



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



Arquiteturas de Classificação Automática de Modulações em Ambientes Impulsivos Baseadas em Características Cicloestacionárias

Tales Vinícius Rodrigues de Oliveira Câmara

Orientador: Prof. Dr. Luiz Felipe de Queiroz Silveira.

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

Número de ordem PPgEEC: D279
Natal, RN, outubro de 2020

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

Câmara, Tales Vinicius Rodrigues de Oliveira.

Arquiteturas de classificação automática de modulações em ambientes impulsivos baseadas em características cicloestacionárias / Tales Vinicius Rodrigues de Oliveira Câmara. - 2021.

117f.: il.

(Doutorado) - Tese Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Natal, 2021.

Orientador: Dr. Luiz Felipe de Queiroz Silveira.

1. Classificação de Modulação - Tese. 2. Análise Cicloestacionária - Tese. 3. Correntropia Cíclica - Tese. I. Silveira, Luiz Felipe de Queiroz. II. Título.

RN/UF/BCZM

CDU 621.3

*A minha família, aos amigos e a
todos que defendem a ciência
brasileira.*

Agradecimentos

Ao Professor Luiz Felipe, pela orientação, paciência e exemplo. "O domínio de uma arte pode ser alcançado através da repetição, e sucessivas correções. Nesse processo, o olhar de um mestre é fundamental".

Aos Professores, Wamberto, Pedro, Joilson e Marcelo, pelas revisões e sugestões que contribuíram para o modelo final deste trabalho.

Aos colegas de laboratório que contribuíram para um ambiente de trabalho, as vezes improdutivo, porém, sempre virtuoso.

Aos amigos, Arthur, Ítalo, Fabrício, Danilo, Pedrão, Lenival, Tiago, Vitor, Maelso, por sanar dúvidas diversas, e pelas discussões que me fizeram enxergar outros pontos relevantes da pesquisa (além de outros contextos). "Se pude enxergar mais longe, é porque me apoiei nos ombros de gigantes".

À minha estimada amiga Clarissa, pelo apoio fundamental no começo do doutorado. "Em algumas ocasiões, um grão de mostarda faz toda a diferença".

À minha querida amiga Jessica, e toda família Bittencourt, pelo carinho, inspiração e por sempre me estimular a buscar soluções criativas. "Para entender um sistema, as vezes, temos que abrir a sua caixa. Mas para inovarmos, para fazermos a diferença, devemos ver além da caixa".

À minha família, em particular para o novo integrante, Apolinho, por dar sentido a caminhada, e pelo suporte durante os desafios. "Todo homem precisa de um norte, pois ainda que não conheça o caminho, saberá qual é a direção certa".

À CAPES, pelo apoio financeiro.

Resumo

O rápido crescimento de aplicações de sistemas de comunicações sem-fio impulsiona a busca por novas estratégias de exploração eficiente do espectro, tal como o uso de rádios cognitivos: um sistema de comunicações inteligente, capaz de se adaptar autonomamente ao canal de comunicação por meio da reconfiguração dos seus parâmetros de operação. Uma importante propriedade dos rádios cognitivos é a capacidade de reconhecer automaticamente o tipo de modulação empregada em um sinal de radiofrequência (RF), o que possibilita a interoperabilidade entre sistemas, a melhoria da eficiência espectral, e viabiliza a vigilância eletrônica. Esse atributo é conhecido como classificação automática de modulações (AMC). Dentre as técnicas de AMC que caracterizam o estado-da-arte, estão as que se baseiam em detecção de padrões obtidos a partir da análise de características cicloestacionárias de segunda ordem. Embora muito difundidas, essas técnicas não são capazes de reconhecer alguns tipos de modulações digitais, tais como a modulação de amplitude em quadratura m -ária (MQAM) e modulação por chaveamento de fase m -ária (MPSK) de alta ordem. Em contra partida, as técnicas de análise cicloestacionárias de ordem superior, usadas para extração de descritores singulares dessas modulações são adequadas apenas para ambientes de comunicação com ruído gaussiano branco aditivo (AWGN). Embora o modelo AWGN seja amplamente empregado para caracterizar canais de comunicação sem-fio, diversos cenários práticos são melhor representados por outros modelos. Um exemplo disso é a comunicação em alta frequência (HF), cujo ambiente apresenta contaminação por ruídos impulsivos. Nesse contexto, duas funções de análise cicloestacionária, a função de autocorrelação cíclica fracionária de ordem inferior (FLOCAF), e a função de correntropia cíclica (CCF), foram eficientemente aplicadas em ambientes impulsivos de comunicações. Contudo, elas foram empregadas para o sensoriamento de um canal impulsivo, um problema de menor complexidade em relação à classificação automática de modulações. Uma vez que na literatura não existe uma solução satisfatória para a classificação automática de modulações de alta ordem em canais com ruído impulsivo, neste trabalho são propostas arquiteturas de AMC baseadas na FLOCAF e na CCF, combinadas com técnicas de classificação por árvore de decisão e por regressão logística. As arquiteturas foram desenvolvidas para o reconhecimento das modulações digitais BPSK, QPSK, 8-QAM, 16-QAM, e 32-QAM, e avaliadas em diferentes contextos de contaminação por ruído aditivo impulsivo, descrito através de uma distribuição alfa-estável. Os resultados mostraram que todas as arquiteturas foram capazes de operar em ambientes impulsivos, sendo as arquiteturas baseadas na CCF as mais eficientes.

Palavras-chave: ruído aditivo alfa-estável, função de correntropia cíclica, função de autocorrelação cíclica fracionária de ordem inferior, ruído impulsivo, reconhecimento automático de modulação, aprendizagem de máquina, regressão logística.

Abstract

The rapid growth of wireless communications systems' applications drives new strategies for efficient exploration of the frequency spectrum, such as cognitive radios. The cognitive radio is an intelligent communications system capable of autonomously adapting to the communication channel by reconfiguring its operating parameters. An essential property of cognitive radios is automatically recognizing the type of modulation employed in a Radio Frequency (RF) signal, thus enabling interoperability between systems, improving spectral efficiency, or even enabling electronic surveillance. This attribute is known as automatic modulation classification (AMC). Techniques based on the detection of patterns obtained from the analysis of cyclostationary characteristics of the second-order compose the state-of-the-art of AMC applications. Although very widespread, these techniques cannot recognize some types of digital high-order modulations, such as M-ary quadrature amplitude modulation (MQAM) and M-ary phase-shift keying (MPSK) modulations. On the other hand, the higher-order cyclostationary analysis techniques used to extract particular descriptors of these modulations have a very high computational cost. Besides, they are only suitable for communication environments with additive white gaussian noise (AWGN). Although the AWGN model is widely used to characterize wireless communication channels, several practical scenarios are better modeled by non-gaussian distributions, such as high frequency (HF) communication, whose environment presents substantial contamination by impulsive noise. In this scenario, two emerging cyclostationary analysis functions, the fractional lower-order cyclic autocorrelation function (FLO-CAF) and the cyclic correntropy function (CCF), were efficiently employed in impulsive communications environments. However, these functions were evaluated on the impulsive channel's spectral sensing, a less complicated problem to the automatic modulation classification. Since there is no satisfactory solution in the literature for the automatic classification of high-order modulations in channels with impulsive noise, this work proposes automatic modulation classification architectures based on FLOCAF and CCF, combined with the decision tree classification and logistic regression techniques. The architectures were developed to recognize digital modulations Binary PSK (BPSK), Quadrature PSK (QPSK), 8-QAM, 16-QAM, and 32-QAM, and evaluated in different contexts of contamination by additive impulsive noise, modeled by alpha-stable distribution. The results showed that all architectures could operate in impulsive environments, with CCF-based architectures being the most efficient.

Keywords: additive alpha-stable noise, cyclic correntropy function, fractional lower-order cyclic autocorrelation function, digital modulations, impulsive noise, automatic modulation recognition, machine learning, logistic regression.

Sumário

Sumário	i
Lista de Figuras	iii
Lista de Tabelas	v
Lista de Abreviaturas	vii
1 Introdução	1
1.1 Contribuições da Tese	4
1.2 Organização da Tese	4
2 Fundamentação Teórica	7
2.1 Ruído Impulsivo	8
2.1.1 Distribuição Alfa-Estável	9
2.2 Análise Cicloestacionária: CAF e FLOCAF	13
2.2.1 Função de autocorrelação cíclica	14
2.2.2 Função de autocorrelação cíclica fracionária de ordem-inferior	15
2.2.3 Síntese da assinatura da FLOCAF	17
2.3 Análise Cicloestacionária por Correntropia	19
2.3.1 Função de Correntropia	19
2.3.2 Função de Correntropia Cíclica	20
2.3.3 Síntese da assinatura da CCF	22
2.4 Estrutura de uma Árvore de Decisão	24
2.5 Classificação por Regressão Logística	25
2.5.1 Regressão linear	25
2.5.2 Regressão logística	28
2.5.3 Classificação multi-classe com regressão logística	31
3 Análise Espectral Cíclica	33
3.1 Análise espectral sem ruído	33
3.2 Análise espectral em canais impulsivos	37
4 Arquiteturas de AMC com Árvore de Decisão	41
4.1 Definição dos <i>Templates</i> das Modulações	42
4.2 Sinal Desconhecido	42
4.3 Coeficiente de Correlação	42

4.3.1	Relação C_{sig}/C_{16QAM}	43
4.4	Resultados de classificação com Árvore de Decisão	44
4.4.1	AMC usando a FLOCAF	44
4.4.2	AMC usando a CCF	49
4.5	Estimação da Taxa de Símbolo	53
4.5.1	Estimação da taxa de símbolos usando a FLOCAF	55
4.5.2	Estimação da taxa de símbolos usando a CCF	55
5	Arquiteturas de AMC com Regressão Logística	57
5.1	Arquitetura de classificação de modulação baseada em análise cicloestacionária e regressão logística	57
5.1.1	Perfil Cíclico Reduzido	58
5.2	Resultado de classificação com Regressão Logística	59
5.2.1	AMC utilizando a FLOCAF com regressão logística	60
5.2.2	AMC utilizando a FLOCAF com perfil-cíclico reduzido e regressão logística	62
5.2.3	AMC utilizando a CCF com regressão logística	64
5.2.4	AMC utilizando a CCF com perfil-cíclico reduzido e regressão logística	66
5.3	Comparação entre as arquiteturas de classificação	68
6	Conclusão	71
6.1	Trabalhos Produzidos	73
6.2	Trabalhos Futuros	73
	Referências Bibliográficas	75
A	Função Característica de uma Variável Aleatória Gaussiana	83
B	Geração Numérica do Ruído Aditivo Alfa-Estável	85
C	Teoria da Informação e Correntropia	87
C.1	Entropia de Shannon	87
C.2	Entropia de Renyi	89
C.3	Estimador de Parzen	90
C.4	Potencial de Informação	91
C.5	Correntropia	93
C.6	Correntropia para Processos Aleatórios	94
D	Interpretação Estatística da Regressão Logística	97

Lista de Figuras

2.1	Sistema de comunicação sem-fio em ambiente impulsivo.	7
2.2	Funções densidades de probabilidade em escala logarítmica para variáveis aleatórias alfa-estáveis (X) com diferentes expoentes característicos α , $\beta = 0$, $\gamma = 1$, e $\delta = 0$. Quanto menor o valor de α mais lento será o decaimento da curva, e portanto, a impulsividade do ruído será maior. . .	11
2.3	Perfis-cíclicos da função de autocorrelação cíclica aplicada a um sinal BPSK, respectivamente sem e com contaminação alfa-estável.	15
2.4	FLOCAF aplicada ao sinal BPSK.	18
2.5	Perfil-cíclico da FLOCAF aplicada ao sinal BPSK.	19
2.6	Função de correntropia cíclica aplicada ao sinal BPSK.	23
2.7	Perfil-Cíclico da Função de correntropia cíclica aplicada ao sinal BPSK. .	23
2.8	Estrutura genérica de uma árvore de decisão.	24
2.9	Função sigmoide $g(z)$	29
2.10	Estimação da função hipótese $h_{\theta}(x)$ a partir da Regressão Logística. . . .	30
3.1	Assinaturas cicloestacionárias do BPSK obtidas pela CAF, FLOCAF e CCF. .	34
3.2	Assinaturas cicloestacionárias do QPSK obtidas pela CAF, FLOCAF e CCF. .	35
3.3	Assinaturas cicloestacionárias do 8-QAM obtidas pela CAF, FLOCAF e CCF.	35
3.4	Assinaturas cicloestacionárias do 16-QAM obtidas pela CAF, FLOCAF e CCF.	36
3.5	Assinaturas cicloestacionárias do 32-QAM obtidas pela CAF, FLOCAF e CCF.	36
3.6	Assinaturas cicloestacionárias, obtidas com a FLOCAF e com a CCF, de sinais modulados contaminados por ruído com $GSNR=10$ dB.	38
3.7	Assinaturas cicloestacionárias, obtidas com a FLOCAF e com a CCF, de sinais modulados contaminados por ruído com $GSNR=0$ dB.	38
4.1	Estrutura de classificação baseada na extração de características cicloestacionárias aplicada a uma árvore de decisão.	41
4.2	Taxa de classificação global de modulação a partir da arquitetura que emprega a FLOCAF, considerando diferentes combinações de momentos fracionários.	45
4.3	Valores médios da taxa de acerto global obtidos sobre faixa de $GSNR$, relacionados a arquitetura que emprega a FLOCAF, considerando diferentes combinações de momentos fracionários.	45

4.4	Definição dos limiares th_{qpsk} e th_{16QAM} , para arquitetura que emprega a FLOCAF.	46
4.5	Curvas de classificação da arquitetura baseada na FLOCAF, parametrizada com $a = 0,3$ e $b = 0,5$	47
4.6	Taxa de classificação global de modulação a partir da arquitetura que emprega a CCF, considerando diferentes tamanhos de <i>kernel</i>	49
4.7	Valores médios da taxa de acerto global obtidos sobre faixa de <i>GSNR</i> , relacionados à arquitetura que emprega a CCF, considerando diferentes tamanhos de <i>kernel</i>	50
4.8	Definição dos limiares th_{qpsk} e th_{16QAM} , para arquitetura que emprega a CCF.	51
4.9	Curvas de classificação da arquitetura baseada na CCF, parametrizadas com $\sigma = 1.2$. Em cada configuração de <i>GSNR</i> foram considerados 1000 testes por modulação.	52
4.10	Assinaturas cicloestacionárias, da modulação QPSK, obtidas pela FLOCAF em três diferentes condições de <i>GSNR</i>	54
4.11	Assinaturas cicloestacionárias, da modulação QPSK, obtidas pela CCF em três diferentes condições de <i>GSNR</i>	54
4.12	Estimação global da taxa de símbolos dos sinais modulados, usando a FLOCAF com $a = 0,3$ e $b = 0,5$	55
4.13	Estimação global da taxa de símbolos dos sinais modulados, usando a CCF com $\sigma = 1,2$	56
5.1	Arquitetura de AMC baseada na extração de características cicloestacionárias aplicada a regressão logística.	58
5.2	Curvas de classificação da arquitetura baseada na FLOCAF ($a = 0,3$ e $b = 0,5$) combinada com regressão logística.	61
5.3	Curvas de classificação da arquitetura baseada no perfil-cíclico reduzido da FLOCAF ($a = 0,3$ e $b = 0,5$) combinada com regressão logística, considerando 1000 testes por modulação em cada condição de <i>GSNR</i>	63
5.4	Curvas de classificação da arquitetura baseada na CCF ($\sigma = 1.2$) combinada com regressão logística.	65
5.5	Curvas de classificação da arquitetura baseada no perfil-cíclico reduzido da CCF ($\sigma = 1.2$) combinada com regressão logística.	67
5.6	Curvas de classificação global de diversas arquiteturas de AMC.	69
C.1	Entropia de Shannon do lançamento de uma moeda	88
C.2	Entropia de Rényi do lançamento de uma moeda para diferentes valores de β	90
C.3	Aplicação do estimador de Parzen sobre um conjunto de dados utilizando kernels gaussianos de tamanhos diferentes.	92

Lista de Tabelas

2.1	Variáveis de entrada e saída de um conjunto de treinamento.	25
2.2	Variáveis de entrada e saída de um conjunto de treinamento de múltiplas características.	27
2.3	Conjunto de treinamento de alvo binário.	28
2.4	Conjunto de treinamento de alvo de múltiplas classes.	31
2.5	Conjunto de treinamento de alvo reestruturado.	32
3.1	Parâmetros de Simulação.	38
4.1	Limiares de classificação da arquitetura que emprega a FLOCAF.	47
4.2	Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0, 3, b = 0, 5, GSNR = 0$ dB).	48
4.3	Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0, 3, b = 0, 5, GSNR = 4$ dB).	48
4.4	Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0, 3, b = 0, 5, GSNR = 10$ dB).	48
4.5	Limiares da arquitetura de AMC baseada na CCF.	51
4.6	Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1, 2, GSNR = 0$ dB).	52
4.7	Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1, 2, GSNR = 4$ dB).	53
4.8	Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1, 2, GSNR = 10$ dB).	53
5.1	Parâmetros de Simulação.	60
5.2	Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 0$ dB).	61
5.3	Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 4$ dB).	62
5.4	Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 10$ dB).	62
5.5	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 0$ dB).	63
5.6	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 4$ dB).	64
5.7	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0, 3, b = 0, 5, GSNR = 10$ dB).	64

5.8	Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 0$ dB).	65
5.9	Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 4$ dB).	66
5.10	Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 10$ dB).	66
5.11	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 0$ dB).	67
5.12	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 4$ dB).	68
5.13	Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 10$ dB).	68
D.1	Banco de dados de alvo binário.	97

Lista de Abreviaturas

AMC	<i>Automatic Modulation Classification</i> (Classificação Automática de Modulação)
AWGN	<i>Additive White Gaussian Noise</i> (Ruído Gaussiano Branco Aditivo)
BPSK	<i>Binary Phase-Shift Keying</i> (Modulação por Chaveamento de Fase Binária)
CAF	<i>Cyclic Autocorrelation Function</i> (Função de Autocorrelação Cíclica)
CCF	<i>Cyclic Correntropy Function</i> (Função de Correntropia Cíclica)
FB	<i>Feature-Based</i> (Classificação Baseada em Análise de Características)
FFT	<i>Fast Fourier Transform</i> (Transformada Rápida de Fourier)
FLOC	<i>Fractional Lower-Order Covariance</i> (Covariância Fracionária de Ordem Inferior)
FLOCAF	<i>Fractional Lower-Order Cyclic Autocorrelation Function</i> (Função de Autocorrelação Cíclica Fracionária de Ordem Inferior)
GSNR	<i>Geometric Signal-to-Noise Ratio</i> (Relação Sinal-Ruído Geométrica)
HF	<i>High Frequency</i> (Alta Frequência)
LB	<i>Likelihood-Based</i> (Classificação Baseada em Análise de Verossimilhança)
PDF	Probability Density Function
QPSK	<i>Quadrature Phase-Shift Keying</i> (Modulação por Chaveamento de Fase em Quadratura)
RF	<i>Radio Frequency</i> (Radiofrequência)
RMS	<i>Root Mean Square</i> (Raiz do Valor Quadrático Médio)
SNR	<i>Signal-to-Noise Ratio</i> (Relação Sinal-Ruído)
SSCCF	<i>Spatial Sign Cyclic Correlation Function</i> (Função Sinal Espacial de Correlação Cíclica)
UHF	<i>Ultra High Frequency</i> (Frequência Ultra Alta)
VHF	<i>Very High Frequency</i> (Frequência Muito Alta)

Capítulo 1

Introdução

Em sistemas de comunicações, a classificação automática de modulação (*automatic modulation classification* - AMC) consiste em identificar o esquema desconhecido da modulação empregado pelo transmissor a partir de um sinal ruidoso captado pelo receptor, sem informação prévia, em um curto período tempo, e com uma taxa de acerto tão alta quanto possível (Zhu & Nandi 2015, Satija et al. 2017). Portanto, AMC é um estágio intermediário que existe entre a recepção e a demodulação do sinal, quando o receptor não possui a informação prévia do esquema de modulação empregado.

As aplicações em AMC tiveram origem no contexto militar, em que a guerra eletrônica, vigilância, e análise de ameaças demandam o reconhecimento de sinais modulados para identificar unidades de transmissão adversárias, com o objetivo de recuperar o sinal interceptado e fazer interferência (Zhu & Nandi 2015). Atualmente, sistemas de AMC também desempenham um importante papel em aplicações civis e comerciais, dando suporte ao desenvolvimento de novos paradigmas de comunicações. Um exemplo disso é o rádio cognitivo, sistema de gerenciamento dinâmico de espectro que possui flexibilidade para ajustar seus parâmetros de transmissão de acordo com as condições do canal (Amjad et al. 2018).

Os algoritmos de AMC podem ser classificados em duas categorias: métodos baseados na análise de verossimilhança (*Likelihood-Based* - LB) e métodos baseados em análise de características (*Feature-Based* - FB) (Zhu & Nandi 2015, Alharbi et al. 2012). Métodos LB fornecem uma solução ótima em termos de taxa de classificação correta. Entretanto, ele é associado a uma complexidade computacional elevada e demanda conhecimento *a priori* da informação estatística do sinal recebido, o que é inviável na prática. Por outro lado, embora sub-ótimo, a abordagem FB não necessita de informações *a priori*, além de ser mais simples de implementar do que a abordagem LB (Zhu & Nandi 2015, Satija et al. 2017, Wu et al. 2017).

Existem várias abordagens de AMC baseadas em extração de características, por exemplo, estimação de amplitude, fase e frequências instantâneas (De Vito et al. 2010, Nandi & Azzouz 1998), transformada wavelet (Avci et al. 2007, Avci & Avci 2008, Malitsos et al. 2007, Park et al. 2008), uso de características estatísticas tais como momentos de alta-ordem e cumulantes (Orlic & Dukic 2010, Wu et al. 2008, Dobre et al. 2005, Xin et al. 2010), análise cicloestacionária (Satija et al. 2017, Ramkumar 2009, Kim et al. 2007, Fehske et al. 2005), e aplicações de medidas de informação tal como correntropia (Fontes

et al. 2012). Particularmente, a análise cicloestacionária pode ser destacada como um método eficiente para a extração de características de sinais em sistemas de comunicações.

Por mais de meio século, a análise cicloestacionária vem sendo usada em vários campos do conhecimento, contudo o maior número de aplicações acontece no âmbito das comunicações (Gardner 1994). Operações típicas em sistemas de comunicações, tais como, modulação, multiplexação e codificação levam a manifestação de momentos estatísticos periódicos nos sinais transmitidos, os quais são então referidos como processos, ou sinais, cicloestacionários (Gardner et al. 2006). Métodos estatísticos clássicos, quando aplicados a sinais cicloestacionários, apenas fornecem uma análise limitada uma vez que consideram que o sinal analisado é estacionário. Contudo, técnicas baseadas em análise cicloestacionária são mais adequadas para processar sinais de comunicações (Fontes et al. 2017).

A análise cicloestacionária permite a extração das características espectrais cíclicas dos sinais de comunicações, também conhecidas como assinaturas cicloestacionárias, e pode ser eficientemente usada em ambientes gaussianos para sensoriamento espectral (Tani et al. 2016, Napolitano 2016), reconhecimento automático de modulações (Satija et al. 2014, Napolitano 2016, Spooner et al. 2017), e estimação de parâmetros do sinal (Napolitano 2016, Chen et al. 2016, Gardner et al. 2006).

Contudo, apesar de ser amplamente usado na literatura para modelar o ruído aditivo em canais de comunicações, a distribuição gaussiana é limitada, pois sistemas sem-fio (como por exemplo comunicações móveis via satélite) são propensos a interferência não-gaussiana (Win et al. 2009, Aalo et al. 2014, de Freitas et al. 2017) e a fontes diversas de ruído impulsivo (Bhatti et al. 2009, Button et al. 2002, Blackard et al. 1993).

Em tais casos, a distribuição alfa-estável modela apropriadamente os canais de comunicações, uma vez que possui propriedades que podem descrever ambientes gaussianos e não-gaussianos (Win et al. 2009, Gonzalez et al. 2006, Shao & Nikias 1993). Entretanto, no cenário não-gaussiano a distribuição alfa-estável não apresenta momentos estatísticos de segunda-ordem e ordem-superior definidos. Portanto, nesse contexto, as análises cicloestacionárias de segunda-ordem e ordem-superior, são inadequadas para a extração de características e análise de sinais.

As principais técnicas utilizadas para o processamento de séries temporais em ambientes impulsivos não-gaussianos são: a análise estatística fracionária de ordem-inferior (Tsihrintzis et al. 1996, Liu & Mendel 2001, Belkacemi & Marcos 2007, Hu et al. 2016); a medida de correntropia (Príncipe 2010, Wang et al. 2017); o estimador sinal-espacial de correlação cíclica (Lundén et al. 2010, Renard et al. 2013, Lunden & Koivunen 2014); análise cicloestacionária fracionária de ordem-inferior (You et al. 2013, Liu et al. 2015); a correntropia cíclica (Luan et al. 2016, Liu et al. 2018).

Entretanto, as abordagens baseadas na estimação sinal-espacial de correlação cíclica e a análise cicloestacionária fracionária de ordem inferior não foram aplicadas a problemas de classificação de modulações. Além disso, não existe uma demonstração matemática que relacione essas duas abordagens.

Por sua vez, a correntropia cíclica foi utilizada com sucesso para o sensoriamento espectral de sinais modulados contaminados por ruído impulsivo (Luan et al. 2016, Fontes et al. 2017). Entretanto, investigações mais aprofundadas são necessárias para avaliar

o uso dessa técnica no reconhecimento de modulações digitais em ambientes impulsivos não-gaussianos, tais como a modulação de amplitude em quadratura m -ária (*M-ary Quadrature Amplitude Modulation* - MQAM) e modulação por chaveamento de fase m -ária (*M-ary Phase Shift Keying* - MPSK) de alta ordem, as quais são amplamente empregadas em inúmeros sistemas de comunicações (Bardyn et al. 2016, Lorincz & Begusic 2006, Stevenson et al. 2009).

Portanto, este trabalho analisa o uso da função de autocorrelação cíclica fracionária de ordem-inferior (*Fractional Lower-Order Cyclic Autocorrelation Function* - FLOCAF), e da função de correntropia cíclica (*Cyclic Correntropy Function* - CCF) para a obtenção de descritores que permitam o reconhecimento das modulações: PSK binária (*Binary PSK* - BPSK), PSK em quadratura (*Quadrature PSK* - QPSK), 8-QAM, 16-QAM e 32-QAM. Além disso, arquiteturas de classificação automática de modulações baseadas nessas funções cicloestacionárias também são desenvolvidas neste trabalho.

Nesse contexto, classificadores automáticos de modulações podem ser avaliados com base em alguns fatores de qualidade, tais como (Zhu & Nandi 2015):

- Acurácia. Também conhecida como precisão do classificador, esse fator é medido considerando o taxa de erro sobre o total de tentativas de classificação. Nessa conjuntura, quanto menor a taxa de erro melhor será o classificador;
- Robustez. Dado que o canal de comunicações pode ser imprevisível, principalmente canais de comunicações sem-fio, os classificadores de modulações precisam ser consistentes em várias configurações (situações) de canal, da condição de comunicações mais favorável a mais adversa;
- Eficiência computacional. O custo computacional pode ser medido em dois aspectos relacionados ao desempenho dos sistemas: um algoritmo complexo de AMC requer um *hardware* mais potente para suportá-lo; um algoritmo complexo poderá demandar um tempo de processamento muito elevado, o que dificulta sua aplicação em sistemas de classificação em tempo real.

Dessa forma, quanto mais eficiente (de menor complexidade computacional para o processamento e para o treinamento), mais robusto, e mais preciso for o algoritmo de classificação, mais relevante será sua contribuição para o estado da arte de sistemas de comunicações inteligentes. Portanto, visando atender a esses critérios, as arquiteturas de classificação propostas neste trabalho, projetadas para o reconhecimento das modulações digitais (BPSK, QPSK, 8-QAM, 16-QAM, e 32-QAM) no cenário de comunicações sem-fio sujeitas a ruído impulsivo, foram baseadas nas funções cicloestacionárias FLOCAF e CCF combinados com estratégias simples e diretas de operação.

Inicialmente, são propostas duas arquiteturas que funcionam como uma árvore de decisão, as quais aplicam critérios de casamento de *templates*¹, e análise de limiares no processo de classificação (as duas arquiteturas se diferenciam apenas pelas funções cicloestacionárias aplicadas, uma utiliza a FLOCAF e a outra utiliza a CCF). Ambas as arquiteturas foram testadas em diferentes condições de canal, variando o nível de ruído impulsivo. Embora as duas tenham apresentado certa robustez à contaminação do ruído impulsivo, a arquitetura baseada na CCF teve a maior taxa de acerto.

¹Esses termo faz referência a padrões, ou características, pré-determinadas pelo sistema de classificação.

Em seguida, com o intuito de melhorar o desempenho na classificação das modulações propostas em ambientes de ruído impulsivo, uma técnica de aprendizagem denominada regressão logística foi combinada com as funções de análise cicloestacionária, FLOCAF e CCF, para o desenvolvimento de novas arquiteturas de AMC. A regressão logística é um dos modelos de aprendizagem de máquina mais simples, o qual permite uma interpretação probabilística dos dados, e em diversos contextos, apresentam resultados de classificação precisos (Ahmed et al. 2020, Chang et al. 2019, Khan et al. 2019, Hosmer Jr et al. 2013).

Durante o processo de elaboração das arquiteturas que utilizam regressão logística, foi desenvolvido a técnica de análise cicloestacionária denominada perfil-cíclico reduzido, a qual contribuiu na redução do número de características do conjunto de treinamento e, por sua vez, resultou em um treinamento menos custoso. As arquiteturas baseadas na regressão logística, quando comparadas entre si, novamente evidenciaram a superioridade da CCF em relação a FLOCAF para a caracterização (extração de descritores) das modulações digitais. Além disso, quando comparadas às arquiteturas baseadas em árvore de decisão, os classificadores que utilizaram regressão logística apresentaram taxas de acerto superiores.

1.1 Contribuições da Tese

Entre as contribuições desse trabalho, se destacam:

- A demonstração matemática de que a função sinal-espacial de correlação cíclica (*Spatial Sign Cyclic Correlation Function* - SSCCF) é um caso particular da FLOCAF;
- Uma análise detalhada dos espectros cíclico das modulações BPSK, QPSK, 8-QAM, 16-QAM, e 32-QAM, obtidos pela FLOCAF e CCF, para ambientes com e sem ruído impulsivo aditivo;
- Proposição de arquiteturas de AMC baseadas em extração de características cicloestacionárias pela FLOCAF e CCF, e estruturadas como árvore de decisão;
- Otimização dos parâmetros da FLOCAF e CCF que maximizam o desempenho das arquiteturas de classificação;
- Um método de estimação da taxa de símbolo dos sinais modulados contaminados por ruído impulsivo, feito a partir dos descritores cíclicos obtidos pela FLOCAF e CCF;
- Proposição de arquiteturas de AMC que combinam a análise cicloestacionária com regressão logística;
- Proposição de uma técnica de otimização do treinamento da regressão logística, denominada análise do perfil-cíclico reduzido.

1.2 Organização da Tese

Esse trabalho é organizado da seguinte forma: o Capítulo 2 desenvolve a fundamentação teórica dos principais conceitos abordados na tese: distribuição alfa-estável, aná-

lise cicloestacionária de segunda ordem, análise cicloestacionária fracionária de ordem inferior, análise cicloestacionária por correntropia, estrutura de classificação de um árvore de decisão, e classificação por regressão logística. O Capítulo 3 contempla a análise espectral cicloestacionária das modulações BPSK, QPSK, 8-QAM, 16-QAM, 32-QAM em ambientes com e sem ruído impulsivo. O Capítulo 4 apresenta arquiteturas de AMC (projetadas para o reconhecimento das modulações digitais abordadas no Capítulo 3), as quais utilizam análise espectral cíclica através da FLOCAF e da CCF, estruturadas como uma árvore de decisão. Além disso, esse Capítulo também introduz um método, baseado em análise cicloestacionária, para a estimação da taxa de símbolo de modulações digitais. O Capítulo 5 apresenta arquiteturas de AMC (projetadas para o reconhecimento das modulações digitais abordadas no Capítulo 3) baseadas na FLOCAF e na CCF combinada com regressão logística. O Capítulo 6 apresenta as conclusões da tese, artigos aceitos e submetidos, além das perspectivas de trabalhos futuros.

Capítulo 2

Fundamentação Teórica

As arquiteturas de AMC propostas neste trabalho são voltadas para sistemas de comunicação sem-fio que operam em canais de ruído impulsivo aditivo. Nesse contexto, é assumido o modelo do receptor super-heteródino, onde a arquitetura de AMC se encontra antes da etapa de demodulação com forme ilustrado na Figura 2.1.

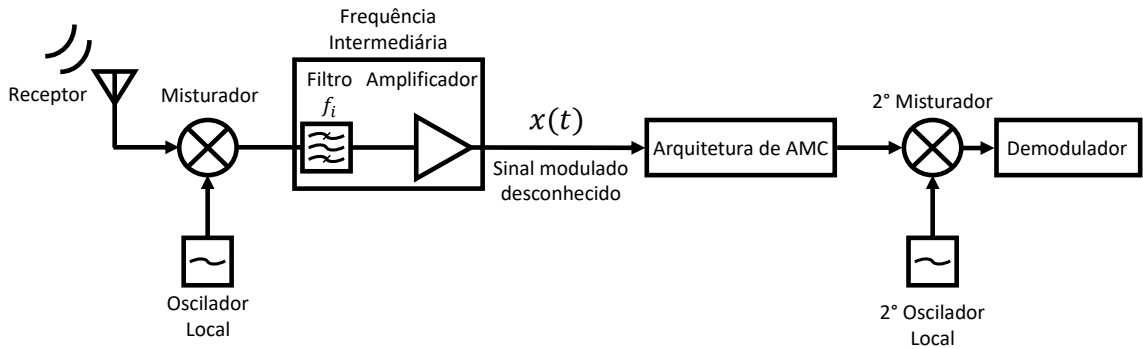


Figura 2.1: Sistema de comunicação sem-fio em ambiente impulsivo.

Por sua vez, o sinal desconhecido avaliado pelo classificador é expresso por

$$x(t) = s(t) + n(t), \quad (2.1)$$

em que, $n(t)$ é o ruído impulsivo aditivo e o sinal modulado é denotado por

$$s(t) = A(t)\cos(2\pi f_i t + \theta(t)). \quad (2.2)$$

O termo $A(t)$ é a envoltória do sinal, $\theta(t)$ é a sua fase e f_i representa uma frequência intermediária ¹. Além disso, as arquiteturas de AMC desenvolvidas são baseadas em análise de características cicloestacionárias, contudo, adotam diferentes estratégias de classificação, uma baseada em árvore de decisão e outra baseada em regressão logística.

Portanto, uma vez que o ambiente impulsivo de comunicação e as arquiteturas de AMC propostas neste trabalho envolvem conceitos teóricos diversos, neste Capítulo são

¹Todos esses parâmetros são definidos de acordo com o formato de modulação adotado para a transmissão.

respectivamente apresentados: os princípios associados de ruídos impulsivos modelados por distribuições alfa-estáveis; a análise das funções cicloestacionárias CAF, FLOCAF, e CCF; a estrutura básica de classificadores por árvore de decisão; e a classificação por regressão logística.

2.1 Ruído Impulsivo

O modelo de ruído gaussiano branco aditivo (*Additive White Gaussian Noise* - AWGN) é amplamente empregado em diversos trabalhos para representar o comportamento do ruído aditivo em sistemas de comunicações sem-fio. Isso se justifica pois o ruído gaussiano, fundamentado pelo teorema do limite central, descreve inúmeros eventos de efeitos práticos na natureza. Por essa razão muitas das técnicas de processamento de sinais aleatórios, empregadas em sistemas de comunicações, se limitam à aplicações do modelo AWGN (Middleton 1999). Além disso, a suposição de gaussianidade também é bastante comum em testes de hipóteses e análise de regressão linear (Charitidou et al. 2015).

Embora o modelo AWGN seja amplamente difundido em diversos cenários da ciência, existem aplicações práticas que são melhor modeladas por distribuições não-gaussianas. É o caso dos ambientes impulsivos, que podem ser descritos por diversos modelos estatísticos. Trabalhar com a suposição de gaussianidade em ambientes impulsivos pode gerar uma degradação expressiva no desempenho dos sistemas avaliados (Zhang et al. 2013).

Ambientes impulsivos podem melhor descrever o comportamento de vários cenários de comunicações sem-fio, desde redes internas (*indoor*) que conectam um número baixo de usuários e que incluem pequenas áreas como escritórios, laboratórios e residências, até redes mais amplas (*outdoor*) que abrangem áreas extensas utilizando faixas de alta frequência (*High Frequency* - HF), frequência muito alta (*Very High Frequency* - VHF) e frequência ultra alta (*Ultra High Frequency* - UHF) (Blackard et al. 1993).

Ruídos impulsivos em redes de comunicações sem-fio *indoor* (em escritórios), podem ser gerados por alguns aparelhos, tais como: fotocopiadoras, impressoras, comutadores de portas de elevador, limpador de piso a gás e compressores de refrigeração (Blackard et al. 1993). Por sua vez, sistemas de comunicações sem-fio *outdoor* podem ter seu desempenho severamente degradado devido a interferência impulsiva de transientes aleatórios, provenientes de descargas elétricas parciais e arcos elétricos (Au et al. 2015, Middleton 1999, Spaulding & Middleton 1977, Blum et al. 1999). Esses eventos resultam em formas de onda que podem ocupar uma ampla faixa de frequências causando distorções na conexão de sistemas de comunicações sem-fio tradicionais (Madi et al. 2010, Ndo et al. 2013, Bhatti et al. 2009).

Ambientes com ruídos atmosféricos, presentes por exemplo em comunicações HF, também podem ser considerados impulsivos, e por isso não são bem representados por modelos puramente gaussianos (Ilizzo & Paura 1981). A comunicação em HF é alvo de várias fontes de impulsividade devido a alta variação das características ionosféricas que mudam com o tempo e com o clima, muitas delas provenientes de tempestades de raios, relâmpagos e até ecos de chuva de meteoros (Alharbi et al. 2012, Turley 2003, Giesbrecht et al. 2003).

Uma série de modelos estatísticos diversos podem ser usados para descrever um ambiente impulsivo, tais como: misturas de gaussianas (Kozick & Sadler 2000a, Kozick & Sadler 2000b), a distribuição gaussiana generalizada (Novey et al. 2010), o modelo bimodal (Gonzalez et al. 2006) e a distribuição alfa-estável (Nikias & Shao 1995, Gonzalez et al. 2006, Pokharel et al. 2009).

Particularmente, a distribuição alfa-estável se destaca por permitir a caracterização de vários cenários de impulsividade, a partir do controle dos parâmetros da distribuição. Desse modo, diversos ambientes podem ser representados por esse modelo, tais como: canais acústicos (*indoor/outdoor*) (Pena et al. 2019, Pena et al. 2020), acústica subaquática, interferências em radares, interferência em sistemas de espalhamento espectral e ondas contínuas (Middleton 1999), problemas de localização de fontes (Liu et al. 2015, Liu et al. 2012, Ma & Nikias 1996), estimação doppler (Belkacemi & Marcos 2007, Tsihrintzis et al. 1996), além de diversos contextos de telecomunicações (Shao & Nikias 1993).

2.1.1 Distribuição Alfa-Estável

A teoria de distribuições estáveis univariada foi essencialmente desenvolvida nas décadas de 1920 e 1930, por Paul Lèvy e Aleksander Yakovlevich Khinchine, e posteriormente detalhada em 1954 por Gnedenko e Kolmogorov (Samorodnitsky & Taqqu 1994). Ela possui um conjunto de definições (as quais são equivalentes entre si), além de teoremas desenvolvidos a partir das mesmas (Samorodnitsky & Taqqu 1994):

Definição 1 Uma variável aleatória X é dita possuir uma distribuição estável se para quaisquer números positivos A e B , existe um número positivo C e um número real D , tal que

$$AX_1 + BX_2 \stackrel{d}{=} CX + D, \quad (2.3)$$

em que, X_1 e X_2 são variáveis independentes e identicamente distribuídas (i.i.d.) a X , e o termo $\stackrel{d}{=}$ denota igualdade de distribuições. Nesse contexto, a variável X é chamada de estritamente estável quando $D = 0$. Além disso, uma variável X será chamada de estável simétrica se sua distribuição for simétrica, isso é, se X possuir a mesma distribuição de $-X$ (Samorodnitsky & Taqqu 1994).

Teorema 1 Para qualquer variável aleatória X , existe um número $\alpha \in (0, 2]$ que garante que o número C , da Equação 2.3, satisfaça a seguinte relação (Samorodnitsky & Taqqu 1994)

$$C^\alpha = A^\alpha + B^\alpha. \quad (2.4)$$

Toda variável aleatória estável X que possui índice α , é chamada de alfa-estável.

Definição 2 Uma variável aleatória X é dita possuir uma distribuição estável se para qualquer inteiro $n \geq 2$, existe um número positivo C_n e um número real D_n , tal que

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} C_n X + D_n, \quad (2.5)$$

em que, $X_1, X_2, \dots, X_n$ são i.i.d em relação a X .

Definição 3 Uma variável aleatória X é dita possuir uma distribuição estável se ela apresentar um domínio de atração, isto é, se existir uma sequência de variáveis aleatórias i.i.d $Y_1, Y_2, \dots, Y_n$, e uma sequência positiva de números d_n e números reais a_n , tal que

$$\frac{Y_1 + Y_2 + \dots + Y_n}{d_n} + a_n \xrightarrow{d} X. \quad (2.6)$$

em que, $\xrightarrow{d}$ denota convergência da distribuição.

Teorema 2 A relação descrita pela Equação 2.6 da Definição 3 pode ser estendida para o limite de $n \rightarrow \infty$ (Shao & Nikias 1993). Em outras palavras, a soma de infinitas (inúmeras) variáveis aleatórias independentes e identicamente distribuídas converge para uma variável aleatória X distribuição estável. Esse teorema é conhecido como teorema do limite central generalizado.

Definição 4 Uma variável aleatória X é dita possuir uma distribuição estável se sua função característica apresentar a seguinte forma (Samorodnitsky & Taqqu 1994)

$$\phi_X(\omega, \alpha, \beta, \gamma, \delta) = \exp \{ \gamma^\alpha [-|\omega|^\alpha + j\omega\Lambda(\omega, \alpha, \beta)] + j\delta\omega \}, \quad (2.7)$$

em que

$$\Lambda(\omega, \alpha, \beta) = \begin{cases} \beta|\omega|^{\alpha-1} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\beta\frac{2}{\pi} \ln|\omega| & \alpha = 1. \end{cases} \quad (2.8)$$

Nesse contexto,

- α é o expoente característico ou índice de estabilidade da distribuição, o mesmo apresentado na Equação 2.4 ($0 < \alpha \leq 2$). Esse parâmetro é particularmente interessante, pois ele indica a velocidade com que as caudas da função densidade de probabilidade (*Probability Density Function* - PDF), associada a $\phi_X(\omega, \alpha, \beta, \gamma, \delta)$, convergem. Quando $\alpha = 2$, as caudas da distribuição têm um decaimento exponencial, quando $\alpha < 2$ as caudas convergem polinomialmente (com decaimento mais lento conforme o expoente característico se torna menor). Esse comportamento é ilustrado na Figura 2.2;
- δ é o parâmetro de localização do centro da distribuição associada a $\phi_X(\omega, \alpha, \beta, \gamma, \delta)$. Quando $\delta = 0$, a distribuição é dita centralizada;
- β é o parâmetro de simetria da PDF da variável alfa-estável X ($-1 \leq \beta \leq 1$). No caso de $\beta = 0$, a distribuição será dita alfa-estável simétrica. Quando $\beta > 0$, a distribuição será assimétrica para a direita, e será assimétrica para a esquerda quando $\beta < 0$;

- γ é o parâmetro de escala (ou dispersão), que indica a abertura da distribuição.

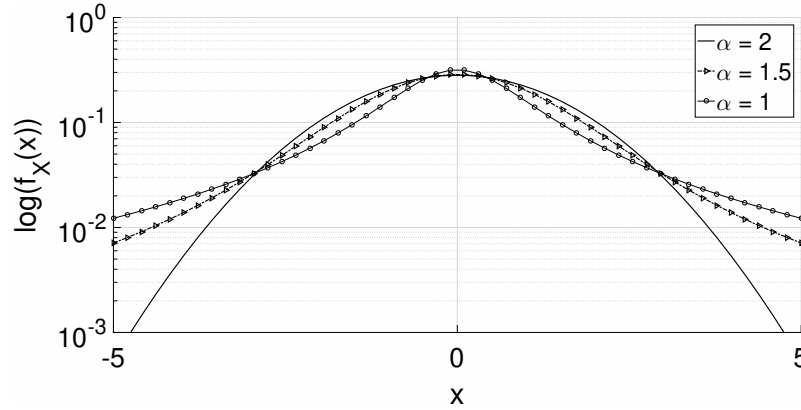


Figura 2.2: Funções densidades de probabilidade em escala logarítmica para variáveis aleatórias alfa-estáveis (X) com diferentes expoentes característicos α , $\beta = 0$, $\gamma = 1$, e $\delta = 0$. Quanto menor o valor de α mais lento será o decaimento da curva, e portanto, a impulsividade do ruído será maior.

Para alguns valores específicos de α e β , a distribuição de uma variável aleatória alfa-estável pode ser reduzida a distribuições de probabilidade conhecidas como a distribuição gaussiana ² ($\alpha = 2$, $\beta = 0$), a distribuição de Cauchy-Lorentz ($\alpha = 1$, $\beta = 0$), e a distribuição de Lèvy ($\alpha = 0,5$, $\beta = 1$). Com exceção desses casos, distribuições alfa-estáveis não possuem expressão fechada para caracterizar suas funções densidade de probabilidade $f_X(x)$. Todavia, elas podem ser calculadas a partir da transformada inversa de Fourier da função característica definida na Equação 2.7 (Shao & Nikias 1993) ³.

Devido à importância da função característica (Equação 2.7) na descrição do comportamento de variáveis aleatórias alfa-estáveis X , sua distribuição $f_X(x)$ é usualmente relacionada aos parâmetros α, β, δ e γ (Samorodnitsky & Taqqu 1994). Desse modo, a seguinte notação pode ser utilizada para indicar que uma variável aleatória X possui distribuição alfa-estável

$$X \sim S_\alpha(\gamma, \beta, \delta).$$

No caso de distribuições alfa-estáveis simétricas e centralizadas ($\beta = \delta = 0$), a representação pode ser feita da seguinte forma:

²Particularmente, a distribuição alfa-estável com $\alpha = 2, \beta = 0, \gamma = \sigma/\sqrt{2}, \delta = \mu$ se reduz a forma gaussiana. A demonstração da convergência é feita no Apêndice A.

³Atualmente diversas *toolbox* voltadas para modelagem estatística de dados copilam um conjunto de métodos numéricos para estimação de distribuições alfa-estáveis (a partir da função característica $\phi_X(\omega, \alpha, \beta, \gamma, \delta)$). O Apêndice B descreve o processo de modelagem da distribuição alfa-estável utilizada neste trabalho.

$$X \sim S_\alpha S.$$

Além disso, variáveis aleatórias alfa-estáveis também apresentam propriedades que permitem a interpretação dos parâmetros α, β, δ e γ da função característica, são elas:

Propriedade 1 Sejam X_1 e X_2 variáveis aleatórias independentes com $X_i \sim S_\alpha(\gamma_i, \beta_i, \delta_i)$, $i = 1, 2$. Então $X_1 + X_2 \sim S_\alpha(\gamma, \beta, \delta)$, com (Samorodnitsky & Taqqu 1994):

$$\gamma = (\gamma_1^\alpha + \gamma_2^\alpha)^{1/\alpha}, \quad (2.9)$$

$$\beta = \frac{\beta_1 \gamma_1^\alpha + \beta_2 \gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}, \quad (2.10)$$

e

$$\delta = \delta_1 + \delta_2. \quad (2.11)$$

Propriedade 2 Seja $X \sim S_\alpha(\gamma, \beta, \delta)$ e seja a uma constante real. Então (Samorodnitsky & Taqqu 1994):

$$X + a \sim S_\alpha(\gamma, \beta, \delta + a). \quad (2.12)$$

Propriedade 3 Seja $X \sim S_\alpha(\gamma, \beta, \delta)$ com $0 < \alpha < 2$. Então (Samorodnitsky & Taqqu 1994):

$$E\{|X|^p\} = \infty, \text{ se } p \geq \alpha, \quad (2.13)$$

e

$$E\{|X|^p\} < \infty, \text{ se } 0 \leq p < \alpha. \quad (2.14)$$

Variáveis aleatórias alfa-estáveis, com $\alpha < 2$, possuem momentos de segunda ordem (e ordem superior) infinitos. Isso significa que muitas técnicas, válidas para o caso gaussiano, não serão efetivas para a análise nesse contexto pois dependem do cálculo da variância da variável aleatória. Além disso, quando $\alpha \leq 1$, distribuições alfa-estáveis possuem $E\{|X|\} = \infty$, fator que compromete o uso do valor esperado (Samorodnitsky & Taqqu 1994). Em contra partida, no caso de $\alpha = 2$, todos momentos estatísticos da distribuição existem (Shao & Nikias 1993).

Ao analisar as definições de estabilidade de uma variável aleatória, pode-se constatar que a distribuição alfa-estável é uma generalização direta da distribuição gaussiana, em outras palavras, a distribuição gaussiana é um caso limitante das distribuições alfa-

estáveis (Shao & Nikias 1993). Isso evidencia que o teorema do limite central é um caso particular do teorema do limite central generalizado definido para variáveis aleatórias estáveis.

A família não-gaussiana das distribuições alfa-estáveis tem como principal característica a cauda de suas PDFs, as quais possuem decaimento mais lento do que as caudas da PDF gaussiana. Por essa razão, pode-se dizer que distribuições alfa-estáveis não-gaussianas possuem caudas longas (ou pesadas). Esse comportamento implica que, os valores mais distantes do centro da distribuição irão ocorrer com uma frequência maior no caso não-gaussiano do que no caso gaussiano.

Devido à característica de cauda longa, além da facilidade de parametrização (que pode ser ajustada para descrever casos diversos), distribuições alfa-estáveis não-gaussianas são usualmente empregadas na modelagem de canais de comunicações sujeitos ao ruído impulsivo de variância indeterminada (infinita) (Chen et al. 2016, de Freitas et al. 2017, Yang et al. 2018). Nesse contexto, o índice de estabilidade controla a impulsividade do modelo. Quanto menor o parâmetro α , mais impulsivo será o canal de comunicação.

Destaca-se ainda que, para canais de comunicações modelados pela distribuição alfa-estável não-gaussiana, a relação sinal-ruído (*Signal-to-Noise Ratio* - SNR) não pode ser usada como um indicador de qualidade de canal, pois uma vez que a variância⁴ do ruído é infinita (Propriedade 3, Equação 2.13), essa medida será sempre igual a zero. Nesses casos, o fator de qualidade será definido pela relação sinal-ruído geométrica (*Geometric Signal-to-Noise Ratio* - GSNR), a qual pode ser calculada como (Mahmood et al. 2012, Chuah et al. 2000)

$$GSNR = \frac{A^2}{2\gamma^2 C_g^{\frac{2}{\alpha}-1}}. \quad (2.15)$$

Em que, A é a raiz do valor médio quadrático (*Root Mean Square* - RMS) do sinal transmitido; $C_g = e^{C_e} \approx 1,78$ é a exponencial da constante de Euler ($C_e \approx 0,5772$) (Gonzalez et al. 2006). No caso gaussiano, C_g garante que a GSNR corresponda a SNR.

Neste trabalho, o ruído impulsivo aditivo alfa-estável é empregado na modelagem do canal de comunicações adotado para análise espectral cíclica apresentada no Capítulo 3, bem como na avaliação das arquiteturas de AMC propostas nos Capítulos 4 e 5.

2.2 Análise Cicloestacionária: CAF e FLOCAF

Muitos processos encontrados na natureza surgem de fenômenos periódicos. Embora não sejam funções periódicas no tempo, tais processos dão origem a dados aleatórios que possuem características estatísticas que variam periodicamente, e são chamados processos cicloestacionários (Gardner 1986). Por exemplo, em aplicações de telecomunicações, telemetria, radar e sonar, a periodicidade estatística dos processos associados a esses sistemas são originados por operações como modulação, multiplexação, e codificação (Gardner et al. 2006).

Por essa razão, a teoria cicloestacionária permite uma análise eficiente dos processos

⁴Quando o sinal possui média nula, sua variância será equivalente a sua potência.

estatísticos nesses cenários. Em particular, no contexto de telecomunicações, a análise cicloestacionária vem sendo usada com sucesso em aplicações como estimação de parâmetros de sinais (Napolitano 2016, Chen et al. 2016, Gardner et al. 2006), sensoriamento espectral (Tani et al. 2016, Napolitano 2016) e reconhecimento automático de modulação (Satija et al. 2014, Napolitano 2016, Spooner et al. 2017).

Nessa seção são apresentados os conceitos das seguintes funções de análise cicloestacionária: a função de autocorrelação cíclica (*Cyclic Autocorrelation Function* - CAF) e a função de autocorrelação cíclica fracionária de ordem-inferior (*Fractional Lower-Order Cyclic Autocorrelation Function* - FLOCAF).

2.2.1 Função de autocorrelação cíclica

Um processo aleatório $X(t)$ é dito cicloestacionário em sentido amplo quando seu valor esperado $E\{X(t)\}$ e sua função de autocorrelação $R_X(t, \tau)$ são ambos periódicos em t com um período T para qualquer valor de τ (Gardner et al. 2006), isto é:

$$E\{X(t+T)\} = E\{X(t)\}, \quad (2.16)$$

$$R_X(t+T, \tau) = R_X(t, \tau), \quad (2.17)$$

em que

$$R_X(t, \tau) = E\{X(t)X^*(t+\tau)\}. \quad (2.18)$$

Assim, em um processo cicloestacionário de segunda-ordem, como a função de autocorrelação é periódica, esta pode ser expandida em termos da série de Fourier como:

$$R_X(t, \tau) = \sum_{\epsilon} R_X^{\epsilon}(\tau) e^{j2\pi\epsilon t}, \quad (2.19)$$

em que $R_X^{\epsilon}(\tau)$ representam os coeficientes da série de Fourier, dados por:

$$R_X^{\epsilon}(\tau) \triangleq \frac{1}{T} \int_{-T/2}^{T/2} R_X(t, \tau) e^{-j2\pi\epsilon t} dt. \quad (2.20)$$

Tais coeficientes definem a CAF, em que ϵ é um parâmetro discreto chamado de frequência cíclica ($\epsilon = n/T, \forall n \in \mathbb{Z}$). A CAF pode ser interpretada como uma função analítica que verifica se um processo $X(t)$ é cicloestacionário avaliando se $R_X^{\epsilon}(\tau)$ é não-nulo ao menos para um $\epsilon \neq 0$. Por outro lado, quando $R_X(t, \tau)$ possui múltiplas frequências fundamentais, $X(t)$ será chamado de processo poli-cicloestacionário (Gardner 1994). Neste caso, a CAF pode ser expressa por (Gardner 1994):

$$R_X^{\epsilon}(\tau) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} R_X(t, \tau) e^{-j2\pi\epsilon t} dt. \quad (2.21)$$

Embora a análise cicloestacionária de segunda-ordem possua várias aplicações, ela não é adequada para processos alfa-estáveis não-gaussianos, uma vez que esses apresen-

tam momento de segunda-ordem infinito (Equação 2.13). Esse aspecto pode ser observado na Figura 2.3, que mostra duas representações gráficas da CAF aplicada a um sinal BPSK em contexto de contaminações diferentes. Tais representações consideram apenas as frequências cíclicas ϵ , e são conhecidas como assinaturas cicloestacionárias, ou perfis-cíclicos⁵.

O perfil-cíclico da Figura 2.3 (a) foi obtido de um sinal BPSK sem contaminação. Por sua vez, o perfil da Figura 2.3 (b) foi obtido a partir do mesmo sinal BPSK contaminado por um ruído aditivo alfa-estável simétrico e centralizado com $GSNR = 10$ dB e $\alpha = 1,5$. Quando o sinal BPSK é contaminado pelo ruído, cujo o momento de segunda ordem é indeterminado, não é possível obter características cicloestacionárias de segunda-ordem.

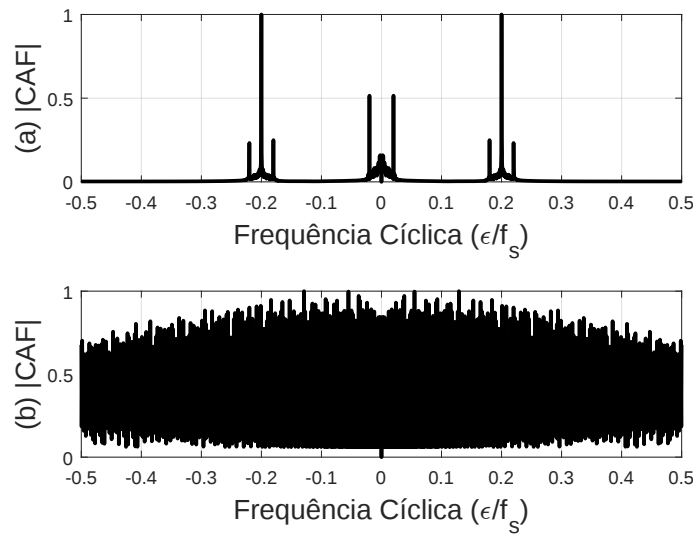


Figura 2.3: Perfis-cíclicos da função de autocorrelação cíclica aplicada a um sinal BPSK, respectivamente sem e com contaminação alfa-estável.

2.2.2 Função de autocorrelação cíclica fracionária de ordem-inferior

Uma vez que sinais contaminados por um ruído aditivo alfa-estável não-gaussiano não podem ser analisados por momentos estatísticos de ordem $p \geq 2$, a caracterização estatística de tais sinais pode ser feita usando momentos estatísticos de ordem-inferior, em particular, aplicando a covariância fracionária de ordem-inferior (*Fractional Lower-Order Covariance* - FLOC), a qual pode ser definida como (Liu et al. 2015):

$$R_X^{a,b}(t, \tau) = E \left\{ \left[X(t) |X(t)|^{(a-1)} \right] \times \left[X^*(t + \tau) |X^*(t + \tau)|^{(b-1)} \right] \right\}, \quad (2.22)$$

⁵O processo de extração de um perfil-cíclico, de um sinal modulado, é detalhado na Seção 2.2.3.

em que, a e b são os parâmetros fracionários de $R_X^{a,b}(t, \tau)$. Diferentes dos momentos de segunda-ordem, a FLOC é finita quando aplicada a distribuições alfa estáveis simétricas sempre que a seguinte relação for atendida (Liu et al. 2015)

$$0 \leq a, b < \alpha/2. \quad (2.23)$$

Quando incorporado à análise cicloestacionária, a FLOC dá origem à função de auto-correlação cíclica fracionária de ordem inferior (FLOCAF), a qual é definida como (Liu et al. 2015)

$$R_X^{\varepsilon, ab}(\tau) \triangleq \frac{1}{T} \int_{-T/2}^{T/2} R_X^{a,b}(t, \tau) e^{-j2\pi\varepsilon t} dt. \quad (2.24)$$

Neste caso, a FLOCAF permite verificar se um dado processo $X(t)$ é cicloestacionário, mesmo quando contaminado por um ruído de variância infinita, avaliando se $R_X^{\varepsilon, ab}(\tau)$ é não-nula para algum $\varepsilon \neq 0$.

Um aspecto interessante da FLOCAF pode ser visto quando seus parâmetros de momentos estatísticos de ordem-inferior (a, b) são zero. Neste caso, a FLOC pode ser representada como

$$R_X^{0,0}(t, \tau) = E \left\{ \frac{X(t)}{|X(t)|} \times \frac{X^*(t + \tau)}{|X^*(t + \tau)|} \right\}. \quad (2.25)$$

Contudo, para evitar a indeterminação quando $X(t)$ e $X(t + \tau)$ forem iguais a zero, $R_X^{0,0}$ deve ser representado em termos da função sinal espacial (Serneels et al. 2006, Lundén et al. 2010), definida por:

$$S(X(t)) = \begin{cases} \frac{X(t)}{|X(t)|} & X(t) \neq 0 \\ 0 & X(t) = 0. \end{cases} \quad (2.26)$$

Portanto, a Equação (2.25) pode ser reescrita como:

$$R_X^{0,0}(t, \tau) = E \{ S(X(t)) S(X^*(t + \tau)) \}. \quad (2.27)$$

Aplicando $R_X^{0,0}(t, \tau)$ à FLOCAF e assumindo que $X(t)$ é um processo poli-cicloestacionário, tem-se:

$$R_X^{\varepsilon, 0, 0}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} E \{ S(X(t)) S(X^*(t + \tau)) \} e^{-j2\pi\varepsilon t} dt, \quad (2.28)$$

e ainda, considerando que $R_X^{\varepsilon, 0, 0}(\tau)$ é poli-cicloergódico, é possível reescrever (2.28) como:

$$R_X^{\varepsilon, 0, 0}(\tau) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} S(X(t)) S(X^*(t + \tau)) e^{-j2\pi\varepsilon t} dt. \quad (2.29)$$

A Equação (2.29) é também conhecida como a função sinal espacial de correlação cíclica (*Spatial Sign Cyclic Correlation Function* - SSCCF), denotada por $R_S(\varepsilon, \tau)$ (Lundén et al. 2010, Lundén & Koivunen 2014).

Assim, demonstra-se aqui que a SSCCF pode ser vista como um caso particular da FLOCAF se os parâmetros a e b forem ambos iguais a zero. Além disso, a SSCCF sem-

pre irá convergir quando aplicada a distribuições alfa-estáveis, independente do valor do expoente característico (α) pois ele sempre atenderá a condição de estabilidade denotada em (2.23). Em outras palavras, essa função é mais robusta à presença de ruído impulsivo do que qualquer outra configuração da FLOCAF. Contudo, considerando todas as configurações possíveis da FLOCAF em termos de a e b , não há garantia de que a SSCCF retornará os melhores descritores cíclicos de um sinal.

2.2.3 Síntese da assinatura da FLOCAF

Computacionalmente, a análise espectral da FLOCAF pode ser feita a partir do seguinte estimador:

$$R_X^{a,b}[m, \kappa] \approx \frac{1}{M} \sum_{m=0}^{M-1} (X[m]|X[m]|^{(a-1)}) \times (X[m+\kappa]|X[m+\kappa]|^{(b-1)}) \quad (2.30)$$

em que, $X[m]$ é um sinal em tempo discreto, M é o número de observações e κ é o deslocamento de tempo discreto. Embora efetivo, o cálculo desse estimador pode ser custoso, pois ele é diretamente relacionado ao número de amostras do sinal. Em outras palavras, quanto maior o valor de M , maior será o tempo de processamento. Assim, a FLOCAF deve ser computada adotando os seguintes passos:

Passo 1. O sinal de entrada $X[m]$ é dividido em L blocos com N amostras. O ajuste de cada bloco pode ser feito utilizando uma janela retangular deslizante de tamanho fixo, desse modo, um dado bloco poderá conter amostras de interseção com o bloco anterior⁶;

Passo 2. Calcula-se o estimador da FLOC, $R_{X_l}^{a,b}[n, \kappa_n]$, para cada bloco l de tamanho N , em que $n = 0, 1, 2, \dots, N-1$ e $l = 1, 2, \dots, L$:

$$R_{X_l}^{a,b}[n, \kappa_n] \approx \frac{1}{N} \sum_{n=0}^{N-1} (X_l[n]|X_l[n]|^{(a-1)}) \times (X_l[n+\kappa_n]|X_l[n+\kappa_n]|^{(b-1)}); \quad (2.31)$$

Passo 3. Calcula-se a média dos blocos da FLOC:

$$R_X^{a,b}[n, \kappa_n] \approx \frac{1}{L} \sum_{l=1}^L R_{X_l}^{a,b}[n, \kappa_n]; \quad (2.32)$$

Passo 4. Para evitar o componente central do espectro, que não possui informação ciclo-estacionária, é necessário centralizar $R_X^{a,b}[n, \kappa_n]$:

⁶A interseção entre blocos é importante para suavizar o resultado da função.

$$R_X^{a,b}[n, \kappa_n] \approx R_X^{a,b}[n, \kappa_n] - \frac{1}{N} \sum_{n=0}^{N-1} R_X^{a,b}[n, \kappa_n]; \quad (2.33)$$

Passo 5. Por fim, o cálculo de $R_X^{\epsilon,ab}[\kappa_n]$ pode ser feito eficientemente usando o algoritmo da transformada rápida de Fourier (*Fast Fourier Transform* - FFT):

$$R_X^{\epsilon,ab}[\kappa_n] \approx \text{fft}\{R_X^{a,b}[n, \kappa_n]\}. \quad (2.34)$$

É importante perceber que o estimador da FLOCAF, $R_X^{\epsilon,ab}[\kappa_n]$, é uma superfície que carrega as informações cicloestacionárias do sinal para cada deslocamento κ_n .

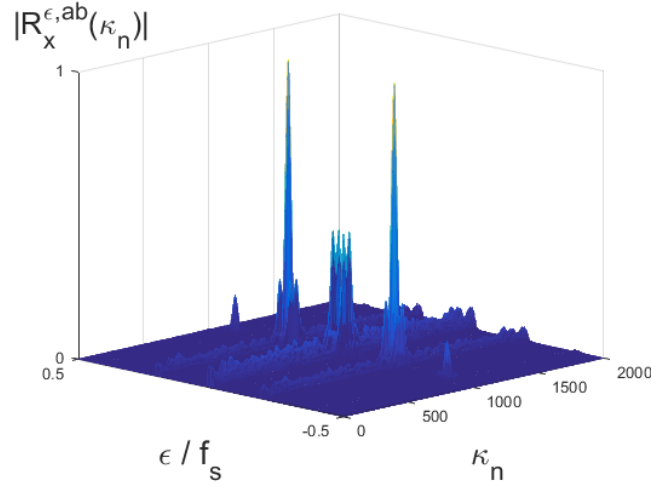


Figura 2.4: FLOCAF aplicada ao sinal BPSK.

A análise dos componentes cíclicos dessa superfície, exige um processamento da ordem de $O(N^2)$. Contudo, os componentes mais relevantes (aqueles que, entre todas as linhas de deslocamento, possuem a maior amplitude) permitem a análise cíclica completa do sinal. O conjunto formado por esses componentes possui apenas N elementos e é denominado perfil-cíclico, ou assinatura cicloestacionária sinal.

Desse modo, a assinatura cicloestacionária do sinal é obtida pela projeção dos maiores componentes cíclicos, dessa superfície, ao longo de κ_n :

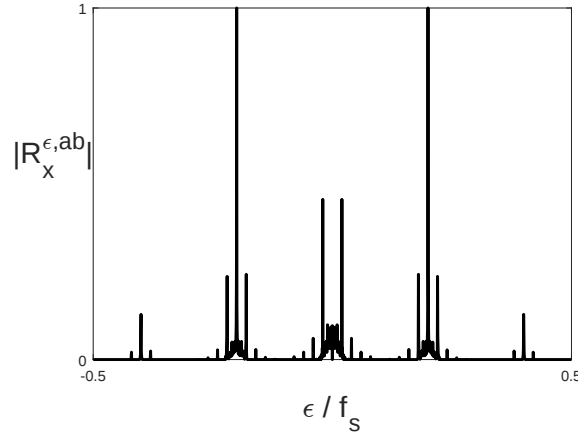


Figura 2.5: Perfil-cíclico da FLOCAF aplicada ao sinal BPSK.

Neste trabalho, as assinaturas cicloestacionárias de diferentes modulações digitais, obtidas pela FLOCAF, são cuidadosamente avaliadas no Capítulo 3. Além disso, as assinaturas cicloestacionárias obtidas pela FLOCAF também são utilizados como descritores de características das modulações digitais empregados nas arquiteturas de AMC propostas nos Capítulos 4 e 5.

2.3 Análise Cicloestacionária por Correntropia

A correntropia é uma função não-linear que indica o grau de relação entre variáveis ou processos aleatórios, é considerada a generalização da correlação, e por isso pode ser definida como uma medida não-linear de similaridade (Príncipe 2010)⁷. Nos últimos anos, essa ferramenta vem se destacando na resolução de problemas que envolvem detecção de padrões, aprendizagem de máquinas, entre outras aplicações na engenharia.

2.3.1 Função de Correntropia

A função de correntropia é uma transformação não-linear que relaciona processos aleatórios distintos a partir da seguinte relação (Príncipe 2010):

$$V_{X,Y}(t_1, t_2) = E\{k(X(t_1), Y(t_2))\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(X(t_1), Y(t_2)) f_{X(t_1), Y(t_2)}(X, Y) dx dy, \quad (2.35)$$

em que, $k(\cdot, \cdot)$ representa qualquer função contínua definida positiva, que relaciona (de forma não-linear) duas séries temporais X_{t_1} e Y_{t_2} (respectivamente equivalentes a $X(t_1)$ e $Y(t_2)$). Essa função é chamada de *kernel*.

O modelo de *kernel* comumente usado para análise de correntropia é o gaussiano

⁷ Os principais conceitos complementares associados a correntropia estão detalhados no Apêndice C.

$(G_\sigma(\cdot, \cdot))$, o qual além de ser uma função contínua definida positiva, é simétrico e invariante a translação. Nesse caso, a função de correntropia é denotada como:

$$V_{X,Y}(t_1, t_2) = E\{G_\sigma(X_{t_1}, Y_{t_2})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_\sigma(X_{t_1}, Y_{t_2}) f_{X_{t_1}, Y_{t_2}}(X, Y) dx dy, \quad (2.36)$$

em que

$$G_\sigma(X_{t_1}, Y_{t_2}) = G_\sigma(X(t_1), Y(t_2)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[X(t_1) - Y(t_2)]^2}{2\sigma^2}\right). \quad (2.37)$$

Nesse contexto, o parâmetro σ é denominado tamanho de *kernel*. Ele possui um importante papel no cálculo da função de correntropia, pois influencia diretamente na análise dos momentos estatísticos da função. Isso fica evidente expandindo a Equação 2.36 em série de Maclaurin

$$V_{X,Y}(t_1, t_2) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E\{(X(t_1) - Y(t_2))^{2n}\}. \quad (2.38)$$

A Equação (2.38) mostra que a função se relaciona a infinitos momentos de ordem par, os quais são inversamente ponderados pelo tamanho do *kernel*. O aumento do tamanho do *kernel* faz com que a contribuição dos momentos de alta ordem decaiam mais rápido, devido ao denominador da divisão que pondera as operações dentro do somatório, então o momento de segunda ordem tende a dominar a correntropia, fazendo com que ela se aproxime da correlação (Príncipe 2010).

Quando aplicado sobre um único processo aleatório, a ferramenta é chamada de função de autocorrentropia e é denotada por:

$$V_X(t, \tau) = E\{G_\sigma(X(t), X(t + \tau))\}. \quad (2.39)$$

2.3.2 Função de Correntropia Cíclica

Assim como no caso da função de autocorrelação, descrito na Seção 2.2.1, a função de autocorrentropia pode ser utilizada na análise de processos cicloestacionários. Desse modo, um processo aleatório $X(t)$ é dito cicloestacionário em sentido amplo quando sua função de autocorrentropia é periódica em t com um período T para qualquer valor de τ (Fontes et al. 2017), isto é

$$V_X(t + T, \tau) = V_X(t, \tau). \quad (2.40)$$

Uma vez que $V_X(t, \tau)$ é periódica, ela pode ser expandida em série de Fourier

$$V_X(t, \tau) = \sum_{\varepsilon} V_X^\varepsilon(\tau) e^{j2\pi\varepsilon t}, \quad (2.41)$$

em que, $V_X^\varepsilon(\tau)$ representa os coeficientes da série, calculados por

$$V_X^\varepsilon(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} V_X(t, \tau) e^{-j2\pi\varepsilon t} dt. \quad (2.42)$$

Esses coeficientes definem a função de correntropia cíclica (*Cyclic Correntropy Function* - CCF) (Fontes et al. 2017), e assim como no caso da CAF, o parâmetro ε representa as frequências cíclicas ($\varepsilon = n/T, \forall n \in \mathbb{Z}$).

Quando $V_X(t, \tau)$ possui múltiplas frequências fundamentais, então $X(t)$ é dito ser um processo poli-cicloestacionário (Fontes et al. 2017). Neste caso, a CCF pode ser definida como

$$V_X^\varepsilon(\tau) \triangleq \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} V_X(t, \tau) e^{-j2\pi\varepsilon t} dt. \quad (2.43)$$

Considerando o operador de média temporal $\langle \cdot \rangle$, equivalente a

$$\langle \cdot \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (\cdot) dt, \quad (2.44)$$

a Equação (2.43) pode ser reestruturada com a seguinte notação

$$V_X^\varepsilon(\tau) = \langle E \{ G_\sigma(X(t), X(t+\tau)) \} e^{-j2\pi\varepsilon t} \rangle. \quad (2.45)$$

Assumindo que $V_X^\varepsilon(\tau)$ é poli-cicloergódico (Fontes et al. 2017, Gardner 1994), é possível representar (2.45) como (Fontes et al. 2017):

$$\begin{aligned} V_X^\varepsilon(\tau) &= \langle G_\sigma(X(t), X(t+\tau)) e^{-j2\pi\varepsilon t} \rangle \\ &= \left\langle \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-[X(t) - X(t+\tau)]^2 - 2\sigma^2 j2\pi\varepsilon t}{2\sigma^2}\right) \right\rangle. \end{aligned} \quad (2.46)$$

Devido ao decaimento exponencial do *kernel* gaussiano mostrado na Equação (2.46), pode-se afirmar que a CCF sempre irá convergir, mesmo quando calculada sobre um processo $X(t)$ sujeito a um ruído impulsivo aditivo. Isso acontece porque o *kernel* gaussiano mapeia a diferença $[X(t) - X(t+\tau)]^2$ para o intervalo $(0, 1]$ ⁸.

Essa característica garante a robustez da função aos *outliers*, oriundos de canais ruidosos, pois eles possuem pouca influência no cálculo de $V_X^\varepsilon(\tau)$ uma vez que apresentam valores elevados que se distanciam do centro do *kernel* gaussiano. Outra característica importante associada à CCF poder ser observada expandindo (2.45) em uma série de Maclaurin

$$\begin{aligned} V_X^\varepsilon(\tau) &= \frac{1}{\sigma\sqrt{2\pi}} \left\langle \sum_{n=0}^{\infty} \frac{(-1)^n}{(2\sigma^2)^n n!} E \{ [(X(t) - X(t+\tau))^2 \right. \\ &\quad \left. + 2\sigma^2 j2\pi\varepsilon t]^n \} \right\rangle. \end{aligned} \quad (2.47)$$

Analisando (2.47), pode-se afirmar que $V_X^\varepsilon(\tau)$ contém informações de infinitos mo-

⁸No caso de deslocamento nulo, os dados serão mapeados para 1. Por sua vez, se a relação $[X(t) - X(t+\tau)]^2$ tende a infinito, a transformação tenderá a zero.

mentos estatísticos associados ao momento de segunda-ordem, os quais são fortemente relacionados ao tamanho do *kernel* (σ). Quanto maior o valor de σ , maior a contribuição dos momentos estatísticos de baixa-ordem no cálculo da CCF. Por outro lado, tamanhos de *kernel* pequenos farão com que os momentos estatísticos de alta-ordem tenham uma maior relevância no cálculo de $V_X^\varepsilon(\tau)$. Devido a essas características, a correntropia cíclica é adequada para extração de informação estatística de alta-ordem de processos cicloestacionários.

2.3.3 Síntese da assinatura da CCF

A função de correntropia cíclica definida na Equação (2.42) pode ser computada da seguinte forma

$$V_X^\varepsilon[\kappa] \approx \frac{1}{M} \sum_{m=0}^{M-1} G_\sigma(X[m], X[m+\kappa]) e^{-j2\pi\varepsilon m}, \quad (2.48)$$

em que, $X[m]$ é um sinal em tempo discreto, M é o número de observações, κ é o deslocamento de tempo discreto, e ε é a frequência cíclica. Contudo, assim como no caso da Seção 2.2.3, o cálculo eficiente da CCF deve ser feito por meio dos seguintes passos:

Passo 1. O sinal de entrada $X[m]$ é dividido em L blocos com N amostras. O ajuste de cada bloco pode ser feito utilizando uma janela retangular deslizante de tamanho fixo;

Passo 2. Calcula-se o estimador da função de correntropia $V_{X_l}[n, \kappa_n]$ para cada bloco l de tamanho N . Nesse caso, $n = 0, 1, 2, \dots, N-1$ e $l = 1, 2, \dots, L$:

$$V_{X_l}[n, \kappa_n] \approx \frac{1}{N} \sum_{n=0}^{N-1} G_\sigma(X_l[n], X_l[n+\kappa_n]); \quad (2.49)$$

Passo 3. Calcula-se a média dos blocos da função de correntropia:

$$V_X[n, \kappa_n] \approx \frac{1}{L} \sum_{l=1}^L V_{X_l}[n, \kappa_n]; \quad (2.50)$$

Passo 4. Para evitar o componente central do espectro, que não carrega informação cíclica, é necessário centralizar $V_X[n, \kappa_n]$:

$$V_X[n, \kappa_n] \approx V_X[n, \kappa_n] - \frac{1}{N} \sum_{n=0}^{N-1} V_X[n, \kappa_n]; \quad (2.51)$$

Passo 5. Por fim, cálculo de $V_X^\varepsilon[\kappa_n]$ pode ser feito eficientemente usando o algoritmo da FFT:

$$V_X^\epsilon[\kappa_n] \approx \text{fft}\{V_X[n, \kappa_n]\}. \quad (2.52)$$

O estimador da CCF, $V_X^\epsilon[\kappa_n]$, retorna uma superfície que carrega as informações ciclo-estacionárias do sinal para cada atraso κ_n .

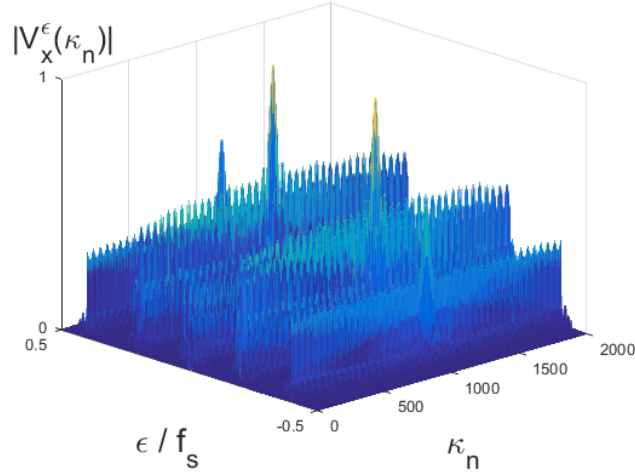


Figura 2.6: Função de correntropia cíclica aplicada ao sinal BPSK.

Após finalizar a estimação da CCF, a assinatura cicloestacionária (perfil-cíclico) do sinal é obtida pela projeção dos maiores componentes cíclicos, dessa superfície, ao longo de κ_n . Portanto a assinatura cicloestacionária do sinal, obtido pela CCF, é denotado por:

$$V_X^\epsilon = \max_{\kappa_n} \{V_X^\epsilon[\kappa_n]\}. \quad (2.53)$$

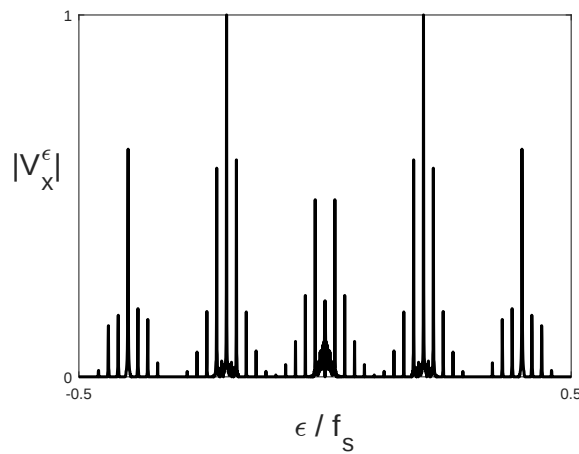


Figura 2.7: Perfil-Cíclico da Função de correntropia cíclica aplicada ao sinal BPSK.

Neste trabalho, as assinaturas cicloestacionárias de diferentes modulações digitais, obtidas pela CCF, são cuidadosamente avaliadas no Capítulo 3, e são utilizados como descritores de características das modulações digitais empregados nas arquiteturas de AMC propostas nos Capítulos 4 e 5.

2.4 Estrutura de uma Árvore de Decisão

No contexto de arquiteturas de classificação automática de modulação (*Automatic Modulation Classification* - AMC), uma árvore de decisão consiste em nós de testes sequenciais dedicados para o reconhecimento do esquema de modulação do sinal de entrada do sistema, através da observação de características particulares do mesmo (Zhu & Nandi 2015).

Uma vez que, em problemas de reconhecimento multi-classe (geralmente) as categorias não são separáveis por um único método, no processo de classificação estruturado em árvore de decisão as modulações devem ser divididas em várias camadas de subgrupos, até que cada modulação possa ser diferenciada das demais (Zhu & Nandi 2015). A Figura 2.8 denota um modelo genérico de árvore de decisão, em que os nós C_i ($i = 1, 2, \dots, n$) da árvore correspondem aos critérios de tomada decisão, e as extremidades dos ramos, também chamados de nós folhas, correspondem as classes possíveis (Coppin 2004).

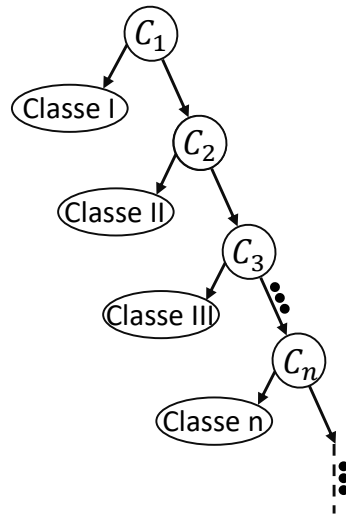


Figura 2.8: Estrutura genérica de uma árvore de decisão.

Para que cada classe seja reconhecida é necessário que os critérios de decisão sejam atendidos em cada nó da árvore. No caso de arquiteturas de AMC, é comum que os nós de tomada de decisão sejam relacionados à análise de limiares, os quais devem ser calculados com base em modelos de sinais sem ruído, ou otimizados para condições de canais específicas (Zhu & Nandi 2015).

Além disso, o ajuste dos limiares de decisão é feito de forma particular, de acordo com a técnica de extração de características do sinal adotada pelo classificador. Neste trabalho,

os descritores das modulações avaliadas são obtidos pelas funções FLOCAF e CCF, e arquiteturas de AMC estruturadas em árvore de decisão são propostas, e detalhadas, no Capítulo 4.

2.5 Classificação por Regressão Logística

Um sistema de classificação capaz de reconhecer automaticamente um tipo de modulação é dito inteligente se conseguir aprender (melhorar seu desempenho) a medida que expande o treinamento da tarefa de classificação. Dessa forma, um sistema inteligente de classificação de modulações utiliza aprendizagem de máquina.

Por sua vez, a aprendizagem de máquina pode ser associada a três aspectos: tarefa; experiência; desempenho. Um algoritmo é dito aprender com uma experiência E em relação a tarefa T com medida de desempenho P , se seu desempenho nas tarefas em T , medido por P , melhorar com a experiência E (Mitchell 1997).

De maneira geral, os problemas associados a aprendizagem de máquinas podem ser divididos em dois campos de análise: aprendizado supervisionado e aprendizado não-supervisionado. Ambas as abordagens trabalham com um conjunto de treinamento.

No aprendizado não-supervisionado, o conjunto de treinamento não possui atributos (rótulos) que permitam a identificação de cada um dos seus elementos. Desse modo, os algoritmos aplicados nessa configuração devem ser capazes de agrupar (clusterizar) os dados automaticamente para separá-los adequadamente. Por sua vez, o aprendizado supervisionado tem o objetivo de, a partir de um conjunto de treinamento que possui características x (variável de entrada) e alvo y (variável de saída), encontrar uma função $h : x \rightarrow y$, de modo que $h(x)$ seja um bom preditor para o valor correspondente y . Nesse contexto, a função $h(\cdot)$ é comumente chamada de hipótese.

Quando a variável alvo, a qual se quer prever com $h(x)$, é contínua, o problema de aprendizado é chamado de problema de regressão. Por outro lado, quando y assumir um número pequeno de valores discretos, o problema de aprendizado é chamado de problema de classificação.

2.5.1 Regressão linear

Para compreender o problema de regressão de forma clara, a Tabela 2.1 é utilizada como exemplo de conjunto de treinamento:

Tabela 2.1: Variáveis de entrada e saída de um conjunto de treinamento.

	Características (x)	Alvo (y)
m	2104	460
	1416	232
	1534	315
	$\vdots$	$\vdots$
	852	178

Em que,

- m é o numero de exemplos de treinamento;
- x' s são as variáveis de entrada, também conhecidas como características;
- y' s são as variáveis de saída, que podem ser chamadas de alvo.

Nesse caso, a notação $(x^{(i)}, y^{(i)})$ corresponde ao i -ésimo exemplo de treinamento. O conjunto de treinamento alimenta o algoritmo de aprendizado, e este retorna a função hipótese $h(\cdot)$. Por sua vez, a função $h(x)$ retorna uma estimativa de y , feita a partir de processos estatísticos (ou métodos numéricos) aplicados às características x . O processo de estimação é chamado de regressão.

A regressão retorna um modelo matemático utilizado para predição do alvo y . Quando este modelo for resultado de combinações lineares⁹ a regressão é dita linear, podendo ser expressa como

$$h_{\theta}(x) = \theta_0 + \theta_1 x \quad (2.54)$$

Para uma boa estimação, devem ser encontrados os valores de θ_0 e θ_1 que minimizam o erro quadrático entre y e $h_{\theta}(x)$. Essa análise deve ser aplicada a cada amostra do conjunto de treinamento, e pode ser expressada da seguinte forma

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2. \quad (2.55)$$

O termo $J(\theta_0, \theta_1)$ é definido como função custo, e neste caso, é denotado pela função de erro quadrático médio. Portanto, para o ajuste adequado da função hipótese é necessário encontrar os valores de θ_0 e θ_1 que minimizam a função custo (a qual também pode ser representada simplesmente por $J(\theta)$).

A função custo irá mudar de comportamento mediante a quantidade de parâmetros θ que a função hipótese possui, bem como, de acordo com o polinômio que é utilizado para descrever $h_{\theta}(x)$. Desse modo, podem haver casos em que $J(\theta)$ possua mais de um ponto de mínimo.

Os pontos de mínimo da função custo serão encontrados a partir de condições iniciais de θ , que devem ser ajustadas gradativamente de modo a minimizar $J(\theta)$. Um dos métodos utilizado pra isso é chamado de gradiente descendente, e pode ser expressado como

$$\theta_j := \theta_j - a \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \quad (2.56)$$

em que, o parâmetro a é conhecido por taxa de aprendizagem.

Embora a função custo, dependendo do polinômio $h_{\theta}(x)$, possa apresentar mínimos locais, sempre que a função hipótese for linear, $J(\theta)$ apresentará apenas o mínimo global. Em outras palavras, sempre que $h_{\theta}(x)$ for linear, a função custo $J(\theta)$ será sempre convexa.

⁹No contexto de aprendizagem de máquina, esse termo compreende funções lineares e afim.

No caso em que o conjunto de treinamento apresenta múltiplas características, a regressão linear retornará uma função hipótese que consiste na combinação linear de todas elas. Por exemplo, para um conjunto de treinamento descrito pela Tabela 2.2, a função $h_{\theta}(x)$ será denotada por

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4. \quad (2.57)$$

Tabela 2.2: Variáveis de entrada e saída de um conjunto de treinamento de múltiplas características.

x_1	x_2	x_3	x_4	y
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
852	2	1	36	178
$\underbrace{\hspace{10em}}_n$				

De maneira geral, a regressão linear aplicada a um conjunto de treinamento, que possui n características, retornará uma função hipótese que pode ser expressa da seguinte forma

$$h_{\theta}(x) = \theta_0 + \sum_{j=1}^n \theta_j x_j \quad (2.58)$$

Essa abordagem apresenta $n + 1$ valores de θ , e é chamada de regressão linear multivariada. Além da representação polinomial, a regressão linear multivariada pode ser representada matricialmente. Nesse caso, θ será denotado da seguinte forma

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}. \quad (2.59)$$

Por sua vez, por conveniência de notação, adota-se $x_0 = 1$. Desse modo, as características x serão representadas matricialmente por

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}. \quad (2.60)$$

Ambas as representações matriciais, de parâmetros e características, assumem valores reais e portanto $\{\theta, x \in \mathbb{R}^{n+1}\}$. Desse modo, a representação da função hipótese será dada simplesmente por

$$h_{\theta}(x) = \theta^T x. \quad (2.61)$$

Nesse contexto, combinando as Equações 2.55 e 2.56, pode-se estabelecer uma forma geral para o cálculo do gradiente descendente de uma regressão linear

$$\theta_j := \theta_j - a \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}. \quad (2.62)$$

Em suma, a regressão linear retorna uma estimativa do alvo y através da combinação linear das características x , ambos de um conjunto de treinamento. Os parâmetros θ utilizados na estimativa partem de uma condição inicial, e são ajustados iterativamente através do cálculo do gradiente descendente descrito na Equação 2.62. Nesse contexto, o índice i percorre os exemplos do conjunto de treinamento (que variam de 1 a m), por sua vez, o índice j faz referência a um parâmetro θ e a uma característica x , os quais podem variar de 0 a n , em que, x_0 sempre será igual a 1.

2.5.2 Regressão logística

Para entender de forma clara um processo de classificação, a Tabela 2.3 será utilizada como exemplo do conjunto de treinamento

Tabela 2.3: Conjunto de treinamento de alvo binário.

	x_1	x_2	x_3	x_4	y
m {	2104	5	1	45	1
	1416	3	2	40	0
	1534	3	2	30	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	852	2	1	36	1
	$\underbrace{\hspace{10em}}_n$				

No conjunto de treinamento da Tabela 2.3, o alvo assume apenas duas classes possíveis, por essa razão a estimação de y corresponde a um problema de classificação binária, e os valores do alvo são comumente representados por 0 ou 1,

$$y \in \{0, 1\}. \quad (2.63)$$

De forma similar ao problema de regressão linear, o conjunto de treinamento utilizado para classificação apresenta m exemplos de treinamento, cada um com n características. Contudo, a função hipótese não será mais definida por $h_{\theta}(x) = \theta^T x$. Uma vez que y assume valores binários, $h_{\theta}(x)$ deverá retornar valores pertencentes ao intervalo $[0, 1]$.

Para isso é necessário fazer uma transformação $g(\cdot)$ sobre $\theta^T x$ que garanta que a função hipótese esteja nesse intervalo. Desse modo, a função hipótese será dada por

$$h_{\theta}(x) = g(\theta^T x), \quad (2.64)$$

em que,

$$g(z) = \frac{1}{1 + e^{-z}}, \quad (2.65)$$

é chamada de função sigmoide ou função logística, e seu comportamento está denotado na Figura 2.9.

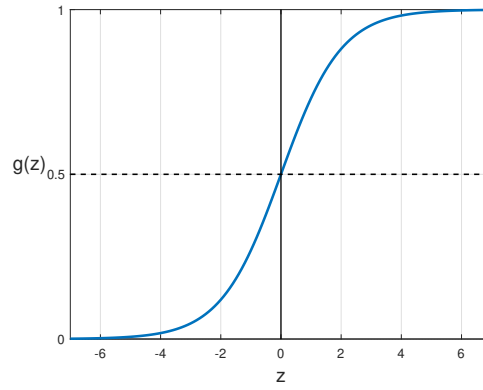


Figura 2.9: Função sigmoide $g(z)$.

Uma vez que a função $g(z)$ retorna valores entre 0 e 1, uma interpretação dada a função hipótese é que ela estima a probabilidade de y assumir 0 ou 1, dado uma entrada x parametrizada por θ . Em outras palavras

$$P(y = 1|x; \theta) = h_{\theta}(x) \quad (2.66)$$

e

$$P(y = 0|x; \theta) = 1 - h_{\theta}(x). \quad (2.67)$$

Desse modo, para a realização de uma classificação discreta entre 0 e 1, a função hipótese pode ser mapeada da seguinte forma

$$\begin{aligned} h_{\theta}(x) = g(\theta^T x) \geq 0,5 &\implies y = 1 \\ h_{\theta}(x) = g(\theta^T x) < 0,5 &\implies y = 0. \end{aligned} \quad (2.68)$$

Isso corresponde a

$$\begin{aligned} \theta^T x \geq 0 &\implies y = 1 \\ \theta^T x < 0 &\implies y = 0. \end{aligned} \quad (2.69)$$

Dessa forma, uma fronteira de decisão pode ser definida a partir de $\theta^T x$, a qual irá

separar uma região que equivale a $y = 0$ da região equivalente a $y = 1$. Por utilizar a função logística para o mapeamento $h : x \rightarrow y$, a estimação da função hipótese será chamada de regressão logística. O processo de regressão logística está esboçado na Figura 2.10.

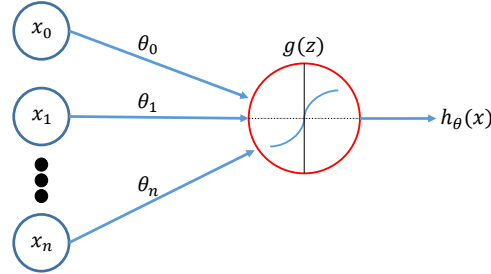


Figura 2.10: Estimação da função hipótese $h_{\theta}(x)$ a partir da Regressão Logística.

No caso da regressão logística, devido a transformação não linear que $h_{\theta}(x)$ aplica sobre x , a função custo $J(\theta)$ só será convexa se for definida como

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)}). \quad (2.70)$$

Em que,

$$\text{Cost}(h_{\theta}(x)) = \begin{cases} -\ln(h_{\theta}(x)) & y = 1 \\ -\ln(1 - h_{\theta}(x)) & y = 0. \end{cases} \quad (2.71)$$

Dado que as condições $y = 0$ e $y = 1$ são complementares, o termo $\text{Cost}(h_{\theta}(x))$ pode ser denotado simplesmente por

$$\text{Cost}(h_{\theta}(x)) = -y \ln(h_{\theta}(x)) - (1 - y) \ln(1 - h_{\theta}(x)), \quad (2.72)$$

por sua vez, aplicando a Equação 2.72 na Equação 2.70 pode-se descrever a função custo da regressão logística da seguinte forma

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \ln(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \ln(1 - h_{\theta}(x^{(i)}))]. \quad (2.73)$$

Conforme $h_{\theta}(x)$ se aproxima de y , $J(\theta)$ será minimizada. Para minimizar $J(\theta)$ aplica-se o método do gradiente descendente definido na Equação 2.56. E uma vez que,

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}, \quad (2.74)$$

o gradiente descendente de uma regressão logística será calculado de forma idêntica ao caso da regressão linear, e portanto, também será descrito pela Equação 2.62.

Em suma, a regressão logística irá estimar uma função hipótese que será utilizada na predição de um alvo y de um conjunto de treinamento, o qual assumirá valores binários¹⁰.

¹⁰Complementar a análise feita nesta seção, uma interpretação estatística da regressão logística é forne-

Nesse contexto, a função hipótese será definida pela aplicação da função sigmoide (Equação 2.65) sobre a combinação linear das características x do conjunto de treinamento.

Para o ajuste adequado de $h_{\theta}(x)$, os valores de θ_j , que minimizam a função custo $J(\theta)$ devem ser encontrados. Para isso, utiliza-se o cálculo do gradiente descendente, método numérico que considera todos os exemplos de treinamento do conjunto de treinamento para o ajuste de cada θ_j .

O gradiente descendente deverá ser repetido até que o ponto de mínimo de $J(\theta)$ seja encontrado, ou até que o critério de parada seja atendido¹¹. Uma vez ajustado os valores dos parâmetros θ_j , a função hipótese poderá ser utilizada como classificador binário de algum elemento que possua um conjunto características x .

2.5.3 Classificação multi-classe com regressão logística

A regressão logística também pode ser empregada em problemas de classificação multi-classe. Para isso, é necessário aplicar a técnica para cada uma das classes individualmente. Para entender de forma clara esse processo, a Tabela 2.4 será utilizada como exemplo de conjunto de treinamento.

Tabela 2.4: Conjunto de treinamento de alvo de múltiplas classes.

x_1	x_2	x_3	x_4	y
2104	5	1	45	10
1416	3	2	40	20
1534	3	2	30	20
2104	5	1	45	10
1367	3	2	40	30
1771	3	2	30	30
700	5	1	45	30
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
852	2	1	36	20

Os alvos do conjunto de treinamento da Tabela 2.4 podem assumir três classes diferentes, as quais são descritas pelos valores: $\{10, 20, 30\}$. Para realizar o reconhecimento da classe representada por 30, que será chamada daqui para frente de classe III, o conjunto de treinamento pode ser reestruturado de acordo com a Tabela 2.5.

Nesse caso, todos os elementos da classe III tiveram seus alvos ajustados para 1, enquanto as demais classes tiveram os alvos ajustados para 0. Isso permite que a regressão logística seja aplicada, retornando uma função hipótese associada a classe III, que será denotada por: $h_{\theta}(x)_{III}$.

De forma geral, esse procedimento pode ser realizado para cada uma das classes, o que resultará em uma função hipótese para cada uma delas ($h_{\theta}(x)_I$ e $h_{\theta}(x)_{II}$). Uma vez definidas as funções hipótese de cada classe, a classificação de um elemento, que possui

cida no Apêndice D.

¹¹ No caso em que $J(\theta)$ seja menor que algum valor pré-determinado.

Tabela 2.5: Conjunto de treinamento de alvo reestruturado.

x_1	x_2	x_3	x_4	y
2104	5	1	45	0
1416	3	2	40	0
1534	3	2	30	0
2104	5	1	45	0
1367	3	2	40	1
1771	3	2	30	1
700	5	1	45	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
852	2	1	36	0

características x , será feita testando-o em cada uma das funções. A escolha da classe será realizada adotando a função hipótese que retorna o valor mais próximo de 1, em outras palavras, será escolhido a classe da função hipótese que retorna a maior probabilidade. Essa abordagem é conhecida como classificação "um contra todos". Neste trabalho, arquiteturas de AMC que utilizam regressão logística são propostas no Capítulo 5.

Uma vez apresentado os princípios teóricos que permeiam a Tese, no próximo Capítulo as análises espectrais cíclicas das modulações BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM, obtidas pela FLOCAF e CCF, serão realizadas.

Capítulo 3

Análise Espectral Cíclica

Neste capítulo, as análises espectrais cíclicas das modulações BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM, são apresentadas. Inicialmente, são realizadas comparações entre as assinaturas cicloestacionárias, dos respectivos sinais modulados (sem ruído), obtidas pelas funções CAF, FLOCAF e CCF. Em seguida, as assinaturas das modulações contaminadas por ruído alfa-estável, obtidas pelas funções FLOCAF e CCF, são avaliadas e comparadas entre si.

3.1 Análise espectral sem ruído

Como discutido anteriormente, mesmo sob contaminação de ruído impulsivo de variância infinita, a FLOCAF e a CCF convergem. Desse modo, ambas as funções podem fornecer assinaturas cicloestacionárias robustas à contaminação por ruído impulsivo. Entretanto, uma análise detalhada das assinaturas de modulações digitais, obtidas por essas funções, não havia sido feita até Câmara et al. 2019.

Portanto, nesta Seção a FLOCAF e a CCF são aplicadas às modulações BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM sem ruído, e ambas funções são avaliadas quanto a capacidade de extração de características cicloestacionárias particulares para cada sinal modulado. Além disso, para uma análise mais ampla, as assinaturas cicloestacionárias das modulações propostas, geradas pela CAF, também são avaliadas.

Os sinais modulados, nos quais as funções cicloestacionárias foram aplicadas, possuem a mesma frequência intermediária f_i , taxa de símbolos f_b e frequência de amostragem f_s . Em cada assinatura obtida, o componente na posição de frequência cíclica $\varepsilon = 0$ foi suprimido por ser comum a todas as assinaturas, e não conter informação cicloestacionária relevante. Além disso, esse componente central é o que carrega a maior quantidade de energia, o que faz com que outros componentes de frequência cíclica sejam mascarados¹. As características cicloestacionárias, extraídas de sinais sem ruído pelas CAF, FLOCAF e CCF, estão representados nas Figuras 3.1 - 3.5.

A Figura. 3.1 mostra as assinaturas cicloestacionárias de um BPSK obtidas, respectivamente, pela CAF, FLOCAF e CCF. Pode-se perceber que, a assinatura gerada pela FLOCAF contém todos os componentes cíclicos apresentados na assinatura feita pela

¹Ao suprimir o componente central, sua energia é distribuída para os demais componentes da assinatura, de modo que componentes de amplitudes menores ganharão mais destaque.

CAF e, além disso, componentes cíclicos adicionais nas posições $\pm 2f_b/f_s$ e $\pm 4f_i/f_s$ foram evidenciados. Por outro lado, a assinatura gerada pela CCF contém todos os componentes proporcionados pela FLOCAF (dentre os quais, alguns apresentaram amplitudes maiores), além de apresentar novos componentes cíclicos nas posições $(2f_i \pm 2f_b)/f_s$, $(4f_i \pm f_b)/f_s$ e $(4f_i \pm 2f_b)/f_s$.

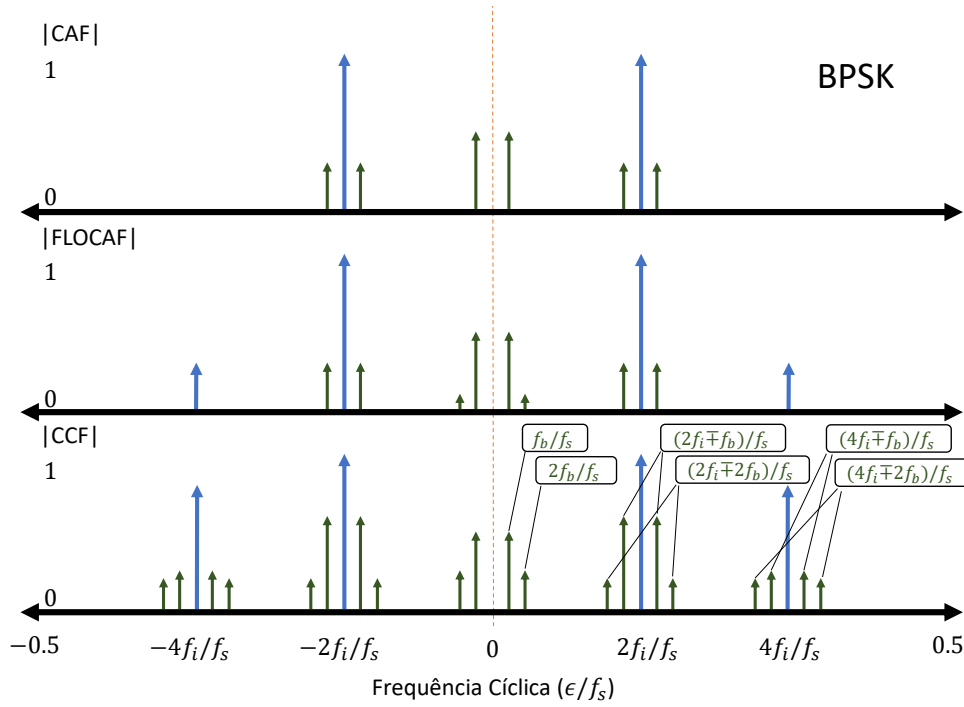


Figura 3.1: Assinaturas cicloestacionárias do BPSK obtidas pela CAF, FLOCAF e CCF.

A Figura 3.2 mostra as assinaturas cicloestacionárias associadas com a modulação QPSK. Mais uma vez, a assinatura gerada pela FLOCAF agrega todos os componentes cíclicos presentes na assinatura da CAF, além de novos componentes nas posições $\pm 2f_b/f_s$, $\pm 4f_i/f_s$ e $\pm(4f_i \pm f_b)/f_s$. Por outro lado, a assinatura gerada pela CCF apresenta todos os componentes cíclicos contidos na assinatura gerada pela FLOCAF, além de componentes nas posições $\pm(2f_i \pm 2f_b)/f_s$ e $\pm(4f_i \pm 2f_b)/f_s$.

Os resultados apresentados nas Figuras de 3.3 a 3.5 mostram que as assinaturas cicloestacionárias das modulações 8-, 16- e 32-QAM possuem comportamentos similares ao descritos para as modulações BPSK e QPSK. Em todos os casos, as assinaturas fornecidas pela FLOCAF contemplam todos os componentes apresentados na assinatura gerada pela CAF, além de introduzir novos descritores. De mesmo modo, as assinaturas geradas pela CCF contêm todos os componentes apresentados pela assinatura gerada pela FLOCAF, além de realçar outras características cicloestacionárias no espectro.

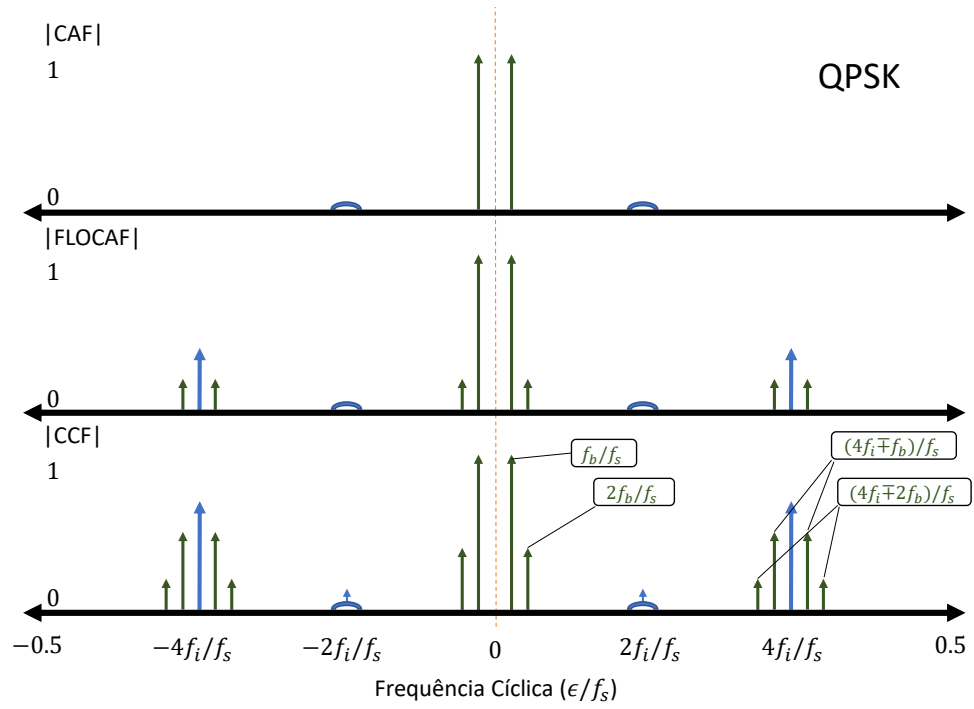


Figura 3.2: Assinaturas cicloestacionárias do QPSK obtidas pela CAF, FLOCAF e CCF.

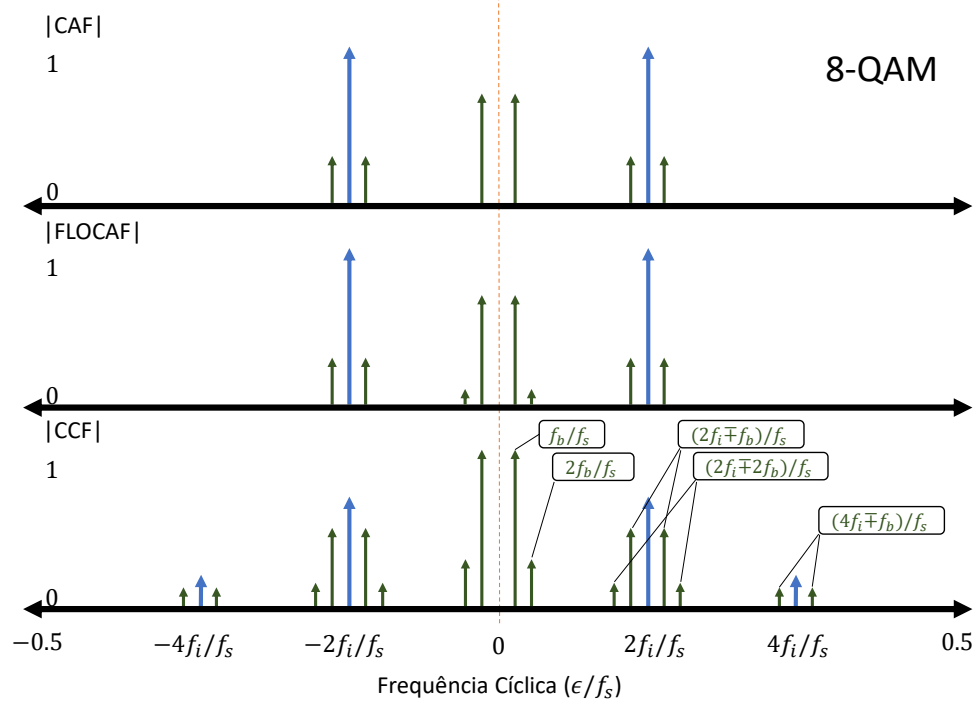


Figura 3.3: Assinaturas cicloestacionárias do 8-QAM obtidas pela CAF, FLOCAF e CCF.

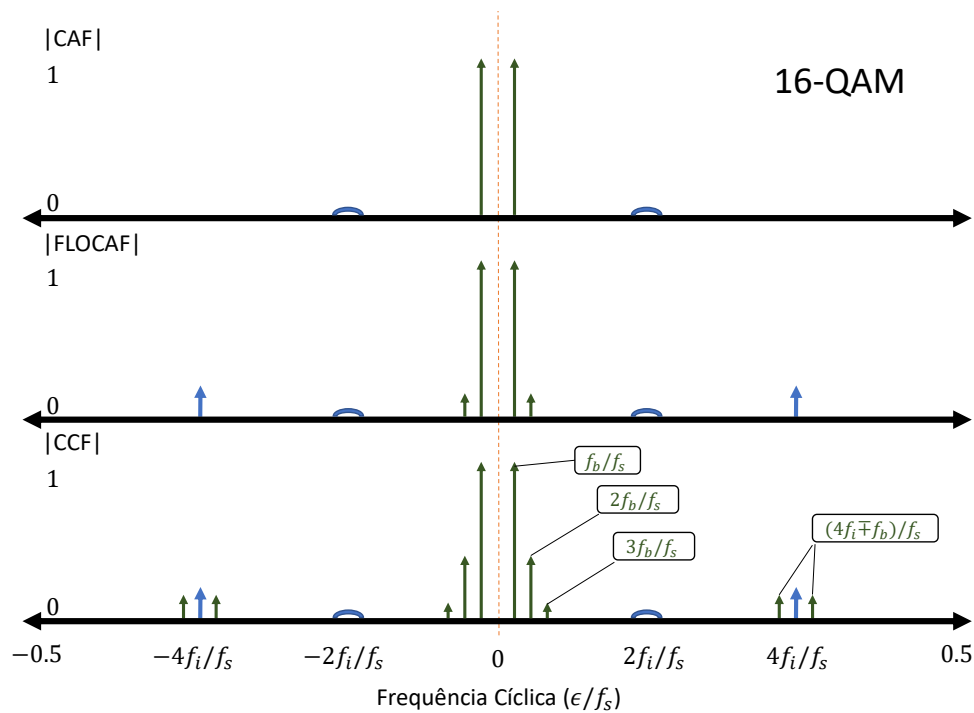


Figura 3.4: Assinaturas cicloestacionárias do 16-QAM obtidas pela CAF, FLOCAF e CCF.

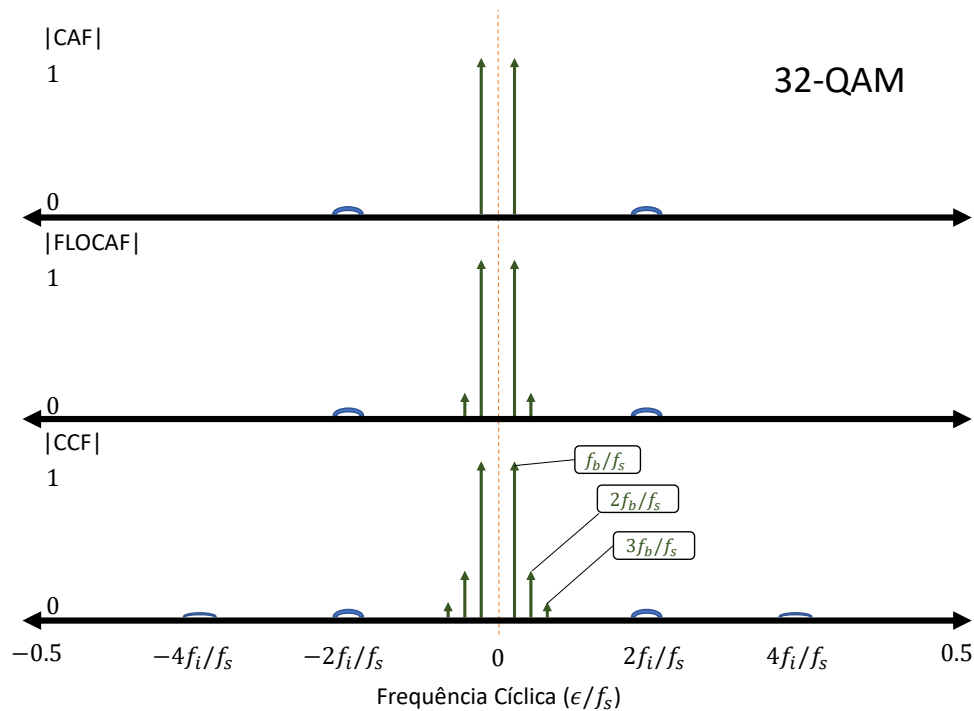


Figura 3.5: Assinaturas cicloestacionárias do 32-QAM obtidas pela CAF, FLOCAF e CCF.

As assinaturas geradas pela CAF mostram que essa função não é efetiva na extração de informações cíclicas capazes de caracterizar adequadamente cada modulação pois modulações distintas apresentam a mesma assinatura. Nesse contexto, por exemplo, as assinaturas associadas ao BPSK e 8-QAM mostradas respectivamente nas Figuras 3.1 e 3.3 possuem o mesmo perfil. Esse comportamento também pode ser visto nas modulações QPSK, 16-QAM e 32-QAM, as quais possuem assinaturas idênticas.

Por outro lado, a FLOCAF é capaz de extrair descritores singulares para cada sinal modulado, embora as assinaturas associadas ao 16-QAM e 32-QAM apresentem diferenças sutis como mostrado respectivamente nas Figuras 3.4 e 3.5. A CCF também é capaz de extrair assinaturas particulares para cada modulação. Além disso, as assinaturas obtidas por essa função contemplam todas as informações cicloestacionárias contidas nas assinaturas geradas pela FLOCAF, além de componentes cíclicos adicionais.

Uma característica interessante associada aos descritores apresentados é que todos os componentes das assinaturas cicloestacionárias são diretamente relacionados com a frequência intermediária, taxa de símbolos e frequência de amostragem dos sinais modulados analisados. Assim, pode-se afirmar que, de acordo com as características físicas de geração do sinal, a FLOCAF e a CCF são capazes de extrair assinaturas singulares para as modulações BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM.

3.2 Análise espectral em canais impulsivos

Nesta Seção, são discutidos os comportamentos das assinaturas cicloestacionárias, obtidas pela FLOCAF e CCF, de sinais modulados sujeitos ruído impulsivo. Para isso, a análise dos sinais BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM, contaminados por um ruído aditivo alfa-estável $S\alpha S$, é realizada.

A Tabela 3.1 apresenta a frequência de amostragem, frequência intermediária, taxa de símbolos, o número de símbolos e o fator de *roll-off* (utilizado no filtro formatador de pulso) empregados na geração de todos as modulações digitais. Além disso, os parâmetros empregados no ajuste da CCF e FLOCAF tais como número de amostras e fator de sobreposição das janelas retangulares usadas na geração das assinaturas, momentos fracionários de ordem inferior e tamanho de *kernel*, também são apresentados.

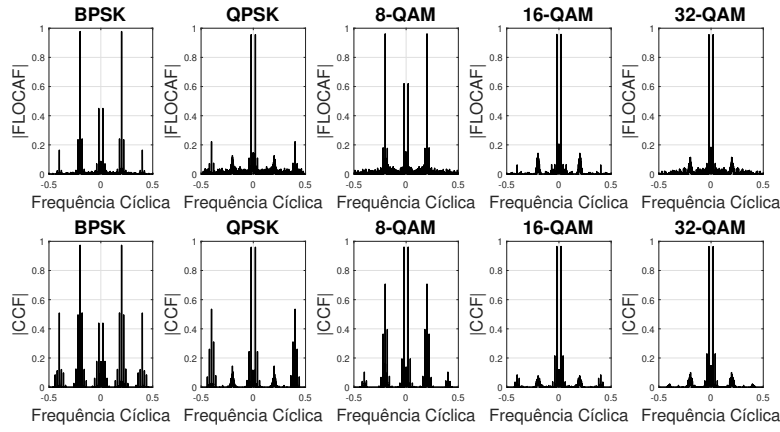
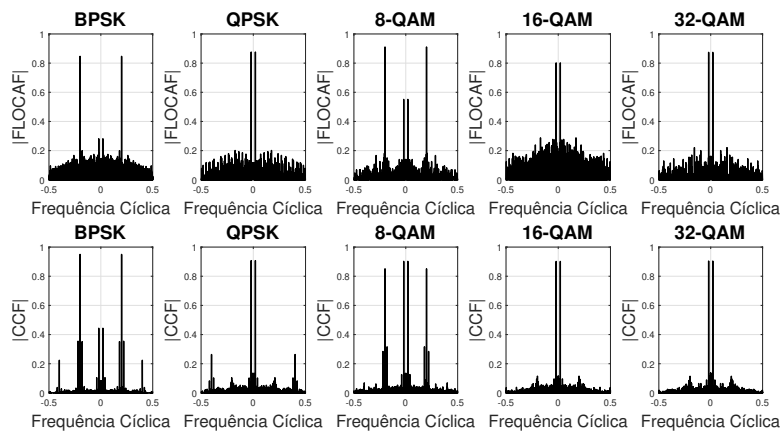
O canal foi modelado como uma distribuição alfa-estável com os seguintes parâmetros: $\alpha = 1,5^2$, $\beta = 0$, $\delta = 0$, e valores de $GSNR$ iguais a 10 dB e 0 dB, respectivamente, os quais estão relacionados com o parâmetro γ da distribuição (Equação 2.15). As assinaturas cicloestacionárias obtidas podem ser vista nas Figuras 3.6 e 3.7.

A Figura 3.6 ilustra as assinaturas das modulações geradas pela FLOCAF e CCF considerando uma contaminação por ruído aditivo impulsivo, descrito por uma distribuição alfa-estável com $GSNR = 10$ dB. Em ambos os casos, as funções fornecem assinaturas únicas para cada uma das modulações.

²Esse valor de expoente característico foi adotado para definir uma distribuição fortemente impulsiva (Liu et al. 2015, Shi et al. 2019)

Tabela 3.1: Parâmetros de Simulação.

Parâmetro	Valor
Frequência de amostragem (f_s)	10 kHz
Frequência Intermediária (f_i)	1 kHz
Taxa de Símbolos (f_b)	200 baud
Número de Símbolos	$n_{sym} = 1000$
Fator de <i>roll-off</i>	$\nu = 0,75$
Número de amostras por janela	$N = 2001$
Fator de sobreposição entre janelas	$Overlap = 0,5$
Ordem-inferior fracionária	$a = 0,5; b = 0,5$
Tamanho do <i>Kernel</i>	$\sigma = 1,1$

Figura 3.6: Assinaturas cicloestacionárias, obtidas com a FLOCAF e com a CCF, de sinais modulados contaminados por ruído com $GSNR=10$ dB.Figura 3.7: Assinaturas cicloestacionárias, obtidas com a FLOCAF e com a CCF, de sinais modulados contaminados por ruído com $GSNR=0$ dB.

A Figura 3.7 representa as assinaturas cicloestacionárias, obtidas pela FLOCAF e CCF, para sinais contaminados pelo ruído aditivo impulsivo, descrito por uma distribui-

ção alfa-estável com $GSNR = 0$ dB. A comparação entre as Figuras 3.6 e 3.7 permite ver que, conforme a intensidade da contaminação aumenta, a degradação das assinaturas cicloestacionárias cresce, o que faz com que alguns componentes sejam mascarados.

Nesse contexto, as assinaturas geradas pela FLOCAF se mostraram muito sensíveis à contaminação. Em outras palavras, os componentes de frequência cíclica de algumas assinaturas em condições de baixa $GSNR$, são significativamente atenuadas, tornando difícil o processo de distinção entre as modulações, tais como, QPSK, 16-QAM e 32-QAM. Por outro lado, as assinaturas geradas pela CCF apresentaram informações cicloestacionárias suficientes para diferenciar apropriadamente as modulações.

Uma vez que as assinaturas, geradas pela FLOCAF e pela CCF, apresentam comportamentos dinâmicos para a variação das condições de $GSNR$, uma comparação mais precisa pode ser feita utilizando uma arquitetura de classificação de modulações. Portanto, empregando essas funções cicloestacionárias, no Capítulo 4 são propostas arquiteturas de AMC estruturadas como árvores de decisão, e no Capítulo 5 são propostas arquiteturas de AMC que utilizam regressão logística.

Capítulo 4

Arquiteturas de AMC com Árvore de Decisão

Como demonstrado no Capítulo 3, a FLOCAF e CCF podem ser usadas para extrair descritores cicloestacionários singulares para cada modulação analisada, mesmo em ambientes de ruído impulsivo. Neste capítulo, duas arquiteturas de AMC, baseadas nas funções mencionadas, são descritas. O fluxograma das arquiteturas está ilustrado na Figura. 4.1.

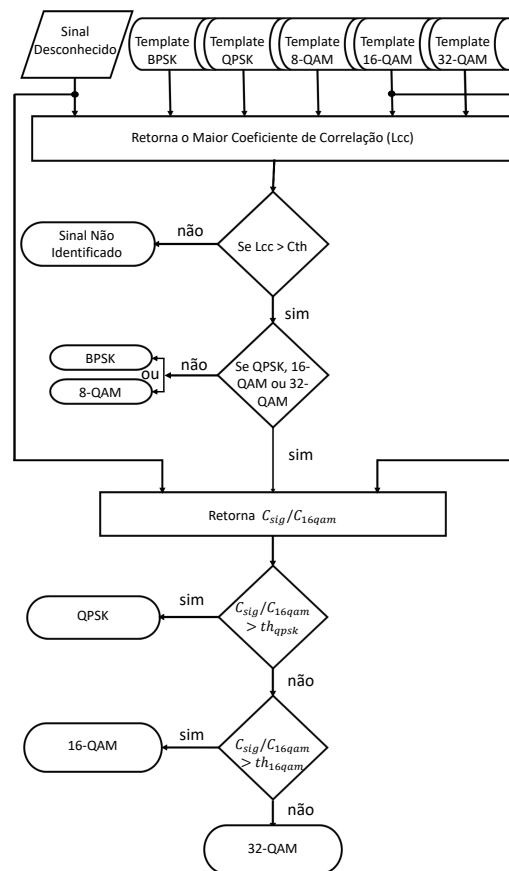


Figura 4.1: Estrutura de classificação baseada na extração de características cicloestacionárias aplicada a uma árvore de decisão.

4.1 Definição dos *Templates* das Modulações

Embora possuam algumas particularidades, ambas as arquiteturas seguem a mesma lógica de operação. Inicialmente elas realizam a comparação entre o perfil cíclico do sinal de entrada e os perfis cíclicos dos sinais modulados sem ruído, os quais serão chamados daqui para frente de *templates*. A principal diferença entre as arquiteturas é a função cicloestacionária usada no cálculo dos perfis cíclicos. A arquitetura baseada na FLOCAF aplica essa função cicloestacionária para gerar os *templates* e os perfis cíclicos do sinal de entrada. A mesma lógica é aplicada no caso da CCF.

4.2 Sinal Desconhecido

O perfil cíclico do sinal de entrada (o qual o esquema de modulação deverá ser classificada) deve ser gerado seguindo o mesmo processo usado nos *templates*. Assim, se a arquitetura usa a FLOCAF, os momentos fracionários de ordem-inferior (a, b) que foram utilizados no cálculo dos *templates* devem ser utilizados também para o cálculo do perfil-cíclico do sinal de entrada. De forma similar, se a arquitetura emprega a CCF, os *templates* e as assinaturas do sinal de entrada devem ser parametrizados com o mesmo tamanho de *kernel* (σ).

4.3 Coeficiente de Correlação

O segundo passo das arquiteturas propostas é comparar o sinal desconhecido com os *templates*. Para isso, uma medida de similaridade, conhecida como coeficiente de correlação de Pearson, foi adotada. Ela é calculada da seguinte forma,

$$\rho_{xy} = \frac{\sum_{k=0}^{N-1} (x[k] - \mu_x)(y[k] - \mu_y)}{\sqrt{\sum_{k=0}^{N-1} (x[k] - \mu_x)^2 \sum_{k=0}^{N-1} (y[k] - \mu_y)^2}} \quad (4.1)$$

em que,

$$\mu_x = \frac{1}{N} \sum_{k=0}^{N-1} x[k]$$

$$\mu_y = \frac{1}{N} \sum_{k=0}^{N-1} y[k]$$

e $x[k]$ e $y[k]$ são, respectivamente, o sinal a ser classificado e o *template*.

O coeficiente de correlação de Pearson é uma técnica de baixa complexidade computacional que retorna valores no intervalo de -1 a 1 , os quais podem ser classificados da seguinte forma: $|\rho| < 0,3$ denota sinais descorrelacionados; $0,3 \leq |\rho| \leq 0,5$ implica

em correlação fraca; $0,5 < |\rho| \leq 0,7$ denota correlação moderada; $0,7 < |\rho| \leq 0,9$ indica correlação forte; $|\rho| > 0,9$ implica em uma correlação muito forte ¹ (Mukaka 2012).

Nas arquiteturas propostas, a assinatura cicloestacionária do sinal de entrada é correlacionada com todos os *templates*, e o maior valor de coeficiente (*Largest correlation coefficient* L_{CC}) entre eles é usado na classificação. Como mostrado na Fig. 4.1, o coeficiente escolhido é comparado com um limiar de correlação previamente definido (*Correlation threshold* C_{th}). Se o maior coeficiente de correlação for menor que C_{th} , então o esquema de modulação do sinal não será reconhecido.

Essa ação contribui para a redução do problema de falso-positivo, que ocorre quando a modulação prevista é impropriamente indicada. Em outras palavras, sob ruído intenso, todas as assinaturas tem seus componentes cíclicos atenuados e isso faz com que o reconhecimento do esquema de modulação se torna muito impreciso, o que muitas vezes resulta em uma escolha enviesada para uma das classes. Desse modo, o limiar diminui o viés de classificação garantindo que a assinatura do sinal de entrada e o *template* estão correlacionados, diminuindo a imprecisão. ².

Quanto maior o limiar de correlação, menor a incidência de falso-positivo. Entretanto, um limiar elevado também contribui para a redução da classificação de verdadeiros-positivos. Em outras palavras, resulta na diminuição da taxa de acerto do classificador. Portanto, para estabelecer um ajuste adequado para a taxa de acerto, e limitar a ocorrência de falsos-positivos, o limiar escolhido foi $C_{th} = 0,7$, uma vez que esse valor garante uma correlação de moderada para forte.

Dessa forma, quando $L_{CC} > C_{th}$, uma decisão sobre a classe do sinal de entrada será tomada. Quando o L_{CC} é associado aos *templates* do BPSK ou 8-QAM, a modulação será atribuída respectivamente às classes BPSK ou 8-QAM. Caso contrário, as arquiteturas seguirão para outra etapa de decisão, para distinguir entre as classes QPSK, 16-QAM e 32-QAM.

4.3.1 Relação C_{sig}/C_{16QAM}

De acordo com as Figuras 3.4 e 3.5, as assinaturas cicloestacionárias do 16-QAM e 32-QAM, geradas tanto pela FLOCAF quanto pela CCF, possuem diferenças sutis. Entre eles, a mais evidente é o componente cíclico que se encontra na posição espectral $4f_i$, o qual aparece na assinatura associada ao 16-QAM, e é ausente na 32-QAM. Além disso, a Figura 3.2 mostra que o elemento na posição espectral $4f_i$, na assinatura do QPSK, também configura a principal diferença entre essa modulação e a 16-QAM. Nesse contexto, esse componente possui maior amplitude na assinatura do QPSK do que na assinatura do 16-QAM.

Portanto, para classificar o sinal de entrada entre QPSK, 16-QAM, e 32-QAM, foi estabelecida uma relação entre o componente cíclico desse sinal na posição $4f_i$ (*Component of the signal* - C_{sig}) e o componente do *template* do 16-QAM (*Component of 16QAM* -

¹O coeficiente de correlação de Pearson quando assume valores negativos implica em variáveis ditas inversamente correlacionadas.

²Em termos práticos, o limiar de correlação do classificador também serve para evitar que um sinal puramente ruidoso seja erroneamente atribuído a uma das classes.

C_{16QAM}) na mesma posição espectral $4f_i$. O critério aplicado para a escolha da classe, a qual o sinal pertence, consiste na comparação da relação $-C_{sig}/C_{16QAM}$ com dois limiares, o limiar do QPSK (*Threshold of 16QAM* - th_{QPSK}) e o limiar do 16-QAM (*Threshold of 16QAM* - th_{16QAM}). Nesse contexto:

- $C_{sig}/C_{16QAM} > th_{QPSK}$, o sinal de entrada será atribuído a classe QPSK;
- $th_{QPSK} > C_{sig}/C_{16QAM} > th_{16QAM}$, o sinal de entrada será classificado como 16-QAM;
- $th_{16QAM} > C_{sig}/C_{16QAM}$, a classe escolhida será a 32-QAM.

Os limiares foram definidos com base em uma análise extensiva da relação C_{sig}/C_{16QAM} , em uma condição de $GSNR$ específica. Esse processo é detalhado na próxima seção.

4.4 Resultados de classificação com Árvore de Decisão

Nesta seção, os desempenhos das arquiteturas de classificação baseadas nas características cicloestacionárias, obtidas pela FLOCAF e CCF, são avaliados em ambientes com ruído impulsivo.

Os ambientes foram caracterizados por um ruído aditivo alfa-estável não-gaussiano com os seguintes parâmetros: expoente característico $\alpha = 1.5$, $\beta = 0$, $\delta = 0$, e $GSNR$ variando na faixa de -10 dB a 10 dB com passo de 2 dB. Os parâmetros de ajuste dos sinais modulados e das assinaturas cicloestacionárias aplicadas nos testes (número de amostras da janela e sobreposição entre janelas) são os mesmos definidos na Tabela 3.1.

4.4.1 AMC usando a FLOCAF

Com o objetivo de melhorar o desempenho da arquitetura de AMC baseada na FLOCAF, uma busca extensiva dos parâmetros fracionários de ordem-inferior (a, b) foi realizada. Assim, a arquitetura foi parametrizada com todas as combinações de a e b , com valores obtidos do conjunto: $\{0; 0, 1; 0, 3; 0, 5\}$.

Esses valores, por sua vez, foram escolhidos para atender aos critérios de convergência apresentados na Equação (2.23). A busca foi realizada da seguinte forma:

- Em cada condição de $GSNR$ avaliada, o sinal de entrada ruidoso foi gerado 100 vezes para cada esquema de modulação, e então foi aplicado a arquitetura;
- Como o classificador apresenta uma taxa de classificação específica para cada modulação, a análise de desempenho foi feita considerando a média das taxas de classificação de cada modulação, denominada taxa de classificação global;
- Finalmente, diferentes curvas de classificação foram geradas, mudando os parâmetros fracionários de ordem-inferior e variando a $GSNR$.

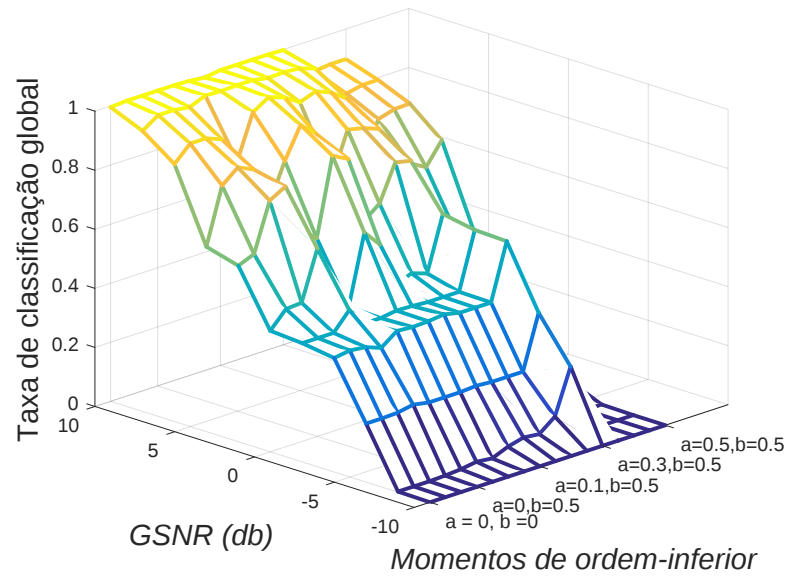


Figura 4.2: Taxa de classificação global de modulação a partir da arquitetura que emprega a FLOCAF, considerando diferentes combinações de momentos fracionários.

A partir da Figura 4.2, pode-se observar que a arquitetura baseada na FLOCAF apresenta comportamentos distintos, em relação à taxa de acerto, para diferentes combinações de a e b , e diferentes valores de $GSNR$. Portanto, para determinar os parâmetros que otimizam o desempenho da arquitetura para uma qualidade média do canal em termos de $GSNR$, uma taxa de acerto média foi calculada considerando a variação de $GSNR$ de -10 dB a 10 dB. O resultado está ilustrado na Figura 4.3.

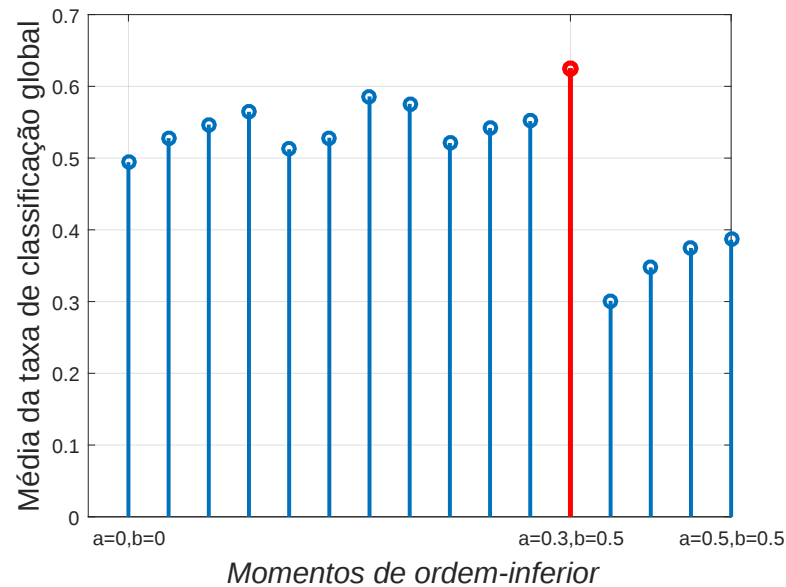


Figura 4.3: Valores médios da taxa de acerto global obtidos sobre faixa de $GSNR$, relacionados a arquitetura que emprega a FLOCAF, considerando diferentes combinações de momentos fracionários.

No caso particular, quando $a = b = 0$, a FLOCAF converge para SSCCF, e portanto, os descritores cicloestacionários serão menos afetados pelo ruído impulsivo. Entretanto, os resultados apresentados na Figura 4.3 indicam que, em média, a arquitetura possui uma maior taxa de acerto na classificação de modulações quando é parametrizada com $a = 0,3$ e $b = 0,5$. Esse resultado sugere que as modulações digitais avaliadas possuem informações cicloestacionárias que são evidenciadas por esses parâmetros fracionários de ordem-inferior.

Após determinar os parâmetros fracionários que otimizam o desempenho médio da arquitetura, uma análise mais detalhada do classificador pode ser realizada, considerando 1000 testes por modulação, para cada valor de $GSNR$.

Os limiares th_{QPSK} e th_{16QAM} foram definidos considerando uma análise exaustiva das assinaturas das modulações QPSK, 16-QAM e 32-QAM, contaminados com ruído impulsivo correspondente a $GSNR = 10$ dB. Para isso, as assinaturas de cada modulação foram geradas 1000 vezes, e o fator C_{sig}/C_{16QAM} foi calculado para cada uma delas. O resultado é apresentado na Figura 4.4.

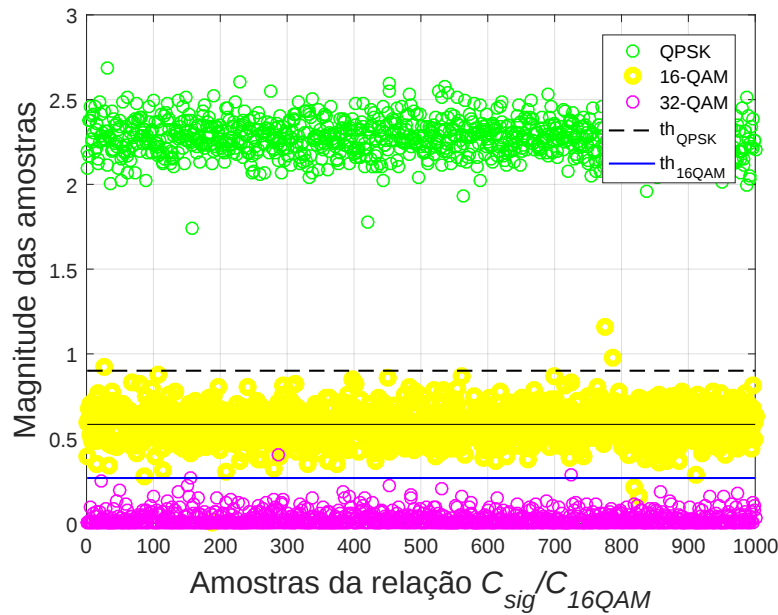


Figura 4.4: Definição dos limiares th_{qpsk} e th_{16QAM} , para arquitetura que emprega a FLOCAF.

A partir da Figura 4.4, verifica-se que as amostras associadas às modulações 16-QAM, QPSK e 32-QAM se concentram em faixas particulares de magnitude. Portanto, os limiares th_{qpsk} e th_{16QAM} foram respectivamente estabelecidos considerando três desvios padrões acima do valor da média de amostras do 16-QAM, e três desvios padrões abaixo dela. Os limiares adotados na arquitetura de AMC baseada na FLOCAF estão apresentados na Tabela 4.1.

Tabela 4.1: Limiares de classificação da arquitetura que emprega a FLOCAF.

Parâmetro	Valor
C_{th}	0,70
th_{QPSK}	0,90
th_{16QAM}	0,26

Após ajustar os parâmetros fracionários de ordem-inferior (a , b), e estabelecer os limiares, a arquitetura foi avaliada considerando 1000 testes por modulação em cada configuração de $GSNR$. O comportamento de classificação está ilustrado na Figura 4.5.

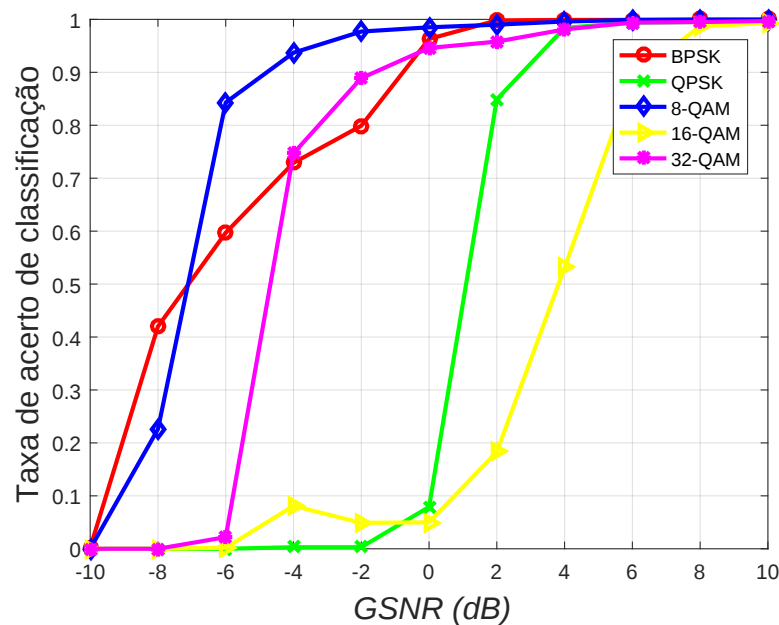


Figura 4.5: Curvas de classificação da arquitetura baseada na FLOCAF, parametrizada com $a = 0,3$ e $b = 0,5$.

A partir da Figura 4.5 pode-se observar que a arquitetura baseada na FLOCAF apresenta desempenhos distintos para cada modulação. A taxa de acerto de classificação alcança valores acima de 90% quando a $GSNR = 0$ dB, para as modulações BPSK, 8-QAM e 32-QAM. Enquanto que para a modulação QPSK, a arquitetura precisa de uma $GSNR = 4$ dB para alcançar essa taxa.

Quando a $GSNR = 8$ dB, a arquitetura alcança uma taxa de acerto próxima de 100% para todas as modulações. Uma vez que o classificador apresenta um comportamento particular em cada condição de $GSNR$, uma análise mais detalhada pode ser feita considerando-se as matrizes de confusão apresentadas nas Tabelas 4.2 a 4.4, para alguns valores de $GSNR$.

Tabela 4.2: Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0,3, b = 0,5$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	963	0	0	0	0
QPSK	0	79	0	0	0
8QAM	31	0	985	0	0
16QAM	0	869	0	50	15
32QAM	0	4	0	919	946
Abaixo de C_{th}	6	48	15	31	39

Tabela 4.3: Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0,3, b = 0,5$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	999	0	0	0	0
QPSK	0	984	0	0	0
8QAM	0	0	996	0	0
16QAM	0	0	0	531	5
32QAM	0	0	0	451	981
Abaixo de C_{th}	1	16	4	18	14

Tabela 4.4: Matriz de confusão da arquitetura que emprega a FLOCAF ($a = 0,3, b = 0,5$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	998	0	2	0
8QAM	0	0	999	0	0
16QAM	0	0	0	993	3
32QAM	0	0	0	2	997
Abaixo de C_{th}	0	2	1	3	0

Os resultados sumarizados indicam que, para baixos níveis de $GSNR$, a modulação QPSK é amplamente confundida com a 16-QAM. Assim, pode-se inferir que para um nível de ruído muito alto, alguns componentes cíclicos oriundos da modulação QPSK são atenuados resultando em uma assinatura que se aproxima mais do *template* da modulação 16-QAM do que da modulação QPSK. Contudo, as matrizes de confusão também indicam que, conforme a $GSNR$ cresce, a acurácia da classificação do QPSK também cresce.

Um comportamento similar também ocorre na classificação da modulação 16-QAM. Para baixos níveis de $GSNR$, essa modulação é amplamente confundida com a modulação 32-QAM. Como no caso anteriormente discutido, conforme a $GSNR$ cresce, a acurácia da classificação da modulação 16-QAM também aumenta. Entretanto, a taxa de acerto dessa modulação cresce mais devagar do que as demais, o que indica que a modulação 16-QAM é mais sensível a contaminação do ruído do que as outras modulações analisadas.

Esses resultados também indicam que, para o critério de classificação adotado nas arquiteturas propostas, as assinaturas cíclicas do BPSK, 8-QAM e 32-QAM são mais robustas a contaminação por ruído impulsivo, uma vez que, mesmo para alguns valores de $GSNR$ negativos, a taxa de acerto para a classificação dessas modulações é elevada.

4.4.2 AMC usando a CCF

No teste da arquitetura de AMC baseada na CCF, de forma similar ao caso da FLO-CAF em relação aos parâmetros fracionários, uma busca extensiva do tamanho do *kernel* (σ) foi realizada de forma a otimizar o desempenho global da arquitetura. A análise foi feita para uma faixa de tamanho de *kernel* 0,1 a 1,4 com passos de 0,1.

O experimento foi realizado da seguinte forma: primeiro, sinais de entrada ruidosos, oriundos de cada um dos esquemas de modulação considerados no trabalho, foram gerados várias vezes para cada condição de $GSNR$ e aplicados à arquitetura. Após isso, a taxa de classificação global foi calculada, considerando 100 testes por modulação, para cada valor de $GSNR$. Por fim, diferentes curvas de classificação foram obtidas para cada tamanho de *kernel* considerando a variação da $GSNR$. A Figura 4.6 ilustra as curvas de classificação global relacionadas a variação de $GSNR$ e a variação do tamanho do *kernel*.

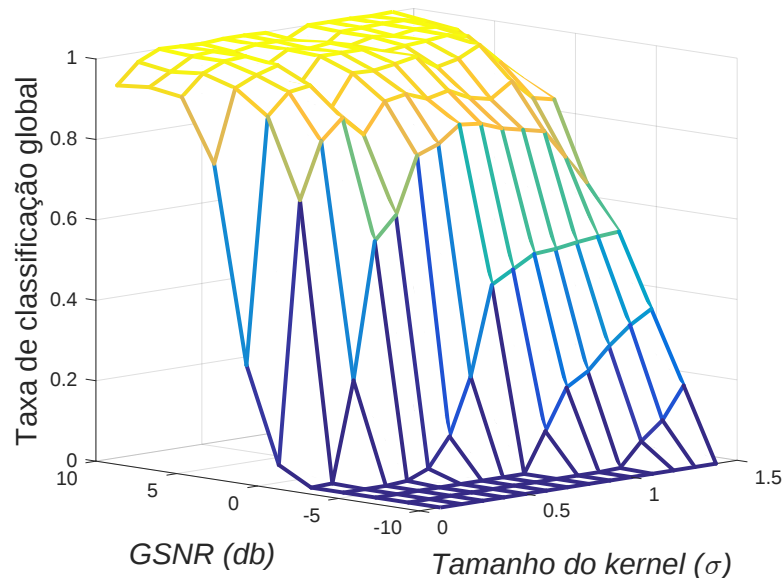


Figura 4.6: Taxa de classificação global de modulação a partir da arquitetura que emprega a CCF, considerando diferentes tamanhos de *kernel*.

Pode-se observar a partir da Figura 4.6 que a arquitetura de AMC baseada na CCF

possui comportamentos distintos de taxa de acerto relacionados a cada tamanho de *kernel*, os quais também variam com a *GSNR*. Portanto, para determinar o valor que otimiza o desempenho da arquitetura em termos da qualidade média do canal, foi avaliada a média da taxa de acerto global relacionada à faixa de *GSNR*.

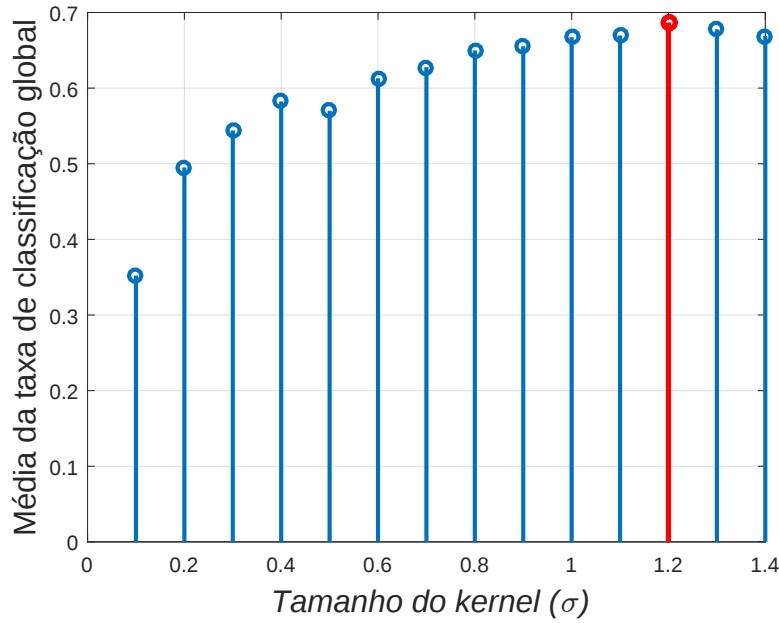


Figura 4.7: Valores médios da taxa de acerto global obtidos sobre faixa de *GSNR*, relacionados à arquitetura que emprega a CCF, considerando diferentes tamanhos de *kernel*.

Os resultados ilustrados na Figura 4.7 indicam que, em média, a arquitetura alcança a maior taxa de acerto quando parametrizada com $\sigma = 1,2$. Assim, após encontrar o tamanho de *kernel* que melhor se adéqua ao classificador, foi realizada uma análise detalhada de desempenho da arquitetura, considerando 1000 testes por modulação para cada valor de *GSNR*.

Para isso, assim como no caso da arquitetura que utiliza a FLOCAF, os valores de th_{QPSK} e th_{16QAM} foram definidos considerando-se a análise das assinaturas das modulações QPSK, 16-QAM, e 32-QAM em um canal α -estável com *GSNR* = 10 dB. As assinaturas das modulações foram geradas 1000 vezes, e o fator C_{sig}/C_{16QAM} foi calculado para cada uma delas. O resultado é apresentado na Figura 4.8.

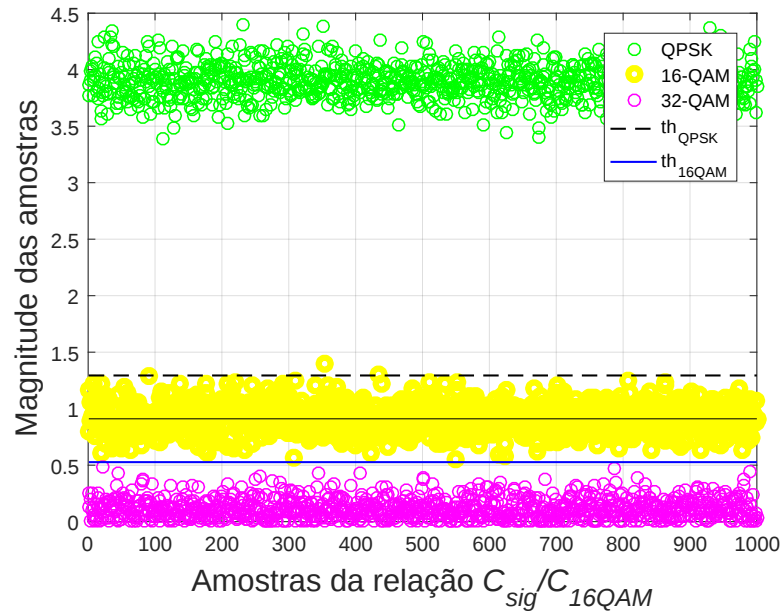


Figura 4.8: Definição dos limiares th_{qpsk} e th_{16QAM} , para arquitetura que emprega a CCF.

Pode-se observar a partir da Figura 4.8, que as amostras associadas à modulação 16-QAM estão em uma faixa particular de magnitude que se encontra entre as faixas referentes às modulações QPSK e 32-QAM. Dessa forma, assim como no caso da arquitetura baseada na FLOCAF, tanto th_{qpsk} e th_{16QAM} foram estabelecidos considerando três desvios padrões acima, e três desvios padrões abaixo do valor da média das amostras relacionadas à modulação 16-QAM. Os valores de limiares adotados na arquitetura de AMC baseada na CCF, estão dispostos na Tabela 4.5.

Tabela 4.5: Limiares da arquitetura de AMC baseada na CCF.

Parâmetro	Valor
C_{th}	0,70
th_{QPSK}	1,29
th_{16QAM}	0,52

Após o ajuste do tamanho do *kernel* (σ), e definidos os limiares, a arquitetura baseada na CCF foi testada, e as taxas de classificação por modulação estão ilustrados na Figura 4.9. A partir desses resultados, observa-se que a arquitetura possui desempenhos diferentes para cada modulação, sendo a classificação dos sinais BPSK a mais precisa, e a classificação dos sinais 16-QAM a mais sensível à contaminação. Ainda assim, a classificação das modulações BPSK, QPSK, 8-QAM, e 32-QAM, alcança taxas de acertos próximos a 100% para $GSNR$ de 0 dB. A classificação de todas as modulações é de aproximadamente 100% quando a $GSNR = 6$ dB.

Para obter uma análise detalhada do classificador, matrizes de confusão foram obtidas para diferentes valores de $GSNR$, e estão dispostas nas Tabelas 4.6 a 4.8. A partir dessas matrizes observa-se que, para condições de $GSNR$ mais baixas, a modulação 16-QAM é

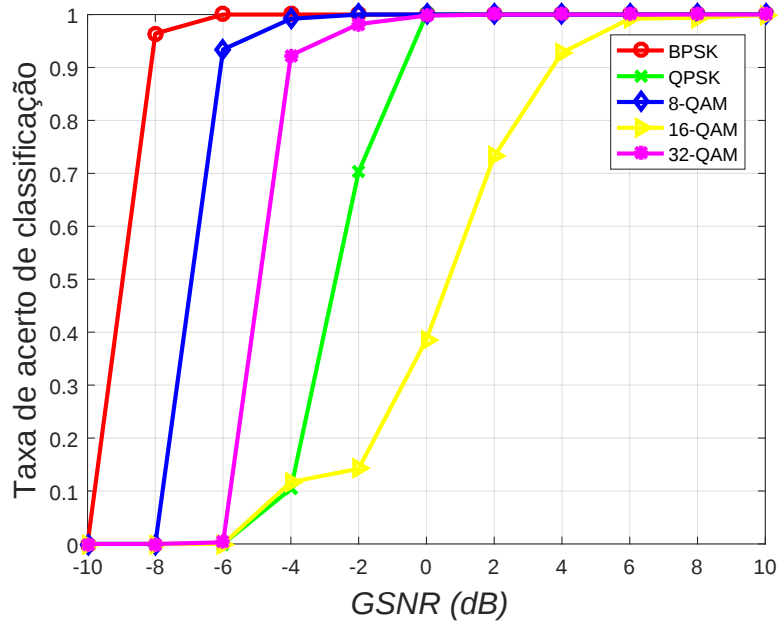


Figura 4.9: Curvas de classificação da arquitetura baseada na CCF, parametrizadas com $\sigma = 1.2$. Em cada configuração de $GSNR$ foram considerando 1000 testes por modulação.

confundida com a modulação 32-QAM. Tal resultado indica que a assinatura da modulação 16-QAM, sobre efeito de níveis elevados de ruído, é mais próxima do *template* da modulação 32-QAM do que do *template* da própria modulação 16-QAM. Portanto, pode-se constatar que a contaminação intensa dos sinais 16-QAM contribui para a atenuação do componente cíclico que diferencia a assinatura da modulação 16-QAM da assinatura do 32-QAM.

Tabela 4.6: Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1, 2$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	2	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	386	2
32QAM	0	0	0	612	998
Abaixo de C_{th}	0	0	0	0	0

Tabela 4.7: Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1,2$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	1	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	928	0
32QAM	0	0	0	71	1000
Abaixo de C_{th}	0	0	0	0	0

Tabela 4.8: Matriz de confusão da arquitetura que emprega a CCF ($\sigma = 1,2$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	1	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	999	0
32QAM	0	0	0	0	1000
Abaixo de C_{th}	0	0	0	0	0

A partir dos resultados apresentados nesta seção, pode-se observar que a arquitetura de AMC baseada na CCF (Figura 4.9) possui um desempenho melhor do que a arquitetura baseada na FLOCAF (Figura 4.5), uma vez que, considerando as mesmas condições de canal, e as mesmas modulações, a taxa de acerto da arquitetura baseada na CCF é maior do que a taxa de acerto da arquitetura baseada na FLOCAF.

4.5 Estimação da Taxa de Símbolo

Como demonstrado na Seção 3.2, ambas as funções, CCF e FLOCAF, são capazes de extrair descritores cicloestacionários das modulações contaminadas pelo ruído impulsivo alfa-estável não-gaussiano. Ainda assim, para $GSNR$ s mais baixas essa tarefa se torna difícil, uma vez que o ruído intenso pode atenuar os componentes das assinaturas cicloestacionárias. Contudo, alguns componentes cíclicos são mais robustos ao efeito do ruído do que outros.

Nas assinaturas cicloestacionárias oriundas da FLOCAF e CCF, o componente associado à taxa de símbolos do sinal modulado, entre todos os outros, é o mais robusto ao ruído impulsivo, como pode ser observado a partir das Figuras 4.10 e 4.11.

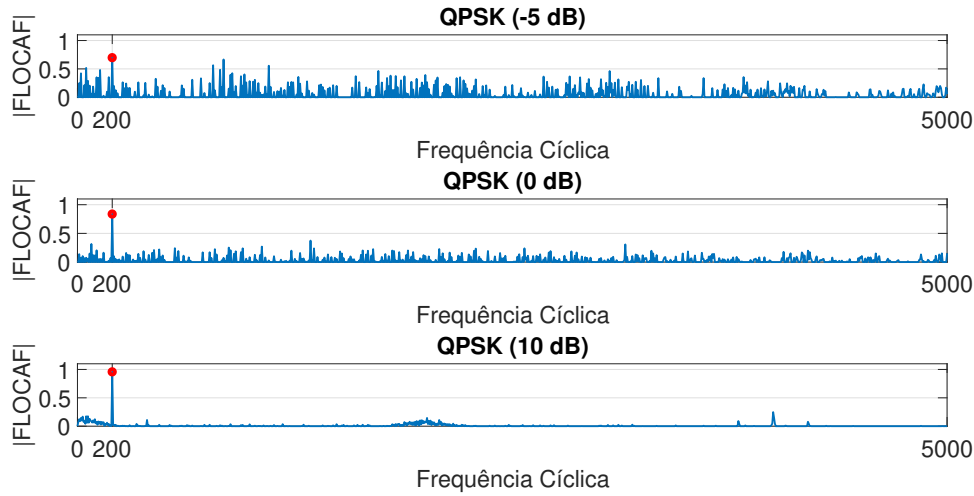


Figura 4.10: Assinaturas cicloestacionárias, da modulação QPSK, obtidas pela FLOCAF em três diferentes condições de $GSNR$.

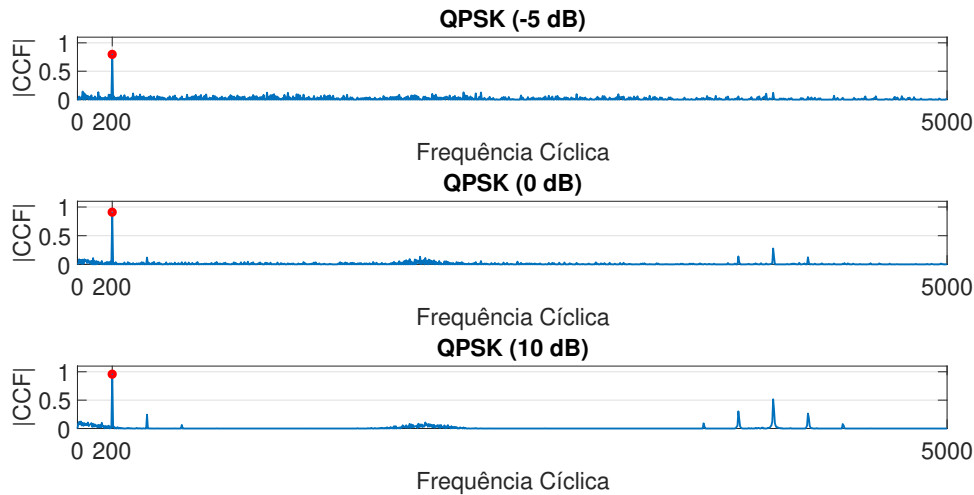


Figura 4.11: Assinaturas cicloestacionárias, da modulação QPSK, obtidas pela CCF em três diferentes condições de $GSNR$.

As Figuras 4.10 e 4.11, ilustram as assinaturas de um QPSK obtidas, considerando-se os parâmetros apresentados na Tabela 3.1, a partir das funções cicloestacionárias FLOCAF e CCF, em diferentes condições de $GSNR$. Em todos os casos, o componente associado à taxa de símbolos do sinal se mantém praticamente inalterado, enquanto os outros componentes são atenuados em baixos valores de $GSNR$.

Assim, uma vez que o componente cíclico associado à taxa de símbolos é robusto à contaminação por ruído impulsivo, além disso, considerando que a frequência intermediária (f_i) é conhecida, e que a taxa de símbolos (f_b) do sinal não excede metade de f_i , então f_b pode ser estimado a partir de assinaturas pelo método descrito a seguir: na faixa de 0 a $f_i/2$, a taxa de símbolos assume o valor da posição espectral do componente cíclico de maior amplitude.

Testes da estimação da taxa de símbolos usando a FLOCAF e a CCF, sob contaminação de ruído impulsivo, são descritos a seguir.

4.5.1 Estimação da taxa de símbolos usando a FLOCAF

Com o objetivo de avaliar a capacidade da FLOCAF, combinada com o método descrito anteriormente, em estimar a taxa de símbolos dos sinais BPSK, QPSK, 8-QAM 16-QAM e 32-QAM contaminados por ruído impulsivo, realizou-se o seguinte experimento: primeiro, sinais modulados foram contaminados com ruído aditivo alfa-estável ($\alpha = 1.5$, $\beta = 0$, $\delta = 0$), considerando-se níveis de $GSNR$ na faixa de -10 dB a 10 dB com passo de 2 dB. Então, uma estimação da taxa de símbolos foi obtida para cada um desses sinais.

A estimação global da taxa de símbolo, que consiste na média entre as taxas de acerto obtidas para cada modulação, foi calculada considerando 1000 testes por modulação, para cada condição de $GSNR$.

A partir da Figura 4.12 pode-se observar que através do uso da FLOCAF é possível obter um elevado nível de precisão na estimação da taxa de símbolo, mesmo para valores de $GSNR$ muito baixos. A estimação correta, que começa acima de 80% para a condição mais baixa de $GSNR$, cresce rapidamente e alcança uma taxa de acerto aproximadamente igual a 100% quando $GSNR = -2$ dB.

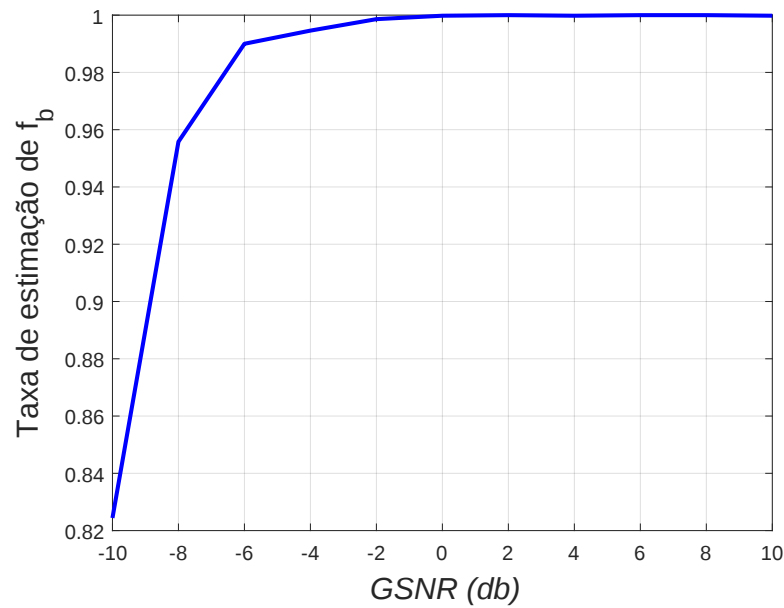


Figura 4.12: Estimação global da taxa de símbolos dos sinais modulados, usando a FLOCAF com $a = 0,3$ e $b = 0,5$.

4.5.2 Estimação da taxa de símbolos usando a CCF

O mesmo cenário usado na análise da estimação da taxa de símbolos pela FLOCAF, foi aplicado nos testes com a CCF: sinais modulados com as modulações avaliadas neste

trabalho foram contaminados com diferentes valores de GSNR. Uma estimativa da taxa de símbolos foi obtida para cada instância do sinal testado. A estimação global da taxa de símbolos foi calculada considerando 1000 testes por modulação, em cada condição de GSNR. O resultado está esboçado na Figura 4.13.

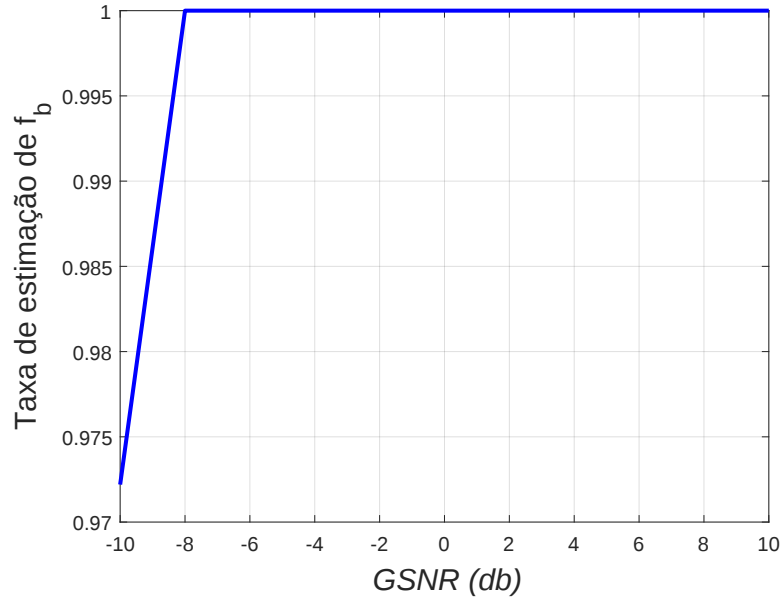


Figura 4.13: Estimação global da taxa de símbolos dos sinais modulados, usando a CCF com $\sigma = 1, 2$.

A partir da Figura 4.13 observa-se que, com a CCF é possível obter uma precisão elevada para a estimação da taxa de símbolo dos sinais modulados, mesmo para valores baixos de GSNR. A estimação correta, que começa acima de 97% na condição mais baixa de GSNR, alcança taxas de acerto de 100% na GSNR = -8 dB. Analisando-se os resultados apresentados nesta seção, observa-se que a CCF consegue ser mais precisa que a FLOCAF na estimação da taxa de símbolos dos sinais modulados.

Neste Capítulo duas arquiteturas de AMC baseadas nas funções FLOCAF e CCF, e estruturadas em árvore de decisão, foram propostas. Além disso, também foi proposto um método de estimação da taxa de símbolo de um sinal modulado baseado na análise das assinaturas cicloestacionárias obtidas pela FLOCAF e CCF. No próximo Capítulo, arquiteturas de AMC baseadas nessas mesmas funções cicloestacionárias, combinadas com regressão logística, são propostas.

Capítulo 5

Arquiteturas de AMC com Regressão Logística

Neste Capítulo, arquiteturas de AMC baseadas nas funções FLOCAF e CCF, combinada com a técnica de regressão logística, são avaliadas. As arquiteturas foram testadas em ambientes caracterizados por um ruído aditivo alfa-estável não-gaussiano com os seguintes parâmetros: $\alpha = 1.5$, $\beta = 0$, $\delta = 0$, e $GSNR$ variando na faixa de -4 dB a 10 dB com passo de 2 dB.

Por fim, os resultados das arquiteturas de AMC baseados em regressão logística foram comparados com os resultados das arquiteturas AMC baseada em árvore de decisão apresentadas no Capítulo anterior.

5.1 Arquitetura de classificação de modulação baseada em análise cicloestacionária e regressão logística

Combinando a análise cicloestacionária com regressão logística é possível realizar uma classificação das modulações digitais: BPSK, QPSK, 8QAM, 16QAM e 32QAM. Para isso, é necessário ajustar o conjunto de treinamento com diversas assinaturas cicloestacionárias de cada modulação, geradas pela FLOCAF ou pela CCF. Nesse contexto, as características do conjunto de treinamento são as assinaturas cicloestacionárias e os alvos são os tipos de modulação descritas acima, representadas respectivamente pelos números inteiros de 1 a 5 .

Uma vez executado o treinamento com as assinaturas cicloestacionárias, obtém-se os parâmetros θ , e por sua vez, as funções hipótese $h_{\theta}(x)$ associadas a cada classe de modulação. Definidas as funções hipótese, a classificação ocorre da seguinte forma:

- inicialmente, a assinatura do sinal de entrada deve ser testada nas funções $h_{\theta}(x)$ de cada sinal modulado;
- então, verifica-se qual hipótese retorna o valor mais alto (hipótese mais provável);
- por fim, o sinal de entrada é atribuído à classe da modulação cuja hipótese é a mais provável.

Um detalhe importante é que a assinatura de entrada do classificador deve ser obtida

pela mesma função cicloestacionária utilizada no conjunto de treinamento. Em outras palavras, se as funções hipótese de cada modulação forem geradas a partir de um conjunto de treinamento de assinaturas obtidas com FLOCAF, a assinatura de entrada do classificador também deve ser obtida pela FLOCAF¹. O processo de classificação completo está apresentado na Figura 5.1.

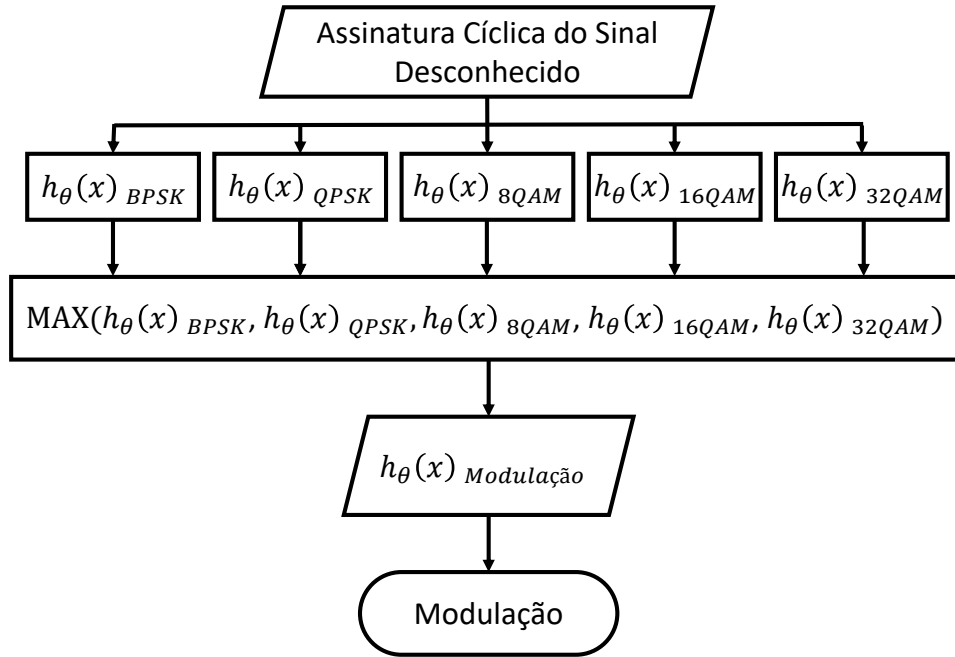


Figura 5.1: Arquitetura de AMC baseada na extração de características cicloestacionárias aplicada a regressão logística.

5.1.1 Perfil Cíclico Reduzido

O elemento de entrada do classificador é a assinatura cicloestacionária do sinal desconhecido, por sua vez, as características de entrada são os componente cíclicos dessa assinatura. Como demonstrado no Capítulo 3, as assinaturas cicloestacionárias geradas pela FLOCAF e CCF são simétricas, e apresentam apenas um conjunto de componente cíclicos não nulos, os quais são mapeados em posições do espectro relacionadas com a taxa de símbolos (f_b), frequência intermediária (f_i), e frequência de amostragem (f_s) do sinal modulado. Tais elementos, são os componentes cicloestacionários de interesse para o reconhecimento de cada modulação.

Cada tipo de modulação possui um arranjo específico dos componentes cicloestacionários de interesse, os quais variam em amplitude e posição de acordo com o tipo de modulação. Desse modo, para reconhecer um sinal modulado, ao invés de verificar todos os elementos de sua assinatura cicloestacionária, basta avaliar os componentes cíclicos de interesse associados a esse sinal. Tal abordagem é chamada de análise do perfil cíclico reduzido.

¹Essa observação também é válida para a CCF.

A análise do perfil cíclico reduzido permite que o treinamento, que define a função hipótese de cada modulação, seja menos custoso pois reduz consideravelmente o número de características do conjunto de treinamento. Para isso, o perfil cíclico reduzido deve ser ajustado para respeitar arranjo componentes cicloestacionários de cada sinal modulado.

Seja M_l um conjunto de componentes cíclicos não-nulos que compõem um descritor de uma modulação, o perfil cíclico reduzido que descreve o conjunto de P modulações distintas será dado por

$$P_{fr} = \bigcup_{l=1}^P M_l. \quad (5.1)$$

Em outras palavras, o perfil cíclico da assinatura de entrada da arquitetura de classificação é definido a partir da união das posições dos componentes cíclicos de interesse de todas as modulações avaliadas. Por exemplo: se uma modulação M_1 apresentar uma assinatura cicloestacionária com componentes não-nulos nas posições $\{f_b, 2f_b, 3f_b\}$, e uma modulação M_2 apresentar uma assinatura com componentes cíclicos não-nulos nas posições $\{f_b, f_i\}$, o perfil-cíclico que contempla as duas modulações deve ser analisado nas posições $\{f_b, 2f_b, 3f_b, f_i\}$.

Considerando a parametrização aplicada nas arquiteturas definidas no Capítulo 4, de $a = 0.3, b = 0.5$ para FLOCAF e $\sigma = 1, 2$ para CCF, os perfis cíclicos reduzidos associados às assinaturas de ambas as funções cicloestacionárias serão definidos avaliando-se os componentes espectrais nas posições:

$$P_{fr} = \{f_b; 2f_b; 2f_i - 2f_b; 2f_i - f_b; 2f_i; 2f_i + f_b; 2f_i + 2f_b; 4f_i - f_b; 4f_i; 4f_i + f_b\}. \quad (5.2)$$

Os testes de classificação foram realizados aplicando-se a análise cicloestacionária à regressão logística, considerando o perfil-cíclico completo e a abordagem de perfil-cíclico reduzido. Os desempenhos de cada arquitetura estão descritos na seção seguinte.

5.2 Resultado de classificação com Regressão Logística

O conjunto de treinamento usado no treinamento de cada arquitetura foi composto de assinaturas cicloestacionária das modulações BPSK, QPSK, 8-QAM, 16-QAM, e 32-QAM, as quais foram geradas em condições de canal de $GSNR$ diferentes, variando de 0 dB a 10 dB com passo de 2 dB. Para cada condição de $GSNR$, 4000 exemplos de assinaturas foram gerados por modulação. Desse modo, cada conjunto de treinamento possui $m = 120000$ exemplos de treinamento, contendo 24000 exemplos por modulação.

Além disso, cada assinatura cicloestacionária é um sinal discreto com 2001 amostras, contudo, como o espectro cíclico é simétrico, apenas os componentes associados as frequências não negativas foram utilizados no treinamento, desse modo, o número de características do conjunto de treinamento foi limitado a $n = 1001$. No caso das archi-

teturas com abordagem de perfil-cíclico reduzido, o número de características é limitado para $n = 10$. Os demais parâmetros utilizados na configuração de sinais modulados, e nas funções cicloestacionárias, estão descritos na Tabela 5.1.

Tabela 5.1: Parâmetros de Simulação.

Parâmetro	Valor
Frequência de amostragem (f_s)	10 kHz
Frequência Intermediária (f_i)	1 kHz
Taxa de Símbolos (f_b)	200 baud
Número de Símbolos	$n_{sym} = 1000$
Fator de <i>roll-off</i>	$\nu = 0,75$
Número de amostras por janela	$N = 2001$
Fator de sobreposição entre janelas	$Overlap = 0,5$
Ordem-inferior fracionária	$a = 0,3; b = 0,5$
Tamanho do kernel	$\sigma = 1,2$

Para avaliar a robustez das arquiteturas, os testes de classificação foram realizados em condições de canal um pouco mais severas do que o treinamento, nesse caso, com a $GSNR$ variando de -4 dB a 10 dB com passo de 2 dB. Em cada valor de $GSNR$, foram utilizados 1000 exemplos de assinaturas para cada modulação. O desempenho de cada arquitetura de classificação está descrito nas subseções seguintes.

5.2.1 AMC utilizando a FLOCAF com regressão logística

A Figura 5.2 ilustra a taxa de acerto da arquitetura baseada na FLOCAF combinada com a regressão logística, em diferentes valores de $GSNR$. No caso das modulações BPSK e 8-QAM, mesmo na condição de $GSNR$ mais baixa, a taxa de acerto é de aproximadamente 100%. As curvas de acerto das demais modulações começam baixas, mas apresentam um crescimento gradativo conforme a $GSNR$ aumenta. Ainda assim, a taxa de acerto da 16-QAM se mostrou bem abaixo das demais modulações. Isso sugere que, essa modulação é mais sensível ao ruído impulsivo.

Para uma análise mais aprofundada do classificador, as matrizes de confusão nas condições de $GSNR$ de 0 dB, 4 dB, e 10 dB foram avaliadas, respectivamente apresentadas da Tabela 5.2 à Tabela 5.4.

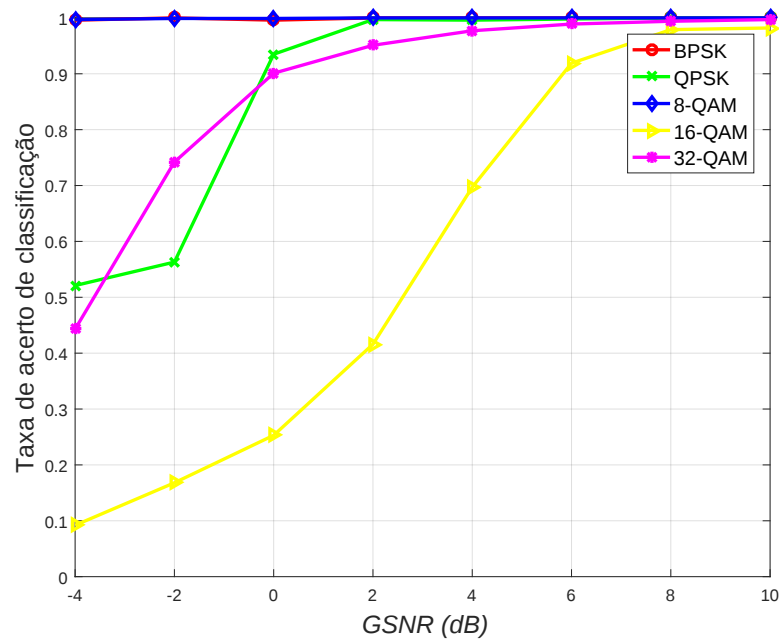


Figura 5.2: Curvas de classificação da arquitetura baseada na FLOCAF ($a = 0,3$ e $b = 0,5$) combinada com regressão logística.

Tabela 5.2: Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	996	0	0	0	0
QPSK	0	935	0	28	22
8QAM	4	0	999	1	0
16QAM	0	59	1	253	77
32QAM	0	6	0	718	901

Os resultados apresentados na matriz de confusão da Tabela 5.2 indicam que, no caso de $GSNR = 0$ dB, a maior parte dos sinais de 16-QAM são equivocadamente classificados como 32-QAM. Na condição de $GSNR = 4$ dB (Tabela 5.3) o classificador melhora o desempenho, mas ainda atribui muitos sinais 16-QAM à classe de 32-QAM, entretanto, para todas as demais modulações têm-se taxa de acerto acima de 90%.

Para a $GSNR = 10$ dB, conforme apresentado na Tabela 5.4, as classificações de todas as modulações tiveram taxa de acerto acima de 98%, em particular, chegando a 100% de acerto para as modulações BPSK, QPSK e 8-QAM.

Tabela 5.3: Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	996	0	19	5
8QAM	0	0	1000	0	0
16QAM	0	4	0	697	18
32QAM	0	0	0	284	977

Tabela 5.4: Matriz de confusão da arquitetura de AMC com FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	16	1
8QAM	0	0	1000	0	0
16QAM	0	0	0	982	2
32QAM	0	0	0	2	997

5.2.2 AMC utilizando a FLOCAF com perfil-cíclico reduzido e regressão logística

A Figura 5.3 ilustra o desempenho da arquitetura baseada na FLOCAF com o perfil cíclico reduzido e regressão logística, em diferentes valores de $GSNR$. Novamente, as modulações BPSK e 8-QAM foram as mais facilmente reconhecidas, apresentando taxa de acerto em torno de 100%, mesmo em condição de $GSNR$ mais severas. Entre as demais modulações, o 16-QAM novamente se mostrou mais sensível ao ruído.

Para realizar uma análise mais detalhada do classificador, matrizes de confusão nas condições de $GSNR$ de 0 dB, 4 dB, e 10 dB foram avaliadas, respectivamente apresentadas da Tabela 5.5 à Tabela 5.7.

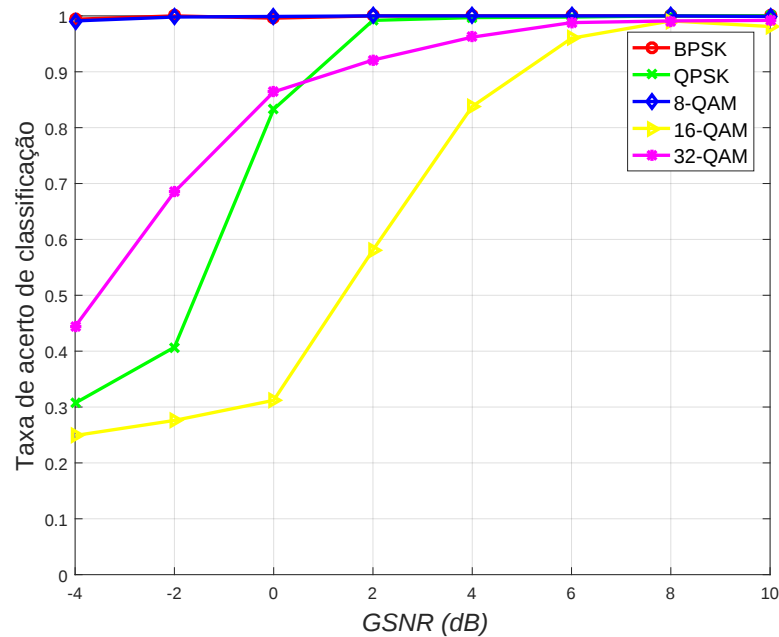


Figura 5.3: Curvas de classificação da arquitetura baseada no perfil-cíclico reduzido da FLOCAF ($a = 0,3$ e $b = 0,5$) combinada com regressão logística, considerando 1000 testes por modulação em cada condição de $GSNR$.

A matriz de confusão descrita na Tabela 5.5, indica que na condição de $GSNR = 0$ dB a modulação 16-QAM é amplamente confundida com a modulação 32-QAM, enquanto as demais modulações apresentam uma taxa de classificação superior a 83%. Por sua vez, a matriz de confusão descrita na Tabela 5.6 mostra que, para a $GSNR = 4$ dB, o classificador consegue identificar corretamente cerca de 83% das modulações 16-QAM, e para as demais modulações a taxa de acerto é superior a 96%.

Na condição de $GSNR = 10$ dB, como apresentado na Tabela 5.7, o classificador alcança uma taxa de acerto superior a 98% para a modulação 16-QAM, taxas de acerto superiores a 99% para as modulações 8-QAM e 32-QAM, e aproximadamente 100% de acerto para as modulações BPSK e QPSK.

Tabela 5.5: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	996	0	0	0	0
QPSK	0	833	1	38	13
8QAM	4	0	999	1	0
16QAM	0	156	0	312	123
32QAM	0	11	0	649	864

Tabela 5.6: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	997	0	12	2
8QAM	0	0	1000	0	0
16QAM	0	1	0	838	36
32QAM	0	2	0	150	962

Tabela 5.7: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da FLOCAF e regressão logística ($a = 0,3$, $b = 0,5$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	1	16	0
8QAM	0	0	999	0	0
16QAM	0	0	0	981	8
32QAM	0	0	0	3	992

Comparando ambas as arquiteturas que empregam a FLOCAF, Figura 5.2 e Figura 5.3, verifica-se que elas possuem comportamentos similares. Contudo, em condições de $GSNR$ negativa, a arquitetura que utiliza o perfil-cíclico completo apresenta taxa de acertos maiores do que a arquitetura baseada em perfil-cíclico reduzido. Ainda assim, a partir da condição de canal correspondente a $GSNR = 0$ dB, o classificador baseado em perfil-cíclico reduzido torna-se mais competitivo, em particular, alcançando uma taxa de acerto do 16-QAM maior do que a taxa apresentada pelo outro classificador que aplica a FLOCAF com perfil-cíclico completo.

5.2.3 AMC utilizando a CCF com regressão logística

A Figura 5.4 descreve, em diferentes condições de canal, o desempenho da arquitetura baseada na CCF combinada com a regressão logística. Além disso, as matrizes de confusão nas condições de $GSNR$ de 0 dB, 4 dB, e 10 dB foram avaliadas para uma análise mais aprofundada do classificador, respectivamente apresentadas da Tabela 5.8 à Tabela 5.10.

As taxas de acerto associadas às modulações BPSK e 8-QAM, por exemplo, são muito próximas de 100%, e isso se mantém para todas as condições de canal avaliadas. As demais modulações apresentam taxas de acerto menores em $GSNR$ s mais baixas. Ainda assim, a cada aumento da $GSNR$, as curvas de desempenho apresentam ganhos consideráveis. Ainda assim, mais uma vez, a curva de taxa de acerto da modulação 16-QAM se mostrou abaixo das curvas das demais modulações na maioria dos valores de $GSNR$

investigados.

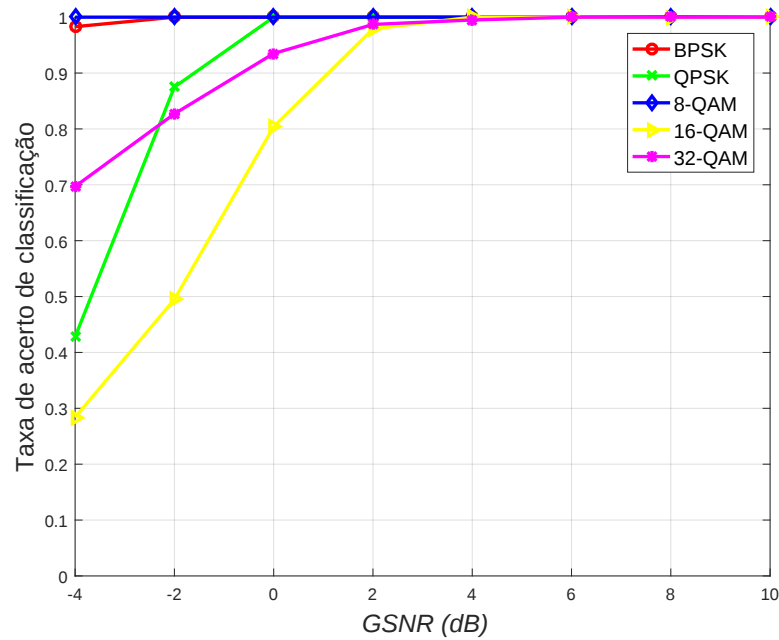


Figura 5.4: Curvas de classificação da arquitetura baseada na CCF ($\sigma = 1.2$) combinada com regressão logística.

Tabela 5.8: Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	807	52
32QAM	0	0	0	193	948

A matriz de confusão descrita na Tabela 5.8, apresenta o comportamento do classificador na condição de canal equivalente a $GSNR = 0$ dB. Nesse cenário, as modulações BPSK, QPSK, e 8-QAM foram classificadas com taxa de acerto de 100%. Contudo, em torno de 20% dos sinais 16-QAM foram equivocadamente atribuídos a classe da modulação 32-QAM. Por sua vez, alguns sinais 32-QAM também foram confundidos com a classe da modulação 16-QAM, porém, com um percentual um pouco maior que 5%.

Tabela 5.9: Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	997	1
32QAM	0	0	0	3	999

Tabela 5.10: Matriz de confusão da arquitetura de AMC com CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	1000	0
32QAM	0	0	0	0	1000

Na condição de $GSNR = 4$ dB, conforme apresentado na Tabela 5.9, a arquitetura classificou equivocadamente apenas 3 sinais 16-QAM, e apenas 1 sinal da modulação 32-QAM. Por sua vez, em uma condição de $GSNR = 10$ dB, conforme apresentado na Tabela 5.10, o classificador identificou corretamente todos os sinais modulados.

5.2.4 AMC utilizando a CCF com perfil-cíclico reduzido e regressão logística

A Figura 5.5 ilustra o desempenho da arquitetura baseada na CCF com o perfil cíclico reduzido e regressão logística, em diferentes valores de $GSNR$. Os resultados mostram que as modulações mais facilmente reconhecidas foram novamente a BPSK e 8-QAM, as quais foram classificadas com taxa de acerto aproximadamente igual a 100% em todas as condições de $GSNR$. Para as demais modulações, as taxas de classificação aumentam com o aumento da $GSNR$, de tal forma que para $GSNR$ positivas todas as modulações foram classificadas com taxa de acerto superior a 83%. Para aprofundar a análise do classificador, matrizes de confusão nas condições de $GSNR$ de 0 dB, 4 dB, e 10 dB foram avaliadas, respectivamente apresentadas da Tabela 5.11 à Tabela 5.13.

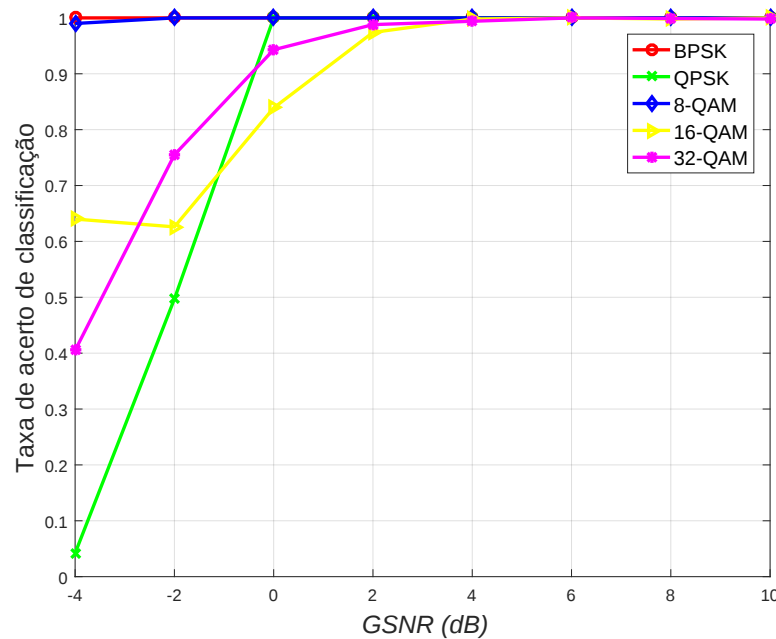


Figura 5.5: Curvas de classificação da arquitetura baseada no perfil-cíclico reduzido da CCF ($\sigma = 1.2$) combinada com regressão logística.

Tabela 5.11: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 0$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	993	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	7	0	837	58
32QAM	0	0	0	163	942

Os resultados de classificação em uma condição de $GSNR = 0$ dB, apresentados na Tabela 5.11, mostram que a arquitetura identificou corretamente todos os sinais BPSK, e 8-QAM, além de alcançar uma taxa de acerto superior a 99% para os sinais QPSK. Ainda nesse cenário, o classificador apresentou cerca de 94% de acerto associado à modulação 32-QAM, e alcançou uma taxa de acerto acima de 83% para a modulação 16-QAM.

Em uma condição de $GSNR = 4$ dB, conforme apresentado na Tabela 5.12, a arquitetura classificou equivocadamente apenas 5 sinais 32-QAM, e acertou na classificação de todos os sinais das demais modulações. Já em uma condição de $GSNR = 10$ dB, conforme apresentado na Tabela 5.13, o classificador errou a classificação de apenas um sinal 32-QAM, alcançando mais uma vez, 100% de acerto para as demais modulações.

Tabela 5.12: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 4$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	1000	5
32QAM	0	0	0	0	995

Tabela 5.13: Matriz de confusão da arquitetura de AMC com perfil-cíclico reduzido da CCF e regressão logística ($\sigma = 1, 2$, $GSNR = 10$ dB).

Modulação Classificada	Modulação Verdadeira				
	BPSK	QPSK	8QAM	16QAM	32QAM
BPSK	1000	0	0	0	0
QPSK	0	1000	0	0	0
8QAM	0	0	1000	0	0
16QAM	0	0	0	1000	1
32QAM	0	0	0	0	999

5.3 Comparação entre as arquiteturas de classificação

Avaliando as arquiteturas de classificação apresentadas neste Capítulo, alguns comportamentos comuns entre elas podem ser destacados, tais como:

- as modulações mais facilmente identificadas, mesmo em condição de $GSNR$ mais baixas, foram a BPSK, e 8-QAM;
- a partir das condições de $GSNR$ não-negativas, todas as arquiteturas apresentaram ganhos significativos na identificação das modulações QPSK e 32-QAM;
- a modulação mais difícil de identificar por todos os classificadores foi a 16-QAM. Isso indica que as assinaturas cicloestacionárias associadas a essa modulação são muito sensíveis ao ruído, o que faz com que ela, muitas vezes, sejam confundidas com as assinaturas da modulação 32-QAM.

Para investigar a arquitetura de melhor desempenho, a taxa de acerto global (média entre as curvas de taxas de acertos das modulações) de cada classificador foi avaliada. Além das arquiteturas propostas nesse Capítulo, as arquiteturas baseadas em árvore de decisão, apresentadas no Capítulo 4, foram utilizadas na comparação. Os resultados estão ilustrados na Figura 5.6:

Entre as arquiteturas avaliadas, a que apresentou o desempenho mais baixo foi a que empregou FLOCAF com árvore de decisão. Esse classificador apresentou uma taxa de acerto abaixo dos demais.

Os classificadores baseados em FLOCAF com regressão logística apresentaram comportamentos similares, entretanto, a superioridade de cada um varia de acordo com a

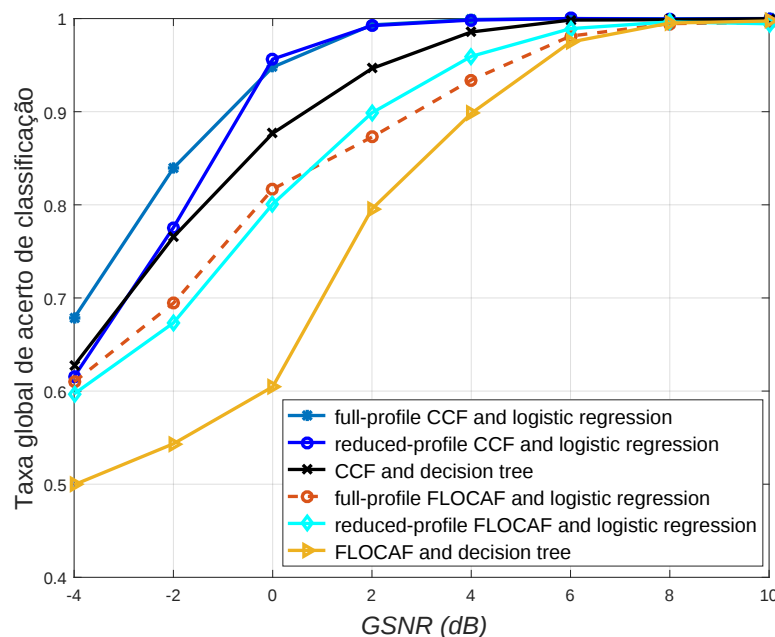


Figura 5.6: Curvas de classificação global de diversas arquiteturas de AMC.

faixa de $GSNR$. Para a faixa de $GSNR$ variando de -4 dB a 0 dB, a arquitetura que utiliza o perfil-cíclico completo foi ligeiramente superior à arquitetura baseada em perfil-cíclico reduzido. Por sua vez, na faixa de $GSNR$ variando de 2 dB a 6 dB a superioridade se inverte, agora com uma sutil vantagem da arquitetura baseada em perfil-cíclico reduzido. Em condições de $GSNR$ de 8 dB a 10 dB, ambas as arquiteturas apresentaram desempenhos similares, com taxa de acerto por volta de 100%.

A arquitetura baseada na CCF combinada com árvore de decisão, apresentou desempenho superior a todas as arquiteturas que empregaram a análise cicloestacionária usando a da FLOCAF. Além disso, no intervalo de $GSNR$ de -4 dB a -2 dB, esse classificador obteve o segundo melhor desempenho entre as arquiteturas. Contudo, a partir de $GSNR = 0$ dB, seu desempenho ficou muito abaixo de ambas as arquiteturas que utilizam a CCF combinada com regressão logística.

As arquiteturas que empregam a CCF com regressão logística, na maioria das condições de $GSNR$ avaliadas, apresentaram desempenhos superiores às demais arquiteturas. Nesse contexto, a arquitetura baseada na análise do perfil-cíclico reduzido, no intervalo de $GSNR$ -4 dB a -2 dB, apresentou desempenho abaixo da arquitetura que emprega a análise do perfil-cíclico completo. Essa, por sua vez, foi a arquitetura que mais vezes identificou corretamente as modulações avaliadas. Ainda assim, a partir da $GSNR = 0$ dB, ambos os classificadores alcançaram taxas de acerto superiores a 95%, e desse ponto em diante, apresentaram aproximadamente o mesmo comportamento. A partir da $GSNR = 2$ dB, ambas as arquiteturas alcançaram taxas de acerto muito próximas a 100%, tendência que se manteve até a condição de $GSNR = 10$ dB.

Esses resultados mostram que as arquiteturas que empregam a análise cicloestacionária usando a CCF, alcançaram desempenhos superiores às arquiteturas que utilizaram a FLOCAF. Essa particularidade reforça que a CCF consegue extrair características ciclo-

estacionárias mais robustas ao ruído impulsivo, do que aquelas obtidas pela FLOCAF.

Além disso, os resultados também demonstram que a técnica de regressão logística, na maioria dos casos, foi muito superior ao método baseado em árvore de decisão proposto no Capítulo 4. Nesse contexto, destaca-se que a estratégia baseada no perfil-cíclico reduzido pois, além de demandar um treinamento menos custoso, é mais adequada quando o canal não é sujeito ao ruído impulsivo muito intenso. Caso contrário, o classificador treinado a partir da análise do perfil-cíclico completo é mais apropriado, pois apresenta uma taxa de classificação mais alta nos valores negativos de GSNR.

Capítulo 6

Conclusão

Neste trabalho, foram propostas arquiteturas de classificação automática de modulações robustas ao ruído impulsivo baseadas na análise cicloestacionária. Nas diferentes arquiteturas apresentadas foram empregadas as funções FLOCAF e CCF, além de estratégias de classificação distintas baseadas em árvore de decisão e regressão logística.

Inicialmente, foi apresentada uma caracterização do ruído aditivo do canal de comunicação considerado no trabalho, modelado pela distribuição alfa-estável. Após isso, foi feita uma descrição dos processos cicloestacionários e as principais funções cicloestacionárias utilizadas no trabalho para a extração de características das modulações consideradas, sendo elas: função de autocorrelação cíclica (CAF), função de autocorrelação cíclica fracionária de ordem-inferior (FLOCAF), função sinal espacial de correlação cíclica (SSCCF), e a função de correntropia cíclica (CCF). Nesse processo, foi verificado que a CAF é ineficiente para extrair informações cicloestacionárias de sinais contaminados por ruído impulsivo, além disso, foi demonstrado que a SSCCF é um caso particular da FLOCAF. Consequentemente, mediante a análise das funções cicloestacionárias, foi verificado que a FLOCAF e a CCF conseguem obter características cicloestacionárias de sinais contaminados por ruído impulsivo.

Após a análise teórica, as funções FLOCAF e CCF foram avaliadas quanto à capacidade de diferenciar as modulações digitais: BPSK, QPSK, 8-QAM, 16-QAM, e 32-QAM. Nesse contexto, foi verificado que ambas as funções são capazes de fornecer descritores cicloestacionários (assinaturas cicloestacionárias) singulares para cada modulação. Tais descritores, que podem ser definidos como conjuntos particulares de frequências cíclicas, são diretamente relacionados a algumas características de geração do sinal, tais como: frequência intermediária (f_i), taxa de símbolos (f_b) e frequência de amostragem (f_s). Além disso, também foi verificado que as assinaturas das modulações digitais, geradas pela FLOCAF e CCF, apresentavam robustez ao ruído impulsivo.

Após avaliar a potencialidade da FLOCAF e CCF para identificar modulações digitais em canais com ruído aditivo impulsivo, duas arquiteturas de classificação automática de modulações foram propostas. As arquiteturas apresentadas utilizaram a análise cicloestacionária para obtenção de descritores dos sinais modulados e seguiram uma árvore de decisão cujos critérios de classificação foram baseados em casamento de *templates*, e análise de limiares. A principal diferença entre os classificadores foram as funções cicloestacionárias empregadas: um utilizava a FLOCAF e o outro a CCF. Ambos os classificadores foram capazes de identificar as modulações propostas em diferentes condições

de *GSNR*. Nesse contexto, a arquitetura que apresentou o melhor desempenho foi a que empregou a CCF.

No processo de testes dos classificadores propostos, foi observado que o ruído impulsivo, quando possui potência muito elevada, pode mascarar alguns componentes cicloestacionários das assinaturas dos sinais modulados, fator que atenua a classificação em *GSNRs* mais baixas. Ainda assim, foi verificado também que, alguns componentes cicloestacionários eram mais robustos do que outros, em particular, o componente associado a taxa de símbolos da modulação foi o mais resistente a contaminação por ruído aditivo impulsivo. Com base nesse aspecto, um método de reconhecimento de taxa de símbolos, a partir de assinaturas cicloestacionárias, foi proposto e testados em diferentes condições de *GSNR*. Os experimentos mostraram que ambas as funções cicloestacionárias foram muito bem-sucedidas na estimação da taxa de símbolos dos sinais modulados. Nesse contexto, mais um vez a estimação de parâmetro baseada na CCF se mostrou superior a estimação baseada na FLOCAF.

Avançando no desenvolvimento de arquiteturas de AMC, voltadas à ambientes impulsivos, novos classificadores que utilizam a análise cicloestacionária foram propostos. Dessa vez, visando alcançar um desempenho superior às arquiteturas anteriormente apresentadas, a técnica de aprendizagem de máquina, denominada regressão logística, foi empregada como método de classificação. Nesse contexto, também foi introduzida a estratégia de análise do perfil-cíclico reduzido. Devido a natureza das assinaturas cicloestacionárias, os componentes espectrais de cada modulação podem ser previamente mapeados uma vez conhecidos os parâmetros f_i , f_b , e f_s do sinal. Essa particularidade permite que, ao invés de avaliar todos os elementos da assinatura cicloestacionária, a análise seja reduzida apenas a posições espectrais de interesse, o que resulta no perfil-cíclico reduzido.

Uma vez que as assinaturas cicloestacionárias compõem o banco de dados utilizado no treinamento do classificador, empregar o perfil-cíclico reduzido implica na redução da complexidade do treinamento. Ainda assim, com o intuito de ampliar a variedade de classificadores, arquiteturas que empregam o treinamento com o perfil-cíclico completo também foram propostas. Desse modo, quatro tipos de sistemas foram desenvolvidos: FLOCAF com perfil-cíclico reduzido; FLOCAF com perfil-cíclico completo; CCF com perfil-cíclico reduzido; CCF com perfil-cíclico completo, todos eles utilizando a regressão logística como classificador.

Posteriormente, os desempenhos de todas as arquiteturas foram comparadas entre si, utilizado a taxa de acerto global de cada classificador. As arquiteturas que utilizaram a FLOCAF como função cicloestacionária, em todos os casos, apresentaram desempenho inferior às arquiteturas que aplicaram a CCF. Esse comportamento reforça que a CCF é mais adequada para discriminar as modulações digitais investigadas neste trabalho, bem como, é mais robusta à contaminação por ruído impulsivo aditivo, em comparação à FLOCAF. Os resultados também mostraram que a classificação por regressão logística, de maneira geral, foi superior à classificação feita a partir da árvore de decisão previamente apresentada.

Por último, foi observado que a estratégia de treinamento baseada na análise do perfil-cíclico reduzido apresentou um comportamento muito próximo da estratégia baseada na

análise do perfil-cíclico completo. Contudo, algumas particularidades podem ser destacadas: em condições de $GSNR$ negativas, as arquiteturas que empregaram o treinamento a partir do perfil-cíclico completo alcançaram uma maior taxa de acerto; em contra partida, a partir de condições de $GSNR$ positivas, as arquiteturas baseadas no perfil-cíclico reduzido apresentaram uma taxa de acerto maior ou igual às arquiteturas que utilizam o treinamento a partir do perfil-cíclico completo. Essa conjuntura demanda uma avaliação de custo-benefício, que pode variar de acordo com a aplicação do classificador: a estratégia baseada no perfil-cíclico reduzido, por resultar em um treinamento menos custoso, é mais adequada quando o canal não for propenso a ruído muito intenso; caso contrário, o classificador treinado a partir da análise do perfil-cíclico completo é mais eficiente.

Em resumo, ambas as funções FLOCAF e CCF podem ser utilizadas com sucesso na caracterização das modulações digitais BPSK, QPSK, 8-QAM, 16-QAM e 32-QAM. Além disso, as diversas arquiteturas de AMC, propostas nesse trabalho, demonstraram que essas funções são robustas ao ruído impulsivo. Ainda nesse contexto, as arquiteturas que empregam a CCF têm melhor desempenho, isso reforça que essa função é mais resistente a contaminação por ruído impulsivo, e consegue caracterizar os sinais modulados melhor do que a FLOCAF. Entre as estratégias de classificação empregadas, as que utilizaram aprendizado de máquina por meio de regressão logística obtiveram desempenho superior à técnicas mais simples, como casamento de *templates* e análise de limiares.

6.1 Trabalhos Produzidos

Durante o desenvolvimento desta Tese, os seguintes trabalhos foram produzidos:

- Artigo em Periódico: **Automatic Modulation Classification Architectures Based on Cyclostationary Features in Impulsive Environments**. Trabalho publicado em 23 de setembro de 2019 pela IEEE Access, DOI: 10.1109/ACCESS.2019.2943300.
- Artigo em Anais de Evento: **Extração de Características Cicloestacionárias por Correntropia Cíclica em Canais com Desvanecimento**. Trabalho publicado em 04 de dezembro de 2020 na Conferência Nacional em Comunicações, Redes e Segurança da Informação (ENCOM 2020), ISBN: 978-65-5941-011-8.
- Propriedade Intelectual: **Extrator de Características de Alta Ordem via Correntropia Cíclica**. Programa de Computador registrado em 03 de novembro de 2020 pelo Instituto Nacional da Propriedade Industrial (INPI), Processo N°: BR512020002317-7.

6.2 Trabalhos Futuros

Como trabalhos futuros, pretende-se estender as aplicações de análise cicloestacionária com a CCF, combinadas com outras técnicas de aprendizagem de máquina, para o desenvolvimento de sistemas de telecomunicações capazes de identificar modulações

digitais de alta ordem, tais como 64-QAM e 128-QAM. Além disso, outros cenários de comunicação sem-fio com desvanecimento e doppler, além do ruído impulsivo aditivo, também serão considerados.

Referências Bibliográficas

- Aalo, Valentine A, Kostas P Peppas, George Efthymoglou, Mohammed Alwakeel & Sami Alwakeel (2014), 'Evaluation of average bit error rate for wireless networks with alpha-stable interference', *Electronics Letters* **50**(1), 47–49.
- Ahmed, Abrar, Ahmad Jalal & Kibum Kim (2020), 'A novel statistical method for scene classification based on multi-object categorization and logistic regression', *Sensors* **20**(14), 3871.
- Alharbi, Hazza, Shoaib Mobien, Saleh Alshebeili & Fahd Alturki (2012), 'Automatic modulation classification of digital modulations in presence of hf noise', *EURASIP Journal on Advances in Signal Processing* **2012**(1), 238.
- Amjad, Muhammad, Mubashir Husain Rehmani & Shiwen Mao (2018), 'Wireless multimedia cognitive radio networks: A comprehensive survey', *IEEE Communications Surveys & Tutorials* **20**(2), 1056–1103.
- Artuso, Alysson Ramos (2012), 'Entropias de shannon e rényi aplicadas ao reconhecimento de padrões', *Revista CIATEC-UPF* **3**(2), 56–72.
- Au, Minh, Basile L Agba & François Gagnon (2015), 'An analysis of transient impulsive noise in a poisson field of interferers for wireless channel in substation environments', *arXiv preprint arXiv:1504.06880*.
- Avci, Engin, Davut Hanbay & Asaf Varol (2007), 'An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition', *Expert Systems with Applications* **33**(3), 582–589.
- Avci, Engin & Derya Avci (2008), 'The performance comparison of discrete wavelet neural network and discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition', *Expert Systems with Applications* **35**(1-2), 90–101.
- Bardyn, Jean-Paul, Thierry Melly, Olivier Seller & Nicolas Sornin (2016), Iot: The era of lpwan is starting now, *em 'European Solid-State Circuits Conference, ESSCIRC Conference 2016: 42nd'*, IEEE, pp. 25–30.
- Belkacemi, Hocine & Sylvie Marcos (2007), 'Robust subspace-based algorithms for joint angle/doppler estimation in non-gaussian clutter', *Signal Processing* **87**(7), 1547–1558.

- Bhatti, Shahzad A, Qingshan Shan, Ian A Glover, Robert Atkinson, Illiana E Portugues, Philip J Moore & Richard Rutherford (2009), Impulsive noise modelling and prediction of its impact on the performance of wlan receiver, *em* 'Signal Processing Conference, 2009 17th European', IEEE, pp. 1680–1684.
- Blackard, Kenneth L., Theodore S. Rappaport & Charles W. Bostian (1993), 'Measurements and models of radio frequency impulsive noise for indoor wireless communications', *IEEE Journal on selected areas in communications* **11**(7), 991–1001.
- Blum, Rick S, Richard J Kozick & Brian M Sadler (1999), 'An adaptive spatial diversity receiver for non-gaussian interference and noise', *IEEE Transactions on Signal Processing* **47**(8), 2100–2111.
- Button, Mark D, John G Gardiner & Ian A Glover (2002), 'Measurement of the impulsive noise environment for satellite-mobile radio systems at 1.5 ghz', *IEEE Transactions on vehicular technology* **51**(3), 551–560.
- Chang, Minwoo, Marc Maguire & Yan Sun (2019), 'Stochastic modeling of bridge deterioration using classification tree and logistic regression', *Journal of Infrastructure Systems* **25**(1), 04018041.
- Charitidou, Efstratia, Dimitris Fouskakis & Ioannis Ntzoufras (2015), 'Bayesian transformation family selection: Moving toward a transformed gaussian universe', *Canadian Journal of Statistics* **43**(4), 600–623.
- Chen, Riqing, Jun Wang, Ruiquan Lin & Xiangning Zhao (2016), 'Spectrum sensing based on nonparametric autocorrelation in wireless communication systems under alpha stable noise', *Mobile Information Systems* **2016**.
- Chuah, TC, BS Sharif & OR Hinton (2000), 'Nonlinear decorrelator for multiuser detection in non-gaussian impulsive environments', *Electronics Letters* **36**(10), 920–922.
- Coppin, Ben (2004), *Artificial intelligence illuminated*, Jones & Bartlett Learning.
- de Freitas, Mauro L, Malcolm Egan, Laurent Clavier, Alban Goupil, Gareth W Peters & Nourddine Azzaoui (2017), 'Capacity bounds for additive symmetric alpha-stable noise channels', *IEEE Transactions on Information Theory* **63**(8), 5115–5123.
- De Vito, Luca, Sergio Rapuano & Maurizio Villanacci (2010), 'Prototype of an automatic digital modulation classifier embedded in a real-time spectrum analyzer', *IEEE Transactions on Instrumentation and Measurement* **59**(10), 2639–2651.
- Dobre, Octavia A, Ali Abdi, Yeheskel Bar-Ness & Wei Su (2005), Selection combining for modulation recognition in fading channels, *em* 'Military Communications Conference, 2005. MILCOM 2005. IEEE', IEEE, pp. 2499–2505.
- Fehske, A, J Gaeddert & Jeffrey H Reed (2005), A new approach to signal classification using spectral correlation and neural networks, *em* 'New Frontiers in Dynamic

- Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on', IEEE, pp. 144–150.
- Fontes, Aluisio IR, Joilson BA Rego, Allan de M Martins, Luiz FQ Silveira & José C Príncipe (2017), 'Cyclostationary correntropy: Definition and applications', *Expert Systems with Applications* **69**, 110–117.
- Fontes, Aluisio IR, Leandro A Pasa, Vicente A de Sousa Jr, Fuad M Abinader Jr, José AF Costa & Luiz FQ Silveira (2012), Automatic modulation classification using information theoretic similarity measures, *em* 'Vehicular Technology Conference (VTC Fall), 2012 IEEE', IEEE, pp. 1–5.
- Gardner, William A (1986), 'Introduction to random processes with applications to signals and systems((book))', *New York, MacMillan Co., 1986, 447*.
- Gardner, William A (1994), Cyclostationarity in communications and signal processing, Relatório técnico, STATISTICAL SIGNAL PROCESSING INC YOUNTVILLE CA.
- Gardner, William A, Antonio Napolitano & Luigi Paura (2006), 'Cyclostationarity: Half a century of research', *Signal processing* **86**(4), 639–697.
- Giesbrecht, James E, Russell Clarke & Derek Abbott (2003), Improved techniques for monitoring the hf spectrum, *em* 'Proceedings of SPIE', Vol. 5274, pp. 112–122.
- Gonzalez, Juan G, Jose L Paredes & Gonzalo R Arce (2006), 'Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals', *IEEE Transactions on Signal Processing* **54**(10), 3839–3851.
- Hosmer Jr, David W, Stanley Lemeshow & Rodney X Sturdivant (2013), *Applied logistic regression*, Vol. 398, John Wiley & Sons.
- Hu, Yaohua, Mingqian Liu, Chaofeng Cao & Bingbing Li (2016), Modulation classification in alpha stable noise, *em* '2016 IEEE 13th International Conference on Signal Processing (ICSP)', IEEE, pp. 1275–1278.
- Ilizzo, Luciano & Luigi Paura (1981), 'Error probability for fading cpsk signals in gaussian and impulsive atmospheric noise environments', *IEEE Transactions on Aerospace and Electronic Systems* (5), 719–722.
- Khan, Muhammad Umar, Sumair Aziz, Muhammad Bilal & Muhammad Bilal Aamir (2019), Classification of emg signals for assessment of neuromuscular disorder using empirical mode decomposition and logistic regression, *em* '2019 International Conference on Applied and Engineering Mathematics (ICAEM)', IEEE, pp. 237–243.

- Kim, Kyouwoong, Ihsan A Akbar, Kyung K Bae, Jung-Sun Um, Chad M Spooner & Jeffrey H Reed (2007), Cyclostationary approaches to signal detection and classification in cognitive radio, *em* 'New frontiers in dynamic spectrum access networks, 2007. DySPAN 2007. 2nd IEEE international symposium on', IEEE, pp. 212–215.
- Kozick, Richard J & Brian M Sadler (2000a), 'Maximum-likelihood array processing in non-gaussian noise with gaussian mixtures', *IEEE Transactions on Signal Processing* **48**(12), 3520–3535.
- Kozick, Richard J & Brian M Sadler (2000b), Robust subspace estimation in non-gaussian noise, *em* 'Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on', Vol. 6, IEEE, pp. 3818–3821.
- Liu, Tao, Tianshuang Qiu & Shengyang Luan (2018), 'Cyclic Correntropy: Foundations and Theories', *IEEE Access* **6**(c), 34659–34669.
- Liu, Tsung-Hsien & Jerry M Mendel (2001), 'A subspace-based direction finding algorithm using fractional lower order statistics', *IEEE Transactions on Signal Processing* **49**(8), 1605–1613.
- Liu, Yang, Tianshuang Qiu & Hu Sheng (2012), 'Time-difference-of-arrival estimation algorithms for cyclostationary signals in impulsive noise', *Signal Processing* **92**(9), 2238–2247.
- Liu, Yang, Tianshuang Qiu & Jingchun Li (2015), 'Joint estimation of time difference of arrival and frequency difference of arrival for cyclostationary signals under impulsive noise', *Digital Signal Processing* **46**, 68–80.
- Lorincz, Josip & Dinko Begusic (2006), Physical layer analysis of emerging ieee 802.11 n wlan standard, *em* 'Advanced Communication Technology, 2006. ICACT 2006. The 8th International Conference', Vol. 1, IEEE, pp. 6–pp.
- Luan, Shengyang, Tianshuang Qiu, Yongjie Zhu & Ling Yu (2016), 'Cyclic correntropy and its spectrum in frequency estimation in the presence of impulsive noise', *Signal Processing* **120**, 503–508.
- Lundén, Jarmo, Saleem A. Kassam & Visa Koivunen (2010), 'Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio', *IEEE Transactions on Signal Processing* **58**(1), 38–52.
- Lunden, Jarmo & Visa Koivunen (2014), 'Spatial sign and rank cyclic detectors', *IEEE Signal Processing Letters* **21**(5), 595–599.
- Ma, Xinyu & Chrysostomos L Nikias (1996), 'Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics', *IEEE Transactions on Signal Processing* **44**(11), 2669–2687.

- Madi, Ghadir, Baptiste Vrigneau, Yannis Pousset, Rodolphe Vauzelle & Basile L Agba (2010), Impulsive noise of partial discharge and its impact on minimum distance-based precoder of mimo system, *em* 'Signal Processing Conference, 2010 18th European', IEEE, pp. 1602–1606.
- Mahmood, Ahmed, Mandar Chitre & Marc Andre Armand (2012), 'Psk communication with passband additive symmetric α -stable noise', *IEEE Transactions on Communications* **60**(10), 2990–3000.
- Maliatsos, Konstantinos, Stavroula Vassaki & Philip Constantinou (2007), Interclass and intraclass modulation recognition using the wavelet transform, *em* 'Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on', IEEE, pp. 1–5.
- Martins, Allan de Medeiros (2005), 'Contribuições aos processos de clustering com base em métricas não-euclidianas'.
- Middleton, David (1999), 'Non-gaussian noise models in signal processing for telecommunications: new methods an results for class a and class b noise models', *IEEE Transactions on Information Theory* **45**(4), 1129–1149.
- Mitchell, Tom M (1997), 'Machine learning. 1997', *Burr Ridge, IL: McGraw Hill* **45**(37), 870–877.
- Mukaka, Mavuto M (2012), 'A guide to appropriate use of correlation coefficient in medical research', *Malawi Medical Journal* **24**(3), 69–71.
- Nandi, Asoke K & Elsayed Elsayed Azzouz (1998), 'Algorithms for automatic modulation recognition of communication signals', *IEEE Transactions on communications* **46**(4), 431–436.
- Napolitano, Antonio (2016), 'Cyclostationarity: New trends and applications', *Signal Processing* **120**, 385–408.
- Ndo, Gaëtan, Fabrice Labeau & Marthe Kassouf (2013), 'A markov-middleton model for bursty impulsive noise: Modeling and receiver design', *IEEE Transactions on Power Delivery* **28**(4), 2317–2325.
- Nikias, Chrysostomos L & Min Shao (1995), *Signal processing with alpha-stable distributions and applications*, Wiley-Interscience.
- Nolan, John P (1997), 'Numerical calculation of stable densities and distribution functions', *Communications in statistics. Stochastic models* **13**(4), 759–774.
- Novey, Mike, Tülay Adali & Anindya Roy (2010), 'A complex generalized gaussian distribution—characterization, generation, and estimation', *IEEE Transactions on Signal Processing* **58**(3), 1427–1433.

- Orlic, VD & ML Dukic (2010), ‘Multipath channel estimation algorithm for automatic modulation classification using sixth-order cumulants’, *Electronics letters* **46**(19), 1348–1349.
- Park, Cheol-Sun, Jun-Ho Choi, Sun-Phil Nah, Won Jang & Dae Young Kim (2008), ‘Automatic modulation recognition of digital signals using wavelet features and svm’, *Advanced Communication Technology* **1**, 387–390.
- Pena, Danilo, Arthur Lima, Vicente Angelo de Sousa Junior, Luiz Felipe Silveira & Allan Medeiros (2020), ‘Robust time delay estimation based on non-gaussian impulsive acoustic channel’, *Journal of Communication and Information Systems* **35**(1), 86–89.
- Pena, Danilo, Carlos Lima, Matheus Dória, Luan Pena, Allan Martins & Vicente Sousa (2019), ‘Acoustic impulsive noise based on non-gaussian models: an experimental evaluation’, *Sensors* **19**(12), 2827.
- Pokharel, Puskal P, Weifeng Liu & Jose C Principe (2009), ‘A low complexity robust detector in impulsive noise’, *Signal Processing* **89**(10), 1902–1909.
- Príncipe, Jose C (2010), *Information theoretic learning: Renyi’s entropy and kernel perspectives*, Springer:Berlin,Germany.
- Ramkumar, Barathram (2009), ‘Automatic modulation classification for cognitive radios using cyclic feature detection’, *IEEE Circuits and Systems Magazine* **9**(2).
- Renard, Julien, Lutz Lampe & Francois Horlin (2013), ‘Spatial sign cyclic-feature detection’, *IEEE Transactions on signal processing* **61**(18), 4521–4531.
- Samorodnitsky, Gennady & Murad S Taqqu (1994), *Stable non-Gaussian random processes: stochastic models with infinite variance*, Vol. 1, CRC press.
- Santamaría, Ignacio, Puskal P Pokharel & José Carlos Principe (2006), ‘Generalized correlation function: definition, properties, and application to blind equalization’, *IEEE Transactions on Signal Processing* **54**(6), 2187–2197.
- Satija, Udit, Barathram Ramkumar & M Sabarimalai Manikandan (2017), ‘A novel sparse classifier for automatic modulation classification using cyclostationary features’, *Wireless Personal Communications* **96**(3), 4895–4917.
- Satija, Udit, MS Manikandan & Barathram Ramkumar (2014), Performance study of cyclostationary based digital modulation classification schemes, *em ‘Industrial and Information Systems (ICIIS), 2014 9th International Conference on’*, IEEE, pp. 1–5.
- Serneels, Sven, Evert De Nolf & Pierre J. Van Espen (2006), ‘Spatial sign preprocessing: A simple way to impart moderate robustness to multivariate estimators’, *Journal of Chemical Information and Modeling* **46**(3), 1402–1409.

- Shao, Min & Chrysostomos L Nikias (1993), 'Signal processing with fractional lower order moments: stable processes and their applications', *Proceedings of the IEEE* **81**(7), 986–1010.
- Shi, Chengzhuo, Zheng Dou & Lin Qi (2019), 'Distortion-constraint-based group sparse channel estimation under α -stable noise', *IEEE Access* **7**, 21392–21399.
- Spaulding, A & David Middleton (1977), 'Optimum reception in an impulsive interference environment—part i: coherent detection', *IEEE Transactions on Communications* **25**(9), 910–923.
- Spooner, Chad M, Apurva N Mody, Jack Chuang & Josh Petersen (2017), Modulation recognition using second-and higher-order cyclostationarity, *em* 'Dynamic Spectrum Access Networks (DySPAN), 2017 IEEE International Symposium on', IEEE, pp. 1–3.
- Stevenson, Carl R, Gerald Chouinard, Zhongding Lei, Wendong Hu, Stephen J Shellhammer & Winston Caldwell (2009), 'Ieee 802.22: The first cognitive radio wireless regional area network standard', *IEEE communications magazine* **47**(1), 130–138.
- Tani, Andrea, Romano Fantacci & Dania Marabissi (2016), 'A low-complexity cyclostationary spectrum sensing for interference avoidance in femtocell lte-a-based networks', *IEEE Transactions on Vehicular Technology* **65**(4), 2747–2753.
- Tsihrintzis, GA, U Tureli & CL Nikias (1996), 'Fractional lower-order statistics-based ambiguity functions for differential delay doppler estimation', *IEE Proceedings-Radar, Sonar and Navigation* **143**(6), 358–365.
- Turley, Michael (2003), Impulsive noise rejection in hf radar using a linear prediction technique, *em* 'Radar Conference, 2003. Proceedings of the International', IEEE, pp. 358–362.
- Wang, Peng, Tian-shuang Qiu, Fu-quan Ren & Ai-min Song (2017), 'A robust doa estimator based on the correntropy in alpha-stable noise environments', *Digital Signal Processing* **60**, 242–251.
- Weron, Aleksander & Rafal Weron (1995), Computer simulation of lévy α -stable variables and processes, *em* 'Chaos—The interplay between stochastic and deterministic behaviour', Springer, pp. 379–392.
- Win, Moe Z, Pedro C Pinto & Lawrence A Shepp (2009), 'A mathematical theory of network interference and its applications', *Proceedings of the IEEE* **97**(2), 205–230.
- Wu, Hsiao-Chun, Mohammad Saquib & Zhifeng Yun (2008), 'Novel automatic modulation classification using cumulant features for communications via multipath channels', *IEEE Transactions on Wireless Communications* **7**(8).

- Wu, Zhilu, Siyang Zhou, Zhendong Yin, Bo Ma & Zhutian Yang (2017), ‘Robust automatic modulation classification under varying noise conditions’, *IEEE Access* **5**, 19733–19741.
- Xin, ZHOU, WU Ying & YANG Bin (2010), ‘Signal classification method based on support vector machine and high-order cumulants’, *Wireless Sensor Network* **2**(01), 48.
- Xu, Jian-Wu, Hovagim Bakardjian, Andrzej Cichocki & Jose C Principe (2008), ‘A new nonlinear similarity measure for multichannel signals’, *Neural Networks* **21**(2), 222–231.
- Yang, Guosheng, Jun Wang, Guoyong Zhang, Qijia Shao & Shaoqian Li (2018), ‘Joint estimation of timing and carrier phase offsets for msk signals in alpha-stable noise’, *IEEE Communications Letters* **22**(1), 89–92.
- You, Guo-hong, Tian-shuang Qiu & Ai-min Song (2013), ‘Novel direction findings for cyclostationary signals in impulsive noise environments’, *Circuits, Systems, and Signal Processing* **32**(6), 2939–2956.
- Zhang, Qiyun, Octavia A Dobre, Yahia A Eldemerdash, S Rajan & R Inkol (2013), ‘Second-order cyclostationarity of bt-scd signals: Theoretical developments and applications to signal classification and blind parameter estimation’, *IEEE Transactions on Wireless Communications* **12**(4), 1501–1511.
- Zhu, Zhechen & Asoke K Nandi (2015), *Automatic modulation classification: principles, algorithms and applications*, John Wiley & Sons.

Apêndice A

Função Característica de uma Variável Aleatória Gaussiana

Uma variável aleatória X cuja a distribuição é Gaussiana, possui a função densidade de probabilidade dada por:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}. \quad (\text{A.1})$$

E uma vez que $f_X(x)$ é uma função contínua, sua função característica é determinada por:

$$\phi_X(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx. \quad (\text{A.2})$$

Então:

$$\phi_X(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{j\omega x} e^{-(x-\mu)^2/2\sigma^2} dx. \quad (\text{A.3})$$

Estabelecendo a relação $y = x - \mu$, tem-se:

$$\phi_Y(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{j\omega(y+\mu)} e^{-y^2/2\sigma^2} dy \quad (\text{A.4})$$

$$\phi_Y(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{j\omega y} e^{-y^2/2\sigma^2} dy \right) e^{j\omega\mu} \quad (\text{A.5})$$

$$\phi_Y(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(y^2 - 2j\omega y\sigma^2)}{2\sigma^2}} dy \right) e^{j\omega\mu} \quad (\text{A.6})$$

$$\phi_Y(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-y(y - 2j\omega\sigma^2)}{2\sigma^2}} dy \right) e^{j\omega\mu} \quad (\text{A.7})$$

Estabelecendo uma nova relação, dada por $m = y - j\omega\sigma^2$, tem-se:

$$\phi_M(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(m+j\omega\sigma^2)(m-j\omega\sigma^2)}{2\sigma^2}} dm \right) e^{j\omega\mu} \quad (\text{A.8})$$

$$\phi_M(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(m^2+\omega^2\sigma^4)}{2\sigma^2}} dm \right) e^{j\omega\mu} \quad (\text{A.9})$$

$$\phi_M(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(m^2)}{2\sigma^2}} dm \right) e^{\frac{-\omega^2\sigma^2}{2}} e^{j\omega\mu} \quad (\text{A.10})$$

$$\phi_M(\omega) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\frac{-(m^2)}{2\sigma^2}} dm \right) e^{\frac{2j\omega\mu - \omega^2\sigma^2}{2}}. \quad (\text{A.11})$$

Uma vez que o termo entre parênteses da Equação A.11 corresponde a integral de uma distribuição Gaussiana (em todo seu suporte), ele possui valor unitário. Portanto, a função característica de uma variável aleatória x de distribuição Gaussiana é dada por:

$$\phi_X(\omega) = e^{\frac{2j\omega\mu - \omega^2\sigma^2}{2}}. \quad (\text{A.12})$$

Considerando a definição da função característica, de uma variável aleatória x de distribuição alfa-estável ($S_\alpha(\beta, \delta, \gamma)$), definida pelas Equações 2.7 e 2.8, representadas respectivamente por:

$$\phi_X(\omega, \alpha, \beta, \gamma, \delta) = \exp \{ \gamma^\alpha [-|\omega|^\alpha + j\omega\Lambda(\omega, \alpha, \beta)] + j\delta\omega \}, \quad (\text{A.13})$$

$$\Lambda(\omega, \alpha, \beta) = \begin{cases} \beta|\omega|^{\alpha-1} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\beta\frac{2}{\pi} \ln|\omega| & \alpha = 1. \end{cases} \quad (\text{A.14})$$

E assumindo os seguintes parâmetros: $\alpha = 2, \beta = 0, \gamma = \frac{\sigma}{\sqrt{2}}, \delta = \mu$. Então:

$$\phi_X(\omega, \sigma, \mu) = \exp \left\{ \left(\frac{\sigma}{\sqrt{2}} \right)^2 (-|\omega|^2 + 0) + j\mu\omega \right\} \quad (\text{A.15})$$

$$\phi_X(\omega, \sigma, \mu) = \exp \left\{ \frac{-\sigma^2\omega^2}{2} + j\mu\omega \right\} \quad (\text{A.16})$$

$$\phi_X(\omega, \sigma, \mu) = \exp \left\{ \frac{2j\omega\mu - \omega^2\sigma^2}{2} \right\}. \quad (\text{A.17})$$

Uma vez que as Equações A.17 e A.12 são equivalentes, pode-se afirmar que: quando parametrizado com $\alpha = 2, \beta = 0, \gamma = \frac{\sigma}{\sqrt{2}}, \delta = \mu$, uma distribuição $S_\alpha(\beta, \delta, \gamma)$ converge para uma distribuição Gaussiana.

Apêndice B

Geração Numérica do Ruído Aditivo Alfa-Estável

Todo o canal de comunicação avaliado neste trabalho foi modelado utilizando um *software* interativo de alto desempenho, voltado para o cálculo numérico, denominado MATLAB (*Matrix Laboratory*). Esse sistema permite a resolução de problemas numéricos diversos, e por possuir uma linguagem de alto nível, suas aplicações são expressas e resolvidas quase exatamente como elas são escritas matematicamente. Essa característica garante à ferramenta a vantagem de utilizar apenas uma fração do tempo que seria gasto para escrever um programa semelhante em linguagem Fortran, Basic ou C.

Nesse contexto, para gerar um sinal contaminado por ruído aditivo alfa-estável de um canal de comunicação impulsivo, foi desenvolvida uma função cujo parâmetros de entrada consistem em: o sinal de comunicação modulado (S_{mod}); os parâmetros α , β , δ da função característica definida na Equação 2.7; e a relação sinal ruído geométrica (GSNR) em dB, definida na Equação 2.15. A função, denominada **aasn**¹, retorna o sinal y contaminado.

$$y = aasn(S_{mod}, GSNR_{dB}, \alpha, \beta, \delta).$$

Para gerar um sinal aleatório de distribuição alfa-estável, a função **aasn** faz uso de uma função local do MATLAB, chamada de *makedist*. Essa função compõe a *Toolbox* de Estatística e Aprendizagem de Máquina do próprio sistema, e tem como finalidade criar um objeto de distribuição de probabilidade (*pd1*), o qual pode ser configurado para atender a diversos tipos de distribuições. No caso particular da distribuição alfa-estável, a seguinte configuração é necessária:

$$pd1 = makedist('Stable', 'alpha', \alpha, 'beta', \beta, 'gam', \gamma, 'delta', \delta);$$

Em que, os termos entre aspas simples são *strings* de ajuste e identificação da distribuição, e o parâmetro γ é calculado a partir da $GSNR_{dB}$ e o S_{mod} , com base na relação estabelecida na Equação 2.15:

¹ Acrônimo do termo *additive alpha-stable noise*.

$$\gamma = \sqrt{\frac{A^2}{2 \times 10^{(GSNR_{dB}/10)} \times Cg^{(\frac{2}{\alpha}-1)}},$$

em que, A é o valor RMS de S_{mod} :

$$A = rms(S_{mod}).$$

Uma vez definido o objeto $pd1$, o sinal ruidoso ($impulsive_{noise}$) de distribuição alfa-estável será gerado a partir da função *random* do MATLAB, a qual será parametrizada da seguinte forma:

$$impulsive_{noise} = random(pd1, size(S_{mod})).$$

Nesse caso, a função *random* é responsável por gerar um sinal pseudoaleatório que segue a distribuição estabelecida por $pd1$, e possui o mesmo tamanho do sinal modulado S_{mod} . Por fim, o sinal contaminado é definido como:

$$y = S_{mod} + impulsive_{noise}.$$

Em resumo, a função **aasn** pode ser estruturada da seguinte forma:

```
1 function y = aasn(Smod, GSNRdB, alpha, beta, delta)
2 A = rms(Smod);
3 gamma = sqrt( A^2 / ( 2 * 10^(GSNRdB/10) * Cg^(2/alpha - 1) ) );
4 pd1 = makedist('Stable', 'alpha', alpha, 'beta', beta, 'gam', gamma, 'delta', delta);
5 impulsiveNoise = random(pd1, size(Smod));
6 y = Smod + impulsiveNoise;
```

Por não ser um *software* livre, o MATLAB não disponibiliza o código fonte, e nem identifica, os métodos utilizados na geração da distribuição alfa-estável. Contudo, a documentação fornecida pelo *software* contém as seguintes referências de métodos numéricos de estimação de distribuições alfa-estáveis: (Nolan 1997, Weron & Weron 1995).

Apêndice C

Teoria da Informação e Correntropia

A correntropia foi criada com base em análises estatísticas e na teoria da informação¹, por isso, este apêndice aborda os conceitos de entropia de Shannon, entropia de Rényi e estimador de Parzen com o objetivo de formar uma linha de desenvolvimento teórico da ferramenta.

C.1 Entropia de Shannon

O conceito de entropia foi formulado inicialmente por Rudolf Clausius por volta de 1850, e em um primeiro momento foi apontada como uma propriedade de sistemas termodinâmicos. Mais tarde, em 1877 Ludwig Boltzmann deu uma interpretação probabilística para a entropia, a qual também foi aplicada à termodinâmica.

Durante seus trabalhos na Bell Telephone, Claude Shannon desenvolveu um novo conceito de entropia, o qual seria usado em uma área até então nunca explorada, a Teoria da Informação (Information Theory - IT) (Artuso 2012). Shannon desenvolveu uma ferramenta matemática com a capacidade de medir a aleatoriedade (nível de incerteza) na transmissão de dados dos sistemas de comunicação de sua época, e inicialmente a nomeou de “missing information” (informação faltante). Entretanto após a sugestão de seu amigo John Von Neumann, mudou o nome da ferramenta para entropia, em virtude de algumas semelhanças com a expressão desenvolvida por Boltzmann (Artuso 2012). Foi em 1948 que então Shannon publicou um artigo apresentando sua ferramenta ao mundo.

A entropia definida por Shannon pode ser simplesmente interpretada como uma medida do grau de informação de uma variável aleatória (Martins 2005). Sua expressão matemática para uma variável aleatória contínua é dada por:

$$H(x) = - \int_d^e f_x(x) \log_b(f_x(x)) dx. \quad (C.1)$$

Em que $f_x(x)$ é a função densidade de probabilidade (PDF) da variável aleatória x , a qual possui um suporte variando de d até e , que define os limites de integração da expressão. Quando aplicada a uma variável aleatória discreta, a entropia será estimada da seguinte forma:

¹O próprio nome correntropia faz referência a correlação e entropia.

$$H(x) = - \sum_{i=1}^n p_i \log_b(p_i). \quad (C.2)$$

Um exemplo tradicionalmente usado para ilustrar a entropia é o experimento que consiste no lançamento de uma moeda. Considerando que tal experimento descreve uma variável aleatória x , que adota p como a probabilidade da moeda assumir “cara” e $(1 - p)$ com a probabilidade da moeda assumir “coroa”, a entropia associada a ela será:

$$H(x) = -(\log_b(p)p + \log_b(1 - p)(1 - p)). \quad (C.3)$$

Como nesse caso x assume valores discretos, sua entropia é calculada pelo somatório das probabilidades ponderadas pelos seus logaritmos. Considerando que p pode assumir valores entre 0 e 1, e adotando a base 2 para os logaritmos da expressão, a entropia associada ao lançamento de uma moeda pode ser avaliado em função de p , como mostra a Figura C.1.

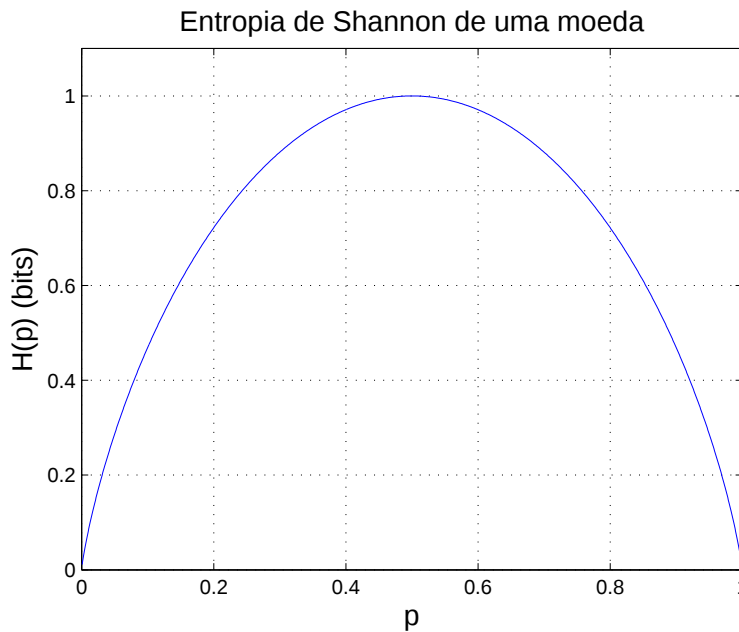


Figura C.1: Entropia de Shannon do lançamento de uma moeda

Um detalhe importante do experimento é que a entropia é medida em bits, isso ocorre devido a base do logaritmo ser 2. Pode-se perceber também que, a situação que implica na maior incerteza sobre a variável aleatória é quando $p = 0.5$, nesse caso a moeda possui chances iguais de assumir qualquer face, e quando isso ocorre a entropia assume valor máximo. A medida que p vai se afastando de 0.5 a entropia diminui, a ponto de ficar nula quando $p = 0$ (no caso em que existe a certeza da moeda assumir coroa) ou quando $p = 1$ (que implica na certeza da moeda assumir cara).

Pode-se dizer que a entropia retorna o nível de informação de uma variável aleatória a partir das probabilidades de seus eventos, que se associam diretamente com a incerteza de cada um ocorrer, por isso outra interpretação dada a ela é: medida do nível de incerteza de uma variável aleatória.

A entropia de Shannon possui diversas aplicações na engenharia tais como codificação de canal, compressão de dados, reconhecimento de padrões, entre outras. Contudo, em uma abordagem prática, o cálculo dessa função pode ser custoso computacionalmente, devido a forma de sua expressão, além de ser necessário o conhecimento prévio do comportamento da PDF dos dados. Devido a esses desafios, outras formas de análise da entropia foram desenvolvidas durante os anos, o que contribuiu para aplicações mais eficientes da ferramenta.

C.2 Entropia de Rényi

Na década de 1950 Alfred Rényi introduziu a família paramétrica de entropia a partir de uma generalização matemática da entropia de Shannon (Príncipe 2010). Rényi desenvolveu uma classe abrangente de medidas de informação que conservava a aditividade de sistemas estatísticos independentes, e respeitava os axiomas de probabilidade de Kolmogorov (Príncipe 2010). A entropia de Rényi aplicado a uma variável aleatória contínua é expressa por

$$H_{\beta}(x) = \frac{1}{1-\beta} \log_b \int_d^e f_x(x)^{\beta} dx. \quad (C.4)$$

Em que β é o parâmetro que determina a ordem da expressão. A entropia de Rényi pode apresentar características variadas para diferentes valores de β . Quando este é igual a 1 por exemplo, pode-se demonstrar que a expressão da entropia de Rényi se torna igual a entropia de Shannon.

Para $\beta = 2$ tem-se a entropia de Rényi de ordem 2, a qual é comumente conhecida como entropia quadrática de Rényi. Para uma variável aleatória contínua, a entropia quadrática de Rényi é dada por

$$H_2(x) = -\log_b \int_d^e f_x(x)^2 dx \quad (C.5)$$

A entropia quadrática de Rényi é particularmente interessante pois pode ser facilmente estimada a partir de dados amostrais, além do que, pode ser calculado de uma forma mais simples que a entropia de Shannon, pois o resultado da integral em sua expressão depende apenas do quadrado da distribuição de probabilidade, enquanto que em Shannon depende da PDF ponderada por seu logaritmo (Martins 2005).

Quando aplicado sobre uma variável aleatória discreta a entropia de Rényi é expressa por

$$H_{\beta}(x) = \frac{1}{1-\beta} \log_b \left(\sum_{i=1}^n p_i^{\beta} \right). \quad (C.6)$$

Reproduzindo o exemplo anterior, lançamento de uma moeda, utilizando a entropia Rényi, pode-se avaliar o comportamento da entropia para diferentes valores de β .

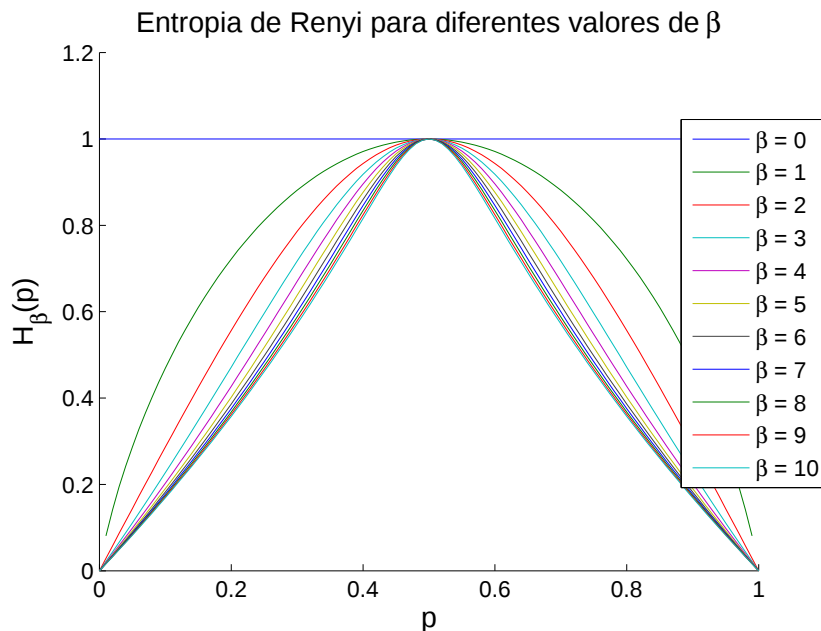


Figura C.2: Entropia de Rényi do lançamento de uma moeda para diferentes valores de β

Avaliando a Figura C.2 pode-se chegar a conclusão de que, com exceção de $\beta = 0$, o comportamento da entropia para diferentes valores de β é bastante consistente, uma vez que ela é maximizada quando $p = 0.5$ e decai conforme p se distancia desse valor. Em outras palavras, a entropia de Rényi pode fazer uma análise tão apurada quanto a entropia de Shannon, com a vantagem de possuir uma forma matemática mais simplificada.

Desse modo, o campo de aplicação da Entropia de Rényi pode ser tão amplo quanto o da entropia de Shannon. Entretanto, abordagens práticas da ferramenta ainda dependem do conhecimento prévio da PDF da variável aleatória medida, o que nem sempre está disponível. Nessas condições, é necessário adotar alguma estratégia para estimar a função densidade de probabilidade dos dados.

C.3 Estimador de Parzen

Estimadores paramétricos são métodos que extraem de um conjunto de dados, uma série de parâmetros que modelam sua função densidade de probabilidade. Métodos não-paramétricos fazem uma estimação da PDF de um conjunto de dados usando amostra dos próprios dados.

Por não precisarem de algoritmos complexos para estimar uma PDF, métodos não-paramétricos apresentam vantagens sobre os modelos paramétricos, entretanto exigem grande processamento computacional (Martins 2005), o qual aumenta de acordo com o tamanho amostral.

O estimador de Parzen é um método não-paramétrico que assume que a densidade de probabilidade de um evento ξ de uma variável aleatória x em uma região do $\mathbb{R}^n$ (definida por sua amostra), é dada por

$$f_x(\xi) = \frac{k(\xi)}{NV}, \quad (\text{C.7})$$

em que N é a quantidade de elementos da amostra, V é o volume (tamanho) da região na qual cada elemento da amostra será avaliada e $k(\cdot)$ é o kernel que avalia os eventos ξ na amostra em diferentes regiões definidas no $\mathbb{R}^n$.

O termo kernel refere-se a uma função unimodal, simétrica (em relação a sua média) e positiva, que denota que, a contribuição da densidade de probabilidade de uma amostra é maximizada no local em que essa amostra se encontra, e decai a medida que se afasta dessa posição (Príncipe 2010). V é um termo muito importante, pois altera diretamente o comportamento da função estimada. Quanto menor seu valor, mais restritivo será a estimação em torno de ξ , e a medida que seu valor cresce a estimação será mais abrangente.

Aplicações que utilizam a função gaussiana como kernel são frequentes, pois ela possui propriedades que facilitam sua manipulação matemática. Nesses casos o kernel será representado por $G_\sigma(\cdot)$, e a estimação a densidade de probabilidade será denotada por:

$$f_x(x) = \frac{1}{N} \sum_{i=1}^N G_\sigma(x - x_i). \quad (\text{C.8})$$

Em que, x_i são os elementos da amostra e $G_\sigma(\cdot)$ é expresso por

$$G_\sigma(\cdot) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2\sigma^2}}. \quad (\text{C.9})$$

O termo σ equivale ao tamanho do kernel. Quanto maior o σ mais suave será a estimativa, caso contrário a estimativa terá um comportamento mais pontual (Martins 2005), como pode ser visto na Figura C.3.

C.4 Potencial de Informação

A importância do estimador de Parzen pode ficar mais clara no caso em que ele é utilizado para definir a função densidade de probabilidade da entropia quadrática de Rényi. Considerando o estimador com kernel gaussiano, a expressão da entropia pode ser analisada substituindo a Equação (C.8) em (C.5):

$$H_2(x) = -\log_b \int_{-\infty}^{\infty} \left(\frac{1}{N} \sum_{i=1}^N G_\sigma(x - x_i) \right)^2 dx \quad (\text{C.10})$$

$$H_2(x) = -\log_b \left(\frac{1}{N^2} \int_{-\infty}^{\infty} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(x - x_i) G_\sigma(x - x_j) dx \right) \quad (\text{C.11})$$

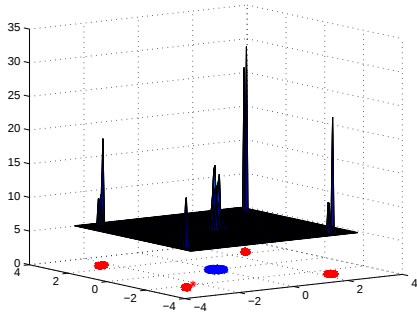
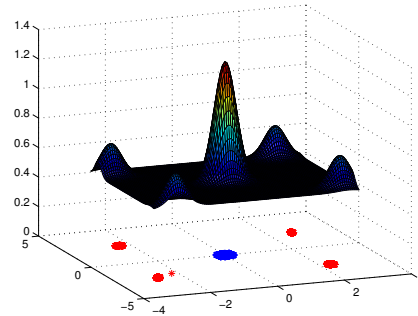
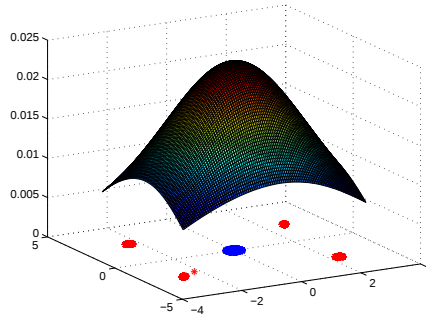
(a) PDF estimada com $\sigma=0.01$ (b) PDF estimada com $\sigma=0.3$ (c) PDF estimada com $\sigma=2$

Figura C.3: Aplicação do estimador de Parzen sobre um conjunto de dados utilizando kernels gaussianos de tamanhos diferentes.

$$H_2(x) = -\log_b \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int_{-\infty}^{\infty} G_{\sigma}(x-x_i) G_{\sigma}(x-x_j) dx \right) \quad (\text{C.12})$$

aplicando a propriedade da convolução de gaussianas ², obtém-se:

$$H_2(x) = -\log_b \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma'}(x_i - x_j) \right) \quad (\text{C.13})$$

A Equação (C.13) é muito importante pois denota uma forma analítica do cálculo da entropia de uma variável aleatória, dependente somente das amostras da própria variável e do tamanho do parâmetro σ' ³. O termo dentro do logaritmo, da expressão (C.13), é chamado de Potencial de Informação (IP):

²A convolução entre duas gaussianas resulta em outra função gaussiana.

³O termo σ' representa o tamanho do kernel resultante da convolução das Gaussianas, o qual é equivalente a $\sigma' = \sqrt{2}\sigma$.

$$IP = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(x_i - x_j). \quad (C.14)$$

O potencial de informação é particularmente interessante, pois trata-se de uma função monotonicamente crescente que, quando maximizada minimiza a entropia. Pode-se perceber que o IP retorna o nível de informação de uma variável aleatória de forma ainda mais simples que a entropia, uma vez que sua expressão decorre apenas de kernels gaussianos, e não depende do logaritmo. O potencial de informação pode ainda ser usado para medir a relação entre variáveis aleatórias distintas. Nesse caso, a medida será chamada de potencial de informação cruzado (CIP) e será denotado por:

$$CIP(x, y) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(x_i - y_j). \quad (C.15)$$

Os conceitos apresentados até aqui (entropia, estimador de Parzen, Potencial de Informação) foram desenvolvidos utilizando o kernel gaussiano como principal função de processamento, o qual também constitui o principal alicerce da correntropia.

C.5 Correntropia

A correntropia é uma transformação não-linear que relaciona duas variáveis aleatórias x e y (Santamaría et al. 2006, Príncipe 2010), e é definida por:

$$V(x, y) = E\{k(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(x, y) f_{x,y}(x, y) dx dy, \quad (C.16)$$

em que $k(\cdot, \cdot)$ é qualquer kernel contínuo definido positivo. Inclusive, no caso em que $k(x, y) = xy$ a correntropia representa a correlação cruzada entre duas variáveis aleatórias (Príncipe 2010).

O modelo de kernel mais difundido na análise da correntropia é o Gaussiano, o qual é simétrico e invariante a translação. Quando parametrizada por esse kernel, a correntropia é definida como:

$$V(x, y) = E\{G_{\sigma}(x - y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{\sigma}(x - y) f_{x,y}(x, y) dx dy. \quad (C.17)$$

Como, em problemas práticos, a PDF conjunta é desconhecida, e apenas um número finito de dados $\{(x_i, y_i)\}_{i=1}^N$ está disponível, a correntropia é calculada pelo seguinte estimador:

$$V(x, y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i - y_i). \quad (C.18)$$

O cálculo da correntropia engloba a análise de vários momentos estatísticos da diferença entre as variáveis analisadas. Isso pode ser observado expandindo a Equação C.18

em série de Maclaurin:

$$V(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E[(x - y)^{2n}]. \quad (\text{C.19})$$

A Equação (C.19) mostra que a correntropia se relaciona a infinitos momentos de ordem par, os quais são inversamente ponderados pelo tamanho do kernel. O aumento do tamanho do kernel faz com que a contribuição dos momentos de alta ordem decaiam mais rápido, devido ao denominador da divisão que pondera as operações dentro do somatório, então o momento de segunda ordem tende a dominar a correntropia, fazendo com que ela se aproxime da correlação (Príncipe 2010). Conforme o tamanho do kernel decai, a correntropia apresenta uma maior contribuição dos momentos de alta ordem.

Assim, o tamanho do kernel funciona como uma janela de observação sobre a variável aleatória, permitindo obter informação sobre os momentos estatísticos avaliados em seu espaço. Quando aplicada sobre uma única variável aleatória, a correntropia recebe o nome de autocorrentropia, e quando aplicada sobre variáveis aleatórias diferentes ela é denominada correntropia cruzada (Xu et al. 2008).

A correntropia cruzada não é centrada em zero, por isso, define-se a correntropia cruzada centralizada ⁴ como (Príncipe 2010):

$$U(x, y) = V(x, y) - CIP(x, y), \quad (\text{C.20})$$

$$U(x, y) = \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x_i - y_i) - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(x_i - y_j). \quad (\text{C.21})$$

E a partir da correntropia cruzada centralizada é possível estabelecer uma medida a similaridade entre variáveis aleatórias com base na seguinte relação:

$$\eta = \frac{U(x, y)}{\sqrt{U(x, x)U(y, y)}}, \quad (\text{C.22})$$

em que, η é conhecido como coeficiente de correntropia, e assim como o coeficiente de correlação, assume valores limitados por $0 < |\eta| \leq 1$ (Príncipe 2010).

C.6 Correntropia para Processos Aleatórios

A teoria da correntropia aplicada à variáveis aleatórias, quando estendida para processos estocásticos, permite uma análise não-linear de séries temporais (Príncipe 2010). Em outras palavras, correntropia também pode ser utilizada para investigar a relação de variáveis aleatórias cujo o comportamento mudam em função do tempo. Nesse caso, a ferramenta será chamada de função de correntropia.

⁴A correntropia cruzada centralizada é a generalização da covariância centralizada.

Particularmente, quando usada para avaliar um único processo aleatório em instantes de tempo diferentes, a ferramenta será denominada função de autocorrentropia, a qual é denotada por:

$$V_x(t, \tau) = E\{G_\sigma(x(t), x(t + \tau))\}, \quad (\text{C.23})$$

em que:

$$G_\sigma(x(t), x(t + \tau)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x(t) - x(t + \tau)]^2}{2\sigma^2}\right). \quad (\text{C.24})$$

Em suma, a correntropia pode ser vista como uma medida de informação, ou técnica de processamento, que pode ser aplicada a variáveis e processos aleatórios, capaz de avaliar infinitos momentos estatísticos de ordem par (Equação C.19), e portanto, pode ser considerada uma generalização da correlação. Além disso, outra grande vantagem dessa ferramenta é o kernel Gaussiano ($G_\sigma(\cdot)$), o qual sempre garante a convergência da função de correntropia, mesmo quando aplicada a dados corrompidos por outliers (impulsividade).

Apêndice D

Interpretação Estatística da Regressão Logística

Na maioria das vezes, problemas de classificação de dados binários não possuem uma regra fechada (perfeita) para definir o comportamento dos alvos y do conjunto de treinamento. Nesses casos uma modelagem estocástica será mais adequada, e descreverá o comportamento de y em termos probabilísticos. Na verdade, o que se espera do modelo estocástico é estimação da distribuição condicional de probabilidade do alvo y , dado um certo conjunto de características x : $P(y|x)$.

Tabela D.1: Banco de dados de alvo binário.

x					y
m {	2104	5	1	45	1
	1416	3	2	40	0
	1534	3	2	30	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	852	2	1	36	1
n					

A distribuição de probabilidade condicional deverá ser estimada a partir de uma função parametrizável $p(x; \Theta)$, aplicada em cada amostra do conjunto de treinamento, de modo que:

$$P(y = y^{(i)} | x = x^{(i)}) = p(x^{(i)}; \Theta), \quad (\text{D.1})$$

em que, Θ é o conjunto de parâmetros de ajuste da distribuição.

Uma vez que o conjunto de treinamento é composto por exemplos independentes, que assumem valores binários, é razoável assumir que os dados podem ser modelados como um processo de Bernoulli, em que cada exemplo é chamado de ensaio (ou experimento) de Bernoulli. Portanto, o processo estocástico do banco de dados pode ser definido pela probabilidade conjunta, ou função de verossimilhança, associada a todos os seus exemplos:

$$L(\Theta) = \prod_{i=1}^m P(y = y^{(i)} | x = x^{(i)}) = \prod_{i=1}^m p(x^{(i)}; \Theta)^{y^{(i)}} (1 - p(x^{(i)}; \Theta))^{1-y^{(i)}}. \quad (D.2)$$

A função de verossimilhança $L(\Theta)$ (ou simplesmente verossimilhança) mede a qualidade do ajuste de um modelo estocástico aplicado a um conjunto de dados para determinados valores de parâmetros Θ . Embora seja formada a partir da distribuição de probabilidade conjunta dos exemplos, ela é vista e utilizada apenas em função dos parâmetros, tratando assim as variáveis aleatórias como fixas nos valores observados.

A maximização da função verossimilhança, conhecida como princípio da máxima verossimilhança, permite achar os valores dos parâmetros Θ que maximizam a distribuição de probabilidade conjunta dos dados a partir de observações dos exemplo de x ($x = x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, \dots, x^{(m)}$). Em outras palavras, o princípio da máxima verossimilhança permite achar os valores de Θ que melhor ajustam a distribuição de probabilidade conjunta dos dados. Contudo, matematicamente é mais vantajoso maximizar o logaritmo da função de verossimilhança, $l(\Theta) = \log(L(\Theta))$, o qual possui o mesmo ponto de otimização de $L(\Theta)$:

$$l(\Theta) = \sum_{i=1}^m [y^{(i)} \log(p(x^{(i)}; \Theta)) + (1 - y^{(i)}) \log(1 - p(x^{(i)}; \Theta))]. \quad (D.3)$$

Além disso, a otimização deverá ser feita para cada valor $\Theta = \theta_0, \theta_1, \theta_2, \dots, \theta_n$, portanto o comportamento de $p(x^{(i)}; \Theta)$ é muito relevante. No caso da regressão logística, os parâmetros Θ se relaciona linearmente com a função logit($\cdot$) de $p(x^{(i)}; \Theta)$, a qual é definida por:

$$\text{logit}(p(x^{(i)}; \Theta)) = \log \left(\frac{p(x^{(i)}; \Theta)}{1 - p(x^{(i)}; \Theta)} \right), \quad (D.4)$$

e equivale a,

$$\log \left(\frac{p(x^{(i)}; \Theta)}{1 - p(x^{(i)}; \Theta)} \right) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \dots + \theta_n x_n^{(i)} = \Theta^T x^{(i)}. \quad (D.5)$$

Invertendo a função $\text{logit}(p(x^{(i)}; \Theta))$, pode-se definir $p(x^{(i)}; \Theta)$ como a função logística de $\Theta^T x^{(i)}$, de modo que:

$$p(x^{(i)}; \Theta) = \frac{1}{1 + e^{-\Theta^T x^{(i)}}}. \quad (D.6)$$

Uma vez definida a relação entre $p(x^{(i)}; \Theta)$ e Θ , para achar os parâmetros que retornam a máxima verossimilhança da distribuição de probabilidade conjunta dos dados, basta diferenciar $l(\Theta)$ em relação a cada componente θ_j e igualar a equação a zero. Desse modo:

$$\frac{\partial l(\Theta)}{\partial \theta_j} = \sum_{i=1}^m \left[y^{(i)} \frac{\partial}{\partial \theta_j} \log \left(\frac{1}{1 + e^{-\Theta^T x^{(i)}}} \right) + (1 - y^{(i)}) \frac{\partial}{\partial \theta_j} \log \left(1 - \frac{1}{1 + e^{-\Theta^T x^{(i)}}} \right) \right]. \quad (\text{D.7})$$

Como

$$\frac{\partial}{\partial \theta_j} \log \left(\frac{1}{1 + e^{-\Theta^T x^{(i)}}} \right) = - \left(\frac{e^{-\Theta^T x^{(i)}}}{1 + e^{-\Theta^T x^{(i)}}} \right) x_j^{(i)}$$

e

$$\frac{\partial}{\partial \theta_j} \log \left(1 - \frac{1}{1 + e^{-\Theta^T x^{(i)}}} \right) = \left(1 - \frac{e^{-\Theta^T x^{(i)}}}{1 + e^{-\Theta^T x^{(i)}}} \right) x_j^{(i)}$$

tem-se que,

$$\frac{\partial l(\Theta)}{\partial \theta_j} = \sum_{i=1}^m \left[-y^{(i)} \left(\frac{e^{-\Theta^T x^{(i)}}}{1 + e^{-\Theta^T x^{(i)}}} \right) x_j^{(i)} + (1 - y^{(i)}) \left(1 - \frac{e^{-\Theta^T x^{(i)}}}{1 + e^{-\Theta^T x^{(i)}}} \right) x_j^{(i)} \right] \quad (\text{D.8})$$

o que resulta em,

$$\frac{\partial l(\Theta)}{\partial \theta_j} = \sum_{i=1}^m \left(y^{(i)} - p(x^{(i)}; \Theta) \right) x_j^{(i)} = - \sum_{i=1}^m \left(p(x^{(i)}; \Theta) - y^{(i)} \right) x_j^{(i)}. \quad (\text{D.9})$$

Portanto, para encontrar os parâmetros θ_j que maximiza a função de verossimilhança basta resolver a seguinte equação:

$$0 = - \sum_{i=1}^m \left(p(x^{(i)}; \Theta) - y^{(i)} \right) x_j^{(i)}. \quad (\text{D.10})$$

Infelizmente, essa equação não possui resolução exata, contudo, vários métodos numéricos podem ser utilizados para achar uma solução aproximada.

Em suma, a regressão logística pode ser vista como o emprego do princípio de máxima verossimilhança sobre o conjunto de dados binários modelados por uma distribuição de Bernoulli, em que os parâmetros Θ se relacionam linearmente com a função $\text{logit}(\cdot)$ do estimador $p(x^{(i)}; \Theta)$.