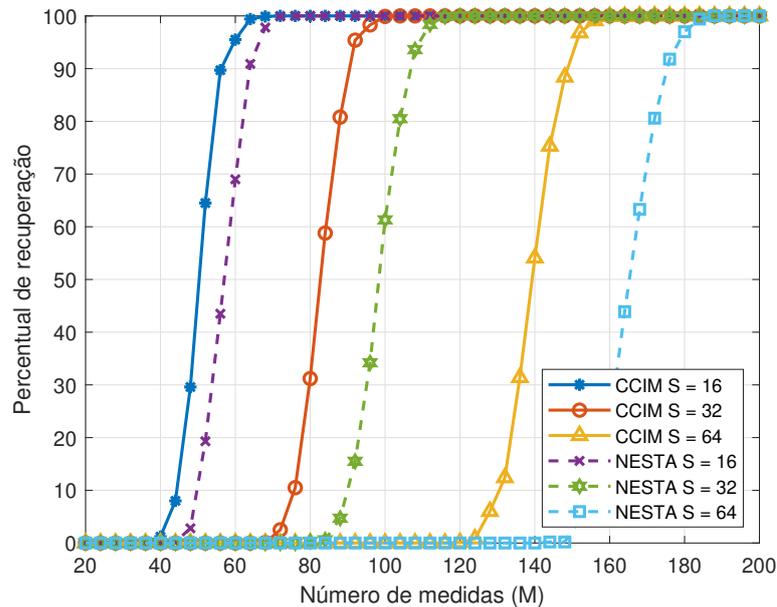


Figura 4.17: Comparativo entre o percentual de reconstruções com sucesso usando a aproximação da norma ℓ_0 com a CCIM com o NESTA, que utiliza a norma ℓ_1 . O vetor esparsos complexo que é recuperado tem tamanho $L = 512$ e S elementos não zero em função do número de medidas (M) realizadas.

4.5.2 ℓ_0 -MCCC

Recentemente, a correntropia foi utilizada como função custo num problema de amostragem compressiva [He et al. 2019], se mostrando uma importante ferramenta para lidar com ruídos não gaussianos. O novo algoritmo foi chamado de ℓ_0 -Maximum Correntropy Criterion (ℓ_0 -MCC). O algoritmo combina o MCC a uma estratégia de aproximação do gradiente da norma ℓ_0 através de um atrator de zeros, que foi primeiramente definido num algoritmo chamado ℓ_0 -Least Mean Square (ℓ_0 -LMS) definido em [Jin et al. 2010].

Ao invés de resolver o problema proposto na Equação (4.10), vamos definir uma nova função custo J , inspirada em [He et al. 2019], como

$$J = -V_{\sigma}^c(B, Y) + \lambda \|\mathbf{w}\|_0, \quad (4.16)$$

onde $V_{\sigma}^c(B, Y)$ é a correntropia complexa entre o sinal desejado e \mathbf{b} , que são as medidas contaminadas com ruído impulsivo, e a saída estimada pelo algoritmo $\mathbf{y} = \mathbf{A}\mathbf{w}$. Vale lembrar que $\mathbf{A} \in \mathbb{C}^{M \times L}$ é a matriz de medição e $\mathbf{w} \in \mathbb{C}^L$ é o vetor esparsa que se deseja reconstruir. λ é um simples parâmetro de regularização que pondera a aproximação do gradiente da norma ℓ_0 de \mathbf{w} .

Minimizar (4.16) significa maximizar a correntropia complexa $V_{\sigma}^c(B, Y)$ que é a similaridade entre B e Y . Uma vez que a similaridade é maximizada, o erro $E = B - Y$ é minimizado. Isso, combinado a minimização da norma ℓ_0 de \mathbf{w} garante a solução mais esparsa [Candes et al. 2006].

Gradiente da Correntropia complexa

Esta tese usa um gradiente estocástico para minimizar (4.16). Isso é feito a partir do uso do cálculo de Wirtinger para encontrar a derivada de $V_{\sigma}^c(\mathbf{b}, \mathbf{y})$ em relação a \mathbf{w}^* . Como a modelagem do problema foi feita de forma diferente da identificação de sistemas, a derivação da regra de atualização precisa ser calculada mais uma vez. Primeiro, reescreve-se $V_{\sigma}^c(\mathbf{b}, \mathbf{y})$ como:

$$V_{\sigma}^c(B, Y) = 1 - \frac{1}{2\pi\sigma^2} \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right), \quad (4.17)$$

onde $e_i = (b_i - y_i) = (b_i - \phi_i \mathbf{w})$ é o escalar da posição i do vetor erro \mathbf{e} . $\phi_i = \mathbf{A}(i, \cdot)$ é um vetor de tamanho $1 \times L$, extraído da i -ésima linha da matriz de medição \mathbf{A} . Usando o cálculo de Wirtinger em (4.17):

$$\frac{\partial V_{\sigma}^c(\mathbf{b}, \mathbf{y})}{\partial \mathbf{w}^*} = \frac{1}{2\pi\sigma^2} \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) \frac{1}{2\sigma^2} \frac{\partial (e_i e_i^*)}{\partial \mathbf{w}^*}, \quad (4.18)$$

onde

$$\begin{aligned}
e_i e_i^* &= (b_i - \phi_i \mathbf{w})(b_i - \phi_i \mathbf{w})^* \\
&= (b_i - \phi_i \mathbf{w})(b_i^* - \phi_i^* \mathbf{w}^*) \\
&= b_i b_i^* - b_i \phi_i^* \mathbf{w}^* - b_i^* \phi_i \mathbf{w} + \\
&\quad + \phi_i \mathbf{w} \phi_i^* \mathbf{w}^*.
\end{aligned} \tag{4.19}$$

Assim,

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = 0 - (b_i \phi_i^*)^T - 0 + (\phi_i \mathbf{w} \phi_i^*)^T, \tag{4.20}$$

que pode ser reescrito como

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = -b_i \phi_i^H + (\phi_i \mathbf{w}) \phi_i^H = -e_i \phi_i^H, \tag{4.21}$$

Então, usando (4.21) em (4.18):

$$\frac{\partial V_\sigma^c}{\partial \mathbf{w}^*} = -\frac{1}{2\pi\sigma^4} \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H. \tag{4.22}$$

Aproximação do gradiente da norma ℓ_0

Utiliza-se da mesma estratégia para aproximação do gradiente da norma ℓ_0 que foi usado em [Jin et al. 2010], e que é comum ao algoritmo ℓ_0 -MCC [He et al. 2019]. Entretanto, uma adaptação deve ser necessária para lidar com números complexos. Seja w_k o k -ésimo elemento do vetor esparsos \mathbf{w} , pode-se o reescrever como $w_k = w_k^{re} + jw_k^{im}$, assim:

$$\nabla \|w_k\|_0 \approx Z_\beta(w_k) = z_\beta(w_k^{re}) + jz_\beta(w_k^{im}), \tag{4.23}$$

onde

$$z_\beta(w) = \begin{cases} \beta^2 w + \beta, & -\frac{1}{\beta} \leq w < 0; \\ \beta^2 w - \beta, & 0 < w \leq \frac{1}{\beta}. \end{cases} \tag{4.24}$$

O termo $z_\beta(w)$ é chamado de atrator e a variável β é um parâmetro livre que controla a zona de atração para pequenos valores no intervalo $[-\frac{1}{\beta}, \frac{1}{\beta}]$. Assim, o produto $\beta^2 w \pm \beta$ controla o quão próximo de zero pode chegar o algoritmo, o que implica diretamente na sua precisão.

Algoritmo ℓ_0 -MCCC

Unindo a derivada de V_σ^c com relação a \mathbf{w}^* , mostrada na Equação (4.22), com a aproximação $\nabla \|\mathbf{w}\|_0$ obtida em (4.23), pode-se usar do gradiente estocástico para reconstruir sinais esparsos complexos usando a seguinte regra de atualização:

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \eta \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H + \eta \lambda Z_\beta(\mathbf{w}_i). \quad (4.25)$$

Assim como no caso real [He et al. 2019], o número de dados pode ser insuficiente para garantir a convergência. Assim, é necessário utilizar a estratégia $\phi_i = \phi_{i+M}$ e $b_{i+M} = b_i$. O seguinte algoritmo resume o processo:

Algorithm 3 ℓ_0 -MCCC

Inicialização:

Escolher passo do gradiente η , largura de *kernel* σ , parâmetro de regularização λ , e região de atuação do atrator β . É usado o contador de iteração $i = 0$. O vetor esparsos é inicializado com zeros $\mathbf{w}_0 = \mathbf{0}$. É definida uma tolerância ε e um número máximo de iterações C .

Computação:

```

1: procedure  $\ell_0$ -MCC:
2:   while ( $i < C$ ) do
3:     % extrair o vetor entrada e o correspondente de saída da matriz de medidas A
     e do vetor de medidas ruidosas  $\mathbf{b}$ :
4:      $r = \text{mod}(i, M) + 1$ ;  $\phi_i = \mathbf{A}(r, \cdot)$ ;  $d_i = b_r$ 
5:     % Calculando a saída
6:      $y_i = \phi_i w_i$ 
7:     % Computando o erro
8:      $e_i = d_i - y_i$ 
9:     % Atualizando o w usando o MCCC
10:     $\mathbf{w}_{i+1} = \mathbf{w}_i + \eta \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H$ 
11:    % Atualizando o w usando a estratégia do atrator
12:     $k = 1$ ;  $\bar{\mathbf{w}} = \mathbf{w}_{i+1}$ ;
13:    for ( $k < L$ ) do
14:       $\bar{w}_k = Z_\beta(\bar{w}_k)$ 
15:       $k = k + 1$ 
16:    end for
17:     $\mathbf{w}_{i+1} = \mathbf{w}_{i+1} + \eta \lambda \bar{\mathbf{w}}$ 
18:    if  $\|\mathbf{w}_{i+1} - \mathbf{w}_i\|^2 < \varepsilon$  then
19:      Parar
20:    end if
21:    % Atualizando o contador
22:     $i = i + 1$ 
23:  end while
24: end procedure

```

O algoritmo acima generaliza tanto o ℓ_0 -LMS (fazendo o tamanho de *kernel* $\sigma \rightarrow \infty$) quanto o ℓ_0 -MCC para o caso complexo.

	Sem ruído	Ruído gaussiano	Ruído impulsivo
NESTA	59.21	17.07	7.25
ℓ_0 -LMS	34.39	23.47	9.51
ℓ_0 -MCCC	34.28	25.16	24.07

Tabela 4.1: Média de 10^3 SER computado para cada algoritmo em 3 diferentes cenários: Sem ruído, com ruído gaussiano e com ruído impulsivo.

Resultados

O SER é usado nessa seção para avaliar o desempenho do algoritmo ℓ_0 -MCCC. Três cenários são usados a princípio no comparativo: Sem ruído, com ruído gaussiano e com ruído impulsivo. A distribuição alfa estável foi usada para gerar tanto o ruído gaussiano quanto o ruído impulsivo. A tabela 4.1 mostra a média de 10^3 valores de SER para cada um desses cenários descritos. Todas as simulações foram feitas usando os seguintes parâmetros: $M = 512$ medidas, um vetor esparso complexo e aleatório \mathbf{x} com tamanho $L = 1024$ e $S = 32$ elementos não nulos. O passo do gradiente utilizado no NESTA foi $\mu = 0.001$, enquanto que o ℓ_0 -LMS e ℓ_0 -MCCC usaram os seguintes parâmetros: passo do gradiente $\mu = 0.0005$, $\lambda = 0.07$, $\beta = 0.85$. O tamanho do *kernel* usado pelo ℓ_0 -MCCC foi $\sigma = 1.83$, e a tolerância de $\varepsilon = 10^{-6}$. O número máximo de iterações escolhido foi $C = 2 \cdot 10^5$.

No cenário sem ruído todos os algoritmos testados foram capazes de reconstruir o vetor esparso. A precisão é um dos pontos fortes do NESTA [Becker et al. 2011], que foi o método que atingiu os maiores níveis de SER nessa configuração, como pode ser visto na Tabela 4.1. Como o ℓ_0 -LMS e ℓ_0 -MCCC usam a mesma estratégia de aproximação do gradiente da norma ℓ_0 , eles atingiram níveis similares de SER, estando limitados numericamente a relação $\beta^2 w \pm \beta$ que compõe a Equação (4.24), responsável pelo gradiente da norma ℓ_0 do vetor esparso.

O ruído gaussiano foi gerado usando a distribuição alfa estável com parâmetros GSNR=20 dB e $\alpha = 2$. A tabela 4.1 mostra que o desempenho do NESTA cai significativamente quando comparado aos outros dois algoritmos testados. A estratégia do atrator de zeros usada tanto pela aproximação do gradiente da norma ℓ_0 do ℓ_0 -LMS e ℓ_0 -MCCC tende a suavizar o efeito do ruído. Nesse cenário, o algoritmo proposto ℓ_0 -MCCC teve um desempenho equivalente ao algoritmo já definido na literatura ℓ_0 -LMS.

Para avaliar o desempenho dos algoritmos sob ruído impulsivo, o parâmetro α da distribuição alfa estável foi mudada de 2 para 1.5, mantendo o nível de GSNR em 20 dB. Essa configuração cria um ambiente de ruído impulsivo, aumentando a probabilidade da aparição de *outliers* nas medições. Um típico resultado do processo de reconstrução dos algoritmos testados pode ser visto na Figura 4.18. A linha superior mostra o resultado para o NESTA, enquanto que a linha do meio trata do ℓ_0 -LMS e, por último, a última linha é referente ao ℓ_0 -MCCC. Já a coluna do lado esquerdo da Figura 4.18 corresponde a parte real dos sinais, enquanto que a coluna do lado direito trata da parte imaginária dos sinais complexos. Como esperado, a presença de ruído impulsivo deteriora o desempenho dos algoritmos cuja restrições são baseadas em métodos de segunda ordem (NESTA e ℓ_0 -

LMS). Em contraste, o método proposto que usa a correntropia complexa na restrição enquanto função custo é capaz de reconstruir um sinal que mais se parece com o sinal original e, por isso, foi o método que atingiu o maior valor de SER nesse ambiente, como comprova a Tabela 4.1.

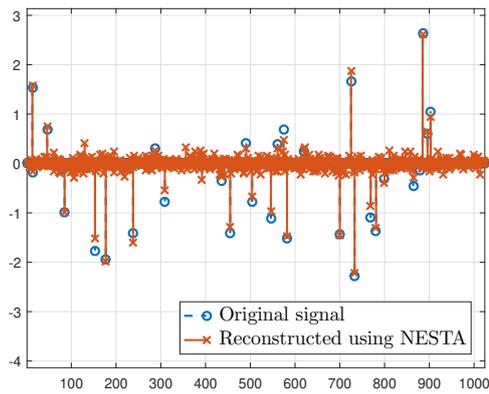
Deseja-se avaliar também o quão robusto é o algoritmo proposto ℓ_0 -MCCC a ruído impulsivo. Sendo assim, 10^3 simulações de Monte Carlo com GSNR no intervalo de $[10, 30]$ dB e um $\alpha = 1.5$ fixo foram realizadas. A média dessas simulações podem ser vistas na Figura 4.19. Os demais parâmetros da simulação são: Comprimento do sinal esparso $L = 1024$, dois níveis de esparsidade $S = \{16, 32\}$, e $M = 512$ medidas. Como pode ser visto na Figura 4.19, o ℓ_0 -MCCC foi o algoritmo que atingiu os maiores valores de SER nessas simulações. Como esperado, independente do algoritmo testado, aumentando o número de elementos não nulos S no vetor esparso implica num aumento da dificuldade da reconstrução.

Um parâmetro importante no problema de amostragem compressiva é o número de medidas realizado. Para testar a variação do método a esse parâmetro, simulações foram realizadas variando o número de medidas M de 64 a 512 no mesmo ambiente de ruído impulsivo com GSNR = 25 dB e $\alpha = 1.5$. A Figura 4.20 mostra a SER média de 10^3 simulações para cada algoritmo e dois níveis de esparsidade $S = \{16, 32\}$. Como pode ser visto na Figura 4.20, o NESTA mostra um bom desempenho principalmente quando menos medidas são usadas. Entretanto, aumentando o número de medidas fez com que o desempenho do método caísse de forma acentuada devido a presença do ruído impulsivo. Por outro lado, tanto o ℓ_0 -LMS quanto o ℓ_0 -MCCC tiveram seus desempenhos incrementados com o aumento do número de medidas. Entretanto, como o ℓ_0 -MCCC usa a correntropia complexa enquanto função custo, ela é capaz de extrair mais informação dos dados do que o ℓ_0 -LMS, justificando os valores de SER maiores.

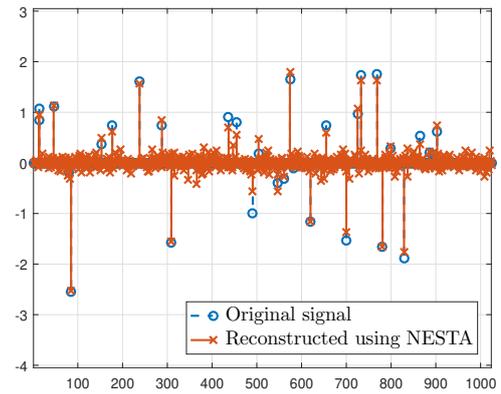
A Figura 4.21 mostra o desempenho do algoritmo ℓ_0 -MCCC em função tanto de α quanto do GSNR, ambos parâmetros de controle do ambiente de ruído impulsivo. Como esperado, o desempenho do algoritmo decai quando a potência do ruído aumenta e ele se torna mais impulsivo.

O tamanho do *kernel* é um parâmetro extremamente importante no desempenho do ℓ_0 -MCCC. Como foi dito, ele é um parâmetro livre e controla a convergência, desempenho em regime e robustez ao ambiente impulsivo. A Figura 4.22 mostra o SER médio do ℓ_0 -MCCC quando submetido a diferentes valores de GSNR e tamanhos de *kernel* σ . Como pode-se notar, o desempenho do algoritmo está intimamente relacionado a escolha do valor do *kernel*, inclusive para mesmos valores de GSNR. Ajustando corretamente o valor da largura do *kernel* faz com que o ℓ_0 -MCCC atinja valores de SER mais altos do que o ℓ_0 -LMS e o NESTA num ambiente de ruído impulsivo.

O desempenho do ℓ_0 -MCCC em função dos parâmetros λ e β é estudado na Figura 4.23. O resultado apresentado é uma média de 10^2 simulações. O objetivo é reconstruir um vetor esparso de tamanho $L = 1024$ e $S = 32$ elementos não nulos usando $M = 512$ medidas. Estas, contaminadas com ruído impulsivo modelado através de uma distribuição alfa estável com GSNR = 20 dB e $\alpha = 1.5$. O parâmetro β na Equação (4.24) controla a região de atração da estratégia de aproximação do gradiente da norma ℓ_0 , enquanto que o parâmetro λ controla a importância da aproximação na atualização do vetor esparso,



(a) NESTA parte real



(b) NESTA parte imaginária

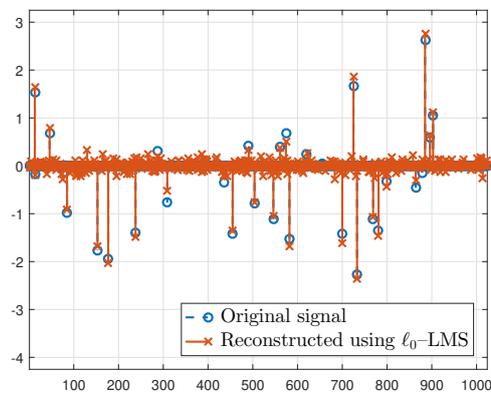
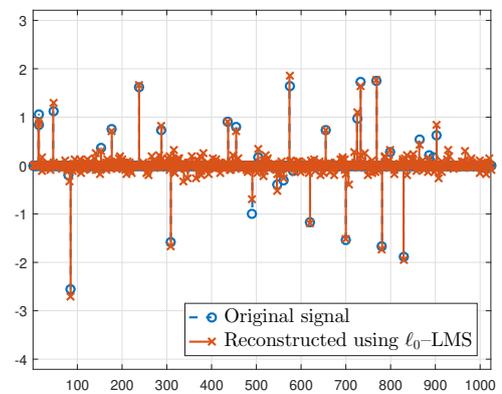
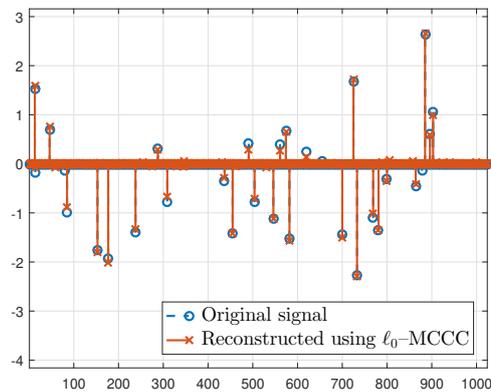
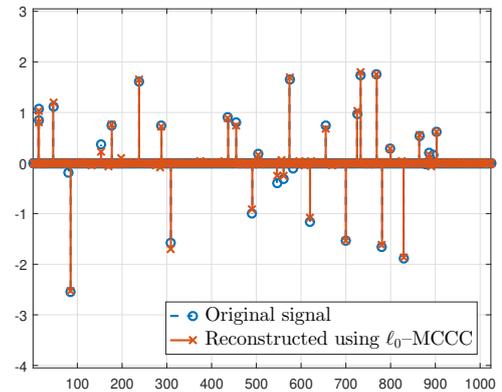
(c) ℓ_0 -LMS parte real(d) ℓ_0 -LMS parte imaginária(e) ℓ_0 -MCCC parte real(f) ℓ_0 -MCCC parte imaginária

Figura 4.18: Resultado típico da reconstrução de um vetor esparso de comprimento $L = 1024$ onde $S = 32$ elementos são não nulos. $M = 512$ medidas contaminadas com ruído impulsivo foram utilizadas ($\alpha = 1.5$, $\text{GNSR} = 20$ dB). (a-b) NESTA alcançou um SER de 7.74 dB; (c-d) ℓ_0 -LMS obteve um SER de 9.73 dB; (e-f) ℓ_0 -MCCC atingiu um SER de 27.70 dB. O ℓ_0 -MCCC, que usa a correntropia complexa como função custo, conseguiu reconstruir um vetor esparso muito mais próximo do vetor original do que os outros algoritmos testados ℓ_0 -LMS e NESTA num ambiente de ruído impulsivo.

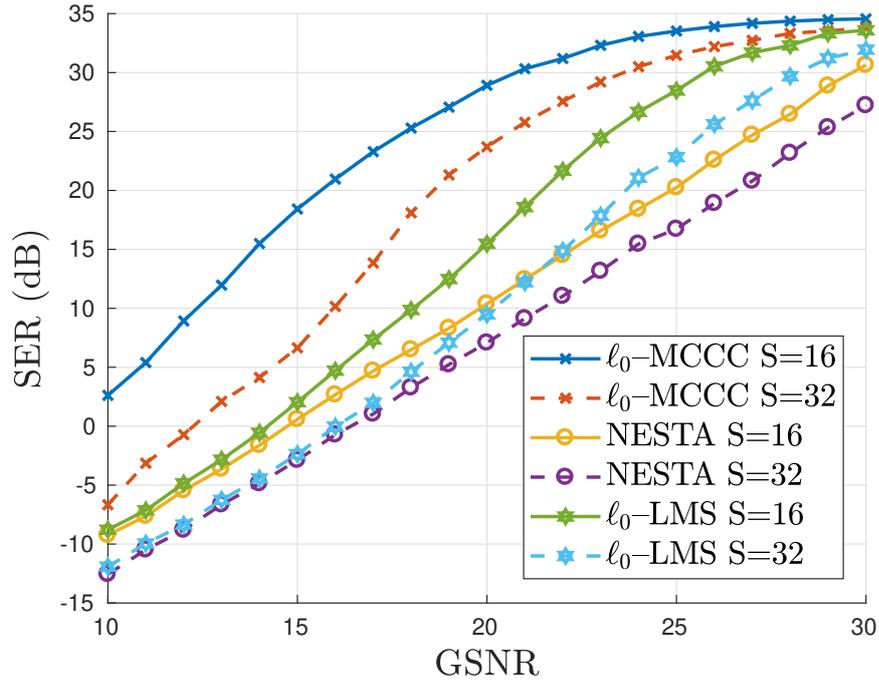


Figura 4.19: SER médio de 10^3 simulações mostrando o desempenho dos algoritmos na reconstrução de um vetor esparso complexo, sobre diferentes valores de GSNR e um $\alpha = 1.5$ fixo. S é o número de elementos não nulos no vetor esparso. Note que, quanto maior for o GSNR, menor a potência do ruído.

como mostra a Equação (4.25). Como pode ser visto na Figura 4.23, existe uma pequena região de λ e β que faz com que o ℓ_0 -MCCC atinja um bom desempenho, o que indica o quão sensível é o algoritmo a escolha desses parâmetros.

Em síntese, o método proposto ℓ_0 -MCCC têm os mesmos parâmetros livres do algoritmo da literatura que resolver o mesmo problema para o caso real, o ℓ_0 -MCC. São eles o tamanho do passo do gradiente μ ; o parâmetro de regularização λ ; a zona de atuação do atrator β ; e a largura de *kernel* σ . Como foi mostrado nessa seção, ao regular os parâmetros livres é possível fazer com que o ℓ_0 -MCCC consiga reconstruir com sucesso o vetor esparso em um ambiente de ruído impulsivo, de forma mais precisa do que os algoritmos clássicos da literatura NESTA e ℓ_0 -LMS.

4.6 Similaridade entre Ângulos

A correntropia convencional necessita de adaptações quando trata de similaridade de dados periódicos como em [Huijse et al. 2012]. Isso fica explícito no problema de similaridade entre ângulos pela não linearidade introduzida entre o primeiro e o último quadrante, implicando em que ângulos que estão muito próximos no círculo, como por exemplo $\zeta = 350^\circ$ e $\theta = 5^\circ$, possam aparentarem estar mais distantes por exemplo, do que os ângulos $\psi = 100^\circ$ e $\Omega = 200^\circ$, já que o erro entre eles é maior, isto é: $|350^\circ - 5^\circ| > |100^\circ - 200^\circ|$.

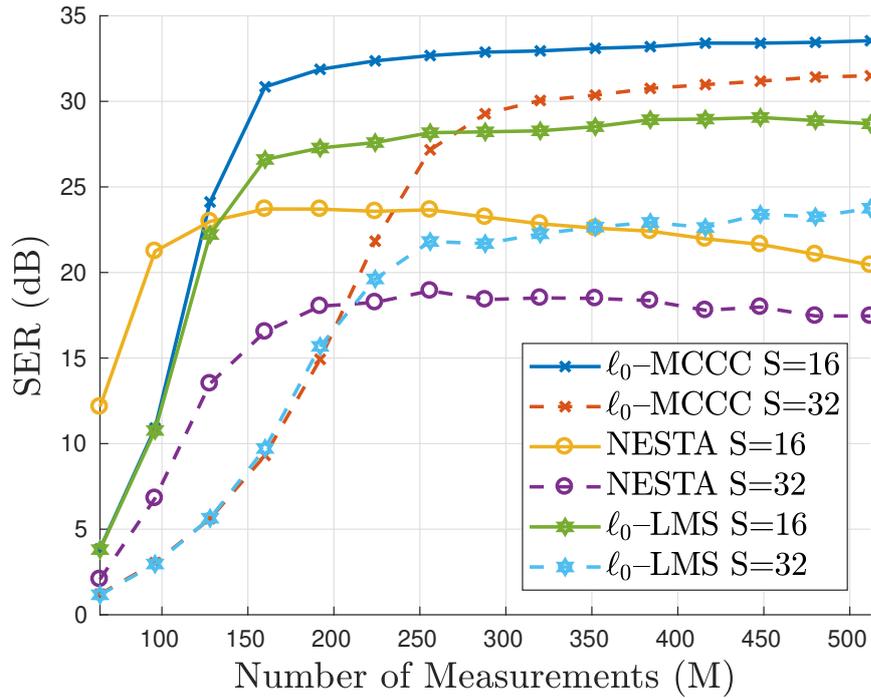


Figura 4.20: SER médio de 10^3 simulações mostrando o desempenho dos algoritmos testados sob múltiplos números de medidas realizadas M e parâmetros do ruído fixo em $\text{GSNR}=25$ dB e $\alpha = 1.5$.

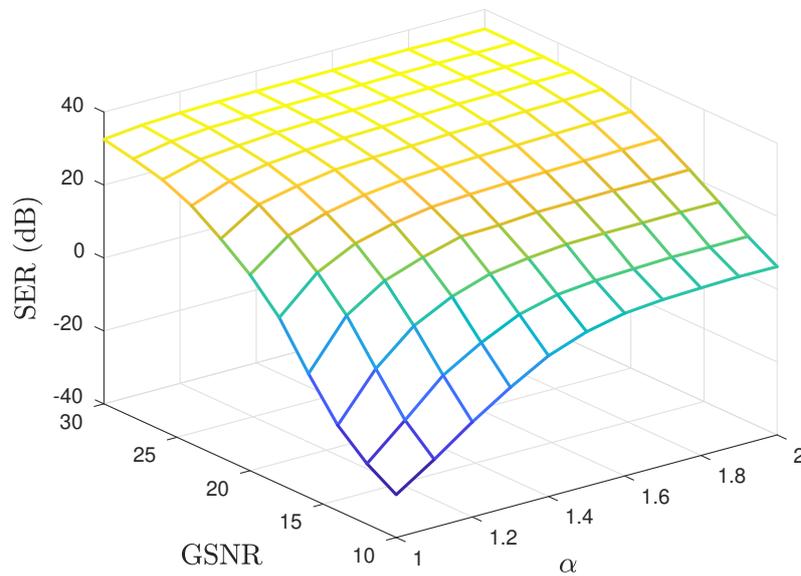


Figura 4.21: SER médio de 10^2 simulações do ℓ_0 -MCCC em função do GSNR e do índice de estabilidade, α , ambos parâmetros de controle da distribuição alfa estável que modela o ruído impulsivo.

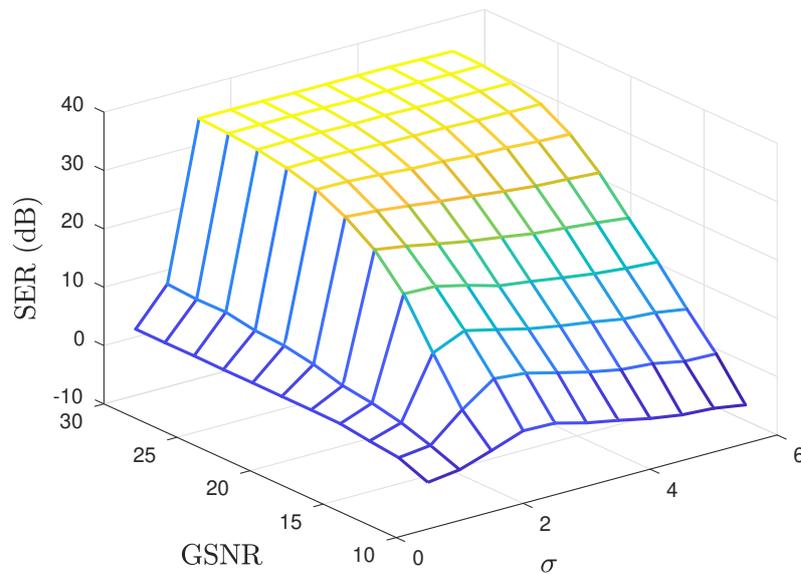


Figura 4.22: SER médio de 10^2 simulações do ℓ_0 -MCCC com $\alpha = 1.5$ e em função da potência do ruído GSNR e do tamanho do *kernel* σ .

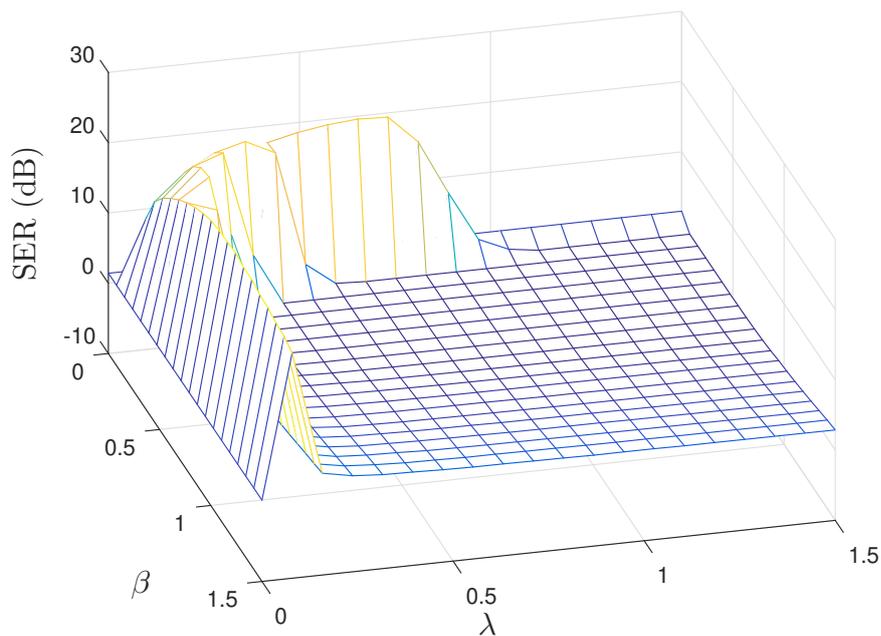


Figura 4.23: Desempenho do ℓ_0 -MCCC variando os parâmetros λ , que pondera a importância da aproximação do gradiente de ℓ_0 , e de β , que controla a zona de atuação do atrator. A Figura é uma média de 10^2 simulações de Monte Carlo.

É possível modelar dados cujo domínio são ângulos dentro do círculo unitário em um número complexo que tem fase igual ao ângulo e módulo unitário. Por exemplo, seja ϕ um

ângulo em radianos, é possível escreve-lo como o complexo equivalente $z = |1| \cdot \exp(j\phi)$. Dessa forma, a correntropia complexa abre a possibilidade da análise de similaridade entre ângulos. A Figura 4.24 mostra os ângulos $\theta = 5^\circ, \Psi = 100^\circ, \Omega = 200^\circ, \theta = 215^\circ$ e $\zeta = 350^\circ$ representados pelos complexos equivalentes $z_1 = (0.9962 + j0.0872), z_2 = (0.9848 - j0.1736), z_3 = (-0.1736 + j0.9848), z_4 = (-0.9397 - j0.3420)$ e $z_5 = (-0.8192 - j0.5736)$, respectivamente.

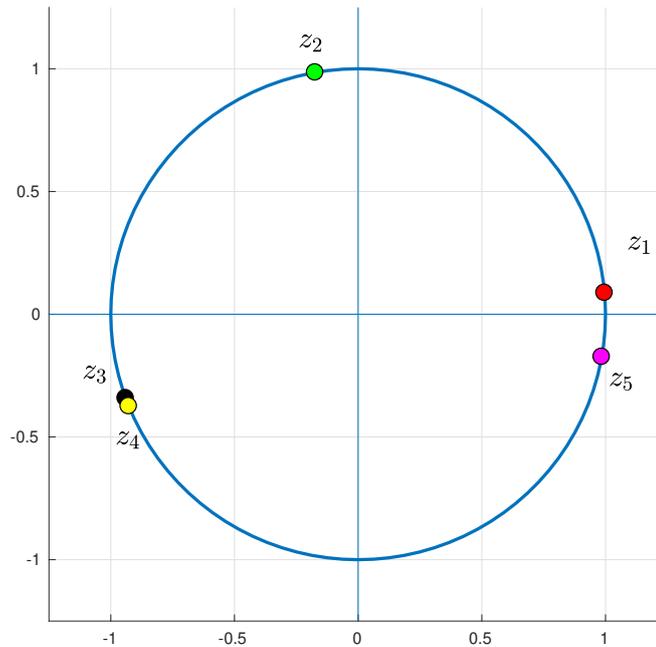


Figura 4.24: Número sobre círculo unitário

Criando vetores aleatórios com distribuição gaussiana com média em cada um dos ângulos e variância 5, como mostra a Figura 4.25, é possível construir o seguinte comparativo mostrado na Tabela 4.2:

$\hat{V}_\sigma(\cdot, \cdot)$	$\angle(\mathbf{z}_1)$	$\angle(\mathbf{z}_2)$	$\angle(\mathbf{z}_3)$	$\angle(\mathbf{z}_4)$	$\angle(\mathbf{z}_5)$
$\angle(\mathbf{z}_1)$	0.3989	0	0	0	0
$\angle(\mathbf{z}_2)$	0	0.3989	0	0	0
$\angle(\mathbf{z}_3)$	0	0	0.3989	0.1191	0
$\angle(\mathbf{z}_4)$	0	0	0.1191	0.3989	0
$\angle(\mathbf{z}_5)$	0	0	0	0	0.3989

Tabela 4.2: Correntropia entre os vetores de ângulos que seguem distribuição gaussiana com média é z_i , com $i = [1, 5]$ e variância 1.

É possível notar que a correntropia entre os vetores ângulo é sempre zero com duas exceções: quando o vetor é comparado com ele mesmo, onde a correntropia atinge o seu va-

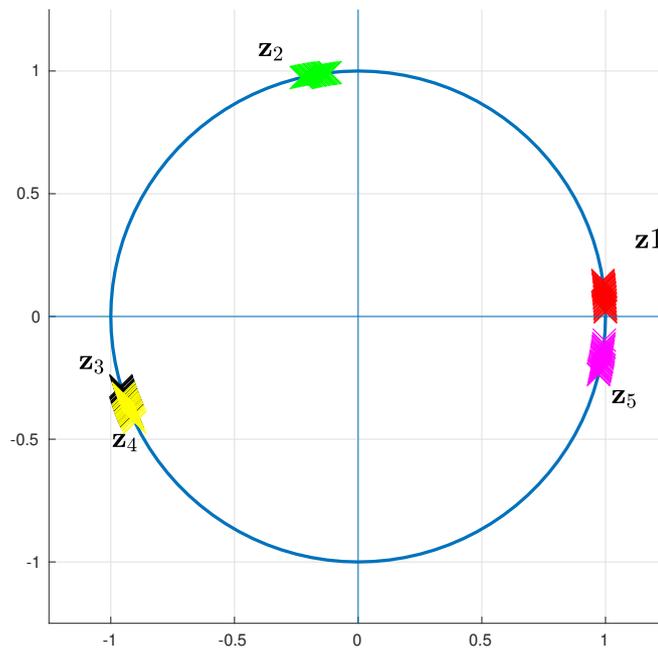


Figura 4.25: Ângulos dentro do círculo unitário

lor máximo $\hat{V}_{\sigma=1}(X, X) = 1/\sqrt{2\pi} = 0.3989$, e $\hat{V}_{\sigma=1}(\angle(\mathbf{z}_3), \angle(\mathbf{z}_3)) = \hat{V}_{\sigma=1}(\angle(\mathbf{z}_3))$, $\angle(\mathbf{z}_3) = 0.1191$, já que os vetores apresentam alguma similaridade dado a proximidade entre as médias (200° e 202°).

Fazendo a mesma análise para com os vetores complexos usando a correntropia complexa, obtêm-se:

Como no caso real, a diagonal principal da Tabela 4.3 é o valor de máximo da correntropia complexa $\hat{V}_{\sigma=1}^c(X, X) = 1/2\pi = 0.1592$. Entretanto, pode-se notar que a correntropia complexa $\hat{V}_{\sigma=1}^c(\mathbf{z}_1, \mathbf{z}_5)$ é próxima do máximo (0.1538), em vetores cujos os ângulos estão no primeiro e no último quadrante, respectivamente, além de ser próximo da máxima também entre \mathbf{z}_3 e \mathbf{z}_4 , mostrando-se assim uma ferramenta adequada para analisar similaridade entre ângulos.

$\hat{V}_{\sigma=1}^c(\cdot, \cdot)$	\mathbf{z}_1	\mathbf{z}_2	\mathbf{z}_3	\mathbf{z}_4	\mathbf{z}_5
\mathbf{z}_1	0.1592	0.0537	0.0223	0.0225	0.1538
\mathbf{z}_2	0.0537	0.1592	0.0493	0.0476	0.0416
\mathbf{z}_3	0.0223	0.0493	0.1592	0.1590	0.0246
\mathbf{z}_4	0.0225	0.0476	0.1590	0.1592	0.0251
\mathbf{z}_5	0.1538	0.0416	0.0246	0.0251	0.1592

Tabela 4.3: Correntropia complexa entre os vetores de números complexos equivalentes aos ângulos da Tabela 4.2.

Capítulo 5

Conclusões

Este trabalho propôs uma nova medida de similaridade entre variáveis aleatórias complexas chamada correntropia complexa. Ela foi desenvolvida baseado na interpretação probabilística que a correntropia convencional possui. Para isso, a correntropia convencional foi estudada e sua teoria foi revisitada para se derivar e definir essa nova medida de similaridade que pode lidar com dados complexos. Além disso, uma breve recapitulação a cerca dos números complexos e do espaço de Hilbert foram apresentados.

Foram estudadas e expostas propriedades da correntropia complexa, mostrando que essa nova medida consegue herdar as características que tornaram a correntropia uma medida de similaridade em franca ascensão com aplicações nos mais diversos campos da engenharia.

O critério de máxima correntropia complexa (MCCC) foi definido, habilitando o uso da correntropia complexa como função custo em problemas de otimização. Uma solução baseada no gradiente ascendente e em ponto fixo são apresentadas, ambas baseadas no cálculo de Wirtinger. Ainda em relação ao MCCC, três estratégias de *kernel* adaptativo foram definidas: AMCCC, SMCCC e VKW-MCCC.

O MCCC foi aplicado a um problema de identificação de sistemas lineares e a um problema de equalização de canal, tendo superado métodos tradicionais como o CLMS, CRLS e o LAD em ambientes de ruído não-gaussiano. As estratégias de *kernel* adaptativo foram confrontadas com a escolha de um valor de *kernel* fixo, mostrando-se estratégias válidas porém não eliminando completamente a dependência de um parâmetro livre.

Foi definida a métrica induzida da correntropia complexa. Ela foi aplicada ao problema de amostragem compressiva mostrando-se mais eficiente do que um algoritmo tradicional da literatura NESTA. Ainda no problema de amostragem compressiva, o MCCC foi utilizado juntamente a um estratégia de aproximação do gradiente da norma ℓ_0 para produzir um novo algoritmo denominado ℓ_0 -MCCC. Este algoritmo foi capaz de superar tanto o NESTA quanto o ℓ_0 -LMS na reconstrução de vetores esparsos complexos em ambientes de ruído impulsivo.

Além disso, mostrou-se como a correntropia complexa pode ser aplicada ao problema de similaridade entre ângulos.

5.1 Lista de publicações

Essa seção enumera as publicações referentes a esta tese e provenientes de colaborações durante o período de doutoramento.

5.1.1 Publicações relacionadas a esta tese

1. (2016) **Guimarães, João P. F.**; Fontes, Aluísio I. R.; Rêgo, Joilson B. A.; Silveira, Luiz F. Q.; Martins, Allan M.; 'Performance evaluation of the maximum correntropy criterion in identification systems'. *IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*
2. (2016) **Guimarães, João P. F.**; Fontes, Aluísio I. R.; Rêgo, Joilson B. A.; Martins, Allan M.; Príncipe, José C.; 'Complex Correntropy: Probabilistic Interpretation and Application to Complex-Valued Data'. *IEEE Signal Processing Letters*
3. (2018) **Guimarães, João P. F.**; Fontes, Aluísio I. R.; Rêgo, Joilson B. A.; Martins, Allan M.; Príncipe, José C.; 'Complex correntropy function: Properties, and application to a channel equalization problem'. *Expert Systems With Applications*
4. (2018) **Guimarães, João P. F.**; Fontes, Aluísio I. R.; Silva, F. ; Martins, Allan M.; von Borries, R.; 'Complex Correntropy Induced Metric Applied to Compressed Sensing with Complex-Valued Data' *IEEE Southwest Symposium on Image Analysis and Interpretation*
5. (2019) Aquino, Manoel B. L.; **Guimarães, João P. F.**; Linhares, Leandro L. S. ; Fontes, Aluísio I. R.; Silva, F. ; Martins, Allan M.; 'Performance evaluation of the maximum complex correntropy criterion with adaptive kernel width update' *EURASIP Journal on Advances in Signal Processing*
6. (2019) **Guimarães, João P. F.**; da Silva, Felipe B.; Fontes, Aluísio I. R.; Silva, F. ; von Borries, R.; Martins, Allan M; 'Complex Correntropy Applied to a Compressive Sensing Problem in an Impulsive Noise Environment' *IEEE Access*

5.1.2 Publicações provenientes de colaborações

1. (2019) Araújo, Ícaro B. Q. ; **Guimarães, João P. F.**; Fontes, Aluísio I. R. ; Linhares, Leandro L. S. ; Martins, Allan M. ; Araújo, Fábio M. U.; 'Narx Model Identification Using Correntropy Criterion in the Presence of Non-Gaussian Noise' *Journal Of Control, Automation and Electrical Systems*

Referências Bibliográficas

- Adali, T. e S. Haykin (2010), *Adaptive Signal Processing: Next Generation Solutions*, Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control, Wiley.
URL: <https://books.google.com.br/books?id=BFyPDAEACAAJ>
- Ahlfors, L.V. (1979), *Complex analysis: an introduction to the theory of analytic functions of one complex variable*, International series in pure and applied mathematics, McGraw-Hill.
URL: <https://books.google.com.br/books?id=2MRuus-5GGoC>
- Akcakaya, M. e V. Tarokh (2010), ‘Shannon-theoretic limits on noisy compressive sampling’, *IEEE Transactions on Information Theory* **56**(1), 492–504.
- Al-Naffouri, Tareq Y. e Ali H. Sayed (2001), ‘Adaptive filters with error nonlinearities: Mean-square analysis and optimum design’, *EURASIP J. Appl. Signal Process.* **2001**(1), 192–205.
URL: <http://dx.doi.org/10.1155/S11110865701000348>
- Alliney, S. e S. A. Ruzinsky (1994), ‘An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation’, *IEEE Transactions on Signal Processing* **42**(3), 618–627.
- Aronszajn, N. (1950), ‘Theory of reproducing kernels’, *Transactions of the American Mathematical Society* **68**(3), 337–404.
URL: <http://dx.doi.org/10.2307/1990404>
- Azam Khalili, Amir Rastegarnia, Bahareh Darvish (2015), ‘A robust adaptive carrier frequency offset estimation algorithm for ofdm’, *American Journal of Signal Processing* **5**(2A), 26–31.
- Baraniuk, R. G. (2007), ‘Compressive sensing [lecture notes]’, *IEEE Signal Processing Magazine* **24**(4), 118–121.
- Becker, S., J. Bobin e E. Candès (2011), ‘Nesta: A fast and accurate first-order method for sparse recovery’, *SIAM Journal on Imaging Sciences* **4**(1), 1–39.
URL: <https://doi.org/10.1137/090756855>
- Bessa, R. J., V. Miranda e J. Gama (2009), ‘Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting’, *IEEE Transactions on Power Systems* **24**(4), 1657–1666.

- Bettens, S., C. Schretter, N. Deligiannis e P. Schelkens (2017), 'Bounds and conditions for compressive digital holography using wavelet sparsifying bases', *IEEE Transactions on Computational Imaging* **3**(4), 592–604.
- Boloix-Tortosa, R., F. J. Payán-Somet, E. Arias de Reyna e J. J. Murillo-Fuentes (2015), Complex kernels for proper complex-valued signals: A review, em 'Signal Processing Conference (EUSIPCO), 2015 23rd European', pp. 2371–2375.
- Bouboulis, P. e S. Theodoridis (2011), 'Extension of wirtinger's calculus to reproducing kernel hilbert spaces and the complex kernel lms', *IEEE Transactions on Signal Processing* **59**(3), 964–978.
- Candes, E. J., J. Romberg e T. Tao (2006), 'Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information', *IEEE Transactions on Information Theory* **52**(2), 489–509.
- Candes, E. J. e M. B. Wakin (2008), 'An introduction to compressive sampling', *IEEE Signal Processing Magazine* **25**(2), 21–30.
- Candès, Emmanuel J. (2008), 'The restricted isometry property and its implications for compressed sensing', *Comptes Rendus Mathématique* **346**(9), 589 – 592.
URL: <http://www.sciencedirect.com/science/article/pii/S1631073X08000964>
- Cao, Z., S. Yu, B. Ouyang, F. Dalglish, A. Vuorenkoski, G. Alsenas e J. C. Principe (2018), 'Marine animal classification with correntropy-loss-based multiview learning', *IEEE Journal of Oceanic Engineering* pp. 1–14.
- Chen, B., L. Xing, B. Xu, H. Zhao, N. Zheng e J. C. Príncipe (2017), 'Kernel risk-sensitive loss: Definition, properties and application to robust adaptive filtering', *IEEE Trans. Signal Process.* **65**(11), 2888–2901.
- Chen, B., L. Xing, J. Liang, N. Zheng e J. C. Príncipe (2014), 'Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion', *IEEE Signal Processing Letters* **21**(7), 880–884.
- Dang, L., B. Chen, S. Wang, Y. Gu e J. C. Príncipe (2019), 'Kernel kalman filtering with conditional embedding and maximum correntropy criterion', *IEEE Transactions on Circuits and Systems I: Regular Papers* pp. 1–13.
- Diniz, P.S.R. (2008), *Adaptive Filtering: Algorithms and Practical Implementation*, Kluwer international series in engineering and computer science, Springer.
- Erdogmus, D. (2002), Information theoretic learning: Renyi's entropy and its application to adaptive system training, Tese de doutorado, Univ. of Florida.
- Fontes, Aluisio I. R., Pedro T. V. Souza, Adriaio D. D. Neto, Allan de M. Martins e Luiz F. Q. Silveira (2014), 'Classification system of pathological voices using correntropy', *Mathematical Problems in Engineering* **2014**(7).

- Hao, Z., X. Xie e Q. Zhang (2015), Robust k-means based active contours for fast inhomogeneity image segmentation, *em* '2015 8th International Congress on Image and Signal Processing (CISP)', pp. 487–492.
- Haykin, S.S. (2002), *Adaptive Filter Theory*, Prentice-Hall information and system sciences series, Prentice Hall.
- He, R., B. Hu, W. Zheng e X. Kong (2011), 'Robust principal component analysis based on maximum correntropy criterion', *IEEE Transactions on Image Processing* **20**(6), 1485–1494.
- He, R., W. Zheng e B. Hu (2011), 'Maximum correntropy criterion for robust face recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(8), 1561–1576.
- He, R., W. Zheng, B. Hu e X. Kong (2011), 'A regularized correntropy framework for robust pattern recognition', *Neural Computation* **23**(8), 2074–2100.
- He, Yicong, Fei Wang, Shiyuan Wang, Jiuwen Cao e Badong Chen (2019), 'Maximum correntropy adaptation approach for robust compressive sensing reconstruction', *Information Sciences* **480**, 381 – 402.
URL: <http://www.sciencedirect.com/science/article/pii/S0020025518309940>
- Horie, R. e E. Aiyoshi (1999), Variable metric gradient projection method and replicator equation, *em* 'Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on', Vol. 3, pp. 515–520 vol.3.
- Huang, F., J. Zhang e S. Zhang (2017), 'Adaptive filtering under a variable kernel width maximum correntropy criterion', *IEEE Trans. Circuits Syst. II, Exp. Briefs* **64**(10), 1247–1251.
- Huijse, P., P. A. Estevez, P. Protopapas, P. Zegers e J. C. Principe (2012), 'An information theoretic algorithm for finding periodicities in stellar light curves', *IEEE Transactions on Signal Processing* **60**(10), 5135–5145.
- Jin, J., Y. Gu e S. Mei (2010), 'A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework', *IEEE Journal of Selected Topics in Signal Processing* **4**(2), 409–420.
- Kim, Nam-Yong (2010), 'Complex-channel blind equalization using cross-correntropy', *Journal of Internet Computing and Services* **11**(5), 19–26.
- Li, R., J. C. Principe, M. Bradley e V. Ferrari (2007), Robust single-trial erp estimation based on spatiotemporal filtering, *em* '2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 5206–5209.
- Lin, F. (2017), Robust compressive reconstruction algorithm for eliminating sampling and measurement impulsive noise, *em* '2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)', Vol. 2, pp. 89–92.

- Liu, W., P. P. Pokharel e J. C. Principe (2007), ‘Correntropy: Properties and applications in non-gaussian signal processing’, *IEEE Transactions on Signal Processing* **55**(11), 5286–5298.
- Liu, Weifeng, P. P. Pokharel e J. C. Principe (2006), Correntropy: A localized similarity measure, *em* ‘The 2006 IEEE International Joint Conference on Neural Network Proceedings’, pp. 4919–4924.
- Mandic, D.P. e V.S.L. Goh (2009), *Complex Valued Nonlinear Adaptive Filters: Non-circularity, Widely Linear and Neural Models*, Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control, Wiley.
- Mathews, J. e R. Howell (2012), *Complex Analysis for Mathematics and Engineering*, International series in mathematics, Jones & Bartlett Learning.
URL: <https://books.google.com.br/books?id=KaYuyyIqnMC>
- Meng, X. e J. Zhu (2019), ‘Bilinear adaptive generalized vector approximate message passing’, *IEEE Access* **7**, 4807–4815.
- Meng, X., S. Wu e J. Zhu (2018), ‘A unified bayesian inference framework for generalized linear models’, *IEEE Signal Processing Letters* **25**(3), 398–402.
- Mercer, James e Andrew Russell Forsyth (1909), ‘Xvi. functions of positive and negative type, and their connection the theory of integral equations’, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **209**(441-458), 415–446.
URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1909.0016>
- Needham, T. (1997), *Visual Complex Analysis*, Clarendon Press.
URL: <https://books.google.com.br/books?id=ay3vAAAAMAAJ>
- Nikias, Chrysostomos L. e Min Shao (1995), *Signal Processing with Alpha-stable Distributions and Applications*, Wiley-Interscience, New York, NY, USA.
- Ogunfunmi, T. e T. Paul (2015), ‘The quaternion maximum correntropy algorithm’, *IEEE Transactions on Circuits and Systems II: Express Briefs* **62**(6), 598–602.
- Papoulis, A. (1984), *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in Electrical Engineering, McGraw-Hill.
URL: <https://books.google.com.br/books?id=3NRQAAAAMAAJ>
- Parzen, Emanuel (1962), ‘On estimation of a probability density function and mode’, *The Annals of Mathematical Statistics* **33**(3), pp. 1065–1076.
URL: <http://www.jstor.org/stable/2237880>
- Principe, J.C. (2010), *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*, Information Science and Statistics, Springer New York.
URL: <https://books.google.com.br/books?id=oJSkBXWctsgC>

- Rudin, W. (1987), *Real and complex analysis*, Mathematics series, McGraw-Hill.
URL: https://books.google.com.br/books?id=Z_fuAAAAMAAJ
- Rêgo, Joilson Batista de Almeida (2014), Extensões multidimensionais para correntropia e suas aplicações em estimativas robustas, Tese de doutorado, UFRN.
- Samorodnitsky, G. e M.S. Taqqu (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Taylor & Francis.
URL: <https://books.google.com.br/books?id=wTTUfYwjksAC>
- Santamaria, I., P.P. Pokharel e J.C. Principe (2006), ‘Generalized correlation function: definition, properties, and application to blind equalization’, *IEEE Transactions on Signal Processing* **54**(6), 2187–2197.
- Schreier, P.J. e L.L. Scharf (2010), *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*, Cambridge University Press.
URL: <https://books.google.com.br/books?id=HBaxLfDsAHoC>
- Seth, S. e J. C. Principe (2008), Compressed signal reconstruction using the correntropy induced metric, *em* ‘2008 IEEE International Conference on Acoustics, Speech and Signal Processing’, pp. 3845–3848.
- Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.
- Singh, Abhishek e Jose C Principe (2010), A closed form recursive solution for maximum correntropy training, *em* ‘Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on’, IEEE, pp. 2070–2073.
- Smith, David Eugene (1925), *History of Mathematics Vol II*, Natural Sciences, Mathematics, Fundamental and general consideration of mathematics, Ginn and Company.
- Sun, H., F. Yang, F. Meng, Z. Zhang, C. Gao e M. Liu (2018), ‘A topographic kirchhoff dynamic focused beam migration method based on compressed sensing’, *IEEE Access* **6**, 56666–56674.
- Sung, C. K., F. de Hoog, Z. Chen, P. Cheng e D. C. Popescu (2017), ‘Interference mitigation based on bayesian compressive sensing for wireless localization systems in unlicensed band’, *IEEE Transactions on Vehicular Technology* **66**(8), 7038–7049.
- Tive, F. H. C. e A. Bouzerdoum (2013), A compressed sensing method for complex-valued signals with application to through-the-wall radar imaging, *em* ‘2013 IEEE International Conference on Acoustics, Speech and Signal Processing’, pp. 2144–2148.
- Vetterli, M., J. Kovačević e V.K. Goyal (2014), *Foundations of Signal Processing*, Cambridge University Press.
URL: <https://books.google.com.br/books?id=5WA5nQAACAAJ>

- Wang, W., J. Zhao, H. Qu, B. Chen e J. C. Principe (2015), A switch kernel width method of correntropy for channel estimation, *em* 'IEEE Int. Joint Conf. Neural Netw. (IJCNN)', pp. 1–7.
- Wang, W., J. Zhao, H. Qu, B. Chen e J. C. Principe (2017), 'Convergence performance analysis of an adaptive kernel width mcc algorithm', *AEU - International Journal of Electronics and Communications* **76**, 71–76.
- Wirtinger, W. (1927), 'Zur formalen theorie der funktionen von mehr komplexen veränderlichen', *Mathematische Annalen* **97**, 357–376.
URL: <http://eudml.org/doc/182642>
- Zhang, C., Y. Guo, F. Wang e B. Chen (2018), Generalized maximum correntropy-based echo state network for robust nonlinear system identification, *em* '2018 International Joint Conference on Neural Networks (IJCNN)', pp. 1–6.
- Zhao, S., B. Chen e J. C. Principe (2012), An adaptive kernel width update for correntropy, *em* 'The 2012 International Joint Conference on Neural Networks (IJCNN)', pp. 1–5.
- Zhu, J., L. Han e X. Meng (2019), 'An amp-based low complexity generalized sparse bayesian learning algorithm', *IEEE Access* **7**, 7965–7976.
- Zhu, Junan e D. Baron (2013), Performance regions in compressed sensing from noisy measurements, *em* '2013 47th Annual Conference on Information Sciences and Systems (CISS)', pp. 1–6.
- Zhu, X. X. e R. Bamler (2011), A fundamental bound for super-resolution — with application to 3d sar imaging, *em* '2011 Joint Urban Remote Sensing Event', pp. 181–184.
- Zhu, X. X. e R. Bamler (2012), 'Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic sar', *IEEE Transactions on Geoscience and Remote Sensing* **50**(1), 247–258.

Apêndice A

Correntropia como integral de f_{XY} na reta $x = y$

Estima-se a Correntropia a partir dos dados $\{(x_i, y_i)\}_{i=1}^N$ como

$$\hat{V}_{\sigma, N}(X, Y) = \frac{1}{N} \sum_{i=1}^N \mathbf{G}_{\sigma}(x_i - y_i), \quad (\text{A.1})$$

onde

$$\mathbf{G}_{\sigma}(x - y) = \frac{1}{\sqrt{2\pi\sigma}} \left(-\frac{(x - y)^2}{2\sigma^2} \right). \quad (\text{A.2})$$

Deseja-se mostrar que:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x, y, \theta) du \Big|_{x=y=u} = \frac{1}{N} \sum_{i=1}^N \mathbf{G}_{\theta\sqrt{2}}(x_i - y_i) \quad (\text{A.3})$$

Prova:

Seja o estimador de Parzen L-dimensional:

$$\hat{f}(x_1, x_2, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)^{\frac{L}{2}}} \exp\left(-\frac{1}{2\theta^2} \left(\sum_{j=1}^L (x_j - x_j^i)^2\right)\right) \quad (\text{A.4})$$

Uma outra forma de escrever esse estimador é:

$$\hat{f}_{X_1, X_2, \dots, X_L}(x^1, x^2, \dots, x^L) = \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^L G_{\sigma}(x^l - x_n^l) \quad (\text{A.5})$$

No caso matricial, para $L = 2$, temos o estimador de Parzen bi-dimensional como:

$$\hat{f}_{XY}(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)^{\frac{2}{2}}} \exp\left(-\frac{1}{2} [x - x_i, y - y_i] \begin{bmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{1}{\theta^2} \end{bmatrix} \begin{bmatrix} x - x_i \\ y - y_i \end{bmatrix}\right) \quad (\text{A.6})$$

Assim:

$$\hat{f}_{XY}(x,y) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)} \exp\left(-\frac{1}{2\theta^2} [(x-x_i)^2 + (y-y_i)^2]\right) \quad (\text{A.7})$$

Deseja-se analisar a probabilidade de que $x = y = u$. Temos que escrever $\hat{f}_{XY}(x,y)$ da seguinte forma:

$$\hat{f}_{XY}(x,y) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)} \exp\left(-\frac{1}{2\theta^2} [(u-x_i)^2 + (u-y_i)^2]\right) \quad (\text{A.8})$$

E em seguida resolver a integral:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x,y) du \Big|_{x=y=u} \quad (\text{A.9})$$

Para resolver essa integral é necessário reescreve-la da seguinte forma:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x,y) du \Big|_{x=y=u} = \int \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)} \exp\left(-\frac{1}{2\theta^2} (a(u-b)^2 + c)\right) du \quad (\text{A.10})$$

$$= \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)} \exp\left(-\frac{1}{2\theta^2} a(u-b)^2\right) \exp\left(-\frac{c}{2\theta^2}\right) du \quad (\text{A.11})$$

Os coeficientes a, b e c são obtidos a partir da técnica de completar quadrados:

$$(u-x_i)^2 + (u-y_i)^2 = a(u-b)^2 + c \quad (\text{A.12})$$

$$2u^2 + x_i^2 + y_i^2 - 2u(x_i + y_i) = au^2 - 2uab + ab^2 + c \quad (\text{A.13})$$

Assim temos que a, b e c valem:

$$a = 2 \quad (\text{A.14})$$

$$b = \frac{(x_i + y_i)}{2} \quad (\text{A.15})$$

$$c = \frac{(x_i - y_i)^2}{2} \quad (\text{A.16})$$

Agora, voltando a integral A.11, temos:

$$= \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi\theta^2)} \exp\left(-\frac{2(u-b)^2}{2\theta^2}\right) \exp\left(-\frac{(x_i-y_i)^2}{2\theta^2}\right) du \quad (\text{A.17})$$

$$= \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}\theta} \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{(u-b)^2}{\theta^2}\right) \exp\left(-\frac{(x_i-y_i)^2}{4\theta^2}\right) du \quad (\text{A.18})$$

Sabe-se que, para um x, μ , e σ genérico temos:

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\theta^2}(-b^2)\right) = 1 \quad (\text{A.19})$$

Fazendo $\theta^2 = 2\sigma^2$, temos $\theta = \sigma\sqrt{2}$, substituindo na equação encontramos:

$$= \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{\sqrt{2\pi\sigma\sqrt{2}}} \right) \left(\frac{1}{\sqrt{2\pi\sigma\sqrt{2}}} \right) \exp\left(-\frac{(u-b)^2}{2\sigma^2}\right) \exp\left(-\frac{(x_i-y_i)^2}{8\sigma^2}\right) du \quad (\text{A.20})$$

Sabendo que a segunda exponencial é independente de u , rearranjamos a equação retirando também as constantes de dentro da integral:

$$= \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-y_i)^2}{8\sigma^2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(u-b)^2}{2\sigma^2}\right) du \quad (\text{A.21})$$

Como mostra a equação A.19, a integral agora vale 1. Assim:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x, y, \sigma) du \Big|_{x=y=u} = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-y_i)^2}{8\sigma^2}\right) \quad (\text{A.22})$$

Pode-se manipular mais uma vez a largura a fim de condicionar a integral. Fazendo $8\sigma^2 = 2\theta'$ ou $2\sigma = \theta'$:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x, y, \theta') du \Big|_{x=y=u} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi\theta'}} \exp\left(-\frac{(x_i-y_i)^2}{2(\theta')^2}\right) \quad (\text{A.23})$$

Onde a θ' se relaciona com a largura original θ pela seguinte relação:

$$\theta' = \theta\sqrt{2} \quad (\text{A.24})$$

Portanto:

$$\int_{-\infty}^{\infty} \hat{f}_{XY}(x, y, \theta) du \Big|_{x=y=u} = \frac{1}{N} \sum_{i=1}^N \mathbf{G}_{\theta\sqrt{2}}(x_i - y_i) \quad (\text{A.25})$$

■

Apêndice B

Análise da Expansão em Série de Taylor da Correntropia com *kernel* Gaussiano

A correntropia é definida como

$$V_{\sigma}(X, Y) = E[K_{\sigma}(X, Y)] \quad (\text{B.1})$$

onde $K_{\sigma}(\cdot)$ é qualquer *kernel* positivo definido com tamanho de *kernel* σ , $E[\cdot]$ é o operador valor esperado e $X, Y \in \mathbb{R}$ duas variáveis aleatórias reais.

O *kernel* mais utilizado é o gaussiano, definido como:

$$K_{\sigma}(x, y) = G_{\sigma}(x - y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right) \quad (\text{B.2})$$

onde σ é o tamanho do *kernel*.

Usando o *kernel* gaussiano da Equação (B.2) na correntropia e fazendo a expansão em série de Taylor, pode-se obter

$$V_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m \sigma^{2m} m!} E[(X - Y)^{2m}] \quad (\text{B.3})$$

Prova:

Em primeiro lugar, assumindo a notação

$$\frac{\partial G_{\sigma}^{[k]}(\delta)}{\partial \delta^{[k]}}$$

é a k -ésima derivada de $G(\delta)$ em relação a δ .

Assim, fazendo $e = X - Y$, é possível dizer que as derivadas do *kernel* gaussiano são:

$$G_{\sigma}^{[1]}(e) = \frac{1}{\sqrt{2\pi\sigma}} \left(-\frac{e}{\sigma^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) \quad (\text{B.4})$$

$$G_{\sigma}^{[2]}(e) = \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{e^2}{\sigma^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{1}{\sigma^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) \quad (\text{B.5})$$

$$G_{\sigma}^{[3]}(e) = \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{e}{(\sigma^2)^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{e^3}{(\sigma^2)^3} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) \quad (\text{B.6})$$

$$G_{\sigma}^{[4]}(e) = \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{3}{(\sigma^2)^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{e^2}{(\sigma^2)^3} \exp\left(-\frac{e^2}{2\sigma^2}\right) + \frac{e^4}{(\sigma^2)^4} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) \quad (\text{B.7})$$

$$G_{\sigma}^{[5]}(e) = \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{10e^3}{(\sigma^2)^4} \exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{e^5}{(\sigma^2)^5} \exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{e}{(\sigma^2)^3} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) \quad (\text{B.8})$$

e assim por diante.

Avaliando o valor das derivadas no ponto $e = 0$, temos:

$$V_{\sigma}(X, Y) = E \left[\frac{G(0)(e-0)^0}{0!} + \frac{G'(0)(e-0)^1}{1!} + \frac{G''(0)(e-0)^2}{2!} + \frac{G'''(0)(e-0)^3}{3!} + \dots \right] \quad (\text{B.9})$$

Avaliando as funções e suas respectivas derivadas quando $e = 0$:

$$G_{\sigma}(e=0) = \frac{1}{\sqrt{2\pi\sigma}} \quad (\text{B.10})$$

$$G_{\sigma}^{[1]}(e=0) = 0 \quad (\text{B.11})$$

$$G_{\sigma}^{[2]}(e=0) = \frac{1}{\sqrt{2\pi\sigma}} \left(-\frac{1}{\sigma^2} \right) \quad (\text{B.12})$$

$$G_{\sigma}^{[3]}(e=0) = 0 \quad (\text{B.13})$$

$$G_{\sigma}^{[4]}(e=0) = \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{3}{(\sigma^2)^3} \right) \quad (\text{B.14})$$

e assim por diante.

Como a função G_{σ} e suas respectivas derivadas tem em comum a constante $\frac{1}{\sqrt{2\pi\sigma}}$, é possível escrever:

$$V_{\sigma}(e) = E \left[\frac{1}{\sqrt{2\pi\sigma}} \left(1 + 0 - \frac{1e^2}{2!(\sigma^2)^1} + 0 + \frac{3e^4}{4!(\sigma^2)^2} + 0 - \frac{15e^6}{6!(\sigma^2)^3} + 0 + \frac{105e^8}{8!(\sigma^2)^4} + 0 - \frac{945e^{10}}{10!(\sigma^2)^5} + \dots \right) \right] \quad (\text{B.15})$$

Sabendo que $E[aX] = aE[X]$, pode-se escrever a Equação (B.15) como:

$$V_{\sigma}(e) = \frac{1}{\sqrt{2\pi\sigma}} E \left[1 + 0 - \frac{1e^2}{2!(\sigma^2)^1} + 0 + \frac{3e^4}{4!(\sigma^2)^2} + 0 - \frac{15e^6}{6!(\sigma^2)^3} + 0 + \frac{105e^8}{8!(\sigma^2)^4} + 0 - \frac{945e^{10}}{10!(\sigma^2)^5} + \dots \right] \quad (\text{B.16})$$

O denominador da sequência dentro do operador valor esperado pode ser reescrito como:

$$1, 2!(\sigma^2)^1, 4!(\sigma^2)^2, 6!(\sigma^2)^3, \dots = (2m)!(\sigma^{2m}), \text{ for } m = 0, 1, 2, \dots \quad (\text{B.17})$$

Analisando os termos não nulos do numerador dentro do operador valor esperado, temos:

$$1, -1e^2, 3e^4, -15e^6, 105e^8, -945e^{10}, \dots \quad (\text{B.18})$$

A sequência 1, 1, 3, 15, 105, 945, ... pode ser obtida pela fórmula:

$$n!! = \frac{(2m)!}{m!2^m} \quad (\text{B.19})$$

que é chamada como fatorial duplo de n , quando n é um número ímpar.

Assim, é possível reescrever a Equação (B.15) como:

$$V_{\sigma}(e) = \frac{1}{\sqrt{2\pi\sigma}} E \left[\sum_{m=0}^{\infty} \frac{(2m)!(X-Y)^{2m}(-1)^m}{m!2^m(2m)!\sigma^{2m}} \right] = \frac{1}{\sqrt{2\pi\sigma}} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m\sigma^{2m}m!} E[(X-Y)^{2m}] \quad (\text{B.20})$$

que completa a prova.

Uma outra forma de realizar a prova é usar a série de Mclaurin da função exponencial.

$$\exp^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (\text{B.21})$$

que converge para todo x . Assim, fazendo $x = -\frac{(X-Y)^2}{2\sigma^2}$, pode-se reescrever a Equação (B.1) usando o *kernel* gaussiano como:

$$V_{\sigma}(X, Y) = E[K_{\sigma}(X, Y)] = E \left[\frac{1}{\sqrt{2\pi\sigma}} \exp \left(-\frac{(X-Y)^2}{2\sigma^2} \right) \right] \quad (\text{B.22})$$

Assim, é possível dizer que

$$V_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} E \left[\exp \left(-\frac{(X-Y)^2}{2\sigma^2} \right) \right] \quad (\text{B.23})$$

mas

$$\exp \left(-\frac{(X-Y)^2}{2\sigma^2} \right) = \sum_{m=0}^{\infty} \frac{\left(-\frac{(X-Y)^2}{2\sigma^2} \right)^m}{m!} = \sum_{m=0}^{\infty} \frac{(-1)^m (X-Y)^{2m}}{m!(2\sigma^2)^m} \quad (\text{B.24})$$

O que nos leva a

$$V_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m \sigma^{2m} m!} E[(X - Y)^{2m}] \quad (\text{B.25})$$

Análise

Vamos analisar a Equação (B.15):

$$V_{\sigma}(e) = \frac{1}{\sqrt{2\pi\sigma}} E \left[1 + 0 - \frac{1e^2}{2!(\sigma^2)^1} + 0 + \frac{3e^4}{4!(\sigma^2)^2} + 0 - \frac{15e^6}{6!(\sigma^2)^3} + 0 + \frac{105e^8}{8!(\sigma^2)^4} + 0 - \frac{945e^{10}}{10!(\sigma^2)^5} + \dots \right]$$

Conforme o tamanho de *kernel* σ cresce, os momentos de ordem superior decaem mais rapidamente devido ao denominador. Dessa forma, os momentos de segunda ordem tendem a dominar a equação, fazendo com que a correntropia se aproxime do valor da correlação (de forma enviesada).

Apêndice C

Correntropia complexa como integral de f_{XYZW} no plano $x = y$ e $z = w$

Assumindo que dados $\{(x_i, y_i, z_i, s_i)_{i=1}^N\}$ são tirados da distribuição conjunta f_{xyzs} , e seja $\hat{f}_{\sigma_{xyzs}}$ a estimação de Parzen como tamanho de *kernel* σ , a correntropia complexa é estimada com tamanho de *kernel* $\sigma' = \sigma\sqrt{2}$ é a integral de $\hat{f}_{\sigma_{xyzs}}$ no plano formado por $x = y$ e $z = s$.

$$\hat{V}_{\sigma'}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{\sigma_{XYZS}}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} \quad (C.1)$$

Prova:

Assumindo duas variáveis aleatórias complexas $C_1 = X + jZ$ e $C_2 = Y + jS$, onde $C_1, C_2 \in \mathbb{C}$, e X, Y, Z, S são variáveis aleatórias reais. A correntropia complexa foi desenvolvida para manter a interpretação probabilística da correntropia. Assim, estimar a correntropia complexa usando o *kernel* gaussiano significa estimar a densidade de probabilidade do evento $C_1 = C_2$, que implica na estimação da da densidade de probabilidade do evento $X = Y$ e $Z = S$. O motivo disso é que, dois números complexos são iguais quando ambas parte real e imaginárias são iguais, logo:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) \delta(x-y) \delta(z-s) dx dy dz ds \quad (C.2)$$

Quando $x = y$ e $z = s$, a equação (C.2) pode ser rescrita como:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 \quad (C.3)$$

que é o lado direito da igualdade da equação (C.1).

Vamos usar a equação 3.11, que estima a correntropia a partir dos dados usando o *kernel* gaussiano:

$$V_{\sigma}^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \quad (C.4)$$

Assim, para completar a prova é preciso obter a equação (3.11) a partir da resolução da integral dupla da equação (C.1).

Primeiramente, substitui-se \hat{f}_{XYZS} pelo estimador de Parzen definido como:

$$\hat{f}_{X^1, X^2, \dots, X^L}(x^1, x^2, \dots, x^L) = \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^L G_{\sigma}(x^l - x_n^l) \quad (C.5)$$

onde

$$G_{\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Assim, usando $L = 4$ na equação (C.5) e substituindo \hat{f}_{XYZS} na equação (C.3):

$$V_{\sigma}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}(u_1 - x_n) G_{\sigma}(u_1 - y_n) G_{\sigma}(u_2 - z_n) G_{\sigma}(u_2 - s_n) du_1 du_2 \quad (C.6)$$

A melhor maneira de resolver essa integral dupla e reescrever a equação (C.6) fazendo:

$$V_{\sigma}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{1}{2\sigma^2}(a(u_1 - b)^2 + a'(u_2 - b')^2 + c)\right) du_1 du_2$$

e então:

$$V_{\sigma}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{1}{2\sigma^2}a(u_1 - b)^2\right) \exp\left(-\frac{1}{2\sigma^2}a'(u_2 - b')^2\right) \exp\left(-\frac{c}{2\sigma^2}\right) du_1 du_2 \quad (C.7)$$

Os coeficientes a, b, a', b' e c são obtidos pela técnica de completar quadrados:

$$(u_1 - x_i)^2 + (u_1 - y_i)^2 + (u_2 - z_i)^2 + (u_2 - s_i)^2 = a(u_1 - b)^2 + a'(u_2 - b')^2 + c \quad (C.8)$$

Desenvolvendo o lado direito da equação (C.8):

$$au_1^2 + a'u_2^2 - 2u_1ab - 2u_2a'b' + ab^2 + a'b'^2 + c$$

E depois o lado esquerdo:

$$2u_1^2 + 2u_2^2 + x_i^2 + y_i^2 + z_i^2 + s_i^2 - 2u_1(x_i + y_i) - 2u_2(z_i + s_i)$$

Assim, pode-se afirmar:

$$a = a' = 2; b = \frac{(x_i + y_i)}{2}; b' = \frac{(z_i + s_i)}{2}; \text{ and } c = \frac{(x_i - y_i)^2 + (z_i - s_i)^2}{2}$$

Usando esses valores na equação (C.7), obtêm-se:

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{(u_1-b)^2}{\sigma^2}\right) \exp\left(-\frac{(u_2-b')^2}{\sigma^2}\right) \exp\left(-\frac{c}{2\sigma^2}\right) du_1 du_2 \quad (C.9)$$

Fazendo $\sigma^2 = 2\theta^2$, pode-se reescrever a equação (C.9) como:

$$= \int \int \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2\pi\theta^2}\right) \left(\frac{1}{2\pi 4\theta^2}\right) \exp\left(-\frac{(u_1-b)^2}{2\theta^2}\right) \exp\left(-\frac{(u_2-b')^2}{2\theta^2}\right) \exp\left(-\frac{c}{4\theta^2}\right) du_1 du_2 \quad (C.10)$$

Mas

$$\left(\frac{1}{\sqrt{2\pi}\theta}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(u_1-b)^2}{2\theta^2}\right) du_1 \left(\frac{1}{\sqrt{2\pi}\theta}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(u_2-b')^2}{2\theta^2}\right) du_2 = 1 \quad (C.11)$$

Logo, a equação (C.10) pode ser escrita como:

$$V_{\sigma}^C(C_1, C_2) = \left(\frac{1}{2\pi 4\theta^2}\right) \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(x_i - y_i)^2 + (z_i - s_i)^2}{8\theta^2}\right)$$

Ou, fazendo $8\theta^2 = (2\sigma')^2$:

$$V_{\sigma}^C(C_1, C_2) = \left(\frac{1}{2\pi(\sigma')^2}\right) \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(x_i - y_i)^2 + (z_i - s_i)^2}{2(\sigma')^2}\right) \quad (C.12)$$

Onde $\sigma' = \sigma\sqrt{2}$, que implica em:

$$V_{\sigma}^C(C_1, C_2) = \frac{1}{N} \sum_{n=1}^N G_{\sigma\sqrt{2}}(x_n - y_n) G_{\sigma\sqrt{2}}(z_n - s_n)$$

O estimador não paramétrico da correntropia complexa usando o *kernel* gaussiano pode ser escrito como:

$$V_{\sigma}^C(C_1, C_2) = E[G_{\sigma}^C(C_1 - C_2)] \quad (C.13)$$

onde

$$G_{\sigma}^C(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (C.14)$$

Ou, de forma alternativa:

$$V_{\sigma}^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \quad (C.15)$$

■

Apêndice D

Análise de convergência

Analisa-se aqui a convergência do algoritmo de gradiente ascendente estocástico que usa o critério de máxima correntropia complexa como função custo. A análise foi baseada na versão real realizada no trabalho [Al-Naffouri e Sayed 2001]. Inicialmente, o algoritmo descrito pela Equação (3.64) pode ser reescrito como:

$$\mathbf{w}_{n+1} = \mathbf{w}_i + \eta f[e_n] \mathbf{x}_n, \quad n \geq 0, \quad (\text{D.1})$$

onde η é o passo do gradiente e $f[e_n]$ é uma função não linear em função do erro $e(i)$, que pode ser escrito como:

$$f[e(i)] = \exp\left(-\frac{e_n e_n^*}{2\sigma^2}\right) e_n^*. \quad (\text{D.2})$$

Assume-se que o sinal desejado d_n pode ser expresso como:

$$d_n = \mathbf{w}_o^H \mathbf{x}_n + v_n, \quad (\text{D.3})$$

onde \mathbf{w}_o é o vetor com os valores ótimos que devem ser estimados. v_n é o ruído aditivo. Dessa forma, a estimação do erro no instante n é dada por:

$$e_n = d_n - \mathbf{w}_n^H \mathbf{x}_n = \mathbf{w}_o^H \mathbf{x}_n - \mathbf{w}_n^H \mathbf{x}_n + v_n. \quad (\text{D.4})$$

Considerando que o vetor erro dos pesos pode ser definido como $\tilde{\mathbf{w}}_n = \mathbf{w}^o - \mathbf{w}_n$, pode-se definir os erros *a priori* e *a posteriori* da seguinte forma:

$$e_a(n) = \tilde{\mathbf{w}}_n^H \mathbf{x}_n, \quad e_p(n) = \tilde{\mathbf{w}}_{n+1}^H \mathbf{x}_n. \quad (\text{D.5})$$

Assim, a regra de atualização dos pesos da Equação (D.1) pode ser reescrita como:

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \eta f[e_n] \mathbf{x}_n. \quad (\text{D.6})$$

Multiplicando os dois lados da Equação D.6 pelo Hermitiano \mathbf{x}_n , e manipulando a Equação D.5, é possível determinar a relação entre os erros $e_a(n)$, $e_p(n)$, e e_n da seguinte forma:

$$e_p(n) = e_a(n) - \eta f^*[e_n] \|\mathbf{x}_n\|^2. \quad (\text{D.7})$$

Para eliminar a não linearidade $f[e_n]$ da Equação D.6, é possível combina-la com a Equação D.7 para obter:

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n - (e_a(n) - e_p(n))^* \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|^2}. \quad (\text{D.8})$$

Como o objetivo de seguir uma abordagem baseada em energia, também realizada no caso real por Al-Naffouri e Sayed (2001), pode-se elevar ao quadrado os dois lados da Equação D.8 para obter:

$$\|\tilde{\mathbf{w}}_{n+1}\|^2 = \left(\tilde{\mathbf{w}}_n - (e_a(n) - e_p(n))^* \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|^2} \right)^H \times \left(\tilde{\mathbf{w}}_n - (e_a(n) - e_p(n))^* \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|^2} \right). \quad (\text{D.9})$$

Depois de manipular a Equação D.9, é possível obter a relação de energia:

$$\|\tilde{\mathbf{w}}_{n+1}\|^2 + \frac{\|e_a(n)\|^2}{\|\mathbf{x}_n\|^2} = \|\tilde{\mathbf{w}}_n\|^2 + \frac{\|e_p(n)\|^2}{\|\mathbf{x}_n\|^2}. \quad (\text{D.10})$$

Usando o operador valor esperado em ambos os lados da Equação D.10 e depois substituindo na Equação D.7, representando o erro *a posteriori* $e_p(n)$.

$$E\left[\|\tilde{\mathbf{w}}_{n+1}\|^2\right] = E\left[\|\tilde{\mathbf{w}}_n\|^2\right] - 2\eta E\left[\text{Re}\{e_a(n)f[e_n]\}\right] + \eta^2 E\left[f^2[e_n]\|\mathbf{x}_n\|^2\right] \quad (\text{D.11})$$

A convergência do algoritmo proposto depende da escolha do passo do gradiente. Assim, a abordagem de Lyapunov é utilizada para obter a convergência em um limite superior no qual $E\left[\|\tilde{\mathbf{w}}_n\|^2\right]$ permanece uniformemente limitado. Analisando a Equação D.11, é possível escrever:

$$\begin{aligned} E\left[\|\tilde{\mathbf{w}}_{n+1}\|^2\right] &\leq E\left[\|\tilde{\mathbf{w}}_n\|^2\right] \\ &\iff -2\eta E\left[\text{Re}\{e_a(n)f[e_n]\}\right] \\ &\quad + \eta^2 E\left[f^2[e_n]\|\mathbf{x}_n\|^2\right] \leq 0 \end{aligned} \quad (\text{D.12})$$

A partir da Equação (D.12), é possível dizer que o passo do gradiente pode ser escolhido da seguinte forma:

$$\eta \leq 2 \frac{E\left[\text{Re}\{e_a(i)f[e_n]\}\right]}{E\left[\|\mathbf{x}_n\|^2\right]}, \quad (\text{D.13})$$

A sequência $E\left[\|\tilde{\mathbf{w}}_n\|^2\right]$ dos pesos da potência do erro vai decrescer e ser limitada inferiormente, o que garante a convergência. Dessa forma, a condição para garantir a convergência pode ser expressa como:

$$\eta \leq 2 \inf_{n \geq 0} \frac{E[\operatorname{Re}\{e_a(n)f[e_n]\}]}{E[f^4[e_n]]^{1/2} E[\|\mathbf{x}_n\|^4]^{1/2}}. \quad (\text{D.14})$$

Assumindo que o filtro é longo o suficiente de tal forma que $e_a(n)$ é uma Gaussiana de média zero e que o processo v_n é independente, é possível definir as seguintes afirmações [Al-Naffouri e Sayed 2001, Chen et al. 2017]:

$$h_G[E[e_a^2(n)]] \triangleq \frac{E[\operatorname{Re}\{e_a(n)f[e_n]\}]}{E[e_a^2(n)]}. \quad (\text{D.15})$$

$$h_C[E[e_a^2(n)]] \triangleq E[f^4[e_n]]. \quad (\text{D.16})$$

Portanto, a condição suficiente para a convergência pode ser estabelecida substituindo as Equações (D.15) e (D.16) na Equação (D.14), o que resulta em:

$$\eta \leq \frac{2}{E[\|\mathbf{x}_n^4\|]^{1/2}} \left(\inf_{n \geq 0} \frac{E[e_a^2(n)] \cdot h_G[E[e_a^2(n)]]}{\sqrt{h_C[E[e_a^2(n)]]}} \right). \quad (\text{D.17})$$

Como todos os termos da Equação (D.17) são funções de $E[e_a^2(i)]$, é possível enfatizar esse aspecto na Equação (D.18) para indicar que a minimização vai ser feita sob os valores de $E[e_a^2(n)]$.

$$\eta \leq \frac{2}{E[\|\mathbf{x}_n^4\|]^{1/2}} \left(\inf_{E[e_a^2(n)]} \frac{E[e_a^2(n)] \cdot h_G[E[e_a^2(n)]]}{\sqrt{h_C[E[e_a^2(n)]]}} \right). \quad (\text{D.18})$$

Por fim, se o passo do gradiente seguir a condição descrita na Equação (D.18), pode-se afirmar que o algoritmo vai convergir.

Apêndice E

Algoritmo do MCCC com Ponto Fixo sem usar o Cálculo de Wirtinger

Relembrando a função custo

$$\begin{aligned} J_{MCCC} &= E[G_{\sigma}(\mathbf{d} - \mathbf{w}^H \mathbf{X})] = E[G_{\sigma}^C(\mathbf{e})] \\ &= \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(d_i - y_i)(d_i - y_i)^*}{2\sigma^2}\right) \\ &= \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right), \end{aligned} \quad (\text{E.1})$$

onde $y_i = \mathbf{w}^H \mathbf{x}_i$, e \mathbf{x}_i é o i -ésima coluna da matriz de entrada \mathbf{X} . Para derivar a solução de ponto fixo é necessário fazer a derivada da função custo (Equação E.1) em relação a w_r e w_i e igualar a zero.

$$\frac{\partial J_{MCCC}}{\partial w_r} = 0 \quad \text{e} \quad \frac{\partial J_{MCCC}}{\partial w_i} = 0 \quad (\text{E.2})$$

O que é equivalente a

$$\frac{\partial J_{MCCC}}{\partial w_r} = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) \frac{\partial \beta}{\partial w_r} = 0 \quad (\text{E.3})$$

e

$$\frac{\partial J_{MCCC}}{\partial w_i} = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) \frac{\partial \beta}{\partial w_i} = 0 \quad (\text{E.4})$$

Vamos chamar β

$$\beta = -\frac{1}{2\sigma^2} (dd^* - d\mathbf{x}^H \mathbf{w} - \mathbf{w}^H \mathbf{x}d^* + \mathbf{w}^H \mathbf{x}\mathbf{x}^H \mathbf{w}) \quad (\text{E.5})$$

e expandir os termos internos, começando com dd^*

$$dd^* = (d_r + jd_i)(d_r - jd_i) = d_r^2 + d_i^2 \quad (\text{E.6})$$

Em seguida, $d\mathbf{x}^H \mathbf{w}$:

$$d\mathbf{x}^H \mathbf{w} = (d_r + jd_i)(\mathbf{x}_r^T - j\mathbf{x}_i^T)(\mathbf{w}_r + j\mathbf{w}_i) \quad (\text{E.7})$$

$$= [(d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T) + j(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)](\mathbf{w}_r + j\mathbf{w}_i) \quad (\text{E.8})$$

$$= (d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)\mathbf{w}_r + j(d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)\mathbf{w}_i + j(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)\mathbf{w}_r + j^2(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)\mathbf{w}_i \quad (\text{E.9})$$

$$= (d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)\mathbf{w}_r + j(d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)\mathbf{w}_i + j(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)\mathbf{w}_r - (d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)\mathbf{w}_i \quad (\text{E.10})$$

Depois, $\mathbf{w}^H \mathbf{x} d^*$

$$\mathbf{w}^H \mathbf{x} d^* = (\mathbf{w}_r^T - \mathbf{w}_i^T)(\mathbf{x}_r + j\mathbf{x}_i)(d_r - jd_i) \quad (\text{E.11})$$

$$= (\mathbf{w}_r^T - \mathbf{w}_i^T)[(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j(\mathbf{x}_i d_r - \mathbf{x}_r d_i)] \quad (\text{E.12})$$

$$= \mathbf{w}_r^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j\mathbf{w}_r^T (\mathbf{x}_i d_r - \mathbf{x}_r d_i) - j\mathbf{w}_i^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) - j^2 (\mathbf{x}_i d_r - \mathbf{x}_r d_i) \quad (\text{E.13})$$

$$= \mathbf{w}_r^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j\mathbf{w}_r^T (\mathbf{x}_i d_r - \mathbf{x}_r d_i) - j\mathbf{w}_i^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + \mathbf{w}_i^T (\mathbf{x}_i d_r - \mathbf{x}_r d_i) \quad (\text{E.14})$$

e, por último $\mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w}$

$$\mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w} = (\mathbf{w}_r^T - j\mathbf{w}_i^T)(\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T)(\mathbf{w}_r + j\mathbf{w}_i) \quad (\text{E.15})$$

$$= (\mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T + \mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T - j\mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T - j\mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T)(\mathbf{w}_r + j\mathbf{w}_i) \quad (\text{E.16})$$

$$= \mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + \mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r + j\mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + j\mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i - j^2 \mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i - j^2 \mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i \quad (\text{E.17})$$

$$= \mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + \mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r + j\mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + j\mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i + \mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i \quad (\text{E.18})$$

Assim, é possível obter as derivadas

$$\frac{\partial d d^*}{\partial \mathbf{w}_r} = 0 \quad (\text{E.19})$$

$$\frac{\partial d \mathbf{x}^H \mathbf{w}}{\partial \mathbf{w}_r} = (d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)^T + j(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)^T = (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j(\mathbf{x}_r d_i - \mathbf{x}_i d_r) \quad (\text{E.20})$$

$$\frac{\partial \mathbf{w}^H \mathbf{x} d^*}{\partial \mathbf{w}_r} = (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j(\mathbf{x}_i d_r - \mathbf{x}_r d_i) \quad (\text{E.21})$$

$$\frac{\partial \mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w}}{\partial \mathbf{w}_r} = 2\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + 2\mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r - j\mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r + j\mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + j\mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i \quad (\text{E.22})$$

$$\frac{\partial \mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w}}{\partial \mathbf{w}_r} = 2\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + 2\mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r \quad (\text{E.23})$$

$$\frac{\partial \beta}{\partial \mathbf{w}_r} = -\frac{1}{2\sigma^2} (0 - [(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j(\mathbf{x}_r d_i - \mathbf{x}_i d_r)] - [(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j(\mathbf{x}_i d_r - \mathbf{x}_r d_i) + 2\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + 2\mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r]) \quad (\text{E.24})$$

$$\frac{\partial \beta}{\partial \mathbf{w}_r} = -\frac{1}{2\sigma^2} (-2\mathbf{x}_r d_r - 2\mathbf{x}_i d_i + 2\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + 2\mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r) \quad (\text{E.25})$$

Substituindo E.45 em E.3, é possível obter

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = -\frac{1}{2\sigma^2} \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) (-2\mathbf{x}_r d_r - 2\mathbf{x}_i d_i + 2\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + 2\mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r) = 0 \quad (\text{E.26})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = -2 \frac{1}{2\sigma^2} \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) (-\mathbf{x}_r d_r - \mathbf{x}_i d_i + \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r) = 0 \quad (\text{E.27})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = -2 \frac{1}{2\sigma^2} \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) (-(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_r) = 0 \quad (\text{E.28})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = 2 \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (-(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_r) = 0 \quad (\text{E.29})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (-(\mathbf{x}_r d_r + \mathbf{x}_i d_i) + (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_r) = 0 \quad (\text{E.30})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_r} = -\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + \sum_{n=1}^N \exp(\beta) (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_r = 0 \quad (\text{E.31})$$

$$\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (\mathbf{x}_r d_r + \mathbf{x}_i d_i) = \sum_{n=1}^N \exp(\beta) (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_r \quad (\text{E.32})$$

$$\mathbf{w}_r = \left[\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (\mathbf{x}_r \mathbf{x}_r^T + \mathbf{x}_i \mathbf{x}_i^T) \right]^{-1} \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (d_r \mathbf{x}_r + d_i \mathbf{x}_i) \quad (\text{E.33})$$

Agora, para encontrar a equação de ponto fixo de \mathbf{w}_i , podemos fazer

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_i} = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) \frac{\partial \beta}{\partial \mathbf{w}_i} = 0 \quad (\text{E.34})$$

$$\frac{\partial \beta}{\partial \mathbf{w}_i} = -\frac{1}{2\sigma^2} \frac{\partial (dd^* - d\mathbf{x}^H \mathbf{w} - \mathbf{w}^H \mathbf{x} d^* + \mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w})}{\partial \mathbf{w}_i} = 0 \quad (\text{E.35})$$

$$d\mathbf{x}^H \mathbf{w} = (d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T) \mathbf{w}_r + j(d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T) \mathbf{w}_i + j(d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T) \mathbf{w}_r - (d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T) \mathbf{w}_i \quad (\text{E.36})$$

$$\frac{\partial \mathbf{x}^H \mathbf{w}}{\partial \mathbf{w}_i} = j(d_r \mathbf{x}_r^T + d_i \mathbf{x}_i^T)^T - (d_i \mathbf{x}_r^T - d_r \mathbf{x}_i^T)^T \quad (\text{E.37})$$

$$\frac{\partial \mathbf{x}^H \mathbf{w}}{\partial \mathbf{w}_i} = -(\mathbf{x}_r d_i - \mathbf{x}_i d_r) + j(\mathbf{x}_r d_r + \mathbf{x}_i d_i) \quad (\text{E.38})$$

$$\mathbf{w}^H \mathbf{x} d^* = \mathbf{w}_r^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + j \mathbf{w}_r^T (\mathbf{x}_i d_r - \mathbf{x}_r d_i) - j \mathbf{w}_i^T (\mathbf{x}_r d_r + \mathbf{x}_i d_i) + \mathbf{w}_i^T (\mathbf{x}_i d_r - \mathbf{x}_r d_i) \quad (\text{E.39})$$

$$\frac{\partial \mathbf{w}^H \mathbf{x} d^*}{\partial \mathbf{w}_i} = +(\mathbf{x}_i d_r - \mathbf{x}_r d_i) - j(\mathbf{x}_r d_r + \mathbf{x}_i d_i) \quad (\text{E.40})$$

$$\frac{\partial \mathbf{w}^H \mathbf{x} d^*}{\partial \mathbf{w}_i} = -(\mathbf{x}_r d_i - \mathbf{x}_i d_r) - j(\mathbf{x}_r d_r + \mathbf{x}_i d_i) \quad (\text{E.41})$$

$$\begin{aligned} \mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w} = \\ \mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r + \mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r - j \mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_r - j \mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_r + \\ + j \mathbf{w}_r^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + j \mathbf{w}_r^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i + \mathbf{w}_i^T \mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{w}_i^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i \end{aligned} \quad (\text{E.42})$$

$$\frac{\partial \mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w}}{\partial \mathbf{w}_i} = 2(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i) \quad (\text{E.43})$$

$$\frac{\partial \beta}{\partial \mathbf{w}_i} = -\frac{1}{2\sigma^2} \frac{\partial (d d^* - d \mathbf{x}^H \mathbf{w} - \mathbf{w}^H \mathbf{x} d^* + \mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w})}{\partial \mathbf{w}_i} \quad (\text{E.44})$$

$$\frac{\partial \beta}{\partial \mathbf{w}_i} = -\frac{1}{2\sigma^2} (+2(\mathbf{x}_r d_i - \mathbf{x}_i d_r) + 2(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i)) \quad (\text{E.45})$$

Usando a Equação (E.45) na Equação (E.4), temos

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_i} = -\frac{1}{2\sigma^2} \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp(\beta) (+2(\mathbf{x}_r d_i - \mathbf{x}_i d_r) + 2(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i)) = 0 \quad (\text{E.46})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_i} = 2 \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) ((\mathbf{x}_r d_i - \mathbf{x}_i d_r) + (\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i)) = 0 \quad (\text{E.47})$$

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}_i} = \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) ((\mathbf{x}_r d_i - \mathbf{x}_i d_r) + (\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i)) = 0 \quad (\text{E.48})$$

$$\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (\mathbf{x}_r d_i - \mathbf{x}_i d_r) + \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) (\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T \mathbf{w}_i) = 0 \quad (\text{E.49})$$

$$-\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{x}_r d_i - \mathbf{x}_i d_r) = \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_i \quad (\text{E.50})$$

$$\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{x}_i d_r - \mathbf{x}_r d_i) = \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T) \mathbf{w}_i \quad (\text{E.51})$$

$$\mathbf{w}_i = \left[\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(d_r \mathbf{x}_i - d_i \mathbf{x}_r) \right] \sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{x}_r \mathbf{x}_r^T \mathbf{w}_i + \mathbf{x}_i \mathbf{x}_i^T) \quad (\text{E.52})$$

Comparativo

Expandindo os termos de fora da função exponencial da Equação (3.67), temos:

$$d^* \mathbf{x} = (d_r - j d_i)(\mathbf{X}_r + j \mathbf{X}_i) \quad (\text{E.53})$$

$$d^* \mathbf{x} = (d_r \mathbf{X}_r + d_i \mathbf{X}_i) + j(d_r \mathbf{X}_i - d_i \mathbf{X}_r) \quad (\text{E.54})$$

e

$$\mathbf{X} \mathbf{X}^H = (\mathbf{X}_r + j \mathbf{X}_i)(\mathbf{X}_r^T - j \mathbf{X}_i^T) \quad (\text{E.55})$$

$$= (\mathbf{X}_r \mathbf{X}_r^T + \mathbf{X}_i \mathbf{X}_i^T) + j(\mathbf{X}_i \mathbf{X}_r^T - \mathbf{X}_r \mathbf{X}_i^T) \quad (\text{E.56})$$

$$= (\mathbf{X}_r \mathbf{X}_r^T + \mathbf{X}_i \mathbf{X}_i^T) + j(\mathbf{X}_r \mathbf{X}_i^T - \mathbf{X}_r \mathbf{X}_i^T) \quad (\text{E.57})$$

$$\mathbf{X} \mathbf{X}^H = (\mathbf{X}_r \mathbf{X}_r^T + \mathbf{X}_i \mathbf{X}_i^T) \quad (\text{E.58})$$

Assim, é possível reescrever a Equação (3.67)

$$\mathbf{w}_r = [G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{X}_r \mathbf{X}_r^T + \mathbf{X}_i \mathbf{X}_i^T)]^{-1} G_{\sigma\sqrt{2}}^C(e_n)(d_r \mathbf{X}_r + d_i \mathbf{X}_i) \quad (\text{E.59})$$

$$\mathbf{w}_i = [G_{\sigma\sqrt{2}}^C(e_n)(\mathbf{X}_r \mathbf{X}_r^T + \mathbf{X}_i \mathbf{X}_i^T)]^{-1} G_{\sigma\sqrt{2}}^C(e_n)(d_r \mathbf{X}_i - d_i \mathbf{X}_r) \quad (\text{E.60})$$

que é equivalente a Equação (E.33), finalizando a demonstração.

Apêndice F

Correntropia Complexa com resposta complexa

O *kernel* da Equação (F.1) faz com que a correntropia complexa tenha uma resposta complexa.

$$\Lambda_{\sigma}(x,y) = \exp\left(-\frac{xx^* - 2xy^* + yy^*}{2\sigma^2}\right) \quad (\text{F.1})$$

Apesar dos produtos xx^* e yy^* serem reais, o produto xy^* dentro da exponencial é complexo. Dessa forma, a única coisa que precisa ser provada é que o *kernel* da Equação (F.1) é positivo definido, habilitando-o para ser usado na correntropia complexa. Devida a extensão dessa prova, ela é mostrada no anexo F.

O uso deste *kernel* faz com que a estimação da correntropia complexa seja $V^c(Q,W) = 1$ quando $Q = W$. Além disso, esse *kernel* tem a propriedade da simetria hermitiana, ou seja $\Lambda_{\sigma}(x,y) = \Lambda_{\sigma}(y,x)^*$.

A seguir mostra-se a prova de que (F.1) é positivo definido:

Considerando as propriedades discutidas na seção 2.2.10 e, seja $x,y \in \mathbb{C}$, deseja-se provar que

$$k_{\sigma}^c(x,y) = \exp\left(-\frac{xx^* - 2xy^* + yy^*}{2\sigma^2}\right) \quad (\text{F.2})$$

é um *kernel* positivo definido. Para isso, podemos reescrever (F.2) como :

$$\begin{aligned} k_{\sigma}^c(x,y) &= \exp\left(-\frac{xx^* - 2xy^* + yy^*}{2\sigma^2}\right) \\ &= \exp\left(\frac{-xx^* + 2xy^* - yy^*}{2\sigma^2}\right) \\ &= \exp\left(-\frac{xx^*}{2\sigma^2}\right) \exp\left(\frac{2xy^*}{2\sigma^2}\right) \exp\left(-\frac{yy^*}{2\sigma^2}\right) \end{aligned} \quad (\text{F.3})$$

Nessa forma, é possível reescrever (F.3) como:

$$k_{\sigma}^c(x,y) = f(x) k'_{\sigma}(x,y) \overline{f(y)} \quad (\text{F.4})$$

onde

$$f(x) = \exp\left(-\frac{xx^*}{2\sigma^2}\right) \quad (\text{F.5})$$

e, conseqüentemente,

$$f(y) = \exp\left(-\frac{yy^*}{2\sigma^2}\right) \quad (\text{F.6})$$

Sabendo que $\overline{\exp(x)} = \exp(\bar{x})$, e possível escrever (F.6) como:

$$\overline{f(y)} = \exp\left(-\frac{y^*y}{2\sigma^2}\right) = \exp\left(-\frac{yy^*}{2\sigma^2}\right) \quad (\text{F.7})$$

Além disso, deve-se ressaltar que:

$$k'_\sigma(x, y) = \exp\left(\frac{2xy^*}{2\sigma^2}\right) = \exp\left(\frac{xy^*}{\sigma^2}\right) \quad (\text{F.8})$$

Sendo assim, para provar que (F.2) é positivo definido, é necessário provar que (F.8) também é. A seção seguinte mostra a prova que $\exp(z)$ é analítica. Esse processo é importante para realizar a prova.

F.1 Prova que $\exp(z)$, com $z \in \mathbb{C}$, é diferenciável:

Seja $z = x + jy$, temos:

$$\exp(z) = \exp(x + jy) = \exp(x) \exp(jy) \quad (\text{F.9})$$

Usando a relação de Euler: $\exp(ja) = \cos(a) + j \sin(a)$, temos:

$$\begin{aligned} \exp(x) \exp(jy) &= \exp(x) (\cos(y) + j \sin(y)) \\ &= \exp(x) \cos(y) + j \exp(x) \sin(y) \end{aligned} \quad (\text{F.10})$$

Para ser diferenciável, $\exp(z)$ deve satisfazer as equações de Cauchy-Riemann, provando ser analítico (holomórfico):

$$\begin{aligned} \frac{\partial u(x, y)}{\partial x} &= \frac{\partial v(x, y)}{\partial y} \\ \frac{\partial v(x, y)}{\partial x} &= -\frac{\partial u(x, y)}{\partial y} \end{aligned} \quad (\text{F.11})$$

Assim, deve-se escrever $\exp(x)$ na forma de $f(x, y) = u(x, y) + jv(x, y)$:

$$\exp(z) = f(x, y) = u(x, y) + jv(x, y) \quad (\text{F.12})$$

onde:

$$u(x, y) = \exp(x) \cos(y) \quad (\text{F.13})$$

e

$$v(x, y) = \exp(x) \sin(y) \quad (\text{F.14})$$

Obtendo as derivadas parciais:

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial(\exp(x) \cos(y))}{\partial x} = \exp(x) \cos(y) \quad (\text{F.15})$$

$$\frac{\partial v(x, y)}{\partial y} = \frac{\partial(\exp(x) \sin(y))}{\partial y} = \exp(x) \cos(y) \quad (\text{F.16})$$

$$\frac{\partial v(x, y)}{\partial x} = \frac{\partial(\exp(x) \sin(y))}{\partial x} = \exp(x) \sin(y) \quad (\text{F.17})$$

$$-\frac{\partial u(x, y)}{\partial y} = -\frac{\partial(\exp(x) \cos(y))}{\partial y} = -(-\exp(x) \sin(y)) = \exp(x) \sin(y) \quad (\text{F.18})$$

Como (F.15) = (F.16) e (F.17) = (F.18), $\exp(z)$ satisfaz as equações de Cauchy-Riemann, provando sua analiticidade. Isso possibilita reescrevê-la usando a sua série de Maclaurin da seguinte forma:

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad (\text{F.19})$$

F.2 Prova que $k'_{\sigma} = \exp(xy^*)$ é positivo-definido

Como $\exp(z)$ é holomórfica, é possível considerar $z = xy^*$ e usar a representação da função exponencial e sua série de Maclaurin (F.19).

$$\exp(xy^*) = \sum_{n=0}^{\infty} \frac{(xy^*)^n}{n!} = 1 + xy^* + \frac{(xy^*)^2}{2!} + \frac{(xy^*)^3}{3!} + \dots \quad (\text{F.20})$$

É fácil ver que (F.20) é uma combinação positiva e de produtos do *kernel* k'_{σ} mostrado em (2.53). Logo, é necessário provar que $k^m_{\sigma}(x, y)$ é positivo-definido para completar a prova.

Prova que $k^m_{\sigma}(x, y) = xy^*$ é positivo definido

Usando (2.52) em $k^m_{\sigma}(x, y)$, obtêm-se:

$$\begin{aligned}
 \sum_{i=1}^n \sum_{k=1}^n x_i y_k^* \xi_i \bar{\xi}_k &\geq 0 \\
 \sum_{i=1}^n \xi_i x_i \sum_{k=1}^n y_k^* \bar{\xi}_k &\geq 0 \\
 \sum_{i=1}^n \xi_i x_i \sum_{i=1}^n x_i^* \bar{\xi}_i &\geq 0
 \end{aligned} \tag{F.21}$$

Por fim, pode-se concluir que:

$$\left| \sum_{i=1}^n \xi_i x_i \right|^2 \geq 0 \tag{F.22}$$

o que completa a prova.