



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Vinícius Alves de Oliveira

Análise do Critério baseado em Correntropia para Equalização Não-Supervisionada com Fontes Pré-Codificadas

Campinas

2018



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Vinícius Alves de Oliveira

Análise do Critério baseado em Correntropia para Equalização Não-Supervisionada com Fontes Pré-Codificadas

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Concentração: Engenharia de Computação.

Orientador: Prof. Dr. Levy Boccato

Co-orientador: Prof. Dr. Denis Gustavo Fantinato

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Vinícius Alves de Oliveira, e orientada pelo Prof. Dr. Levy Boccato

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

OL4a Oliveira, Vinícius Alves de, 1991-
Análise do critério baseado em correntropia para equalização não-supervisionada com fontes pré-codificadas / Vinícius Alves de Oliveira. – Campinas, SP : [s.n.], 2018.

Orientador: Levy Boccato.
Coorientador: Denis Gustavo Fantinato.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais. 2. Aprendizado de máquina. 3. Teoria da Informação. I. Boccato, Levy, 1986-. II. Fantinato, Denis Gustavo, 1985-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Analysis of the correntropy-based criterion for blind equalization with precoded sources

Palavras-chave em inglês:

Signal processing

Machine learning

Information theory

Área de concentração: Engenharia de Computação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora:

Levy Boccato [Orientador]

Murilo Bellezoni Loiola

Romis Ribeiro de Faissol Attux

Data de defesa: 30-07-2018

Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA – DISSERTAÇÃO DE MESTRADO

Candidato: Vinícius Alves de Oliveira RA 190716

Data de Defesa: 30 de julho de 2018

Título da dissertação: “Análise do Critério baseado em Correntropia para Equalização Não-Supervisionada com Fontes Pré-Codificadas”

Prof. Dr. Levy Boccato (Presidente, FEEC/UNICAMP)

Prof. Dr. Murilo Bellezoni Loiola (CECS/UFABC)

Prof. Dr. Romis Ribeiro de Faissol Attux (FEEC/UNICAMP)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no processo de vida acadêmica do aluno.

*Aos meus pais,
AluÍzio e Cezamar.*

Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus, por ter me dado forças para conseguir realizar esta grande conquista na minha vida.

Agradeço também aos meus pais, Aluízio e Cezamar, pelo amor incondicional e especialmente por todo o esforço e dedicação que eles tiveram para me oferecer uma boa formação; por todo o apoio, paciência e compreensão durante essa minha jornada. Esse importante passo em minha vida só foi possível graças a eles. Ao meu irmão, Douglas, pelo companheirismo e ajuda durante esse período. Sou muito grato a Deus por tê-los na minha vida.

À Maria Luisa, pelo grande amor e carinho demonstrados em todos os momentos; pela excepcional companhia, pela paciência, sabedoria, compreensão e grande incentivo para conseguir seguir em frente e alcançar os meus sonhos.

Ao Prof. Levy Boccato, que, além de ser um ótimo orientador, foi muito além das suas atribuições convencionais ao estar sempre presente e por me instruir em vários aspectos que vão além da vida acadêmica. Sou muito grato por todo o tempo em que dedicou a mim, sempre demonstrando grande sabedoria e competência para me guiar, mesmo antes do início do meu mestrado.

Ao meu co-orientador, Prof. Denis Fantinato, por todo o seu apoio e por todas as suas contribuições, que além de enriquecerem o meu trabalho, foram muito inspiradoras e marcantes para a minha formação.

Ao Prof. Daniel Guerreiro, por ter ampliado os meus horizontes e por ter me inspirado a ingressar no mestrado.

À todos os amigos do DSPcom, por toda a companhia e por toda a convivência que tornaram o meu trabalho muito mais agradável.

Aos amigos da graduação (Engenharia de Redes de Comunicação 2/2015 - ENE/UnB).

À CAPES, pelo apoio financeiro.

“Dreams are made possible if you try”
Terry Fox

Resumo

Dentro da área de processamento de sinais, o problema de equalização de canais tornou-se um tópico de pesquisa muito relevante devido a seu vasto horizonte de aplicações em, por exemplo, telecomunicações, processamento sísmico e medicina. De uma forma geral, as técnicas e métodos desenvolvidos para este problema baseiam-se no pressuposto de que a fonte do sinal é estatisticamente independente no tempo. No entanto, a ocorrência de fontes dotadas de dependência temporal (como sinais de áudio e vídeo) é significativa, e ainda há a necessidade de uma análise mais profunda dos métodos de equalização para este contexto. De fato, fontes com dependência temporal exigem um processamento mais profundo dos dados, de forma que a informação contida na estrutura temporal dos dados seja aproveitada para desempenhar a equalização.

Neste contexto, emerge uma abordagem promissora que envolve o uso de uma medida chamada correntropia. Esta medida, pertencente ao campo de aprendizado baseado na Teoria da Informação, é capaz de explorar tanto a informação estatística quanto a estrutura temporal dos sinais envolvidos e, por isso, vem sendo amplamente utilizada em equalização de canais com fonte dotada de estrutura temporal.

Baseando-se nisso, neste trabalho, realiza-se uma análise detalhada do critério baseado em correntropia para equalização cega, dedicando especial atenção ao efeito dos parâmetros ajustáveis para estimação. Mais detalhadamente, investiga-se a relação entre as versões estimada e analítica da correntropia, em que, neste último caso, alcança-se a derivação analítica para alguns cenários de interesse, o que é outra contribuição deste trabalho. Os resultados experimentais indicam algumas relações entre os principais parâmetros e o desempenho alcançável.

Palavras-chaves: Filtragem adaptativa; Equalização Cega; Aprendizado baseado na Teoria da Informação; Correntropia.

Abstract

Within the area of signal processing, the channel equalization problem has become a very relevant research topic due to its wide range of applications, for example, in telecommunications, in seismic imaging and in medicine. In general, the techniques and methods developed for this problem are based on the assumption that the signal provided by the source is statistically independent in time. However, the occurrence of sources with temporal dependence (such as audio and video signals) is significant and still requires a more in-depth analysis of the equalization methods for this context. In fact, temporally structured sources require deeper processing of the data, so that the information contained in the temporal structure of the data is explored to perform the channel equalization.

In this context, a promising approach emerges, which involves the use of a measure called correntropy. This measure comes from the field of the Information Theoretic Learning (ITL), and is capable of exploring both the statistical information and the temporal structure of the signals involved and, therefore, it has been widely used in channel equalization with sources with temporal structure.

Based on this, in this work, a detailed analysis of the criterion based on correntropy for blind equalization is performed, giving particular attention to the effect of the adjustable parameters for estimation. In more detail, the relationship between the estimated and analytical versions of the correntropy is investigated, in which, in the latter case, the analytical derivation is reached for some scenarios of interest, which is another contribution of this work. The experimental results indicate some relationships between the main parameters and the attainable performance.

Keywords: Adaptive filtering; Blind channel equalization; Information-theoretic learning; Correntropy.

Lista de ilustrações

Figura 1 – Diagrama de blocos apresentando o modelo de um sistema de comunicação.	22
Figura 2 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação.	26
Figura 3 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização supervisionada.	28
Figura 4 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas. . .	33
Figura 5 – PDF estimadas por meio do método da Janela de <i>Parzen</i>	37
Figura 6 – Processos aleatórios brancos com diferentes distribuições de entrada. .	42
Figura 7 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas. . .	50
Figura 8 – Distribuições - $u(n)$ e $s(n)$	52
Figura 9 – PMF de $s'(n) = s(n) - s(n-1)$	53
Figura 10 – Formato da PDF da fonte contínua para $\sigma_g^2 = 0,1$	53
Figura 11 – Superfícies das funções custo analítica e estimada para $\sigma = 0,175$	64
Figura 12 – Superfícies das funções custo analítica e estimada para $\sigma = 0,9$	64
Figura 13 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,175$	65
Figura 14 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,9$	65
Figura 15 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$	66
Figura 16 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,175$	66
Figura 17 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,9$	67
Figura 18 – ISI como uma função da largura do <i>kernel</i> σ	68
Figura 19 – ISI como uma função do número de amostras L	69
Figura 20 – Perfil de correntropia como uma função do atraso m	70
Figura 21 – Superfícies das funções custo analítica e estimada para $\sigma = 0,225$	71
Figura 22 – Superfícies das funções custo analítica e estimada para $\sigma = 0,95$	72
Figura 23 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,225$	72
Figura 24 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,9$	73

Figura 25 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$	73
Figura 26 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,225$	74
Figura 27 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,95$	74
Figura 28 – ISI como uma função da largura do <i>kernel</i> σ	75
Figura 29 – Superfícies das funções custo analítica e estimada para $\sigma = 0,05$	77
Figura 30 – Superfícies das funções custo analítica e estimada para $\sigma = 2,2$	77
Figura 31 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,05$	78
Figura 32 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 2,2$	78
Figura 33 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$	79
Figura 34 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,05$	79
Figura 35 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 2,2$	80
Figura 36 – ISI como uma função da largura do <i>kernel</i> σ	81
Figura 37 – Superfícies das funções custo analítica e estimada para $\sigma = 0,25$	82
Figura 38 – Superfícies das funções custo analítica e estimada para $\sigma = 1,525$	82
Figura 39 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,25$	83
Figura 40 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 1,525$	83
Figura 41 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$	84
Figura 42 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,25$	84
Figura 43 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 1,525$	85
Figura 44 – ISI como uma função da largura do <i>kernel</i> σ	86
Figura 45 – Superfícies das funções custo analítica e estimada para $\sigma = 0,01$	87
Figura 46 – Superfícies das funções custo analítica e estimada para $\sigma = 1,75$	88
Figura 47 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,01$	88
Figura 48 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 1,75$	89

Figura 49 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$ (SNR = 30 dB).	89
Figura 50 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,01$	90
Figura 51 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 1,75$	90
Figura 52 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 12 dB).	91
Figura 53 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB).	92
Figura 54 – ISI como uma função da largura da SNR (dB) ($\sigma = 1,2$).	92
Figura 55 – ISI como uma função da largura do <i>kernel</i> σ para $\sigma_g^2 = 0,5$ (SNR = 30 dB).	93
Figura 56 – Superfícies das funções custo analítica e estimada para $\sigma = 0,1$	94
Figura 57 – Superfícies das funções custo analítica e estimada para $\sigma = 2$	94
Figura 58 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,1$	95
Figura 59 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 2$	95
Figura 60 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$ (SNR = 30 dB).	96
Figura 61 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,1$	96
Figura 62 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 2$	97
Figura 63 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 12 dB).	98
Figura 64 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB).	98
Figura 65 – ISI como uma função da largura do <i>kernel</i> σ para $\sigma_g^2 = 0,5$ (SNR = 30 dB).	99
Figura 66 – Desempenho em termos de ISI (em dB).	100
Figura 67 – Desempenho em termos de $J_{Wiener}(\mathbf{w})$	101
Figura 68 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB) para o pré-codificador com função de transferência $P(z) = 1 + 0.4z^{-1} - 0.6z^{-2}$	103
Figura 69 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB) para as normas L_2 e L_1	106
Figura 70 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB) para as normas L_2 e L_∞	106
Figura 71 – ISI como uma função da largura do <i>kernel</i> σ (SNR = 30 dB) para as normas L_1 e L_∞	107

Figura 72 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas. . . 115

Lista de tabelas

Tabela 1	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	67
Tabela 2	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	70
Tabela 3	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	74
Tabela 4	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	80
Tabela 5	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	85
Tabela 6	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	90
Tabela 7	– ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr}	97
Tabela 8	– Desempenhos em ISI (em dB) e $J_{Wiener}(\mathbf{w})$ associados com as soluções ótimas de J_{corr} e \hat{J}_{corr} para $\sigma = 1$	101

Sumário

1	Introdução	17
1.1	Motivação	17
1.2	Objetivos	18
1.3	Organização da Dissertação	19
1.4	Produções Acadêmicas	19
I	Fundamentos	21
2	Equalização de Canais de Comunicação	22
2.1	Fundamentos do Problema	22
2.1.1	Fonte de Informação	22
2.1.2	Modelagem Matemática	23
2.1.3	Estados do Canal	24
2.1.4	O Problema da Equalização	25
2.1.5	Equalização Supervisionada	28
2.1.6	Algoritmos Adaptativos	29
2.1.7	Equalização Não-Supervisionada	30
2.1.7.1	Algoritmos de Godard	31
2.1.7.2	O Critério do Módulo Constante	32
2.2	Problema de Equalização de Canais para Fontes Pré-Codificadas	33
2.3	Conclusão	34
3	Aprendizado Baseado na Teoria da Informação	35
3.1	Estimadores de Densidade Não-Paramétricos	35
3.2	Entropia de Rényi	38
3.3	Definição de Correntropia	40
3.3.1	Propriedades	42
3.4	Os Critérios Baseados na Teoria da Informação	44
3.4.1	Critério de Mínima Entropia do Erro	44
3.4.2	Critério de Máxima Correntropia do Erro	46
3.4.3	Critério Baseado em Correntropia para Equalização Cega	46
3.5	Conclusão	48
II	Contribuições	49
4	Correntropia Analítica	50
4.1	Fonte Gaussiana	51
4.2	Fonte BPSK (sem ruído)	51

4.3	Fonte Contínua (com ruído)	53
4.3.1	Correntropia da Saída do Equalizador	54
4.3.2	Correntropia da Fonte	56
4.4	Conclusão	57
5	Experimentos Computacionais e Resultados	59
5.1	Evolução Diferencial	60
5.2	Fonte BPSK (sem ruído)	62
5.2.1	Canal de fase mínima	63
5.2.2	Canal de fase máxima	70
5.3	Fonte Gaussiana	76
5.3.1	Canal de fase mínima	76
5.3.2	Canal de fase máxima	81
5.4	Fonte Contínua (com ruído)	86
5.4.1	Canal de fase mínima	87
5.4.2	Canal de fase máxima	93
5.5	Análise do Efeito do Pré-Codificador	99
5.5.1	Pré-codificador escolhido com base no desempenho em ISI	102
5.6	Emprego de Outras Normas no Casamento da Correntropia	103
	Conclusão	108
	Referências	111
	Anexos	114
	ANEXO A Solução de Wiener para o Sistema com Pré-codificador	115

1 Introdução

A área de processamento digital de sinais lida com o desafio de extrair medidas de interesse a partir de uma coleção de informações que frequentemente apresenta distorções, tais como aquelas resultantes de interferências entre sinais e/ou da presença de ruído, considerando o mínimo de informação disponível acerca dos sinais originais e das características específicas do processo responsável pelas distorções. Um exemplo de tal desafio está relacionado ao contexto do problema de equalização de canais de comunicação, no qual o objetivo é projetar um filtro no receptor, denominado equalizador, que cancele ou amenize os efeitos do canal empregado na transmissão, especialmente o fenômeno conhecido como interferência intersimbólica (ISI, do inglês *Intersymbol Interference*), permitindo, assim, a recuperação da informação transmitida. O problema de equalização de canais está relacionado a um vasto conjunto de aplicações da ideia de desconvolução, tais como: astronomia, biomedicina, voz, radar, sonar e aplicações sísmicas (ROMANO *et al.*, 2011). Porém, o contexto natural deste problema, que será abordado neste trabalho, nos remete a sistemas de comunicação.

O filtro equalizador, portanto, deve atuar de maneira inversa ao canal no tocante às distorções inseridas no sinal transmitido, de maneira que sua saída deve, idealmente, corresponder a uma versão perfeita deste sinal, a menos de um ganho e um atraso. O projeto de um equalizador abrange um conjunto de escolhas que são responsáveis pelo processo de filtragem: 1) uma estrutura de filtragem; 2) um critério de equalização adequado; e 3) um método de otimização para ajustar os parâmetros do filtro. No presente trabalho, daremos uma maior ênfase para a segunda escolha, isto é, para o estudo e análise de critérios de filtragem, considerando o cenário em que não se dispõe de uma sequência de treinamento, o que configura a versão não-supervisionada ou cega do problema de equalização (ROMANO *et al.*, 2011; HAYKIN, 2013).

1.1 Motivação

Vários critérios e algoritmos têm sido propostos com base nos teoremas de Benveniste-Goursat-Ruget (BGR) e de Shalvi-Weinstein (SW) para o problema de equalização não-supervisionada (ROMANO *et al.*, 2011; HAYKIN, 2013). Contudo, quando o sinal de interesse é dotado de uma estrutura temporal, muitas das técnicas não-supervisionadas já propostas podem experimentar uma degradação em seu desempenho ou mesmo não podem ser diretamente aplicáveis (FANTINATO *et al.*, 2017). Esse tipo de cenário pode emergir devido às características naturais dos sinais envolvidos, por exemplo, no contexto de voz, áudio e sinais de vídeo, ou mesmo com o emprego de esquemas de codi-

ificação no transmissor, os quais propositalmente inserem redundâncias no sinal antes da sua transmissão.

Interessantemente, a área de Aprendizado Baseado na Teoria da Informação (ITL, do inglês *Information Theoretic Learning*) fornece um conjunto de medidas e conceitos que se mostram capazes de realizar uma extração mais rica e efetiva das propriedades estatísticas disponíveis nos sinais de interesse. Dentre este conjunto de medidas, destaca-se uma medida de correlação generalizada, conhecida como correntropia, a qual serve de base para a definição de um critério de equalização não-supervisionada com fontes que não são independentes e identicamente distribuídas (i.i.d.), que representa o estado da arte na literatura (PRINCIPE, 2010; SANTAMARÍA *et al.*, 2006). Em virtude destes fatos, este trabalho apresenta uma análise do critério baseado em correntropia para equalização cega considerando a presença de fontes dotadas de dependência temporal, as quais são geradas por meio de um estágio linear de pré-codificação.

1.2 Objetivos

Como já destacado, o critério baseado na medida de correntropia desponta como uma opção particularmente atraente para o projeto de equalizadores quando as fontes envolvidas passam por um estágio de pré-codificação. No entanto, algumas questões com relação às reais vantagens e desvantagens deste critério, bem como à sua eficiência em termos teóricos e práticos, ainda carecem de estudos mais aprofundados. Neste contexto, este trabalho de mestrado visa realizar uma análise extensiva do critério baseado em correntropia para o problema de equalização não-supervisionada. Para isto, foi realizada a derivação formal da expressão analítica da correntropia associada ao sinal gerado na saída de um equalizador com resposta finita ao impulso (FIR, do inglês *finite impulse response*) para fonte discreta e para fonte contínua (com ruído), ambas pré-codificadas, possibilitando assim a investigação do critério correspondente na sua forma ideal.

Adicionalmente, foi obtida a expressão exata da solução de Wiener para o sistema com pré-codificador (em inglês, *precoder*), a qual servirá de referência para as soluções de equalização encontradas com o critério baseado em correntropia. Outro aspecto importante deste trabalho refere-se à análise da influência da largura do *kernel* (σ) sobre o critério baseado em correntropia, considerando tanto a sua versão analítica como a sua versão estimada. Com isto, pretende-se realizar um amplo estudo comparativo entre os comportamentos das versões estimada e analítica do critério, bem como avaliar os desempenhos associados a cada equalizador.

1.3 Organização da Dissertação

Esta dissertação está organizada da seguinte maneira:

- Capítulo 2: é feita uma apresentação dos conceitos fundamentais e dos desafios referentes ao problema de equalização de canais de comunicação, quando o sinal transmitido é composto por amostras independentes, tanto no contexto supervisionado quanto no caso cego. Por fim, é apresentado o problema de equalização de canais para fontes pré-codificadas, onde o sinal da fonte apresenta dependência temporal.
- Capítulo 3: foram apresentados alguns conceitos e fundamentos referentes ao aprendizado baseado na Teoria da Informação, bem como os seus principais critérios que são empregados tanto na vertente supervisionada quanto na vertente não-supervisionada do problema de equalização de canais. É neste capítulo que é introduzida a medida da correntropia, cujas propriedades são de fundamental importância para este trabalho.
- Capítulo 4: são apresentadas as derivações teóricas da correntropia analítica para fontes BPSK, fontes Gaussianas e para fontes Contínuas (com ruído).
- Capítulo 5: O critério baseado em correntropia para equalização não-supervisionada é comparado nas suas versões analítica e estimada em termos do ajuste de seus parâmetros, superfície de erro e desempenho de equalização.
- Por fim, são apresentadas as conclusões gerais sobre o trabalho e as perspectivas para trabalhos futuros.
- Anexo A: é apresentada a derivação formal da solução de Wiener para o sistema com pré-codificador.

1.4 Produções Acadêmicas

Durante o período do mestrado, dois trabalhos diretamente ligados à temática desta dissertação foram publicados:

- V. A. Oliveira, D. G. Fantinato, R. Ferrari, R. Attux, L. Boccato, “Analysis of the Correntropy-Based Criterion for Blind Equalization with Precoded Sources,” *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBrT2017*, São Pedro, SP.
- V. A. Oliveira, D. G. Fantinato, L. Boccato, “Análise do Critério Baseado em Correntropia para Equalização Não-Supervisionada com Fontes Gaussianas Precodi-

ficadas,” *X Encontro dos Alunos e Docentes do Departamento de Engenharia de Computação e Automação Industrial (2017) - X EADCA*, Campinas, SP.

Além disso, outros dois trabalhos foram publicados em parceria com pesquisadores ligados ao Laboratório de Processamento de Sinais para Comunicações (DSPCom):

- L. S. Uribe, C. S. Filho, V. B. Olivatto, V. A. Oliveira, D. Soriano, L. Boccato, G. Castellano, R. Attux, “A Comparative Analysis of Correlation and Correntropy in Graph-Based Brain Computer Interfaces,” *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBrT2017*, São Pedro, SP.
- M. E. da Silva, V. A. Oliveira, K. Nose-Filho, L. Boccato, J. M. T. Romano, “Estratégias Evolutivas para o Problema de Desconvolução Sísmica,” *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBrT2017*, São Pedro, SP.

Em especial, destacamos o primeiro destes trabalhos, que trouxe uma aplicação da noção de correntropia para a análise de sinais cerebrais em interfaces cérebro-computador. Tal perspectiva se mostrou promissora e deu origem a um artigo submetido à *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, e que se encontra em fase de revisão.

Parte I

Fundamentos

2 Equalização de Canais de Comunicação

Sistemas de comunicação são projetados com a finalidade de assegurar que uma informação de interesse seja transmitida de um ponto a outro (e.g., da estação rádio-base para um aparelho celular) de forma confiável e garantindo um nível de fidelidade que seja compatível com as exigências da aplicação. Em um contexto prático, o fluxo de comunicação se origina a partir de um nó, denominado transmissor, que é responsável por enviar informações de interesse para um outro nó, denominado receptor. Usualmente, o canal é responsável por introduzir algum tipo de distorção sobre o sinal que ele transmite $s(n)$, a qual é modelada em termos de um sistema cuja entrada é $s(n)$ e cuja saída é o sinal $r(n)$ que chega ao receptor já contendo distorções, como mostra a Figura 1.

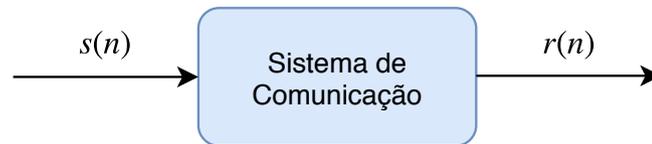


Figura 1 – Diagrama de blocos apresentando o modelo de um sistema de comunicação.

2.1 Fundamentos do Problema

Na prática, o fluxo de informação em um sistema de comunicação se dá através de um canal, o qual representa o meio físico que estabelece a conexão entre o transmissor e o receptor, e que, dependendo de suas características, pode introduzir distorções na informação enviada e, conseqüentemente, prejudicar a interpretação correta da mensagem que chega ao receptor.

A partir deste contexto, surge a necessidade de se projetar e empregar um filtro no receptor, denominado equalizador, cuja função é reverter as eventuais distorções sofridas pelo sinal transmitido, a fim de possibilitar a recuperação da informação transmitida. Esta seção apresenta os fundamentos do problema de equalização de canais de comunicação, descrevendo os principais conceitos e elementos teóricos relacionados ao problema.

2.1.1 Fonte de Informação

A fonte de informação é responsável por gerar o sinal que contém as informações a serem transmitidas, podendo ser discreta ou contínua. Uma hipótese amplamente aceita é a de que a sequência de símbolos gerada, definida por $s(n)$, consiste de variáveis aleatórias (VAs) independentes e identicamente distribuídas (i.i.d.) que respeitam

uma determinada função densidade de probabilidade (PDF, do inglês *Probability Density Function*), para o caso contínuo, ou uma determinada função massa de probabilidade (PMF, do inglês *Probability Mass Function*) para o caso discreto, que, frequentemente, está associada a um alfabeto finito $\mathbb{A} = \{s_i, i = 1, \dots, S\}$, onde S é o número de símbolos do alfabeto. Tal modelo de fonte surge, por exemplo, no contexto de modulações digitais, como nos esquemas PSK (do inglês *Phase Shift-Keying*) e QAM (do inglês *Quadrature Amplitude Modulation*) (BARRY *et al.*, 2003). Entretanto, pode ser também considerada uma fonte de informação dotada de uma estrutura temporal, i.e., composta por amostras estatisticamente dependentes (SANTAMARÍA *et al.*, 2006). Esse tipo de cenário pode emergir devido às características naturais dos sinais envolvidos, por exemplo, no contexto de voz, áudio e sinais de vídeo, ou mesmo por meio do emprego de esquemas de codificação no transmissor, os quais introduzem redundâncias no sinal antes da sua transmissão. O uso do pré-codificador representa um exemplo possível de modelo responsável por produzir dados com dependência temporal.

2.1.2 Modelagem Matemática

As modificações sofridas pelas amostras transmitidas estão diretamente relacionadas a um modelo definido para o canal de comunicação, o qual pode ser de natureza tanto linear quanto não-linear, podendo ser também responsável por inserir diferentes níveis de incerteza/aleatoriedade nas amostras recebidas. Para este trabalho, serão considerados os seguintes efeitos nocivos do canal de comunicação:

- (a) Interferência inter-simbólica (ISI, em inglês *intersymbol interference*): fenômeno produzido pelo espalhamento temporal dos pulsos transmitidos devido à limitação da largura de banda do canal e que também surge devido ao efeito multipercurso em sistemas de comunicação sem fio (BARRY *et al.*, 2003).
- (b) Ruído: decorrente de perturbações aleatórias no sinal recebido devido a fatores como agitação térmica e interferências eletromagnéticas.

De uma maneira usual, um modelo linear é empregado para descrever um canal de comunicação (HAYKIN, 2013). Neste caso, a ISI é representada por meio de uma combinação linear das amostras do sinal transmitido, de acordo com a seguinte

equação¹:

$$\begin{aligned}
 x(n) &= h_0^* s(n) + h_1^* s(n-1) + \dots + h_{D-1}^* s(n-D+1) \\
 &= [h_0^* \ h_1^* \ \dots \ h_{D-1}^*] \cdot \begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s(n-D+1) \end{bmatrix} \\
 &= \mathbf{h}^H \mathbf{s}(n),
 \end{aligned} \tag{2.1}$$

onde os parâmetros h_j^* representam os coeficientes do canal, D é o comprimento do canal, $(*)$ denota a operação de convolução discreta no tempo e $(\cdot)^H$ denota o operador complexo conjugado transposto. A Equação (2.1) mostra que a ISI pode ser modelada por um filtro FIR, cuja função de transferência pode ser dada como:

$$H(z) = \sum_{j=0}^{D-1} h_j^* z^{-j}. \tag{2.2}$$

O ruído por sua vez é geralmente modelado matematicamente como um processo estacionário aditivo, branco e com distribuição Gaussiana (AWGN, do inglês *Additive White Gaussian Noise*) de média nula e variância $\sigma_{\eta_n}^2$. Como o próprio nome indica, tal modelo considera a existência de um sinal aleatório que se soma ao resultante da interferência inter-simbólica. Sendo assim, o sinal que está sendo efetivamente recebido é definido por

$$r(n) = x(n) + \eta(n), \tag{2.3}$$

onde $\eta(n)$ representa o ruído no instante n . Observe que o ruído em questão, que pode ser resultante de interferências e de ruído térmico, é modelado por um processo estacionário AWGN. Entretanto, em alguns cenários menos convencionais, outros modelos podem ser adotados para o ruído, como um ruído do tipo impulsivo (ROMANO *et al.*, 2011; PRINCIPE, 2010).

2.1.3 Estados do Canal

Os estados do canal correspondem às possíveis combinações que o sinal recebido pode assumir na ausência de ruído. Considerando um canal linear, definido pelo conjunto de coeficientes $h_0^*, h_1^*, \dots, h_{D-1}^*$, os estados do canal com dimensão K , denotados por $\xi_i \in \mathbb{C}^K$, são obtidos exclusivamente a partir do efeito da ISI. Dessa maneira, eles são relacionados diretamente com os próprios vetores do sinal recebido $\mathbf{x}(n) = [x(n) \ \dots \ x(n-K+1)]^T$ e, portanto, sem a adição de ruído. Com base na Equação (2.1),

¹ Note que está sendo empregado um canal invariante no tempo e com memória finita, ou seja, que é do tipo FIR.

temos que:

$$\xi_{\mathbf{i}} = \mathbf{x}_{\mathbf{i}}(n) = \begin{bmatrix} h_0^*s(n) + h_1^*s(n-1) + \dots + h_{D-1}^*s(n-D+1) \\ h_0^*s(n-1) + h_1^*s(n-2) + \dots + h_{D-1}^*s(n-D) \\ \vdots \\ h_0^*s(n-D+1) + h_1^*s(n-D) + \dots + h_{D-1}^*s(n-D-K+2) \end{bmatrix}. \quad (2.4)$$

Definindo a matriz de convolução do canal

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \dots & h_{D-1} & 0 & \dots & 0 & \dots & 0 \\ 0 & h_0 & \dots & h_{D-2} & h_{D-1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & \dots & \dots & 0 & h_0 & \dots & h_{D-1} \end{bmatrix}, \mathbf{H} \in \mathbb{C}^{K \times (K+D-1)} \quad (2.5)$$

e o vetor $\mathbf{s}_i(n) = [s(n) \dots s(n-D-K+2)]^T$ contendo uma possível sequência de $(D+K-1)$ símbolos da fonte, cada estado do canal pode ser determinado a partir da seguinte expressão:

$$\xi_{\mathbf{i}} = \mathbf{H}^* \mathbf{s}_i(n). \quad (2.6)$$

É importante ressaltar que a noção de estados do canal é válida quando se considera que o sinal transmitido pertence a um alfabeto finito como, por exemplo, no âmbito da modulação digital 2-PAM (ou BPSK), a qual está associada ao alfabeto binário $\{+1, -1\}$. O número total de estados possíveis é dado pela relação $N_{\text{estados}} = S^{D+K-1}$.

2.1.4 O Problema da Equalização

No contexto do problema de equalização de canais de comunicação, o objetivo é projetar um filtro no receptor, denominado equalizador, que seja capaz de mitigar os efeitos nocivos do canal empregado na transmissão, de maneira especial o fenômeno conhecido como ISI, permitindo, assim, a recuperação da informação transmitida para que a qualidade da informação reconstruída atinja um patamar mínimo requerido pela aplicação. Em outras palavras, a saída do filtro equalizador $y(n)$ deve ser tão parecida com o sinal original produzido pela fonte ($s(n)$) quanto possível. Matematicamente, deseja-se que:

$$y(n) = ks(n - \alpha), \quad (2.7)$$

onde $y(n)$ denota a saída do equalizador, k é um fator de escala e α é um atraso de equalização arbitrário. Quando a relação estabelecida na Equação (2.7) é satisfeita, é possível afirmar que o equalizador compensou totalmente o efeito do canal, eliminando a ISI. A Figura 2 mostra os principais elementos que compõem o problema de equalização de canais de comunicação.

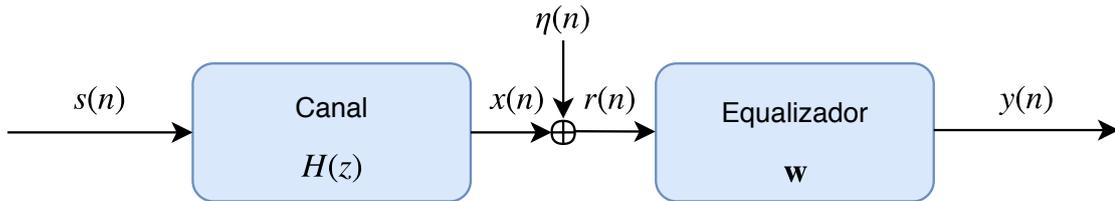


Figura 2 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação.

É fundamental que o filtro equalizador escolhido seja dotado de uma estrutura que seja capaz de cumprir essa tarefa de maneira satisfatória, uma vez que é importante estabelecer um compromisso entre potencial de aproximação e simplicidade prática. Um bom candidato, que representa uma boa opção para lidar com a maioria das aplicações práticas, é um filtro do tipo FIR.

Neste sentido, o filtro equalizador deve atuar de maneira inversa ao papel desempenhado pelo canal. Em outras palavras, deseja-se que o equalizador seja capaz de neutralizar a interferência intersimbólica, bem como de lidar com o ruído presente no sinal recebido, de modo que sua saída deve corresponder, idealmente, a uma versão sem distorções do sinal transmitido. Considerando um equalizador do tipo FIR, cujos coeficientes são dados por $\mathbf{w} = [w_0^* \ w_1^* \ \dots \ w_{K-1}^*]^T$, o sinal de saída $y(n)$ é produzido de acordo com a seguinte expressão:

$$y(n) = \mathbf{w}^H \mathbf{r}(n), \quad (2.8)$$

onde $\mathbf{r}(n) = [r(n) \ \dots \ r(n-K+1)]^T$ é o vetor do sinal recebido.

Esse conceito pode ser melhor entendido quando olhamos a resposta combinada canal-equalizador $c(n) = \sum_{i=-\infty}^{\infty} w_i^* h^*(n-i)$, que representa a convolução entre o canal $h(n)$ e o vetor de coeficientes \mathbf{w} do filtro. Supondo agora que não há a presença do ruído aditivo $\eta(n)$ e considerando, hipoteticamente, um comprimento infinito tanto para a parte causal como não-causal, podemos escrever $y(n)$ como

$$\begin{aligned} y(n) &= \mathbf{w}^H \mathbf{x}(n) = \sum_{i=-\infty}^{\infty} w_i^* x(n-i) \\ &= \sum_{i=-\infty}^{\infty} w_i^* \left(\sum_{j=-\infty}^{\infty} h^*(j) s(n-i-j) \right), \end{aligned} \quad (2.9)$$

tornando $j = l - i$, temos que

$$\begin{aligned} y(n) &= \sum_{i=-\infty}^{\infty} w_i^* \left(\sum_{l=-\infty}^{\infty} h^*(l-i) s(n-l) \right) \\ &= \sum_{l=-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} w_i^* h^*(l-i) \right) s(n-l) \\ &= \sum_{l=-\infty}^{\infty} c(l) s(n-l) = \mathbf{c}^H \mathbf{s}(n). \end{aligned} \quad (2.10)$$

Desse modo, tal formulação é bastante atraente por fornecer uma relação direta entre a saída do equalizador $y(n)$ e o sinal de entrada $s(n)$. Essa formulação nos dá uma noção de que o equalizador só conseguirá atingir uma condição de equalização perfeita se $c(n)$ for proporcional a $\delta(n - \alpha)$, sendo $\delta(\cdot)$ a função impulso discreta. Com o objetivo de se obter a relação dada pela Equação (2.7), o sistema combinado canal-equalizador deve ser representado por um vetor de coeficientes em que apenas o elemento na posição α seja não-nulo e igual a k , isto é:

$$\mathbf{c} = [0 \quad \cdots \quad 0 \quad k \quad 0 \quad \cdots \quad 0]^T. \quad (2.11)$$

Esta situação – de perfeito cancelamento do canal – é conhecida como *Zero-Forcing* (ZF) (HAYKIN, 2013). É importante mencionar que essa condição nem sempre poderá ser satisfeita, uma vez que os equalizadores empregados podem apresentar limitações estruturais que inviabilizem a inversão do canal.

A estrutura escolhida para o equalizador deve ser suficientemente flexível para lidar com as distorções produzidas pelo canal de comunicação, ou seja, com a influência das amostras passadas do sinal transmitido sobre a saída observada do canal e com perturbações aleatórias associadas ao ruído aditivo. Além disso, para que o processo de equalização seja bem-sucedido, uma vez definida a estrutura do equalizador, deve-se realizar o ajuste de seus parâmetros. Esta tarefa envolve: (1) a formulação de um critério a ser otimizado, que, por sua vez, requer a definição de uma função custo, isto é, de uma métrica capaz de avaliar o desempenho de possíveis soluções candidatas para os parâmetros do equalizador; e (2) a adoção de uma estratégia eficiente para realizar, de fato, a otimização da função custo definida. O projeto do equalizador pode, então, ser visto como um problema de otimização, que requer a adoção de uma abordagem eficiente para desempenhar tal tarefa.

Em suma, é possível constatar que o problema de equalização engloba três escolhas fundamentais: (i) a estrutura de filtragem; (ii) o critério de equalização a ser otimizado; (iii) um método ou algoritmo para a otimização dos parâmetros do equalizador. Do ponto de vista da quantidade de informação *a priori* disponível a respeito do sinal de interesse, emergem duas abordagens possíveis: quando uma sequência de amostras do sinal transmitido ($s(n)$) se encontra à disposição durante o processo de otimização, podendo servir de referência para a saída do equalizador, temos o cenário supervisionado do problema de equalização. Caso contrário, consideramos o cenário não-supervisionado (ou cego) de equalização (ROMANO *et al.*, 2011; HAYKIN, 2013), que será abordado na sessão seguinte.

2.1.5 Equalização Supervisionada

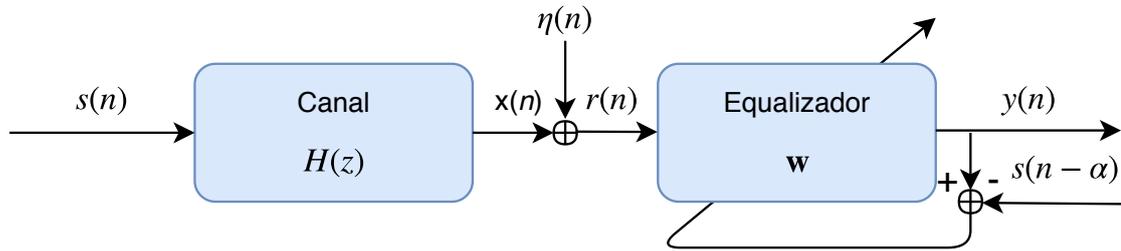


Figura 3 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização supervisionada.

No cenário supervisionado, o processo de adaptação do equalizador ocorre durante uma determinada etapa, designada como período de treinamento, na qual a sequência conhecida de símbolos da fonte é transmitida e, ao mesmo tempo, é utilizada no receptor como referência para a saída $y(n)$ do equalizador, como mostra a Figura 3. Seguindo essa abordagem, emerge uma possibilidade natural de buscar uma medida de proximidade entre a saída do equalizador ($y(n)$) e o sinal de referência ($s(n - \alpha)$), com atraso α constante. Uma opção comum seria a adoção da média das distâncias Euclidianas entre as amostras de ambos os sinais. Podemos considerar ainda que a diferença entre $s(n - \alpha)$ e $y(n)$ pode ser vista como um sinal de erro $e(n)$, convergindo assim para uma função custo denominada erro quadrático médio (MSE, do inglês *Mean Squared Error*)

$$J_{Wiener}(\mathbf{w}) = E[|s(n - \alpha) - y(n)|^2] = E[|e(n)|^2], \quad (2.12)$$

onde $E[\cdot]$ representa o operador de esperança estatística. O critério associado à função custo em (2.12) representa um dos alicerces da teoria de filtragem ótima, sendo designado como critério de Wiener (HAYKIN, 2013). Partindo do ponto de vista analítico, a simplicidade na sua formulação, combinada com o fato de se ter uma função quadrática, faz com que $J_{Wiener}(\mathbf{w})$ possua um ponto de mínimo global com solução fechada, a qual serve de referência para as soluções obtidas por outras abordagens de equalização.

Considerando os sinais envolvidos no sistema como sendo processos estocásticos estacionários, é possível definir medidas de segunda ordem que se relacionam diretamente com a função custo em (2.12). Considerando o vetor de entrada do filtro FIR de ordem K dado por $\mathbf{r}(n) = [r(n) \ r(n-1) \ \dots \ r(n-K+1)]^T$, temos que a matriz de autocorrelação \mathbf{R}_r pode ser escrita como

$$\mathbf{R}_r = E[\mathbf{r}(n)\mathbf{r}^H(n)] = \begin{bmatrix} \rho(0) & \rho(1) & \dots & \rho(K) \\ \rho(-1) & \rho(0) & \dots & \rho(K-1) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(-K) & \rho(-K+1) & \dots & \rho(0) \end{bmatrix}, \quad (2.13)$$

onde $\rho(i) = E[r(n)r^*(n-i)]$ é a função de autocorrelação. Podemos definir também o vetor de correlação cruzada entre $\mathbf{r}(n)$ e $s(n-\alpha)$ da seguinte forma:

$$\mathbf{p}_{\mathbf{x}d} = E\{\mathbf{r}(n)s(n-\alpha)\}. \quad (2.14)$$

A partir destas definições, considerando $|e(n)|^2 = e(n)e^*(n)$ é possível expandir a expressão da função custo MSE em (2.12) da seguinte forma:

$$\begin{aligned} J_{Wiener}(\mathbf{w}) &= E[|e(n)|^2] = E[e(n)e^*(n)] \\ &= E\left[\left(s(n-\alpha) - \mathbf{w}^H \mathbf{r}(n)\right) \left(s^*(n-\alpha) - \mathbf{r}(n)^H \mathbf{w}\right)\right] \\ &= E[s(n-\alpha)s^*(n-\alpha)] - \mathbf{w}^H E[\mathbf{r}(n)s^*(n-\alpha)] - E[s(n-\alpha)\mathbf{r}^H(n)] \mathbf{w} \\ &\quad + \mathbf{w}^H E[\mathbf{r}(n)s(n-\alpha)] \mathbf{w} \\ &= \sigma_s^2 - \mathbf{p}_{\mathbf{x}s}^H \mathbf{w} - \mathbf{w}^H \mathbf{p}_{\mathbf{x}s} + \mathbf{w}^H \mathbf{R}_{\mathbf{r}} \mathbf{w}, \end{aligned} \quad (2.15)$$

sendo σ_s^2 denota a variância de $s(n-\alpha)$. A Equação (2.15) nos mostra que a função custo J_{Wiener} é uma função de segunda ordem em relação ao vetor de pesos \mathbf{w} . Considerando $\mathbf{R}_{\mathbf{r}}$ como sendo uma matriz definida positiva, é garantida a existência de um único ponto de mínimo para o parabolóide formado por $J_{Wiener}(\mathbf{w})$.

A solução que minimiza a Equação (2.12) corresponde à famosa solução de Wiener, dada pela expressão:

$$\mathbf{w}_o = \mathbf{R}_{\mathbf{r}}^{-1} \mathbf{p}_{\mathbf{x}s}, \quad (2.16)$$

onde \mathbf{w}_o representa o vetor ótimo de pesos no sentido de mínimo erro quadrático médio. Essa solução atua, indiretamente, no sentido de mitigar os efeitos nocivos do canal de comunicação, em especial a ISI.

2.1.6 Algoritmos Adaptativos

Embora a solução de Wiener seja conhecida, também podemos lançar mão de métodos iterativos para o ajuste dos coeficientes do equalizador. Isto pode ser particularmente interessante quando não se tem, a priori, todos os dados à disposição, em um contexto real onde o canal de comunicações é desconhecido (o que inviabiliza o cálculo exato da esperança estatística) e, principalmente, quando os sinais envolvidos não são estacionários. O intuito agora é acompanhar as variações do canal e, portanto, não mais chegar em uma solução ótima estática.

Os algoritmos adaptativos são conhecidos por realizarem o processo de otimização de maneira iterativa e apresentando um custo computacional reduzido. Esses algoritmos também são famosos por serem robustos e permitirem a operação em ambientes dinâmicos e diversos. Dentro deste contexto, podemos destacar os algoritmos *Least-Mean-Square* (LMS) e *Recursive Least-Squares* (RLS) (ROMANO *et al.*, 2011), que são capazes de produzir uma estimativa de \mathbf{w}_o em tempo real.

2.1.7 Equalização Não-Supervisionada

O problema de equalização se torna mais desafiador quando não é possível dispor de uma sequência de referência para o ajuste dos coeficientes do filtro, o que configura a sua versão não-supervisionada ou cega. Por este motivo, surge então a necessidade de se empregar outro tipo de informação, e uma das opções encontradas foi a de recorrer, implicitamente ou explicitamente, a estatísticas de ordem superior acerca do sinal de informação que se deseja recuperar (HAYKIN, 2013; ROMANO *et al.*, 2011). Sob uma perspectiva prática, assume-se empregar algum conhecimento estatístico prévio sobre o sinal da fonte $s(n)$. Dois resultados fundamentais que sustentam essa abordagem e que estabelecem as condições estatísticas necessárias para uma perfeita equalização surgem dos trabalhos de Benveniste-Goursat-Ruget (BGR) e de Shalvi-Weinstein (SW) que serão apresentados a seguir.

O Teorema de Benveniste-Goursat-Ruget (BGR)

Em termos simples, o teorema BGR (ROMANO *et al.*, 2011) pode ser enunciado como:

Teorema 1 (Teorema BGR). *Considere que um sinal transmitido seja composto por amostras i.i.d. de uma fonte não-Gaussiana, e que tanto o canal $H(z)$ quanto o equalizador $W(z)$ sejam filtros lineares invariantes no tempo, em um cenário sem a presença de ruído. Sob estas condições, se a PDF da fonte $f_S(v)$ e a PDF da saída do equalizador $f_Y(v)$ forem iguais, pode-se afirmar então que o canal foi perfeitamente equalizado.*

Este teorema assegura que a equalização pode ser obtida utilizando apenas as propriedades estatísticas sobre os sinais. Entretanto, é necessário ter o conhecimento sobre a distribuição do sinal da fonte. O teorema BGR prova que se a PDF do sinal na saída do equalizador for igual à PDF do sinal da fonte, então pode-se afirmar que o canal foi devidamente equalizado. A limitação sobre distribuições Gaussianas recai sobre o fato de que sinais gaussianos, quando submetido a um processo de filtragem linear, continuam gaussianos após a filtragem e, portanto, nenhuma informação além da variância pode ser extraída (ROMANO *et al.*, 2011).

O Teorema de Shalvi-Weinstein (SW)

Em 1990, Shalvi e Weinstein propuseram condições menos restritivas para que uma equalização perfeita pudesse ser satisfeita, o que pode ser visto como um refinamento do que foi estabelecido pelo teorema BGR.

Teorema 2 (Teorema SW). *Considere que um sinal transmitido seja composto por amostras i.i.d. de uma fonte não-Gaussiana, e que tanto o canal $H(z)$ quanto o equalizador*

$W(z)$ sejam filtros lineares invariantes no tempo, em um cenário sem a presença de ruído. Sob estas condições, se $E[|s(n)|^2] = E[|y(n)|^2]$ e alguma outra estatística não-nula de ordem superior (EOS) forem iguais, pode-se afirmar então que o canal foi perfeitamente equalizado.

Este teorema assegura que a equalização perfeita pode ser alcançada igualando-se apenas as estatísticas de segunda ordem e de alguma outra EOS do sinal transmitido e do sinal recebido (ROMANO *et al.*, 2011).

Vários critérios e algoritmos têm sido propostos ou, de certa forma, ganharam um maior entendimento a partir de ambos os teoremas para o problema da equalização não-supervisionada, e.g., o algoritmo de decisão direta (DD), o algoritmo de Sato, os algoritmos de Godard e, um caso específico da família de algoritmos de Godard: o algoritmo de módulo constante (em inglês, *constant modulus algorithm*, CMA) (ROMANO *et al.*, 2011; HAYKIN, 2013).

2.1.7.1 Algoritmos de Godard

As técnicas de equalização cega propostas por Godard (GODARD, 1980) são baseadas em um critério que quantifica a dispersão do sinal equalizado relativamente a um valor pré-determinado sem que seja considerada a fase, de modo a assegurar a independência do processo de desconvolução em relação à estimação de parâmetros de uma portadora. Esta abordagem permite que modulações complexas do tipo QAM sejam contempladas. A função custo, também conhecida como "dispersão de ordem p ", assume a forma:

$$J_{\text{Godard}}(\mathbf{w}, p) = E \left[(|y(n)|^p - R_p)^2 \right], \quad (2.17)$$

em que p é um inteiro positivo e R_p é uma constante que controla o ganho das soluções obtidas, definida por:

$$R_p = \frac{E \left[|s(n)|^{2p} \right]}{E \left[|s(n)|^p \right]}. \quad (2.18)$$

Calculando o gradiente da função custo dada pela Equação (2.17) e lançando mão da aproximação estocástica clássica (HAYKIN, 2013) – que substitui a esperança matemática por uma estimativa instantânea –, Godard obteve a expressão de atualização dos coeficientes do equalizador que, posteriormente, recebeu o seu nome:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu y^*(n) |y(n)|^{p-2} (|y(n)|^p - R_p) \mathbf{x}(n), \quad (2.19)$$

onde μ representa o passo de adaptação. Em particular, o caso em que $p = 2$ é de grande interesse, uma vez que a sua formulação matemática tornou-se mais conveniente e os resultados apresentados foram mais satisfatórios e atraentes no tocante à robustez e convergência nos casos simulados. Foi constatado, ainda para este caso, que a função custo resultante apresentava um menor número de mínimos locais indesejáveis, o que contribuiu

para sua maior aceitação e emprego em aplicações práticas. Entretanto, por possuir uma função custo não convexa, a inicialização adequada dos coeficientes do filtro \mathbf{w} se torna um elemento crucial para assegurar a convergência do algoritmo para uma solução adequada.

É possível observar, para certas modulações, que o algoritmo de Godard penaliza desvios da saída do equalizador $y(n)$ com relação a um módulo constante. Esse fato contribuiu para que a função custo fosse chamada de critério do módulo constante (CM) (TREICHLER; AGEE, 1983) e o algoritmo dele derivado, chamado de algoritmo do módulo constante (CMA, do inglês *Constant Modulus Algorithm*).

2.1.7.2 O Critério do Módulo Constante

Como foi visto anteriormente, o caso específico para $p = 2$ na família de critérios de Godard, definidos em (2.17), resulta na função custo do critério CM:

$$J_{CM}(\mathbf{w}) = E \left[\left(|y(n)|^2 - R_2 \right)^2 \right], \quad (2.20)$$

enquanto a constante R_2 , que controla o ganho das soluções obtidas, passa a ser definida como:

$$R_2 = \frac{E \left[|s(n)|^4 \right]}{E \left[|s(n)|^2 \right]}. \quad (2.21)$$

O algoritmo correspondente a essa nova formulação, definido anteriormente como CMA, pode ser expresso da seguinte forma:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu y^*(n) \left(|y(n)|^2 - R_2 \right) \mathbf{x}(n). \quad (2.22)$$

O CMA é capaz de caracterizar o nível de interferência intersimbólica na saída do equalizador sem recuperar a informação de fase da portadora. Tal comportamento é justificado pelo fato de a formulação da sua função custo ser baseada exclusivamente na amplitude do sinal recebido. Apesar disso, a convergência para os mínimos da função custo CM pode ter uma associação com as soluções fornecidas pelo critério de Wiener (SUYAMA *et al.*, 2003).

É pertinente mencionar que existem também outras abordagens para o problema de equalização cega que não se baseiam diretamente nos teoremas BGR e SW, e que não necessariamente exploram as estatísticas de ordem superior dos sinais para realizar a tarefa. Em particular, destacamos a possibilidade de se empregar filtros de erro de predição (FEPs) ou o uso de restrições, como, por exemplo, no filtro de mínima variância com restrições (ROMANO *et al.*, 2011).

O emprego de filtros de erro de predição é aplicável ao problema de equalização cega toda vez que o sinal desejado for composto de amostras i.i.d. e o canal for linear. O filtro preditor é então projetado, de modo que o erro de predição denotado por $e(n) = r(n) - \hat{r}(n)$, onde $\hat{r}(n)$ representa uma estimativa do sinal recebido, atinja uma condição

de mínimo erro quadrático médio. Considerando a informação disponível nas amostras do sinal recebido $r(n-1), \dots, r(n-N_p)$, a melhor estimativa que pode ser feita pelo filtro preditor é $e_{opt}(n) = h_0^*s(n) + \eta(n)$, dado que $s(n)$ não pode fazer parte de $\hat{r}(n)$, uma vez que as amostras do sinal da fonte são, por hipótese, i.i.d. Por fim, o erro de predição ótimo fornece uma estimativa de $s(n)$, a menos de um ruído aditivo e um fator de escala / desvio de fase devido a contribuição de h_0 . Portanto, o filtro de erro de predição atua como um equalizador não-supervisionado (FERRARI *et al.*, 2008).

2.2 Problema de Equalização de Canais para Fontes Pré-Codificadas

O problema de equalização de canais de comunicação para fontes pré-codificadas pode ser representado pelo diagrama da Figura 4.

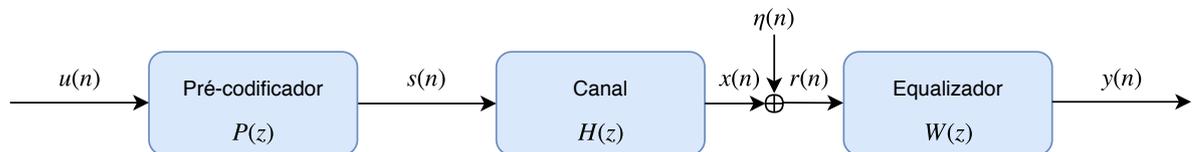


Figura 4 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas.

Diferentemente da abordagem clássica (ROMANO *et al.*, 2011), considera-se a existência de um pré-codificador $P(z)$, representado pelo seu vetor de coeficientes $\mathbf{p} = [p_0 \dots p_{R-1}]^T$, o qual recebe na entrada um sinal i.i.d. $u(n)$, e produz o sinal da fonte $s(n)$, composto por amostras estatisticamente dependentes e, portanto, dotado de uma estrutura temporal. Para este cenário, as condições apresentadas anteriormente não são diretamente aplicáveis, e muitas das técnicas não-supervisionadas já propostas podem experimentar uma degradação em seu desempenho (FANTINATO *et al.*, 2017). O efeito do pré-codificador pode estar relacionado com esquemas de codificação de canal com o objetivo de diminuir a taxa de erro de símbolo e/ou de bit durante a transmissão (LIN; COSTELLO, 2001).

A estrutura temporal dos sinais pode ser empregada de uma maneira benéfica, uma vez que ela carrega informação valiosa sobre a fonte. Neste sentido, a sua manipulação efetiva possibilita que novas abordagens capazes de extrair a informação da dependência temporal possam ser empregadas. Estas perspectivas abrem novos horizontes para que a própria dependência temporal presente nas amostras de interesse possa ser empregada no processo de equalização do canal. É importante mencionar que o ajuste do pré-codificador deve permitir com que as novas abordagens capazes de extrair a informação da dependência temporal sejam efetivas, ou seja, o pré-codificador deve ser escolhido de modo a favorecer o processo de equalização do canal.

2.3 Conclusão

Neste capítulo, apresentamos, inicialmente, os conceitos fundamentais e os desafios referentes ao problema de equalização de canais de comunicação, quando o sinal transmitido é composto por amostras i.i.d., tanto no contexto supervisionado quanto no caso cego. Por fim, apresentamos o problema de equalização de canais para fontes pré-codificadas, onde o sinal da fonte apresenta dependência temporal, ou seja, ele não corresponde a um processo aleatório com amostras i.i.d..

Nesse contexto, surge a necessidade de se recorrer a critérios de otimização que não somente sejam capazes de estender a análise das EOS através de um emprego mais extensivo das densidades de probabilidade dos sinais envolvidos, mas também consigam explorar a informação da estrutura temporal agora presente nas amostras de interesse. A área de Aprendizado Baseado na Teoria da Informação surge como uma alternativa promissora, pois provê um conjunto de técnicas que trazem à tona noções derivadas da teoria da informação e que buscam realizar a extração e o uso mais efetivo das informações estatísticas disponíveis acerca dos sinais de interesse, conforme veremos no próximo capítulo.

3 Aprendizado Baseado na Teoria da Informação

O estudo de critérios e técnicas capazes de extrair as informações estatísticas presentes nos sinais subjacentes do modo mais completo possível, explorando conceitos e medidas oriundos da teoria da informação (COVER; THOMAS, 2006), como a entropia e a informação mútua, constitui a principal motivação da área de pesquisa conhecida como aprendizado baseado na teoria da informação (ITL, em inglês *information-theoretic learning*). A importância destas entidades se deve à sua estrutura probabilística, que permite uma caracterização estatística mais extensiva do que aquela oferecida por momentos específicos de segunda ordem ou de ordem mais elevada. Neste sentido, a sua utilização em aplicações de processamento de sinais e, particularmente, no problema de equalização não-supervisionada de canais é bastante promissora (ROMANO *et al.*, 2011; HAYKIN, 2013). De uma maneira particular, este tópico consiste no tema principal do presente trabalho, cujas propriedades serão cruciais quando empregadas na extração das informações estatística e temporal a partir de sinais de interesse.

O desenvolvimento da teoria da informação se estabeleceu como consequência direta de uma medida definida por Shannon: a entropia, a qual expressa uma noção de incerteza e, sobretudo, da informação que está contida nos sinais aleatórios. No contexto de equalização de canais, a noção de entropia e de comparação entre variáveis aleatórias é explorada com o intuito de cancelar as distorções produzidas pelo canal. Apesar da sólida fundamentação do conceito de entropia produzido pelo trabalho de Shannon (SHANNON, 1948), ele não é facilmente aplicado em um cenário no qual o aprendizado acontece com base em um conjunto limitado de amostras. Esta dificuldade reside na impossibilidade de se calcular a entropia de uma variável sem o conhecimento prévio de sua PDF ou PMF.

Neste cenário, o uso de estimadores não-paramétricos de densidades, como os métodos de kernel associados ao janelamento de Parzen (PARZEN, 1962), combinados com a definição dada por Alfred Rényi (RÉNYI, 1961) ao conceito de entropia abrem um caminho alternativo e possível para o usufruto das vantagens inerentes às medidas de teoria da informação. Esta perspectiva será detalhada na sequência do texto.

3.1 Estimadores de Densidade Não-Paramétricos

A obtenção de uma PDF exata associada a uma fonte de sinal aleatório, idealmente, requer o acesso a uma sequência infinita ou muito grande de amostras, com o intuito de se obter a sua descrição teórica mais fiel e completa possível. Entretanto, em

um contexto real, onde se tem acesso a um número limitado de amostras, torna-se necessário recorrer a métodos efetivos que sejam capazes de estimar a PDF, ou seja, que sejam capazes de extrair o máximo de informação possível das amostras a fim de caracterizar o processo em questão. Um exemplo deste tipo de abordagem é a estimação não-paramétrica de Parzen, que usa como base os métodos de *kernel* (PRINCIPE, 2010).

Os métodos de *kernel* apresentam uma grande vantagem em sua utilidade devido à possibilidade de se resolver problemas não-lineares no próprio espaço dos dados de entrada, denominado \mathfrak{R}^m . Tem-se aqui o então chamado “truque do *kernel*” (do inglês *kernel trick*), permitindo que se substitua um produto interno $\langle \cdot, \cdot \rangle$ num espaço de dimensão potencialmente elevada q , denominado espaço de características \mathfrak{R}^q (do inglês *feature space*), pela aplicação da função *kernel* no espaço original dos dados de dimensão m . Para ilustrar melhor este conceito, pode-se supor um mapeamento não-linear $\phi(\cdot) : \mathfrak{R}^m \rightarrow \mathfrak{R}^q$. A partir deste mapeamento $\phi(\cdot)$, é possível provar que existe uma função *kernel* bivariada $\kappa(\mathbf{x}_i, \mathbf{x}_j)$, definida positiva e simétrica, ou seja, que satisfaz as condições de Mercer (CORTES; VAPNIK, 1995) e que estabelece a seguinte relação de maneira implícita:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j). \quad (3.1)$$

Interessantemente, essa abordagem elegante nos permite, por meio das funções *kernel*, realizar operações de produto interno $\langle \cdot, \cdot \rangle$ num espaço de características, sem sequer ter o conhecimento explícito da função $\phi(\cdot)$. Entretanto, esse mapeamento está diretamente ligado ao tipo de função *kernel* escolhida e do seu ajuste. Esta importante propriedade das funções *kernel* é fortemente explorada pelas máquinas de vetores-suporte (SVMs, em inglês *Support Vector Machines*) e pelas metodologias derivadas dela, como *Support Vector Regression* (STEINWART; CHRISTMANN, 2008).

Matematicamente, a estimativa da janela de *Parzen* da PDF de uma variável aleatória (VA) X , que emprega como base uma função *kernel* arbitrária, é dada por:

$$\hat{p}_X(x) = \frac{1}{L} \sum_{i=1}^L \kappa_\sigma(x - x_i), \quad (3.2)$$

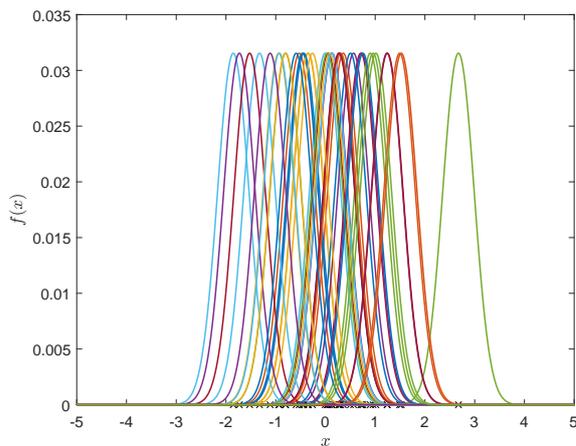
onde temos L amostras i.i.d. da variável aleatória X , $\kappa_\sigma(\cdot)$ é uma função *kernel* arbitrária e σ é a largura do *kernel*. Lembrando que a função *kernel*, além de satisfazer as condições de Mercer (SANTAMARÍA *et al.*, 2006), deve também se adequar às distribuições dos dados e, portanto, precisa respeitar às três seguintes restrições:

1. $\kappa(x) \geq 0$;
2. $\int_{\mathfrak{R}} \kappa(x) dx = 1$;
3. $\lim_{x \rightarrow \infty} |x\kappa(x)| = 0$.

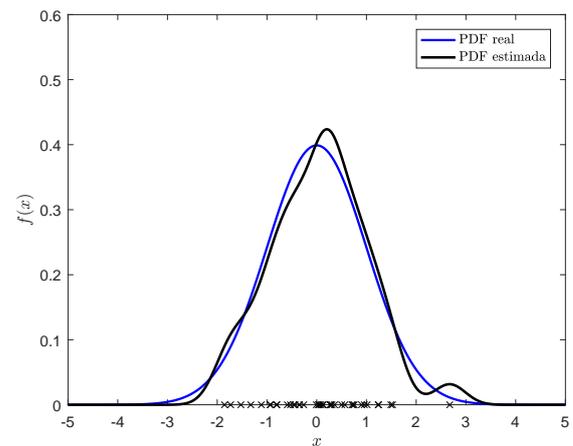
Um exemplo emblemático de função *kernel* que será utilizada neste trabalho, particularmente por possuir propriedades matemáticas bem conhecidas e que serão vitais em nossas derivações, é a função Gaussiana, denotada por $G_\sigma(\mathbf{x}_i - \mathbf{x}_j)$:

$$G_\sigma(\mathbf{x}_i - \mathbf{x}_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (3.3)$$

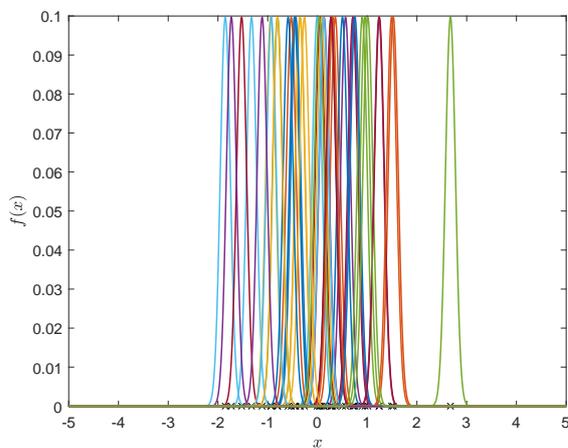
onde σ define a largura do *kernel*. Além das propriedades básicas de uma função *kernel*, a Gaussiana também é simétrica e invariante a translações.



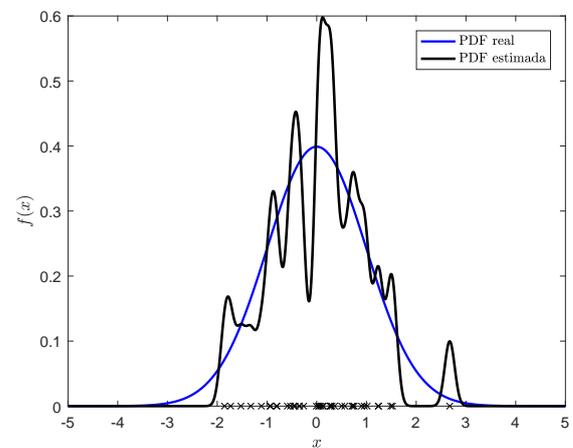
(a) Gaussianas ($\sigma^2 = 0, 1$).



(b) PDF estimada ($\sigma^2 = 0, 1$).



(c) Gaussianas ($\sigma^2 = 0, 01$).



(d) PDF estimada ($\sigma^2 = 0, 01$).

Figura 5 – PDF estimadas por meio do método da Janela de *Parzen*.

A fim de proporcionar um melhor entendimento do problema do ajuste da largura do *kernel* para a estimação de PDF por meio do método da janela de *Parzen*, considere um exemplo simples no qual a PDF verdadeira a ser aproximada corresponde a uma distribuição Gaussiana de média nula e variância unitária. São empregadas $L = 40$ amostras i.i.d. da variável aleatória X . Com o intuito de estimar a PDF original, foram computadas duas PDFs aproximadas por meio do método da janela de *Parzen*

considerando $L = 40$ amostras da variável aleatória X e *kernel* Gaussiano com diferentes larguras do *kernel*: $\sigma = \sqrt{0,1}$ e $\sigma = \sqrt{0,01}$. As aproximações geradas pelo método nos dois casos são exibidas nas Figuras 5b e 5d, respectivamente. Note que nas Figuras 5a e 5c, respectivamente, são apresentadas as Gaussianas posicionadas sobre as amostras da variável aleatória X e ponderadas por L , que são combinadas para formarem as PDFs estimadas para ambos os casos. É possível observar, na Figura 5b, que a PDF estimada para $\sigma^2 = 0,1$ é muito próxima da PDF original. Por sua vez, a PDF estimada para $\sigma^2 = 0,01$ é muito diferente da PDF original. Este exemplo indica que a escolha da largura do *kernel*, para o problema de estimação de PDF, deve se adequar à PDF que se deseja aproximar e ao número de amostras L que são empregadas na sua estimativa.

3.2 Entropia de Rényi

A partir desse contexto, surge o conceito da entropia de Rényi, o qual torna mais direta a obtenção de estimadores não-paramétricos que são particularmente úteis para a adaptação de sistemas, uma vez que, em um contexto real, o emprego direto da entropia de Shannon se esbarra em uma maior complexidade computacional e em dificuldades relativas à diferenciação no caso de fontes contínuas. A entropia de Rényi de uma variável aleatória X com PDF $p_X(x)$ é definida por:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int p_X^\alpha(x) dx. \quad (3.4)$$

onde $\alpha \geq 0$. No caso limite em que $\alpha \rightarrow 1$, temos que a entropia de Rényi se torna equivalente à entropia de Shannon (PRINCIPE, 2010). O argumento do logaritmo na Equação (3.4), denominado potencial de informação (em inglês, *information potential*, IP) de ordem α , pode ser definido como

$$IP_\alpha(X) = \int p_X^\alpha(x) dx = \int p_X^{\alpha-1}(x) p_X(x) dx = E_X [p_X^{\alpha-1}(X)]. \quad (3.5)$$

Considerando $\alpha = 2$, temos a entropia quadrática de Rényi:

$$H_2(X) = -\log \int p_X^2(x) dx, \quad (3.6)$$

de modo que, neste caso, o potencial de informação é correspondente ao valor esperado da PDF de X . A formulação da entropia quadrática de Rényi é bastante conveniente devido a: (1) maior simplicidade de estimação; e (2) possibilidade de derivar algoritmos baseados em gradiente para critérios baseados nesta medida (PRINCIPE, 2010). Dessa maneira, empregando funções do tipo *kernel* gaussiano e substituindo a estimativa de PDF, definida na Equação (3.2), na definição da entropia quadrática de Rényi expressa

em (3.6), obtemos:

$$\begin{aligned}
\hat{H}_2(X) &= -\log \int \left(\frac{1}{L} \sum_{i=1}^L G_\sigma(x - x_i) \right)^2 dx \\
&= -\log \frac{1}{L^2} \int \left(\sum_{i=1}^L \sum_{j=1}^L G_\sigma(x - x_j) \cdot G_\sigma(x - x_i) \right) dx \\
&= -\log \frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L \int G_\sigma(x - x_j) \cdot G_\sigma(x - x_i) dx.
\end{aligned} \tag{3.7}$$

A Equação (3.7) pode ser simplificada graças a uma propriedade interessante da função Gaussiana: o resultado da integral do produto de duas Gaussianas é igual ao valor da Gaussiana calculada na diferença entre os argumentos e cuja variância é a soma das variâncias originais. A partir deste resultado, obtem-se o estimador da entropia quadrática de Rényi (PRINCIPE, 2010):

$$\hat{H}_2(X) = -\log \left(\frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L G_{\sigma\sqrt{2}}(x_j - x_i) \right). \tag{3.8}$$

É possível observar na Equação (3.8) que o potencial de informação, que corresponde a um valor escalar dado pelo argumento do logaritmo, pode ser estimado diretamente a partir dos dados amostrados de acordo com a seguinte equação:

$$\hat{IP}_2(X) = \frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L G_{\sigma\sqrt{2}}(x_j - x_i). \tag{3.9}$$

Assim, é possível estabelecer conexões com os estimadores tradicionais de média e variância, uma vez que o IP é agora expresso por uma função de pares de amostras sobre o conjunto de dados disponíveis, em vez de depender do formato da PDF (PRINCIPE, 2010).

A partir da Equação (3.8), temos que o desvio padrão da Gaussiana é um parâmetro livre que depende de ajuste. Dessa forma, os valores estimados de entropia são influenciados diretamente pela escolha deste parâmetro, comumente referenciado como largura de *kernel*. O ajuste da largura do *kernel* representa um problema crucial em estimação de densidades de probabilidade (SILVERMAN, 1986). O estimador da janela de *Parzen* é considerado como sendo assintoticamente não-polarizado² para a largura do *kernel* quando $L \rightarrow \infty$. Entretanto, em um contexto prático, quando se dispõe de um número limitado de amostras, a largura do *kernel* deve ser ajustada como um *trade-off* entre o viés do estimador e a variância (EMMERT-STREIB; DEHMER, 2009). Uma largura de *kernel* menor, geralmente, favorece um pequeno viés de estimação ao custo de apresentar uma maior variância. Existem vários critérios propostos para o ajuste da

² Um estimador não polarizado é dado por uma estatística amostral na qual o valor esperado é igual ao parâmetro a ser estimado.

largura do *kernel* na literatura (SILVERMAN, 1986), especialmente adequados para dados de pequena dimensão.

É possível estabelecer conexões interessantes entre o aprendizado baseado na teoria da informação e os métodos de *kernel* (EMMERT-STREIB; DEHMER, 2009). Primeiro, a estimativa da janela de *Parzen* da PDF de uma variável aleatória satisfaz as condições de Mercer, e então pode ser vista como o produto interno em algum espaço de características \mathfrak{R}^q induzido. Também é possível interpretar a quantificação baseada na entropia de Rényi em termos de medidas contidas no espaço de características. Substituindo os *kernels* gaussianos na Equação (3.8) pelo produto interno no espaço de características expresso em (3.1), chegamos na seguinte expressão alternativa

$$\begin{aligned}\hat{H}_2(X) &= -\log \left(\frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \right) \\ &= -\log \left\langle \frac{1}{L} \sum_{i=1}^L \phi(\mathbf{x}_i), \frac{1}{L} \sum_{i=1}^L \phi(\mathbf{x}_i) \right\rangle \\ &= -\log (\mathbf{m}^T \mathbf{m}) = -\log \|\mathbf{m}\|^2\end{aligned}\tag{3.10}$$

onde \mathbf{m} representa o valor médio dos dados transformados no espaço de características \mathfrak{R}^q induzido. Este resultado mostra que a noção de quantificação presente na entropia quadrática de Rényi pode estar relacionada com a norma Euclidiana quadrática do valor médio dos dados transformados, calculado no espaço de características.

3.3 Definição de Correntropia

Uma entidade significativa em ITL é a medida conhecida como correntropia, que pode ser vista como uma função de correlação generalizada, capaz de quantificar a similaridade entre duas variáveis aleatórias e que apresenta relações interessantes com entropia quadrática de Rényi e com a estimativa de PDF por meio da janela de *Parzen* (PRINCIPE, 2010). Além de levar em consideração a distribuição estatística dos sinais envolvidos, a correntropia é capaz de englobar a sua estrutura temporal, que é particularmente útil quando se lida com sinais estatisticamente dependentes. A correntropia cruzada entre duas VAs X e Y pode ser definida como:

$$\begin{aligned}V(X, Y) &= E_{XY} [\kappa_\sigma(X, Y)] \\ &= \int \int \kappa_\sigma(x, y) f_{X,Y}(x, y) dx dy,\end{aligned}\tag{3.11}$$

onde $\kappa_\sigma(\cdot)$ denota uma função *kernel*, $f_{X,Y}(\cdot)$ representa a função densidade de probabilidade conjunta (PDF) de X e Y , $E[\cdot]$ é o operador de esperança estatística e σ é a largura do *kernel*. Usualmente, assume-se uma função *kernel* Gaussiana. Outra definição possível para a correntropia surge quando as variáveis aleatórias envolvidas estão associadas a um

mesmo processo aleatório, porém em instantes de tempo diferentes. Neste contexto, esta medida de similaridade é denominada autocorrentropia e é definida pela expressão:

$$\begin{aligned} V(m) &= E_{Y_n Y_{n-m}} [\kappa_\sigma(Y_n, Y_{n-m})] \\ &= \int \int \kappa_\sigma(y_n, y_{n-m}) f_{Y_n, Y_{n-m}}(y_n, y_{n-m}) dy_n dy_{n-m}, \end{aligned} \quad (3.12)$$

onde $f_{Y_n, Y_{n-m}}(\cdot)$ representa a PDF conjunta de Y_n e Y_{n-m} , que, dentro do contexto de equalização não-supervisionada, são versões atrasadas do mesmo sinal e m é o atraso de tempo entre as amostras.

Em cenários práticos, em que a PDF conjunta é desconhecida e apenas um número limitado de amostras $\{(y_n(i), y_{n-m}(i))\}_{i=0}^L$ está disponível, uma média amostral é empregada para aproximar o cálculo da esperança estatística em (3.12), resultando em:

$$\hat{V}_Y(m) = \frac{1}{L - m + 1} \sum_{n=m}^L G_\sigma(y(n) - y(n-m)), \quad (3.13)$$

onde L é o tamanho da janela ou o número de amostras e $G_\sigma(x)$ é o *kernel* Gaussiano com largura de *kernel* σ . É importante ressaltar que o estimador de correntropia estabelece relevantes conexões com outros critérios em ITL (FANTINATO *et al.*, 2017). Usando a expansão em série de Taylor, em torno de 0, para o *kernel* Gaussiano, a correntropia pode ser reescrita como (PRINCIPE, 2010):

$$V_\sigma(m) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{j=0}^{\infty} \frac{(-1)^j}{2^j \sigma^{2j} j!} E[(Y_n - Y_{n-m})^{2j}]. \quad (3.14)$$

Essa expressão motiva o uso da correntropia, uma vez que ela implicitamente engloba vários momentos estatísticos de ordem superior, além do momento de segunda ordem, podendo ser úteis no problema de equalização não-supervisionada, à luz dos teoremas BGR e SW. Observe que a largura do kernel (σ) surge como um fator que pondera de maneira diferente cada momento estatístico. Devido ao operador de esperança estatística $E[\cdot]$, o ajuste da largura do *kernel* em correntropia é diferente do ajuste da largura do *kernel* para o problema de estimação de PDFs (LIU *et al.*, 2007). Em outras palavras, o problema de ajustar a largura do *kernel* em correntropia não representa um fator capaz de comprometer os resultados obtidos como ocorre no problema de estimação de PDFs. Independentemente da largura do *kernel*, a correntropia ainda será capaz de exibir as suas propriedades como medida de similaridade.

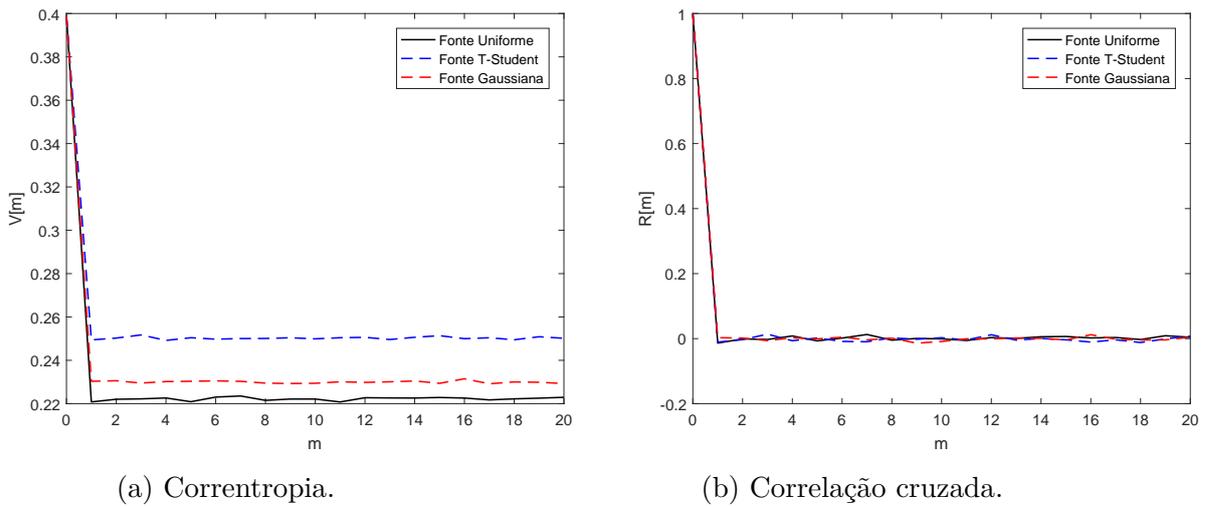


Figura 6 – Processos aleatórios brancos com diferentes distribuições de entrada.

A fim de proporcionar um melhor entendimento do comportamento e das características da correntropia como uma função de autocorrelação generalizada, são considerados três processos aleatórios brancos de variância unitária, de média nula (X, Y e Z) com diferentes distribuições: Uniforme, T-Student e Gaussiana. Para todos os processos, foram computadas a função de correntropia $V[m]$, onde m denota o intervalo de tempo entre duas observações, e a função de autocorrelação $R[m]$, que são exibidas nas Figuras 6a e 6b, respectivamente. A função de autocorrentropia foi estimada usando $L = 30000$ amostras e foi adotado $\sigma = 1$ como largura do *kernel* gaussiano. É possível observar, na Figura 6a, que o valor médio da correntropia não é o mesmo para diferentes distribuições de fonte, mas a função de autocorrelação é basicamente a mesma para todos, como mostrado na Figura 6b. De fato, as diferenças entre distribuições de fontes distintas são mais evidentes quando a forma geral de $V[m]$ é considerada. Este exemplo indica que a correntropia é capaz de extrair informações mais efetivas para a distinção de diferentes processos do que a medida de autocorrelação convencional.

3.3.1 Propriedades

Algumas propriedades importantes da Correntropia foram definidas em (SANTAMARÍA *et al.*, 2006) e são apresentadas a seguir:

- Propriedade 1: Para qualquer *kernel* simétrico definido positivo $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ definido em $\mathfrak{R} \times \mathfrak{R}$, a correntropia está associada a um Espaço de Hilbert Reproduzido por *kernel* (RKHS, em inglês *Reproducing Kernel Hilbert Space*).

Demonstração: Uma vez que $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ é simétrico, é correto afirmar que $V(X, Y)$ é também simétrico. Agora, uma vez que $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ é definido positivo, para qualquer conjunto de n pontos $\{x_1 \dots x_n\}$ e para qualquer conjunto não completamente

nulo de números reais $\{a_1 \dots a_n\}$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \kappa(\mathbf{x}_i, \mathbf{x}_j) > 0. \quad (3.15)$$

Pode-se também afirmar que para qualquer função estritamente positiva $g(\cdot, \cdot)$ de duas variáveis aleatórias X e Y , $E[g(X, Y)] > 0$. Temos então que

$$E \left[\sum_{i=1}^n \sum_{j=1}^n a_i a_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \right] > 0 \quad (3.16)$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \kappa(\mathbf{x}_i, \mathbf{x}_j) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j V(i, j) > 0$$

Assim, $V(X, Y)$ é um operador tanto simétrico quanto definido positivo. Sendo assim, é possível concluir que a correntropia está associada a um RKHS.

- Propriedade 2: A correntropia é uma função simétrica, ou seja $V(X, Y) = V(Y, X)$.
- Propriedade 3: A correntropia é uma função positiva e limitada, e temos que para um *kernel* gaussiano $0 \leq V(X, Y) \leq 1/\sigma\sqrt{2\pi}$. Logo, temos que a correntropia atinge o seu valor máximo se, e somente se, $X = Y$.
- Propriedade 4: Para um *kernel* gaussiano, a correntropia pode ser vista como uma soma ponderada de todos os momentos de ordem par da variável aleatória $X - Y$. Olhando para a equação (3.14), é possível observar que a largura do *kernel* atua como um termo de ponderação para os momentos de segunda ordem ($j = 1$) e de ordem superior. Com $\sigma > 1$, os momentos de ordem mais elevada tendem a ser menos predominantes e decair de um modo mais rápido devido ao denominador da Equação (3.14).
- Propriedade 5: Quando a largura do *kernel* tende a 0, o valor da correntropia se aproxima do valor da probabilidade $p(X = Y)$, que é dado por:

$$\begin{aligned} \lim_{\sigma \rightarrow 0} V(X, Y) &= \lim_{\sigma \rightarrow 0} \iint G_\sigma(x - y) p_{XY}(x, y) dx dy \\ &= \iint \delta(x - y) p_{XY}(x, y) dx dy \\ &= \int p_{XY}(x, x) dx. \end{aligned} \quad (3.17)$$

- Propriedade 6: Seja uma PDF conjunta $p_{XY}(x, y)$, e $\hat{p}_{XY}(x, y)$ é a sua estimativa de Parzen com largura de *kernel* σ . A correntropia estimada com largura de *kernel* $\sigma\sqrt{2}$ é a integral de $\hat{p}_{XY}(x, y)$ ao longo da reta $x = y$, dada por

$$\hat{V}_{\sigma\sqrt{2}}(X, Y) = \int \hat{p}_{X, Y, \sigma}(x, y)|_{x=y=u} du \quad (3.18)$$

- Propriedade 7: Se X e Y são variáveis aleatórias independentes com distribuições de probabilidade $p_X(x)$ e $p_Y(y)$ de modo que a distribuição de probabilidade conjunta seja $p_{XY}(x, y) = p_X(x)p_Y(y)$, temos que:

$$V(X, Y) = \langle E[\phi(X)], E[\phi(Y)] \rangle_{H_\kappa}, \quad (3.19)$$

onde $\langle \cdot, \cdot \rangle_{H_\kappa}$ representa operações de produto interno no RKHS definido pelo *kernel*. Essa propriedade é importante por fornecer uma interpretação da correntropia em termos de ITL quando as VAs envolvidas são independentes, como foi observado na Equação (3.10).

- Propriedade 8: A correntropia pode ser vista como uma simplificação da estatística de segunda ordem dos dados projetados no RKHS definido por uma função *kernel* $\kappa(\mathbf{x}, \mathbf{y})$.

Assumindo-se a dimensão do espaço de características q e que o mapeamento no espaço do *kernel* seja expresso pela função $\phi(\mathbf{x}) = [\Phi_1(\mathbf{x}) \Phi_2(\mathbf{x}) \dots \Phi_q(\mathbf{x})]$. O momento estatístico de segunda ordem entre $\phi(\mathbf{x})$ e $\phi(\mathbf{y})$ é expresso pela seguinte matriz de correlação:

$$\mathbf{R}_{XY} = E[\phi(\mathbf{x})\phi(\mathbf{y})^T] = \begin{bmatrix} E[\Phi_1(\mathbf{x})\Phi_1(\mathbf{x})] & \dots & E[\Phi_1(\mathbf{x})\Phi_q(\mathbf{x})] \\ \vdots & \ddots & \vdots \\ E[\Phi_q(\mathbf{x})\Phi_1(\mathbf{x})] & \dots & E[\Phi_q(\mathbf{x})\Phi_q(\mathbf{x})] \end{bmatrix}. \quad (3.20)$$

Por fim, a partir da Equação (3.19), temos que

$$V(X, Y) = E[\phi(\mathbf{x})\phi(\mathbf{y})^T] = tr(\mathbf{R}_{XY}), \quad (3.21)$$

onde $tr(\cdot)$ denota o traço de uma matriz.

3.4 Os Critérios Baseados na Teoria da Informação

Vários critérios e algoritmos de otimização têm sido propostos com base no arcabouço de medidas e conceitos provenientes de ITL, uma vez que o emprego de estimadores não-paramétricos ampliou novos horizontes no contexto de sistemas adaptativos com a promessa de prover uma caracterização estatística mais extensiva e completa dos sinais de interesse. Alguns destes critérios e algoritmos serão detalhados a seguir.

3.4.1 Critério de Mínima Entropia do Erro

Considere que se deseja ajustar os parâmetros $\mathbf{w} = [w_1, \dots, w_M]$ de um sistema adaptativo linear dentro do paradigma de equalização supervisionada. Partindo dos conceitos de entropia de Rényi e de estimação de PDF baseada em funções do tipo *kernel*, o

critério de mínima entropia do erro (MEEC, do inglês *minimum error entropy criterion*) é proposto como (ERDOGMUS; PRINCIPE, 2002; PRINCIPE, 2010):

$$\begin{aligned} \text{MEEC} : \min_{\mathbf{w}} \hat{H}(e(n)) &= \max_{\mathbf{w}} \hat{\text{IP}}(e(n)) \\ \text{s.a. } e(n) &= d(n) - y(n) \\ E\{e(n)\} &= 0. \end{aligned} \quad (3.22)$$

Este critério busca remover o máximo de incerteza possível associada ao sinal do erro. Portanto, a PDF do erro seria expressa, idealmente, por uma função delta de Dirac ($\delta(\cdot)$), o que significaria que a incerteza sobre $e(n)$ foi totalmente removida ou que toda a informação presente no sinal do erro foi incorporada pelo sistema adaptativo.

Algoritmo *Stochastic Gradient for Minimum Error Entropy* (MEE-SIG)

Em um cenário adaptativo, onde se tem acesso apenas a um número limitado de amostras que são atualizadas instantaneamente, torna-se necessária a adoção de alguma estratégia que seja capaz de atualizar os parâmetros do sistema adaptativo em tempo real. Surge então a ideia de substituir a esperança estatística por seu valor instantâneo, conhecida como aproximação estocástica (ROBBINS; MONRO, 1951), que também é explorada no algoritmo LMS (WIDROW; HOFF, 1988). Partindo da ideia da aproximação estocástica, a expressão do IP pode ser aproximada pelo argumento do operador esperança estatística da expressão do IP – $\text{IP}_\alpha(X) = E_X [p_X^{\alpha-1}(X)]$ – e substituindo $p_X^{\alpha-1}(X)$ pela estimativa da janela de *Parzen* computada sobre as L_j amostras mais recentes no instante n , o potencial de informação estimado se torna

$$\hat{\text{IP}}_\alpha(e(n)) \approx \left(\frac{1}{L_j} \sum_{i=n-L_j}^{n-1} \kappa_{\sigma_\kappa}(e(n) - e(i)) \right)^{\alpha-1}. \quad (3.23)$$

Derivando o potencial de informação com respeito a um coeficiente w_k do filtro, no caso particular do IP quadrático ($\alpha = 2$) e considerando *kernels* gaussianos, obtém-se o gradiente estocástico (PRINCIPE, 2010), dado por:

$$\frac{\partial \hat{\text{IP}}_2(e(n))}{\partial w_k} = \frac{1}{\sigma^2 L_j} \left[\sum_{i=n-L_j}^{i=n-1} G_{\sigma\sqrt{2}}(e(n) - e(i)) \cdot (e(n) - e(i)) \cdot (x_k(n) - x_k(i)) \right]. \quad (3.24)$$

Com este resultado, os coeficientes do filtro são atualizados na direção definida pelo gradiente a cada nova amostra (PRINCIPE, 2010):

$$w_k(n+1) = w_k(n) + \mu \frac{\partial \hat{\text{IP}}_2(e(n))}{\partial w_k}, \quad (3.25)$$

onde $k = 1, \dots, K$ e μ define o tamanho do passo de adaptação.

3.4.2 Critério de Máxima Correntropia do Erro

Outro critério que possui qualidades semelhantes ao MEEC, mas que envolve operações mais simples, explora o conceito de correntropia cruzada entre duas variáveis aleatórias X e Y , cuja a expressão é dada em (3.11) (LIU *et al.*, 2007; PRINCIPE, 2010).

Logo, surge o critério de máxima correntropia do erro (MCC, do inglês *maximum correntropy criterion*), expresso na Equação (3.26), cujo objetivo é maximizar a medida de similaridade entre a saída do equalizador $y(n)$ e o sinal desejado $d(n)$ (PRINCIPE, 2010). Em outras palavras, o objetivo é maximizar o número de amostras com pequenos desvios entre $y(n)$ e $d(n)$. Interessantemente, é possível verificar que a correntropia entre duas variáveis aleatórias X e Y equivale à estimativa da janela de Parzen da PDF da variável $E = X - Y$ na origem, i.e., $\hat{I}P(d(n), y(n)) = \hat{p}_E(0)$ (PRINCIPE, 2010). Do ponto de vista da PDF estimada da variável E , o critério visa maximizar a probabilidade do sinal de erro ser nulo. Formalmente, o MCC é definido da seguinte forma:

$$\mathbf{MCC} : \max_{\mathbf{w}} \hat{V}(d(n), y(n)) = \max_{\mathbf{w}} \hat{p}_E(0). \quad (3.26)$$

Uma grande motivação para o emprego de um critério adaptativo baseado em correntropia reside na forma com que ela atenua exponencialmente a contribuição dos pontos distantes da reta $X = Y$. Isso se dá devido à forma da função *kernel*, o que lhe confere maior imunidade à presença de *outliers* (PRINCIPE, 2010).

Algoritmo *Stochastic Information Gradient for Maximum Correntropy Criterion* (MCC-SIG)

O objetivo aqui é maximizar a correntropia entre a saída do filtro $y(n) = \mathbf{w}^T \mathbf{x}(n)$ e o sinal desejado $d(n)$. Substituindo a PDF do sinal erro pela estimativa de *Parzen* tomada sobre L amostras e derivando a Equação (3.2) com respeito a um coeficiente w_k do filtro, obtem-se a direção do ajuste a ser realizado (PRINCIPE, 2010):

$$\frac{\partial \hat{V}(d(n), y(n))}{\partial w_k} = \frac{1}{\sigma_k^2 L} \sum_{i=1}^L G_\sigma(e(i)) e(i) x_k(i). \quad (3.27)$$

Os coeficientes do filtro são atualizados na direção definida pelo gradiente a cada nova amostra conforme (PRINCIPE, 2010):

$$w_k(n+1) = w_k(n) + \mu \frac{\partial \hat{V}(d(n), y(n))}{\partial w_k}, \quad (3.28)$$

onde $k = 1, \dots, K$ e μ define o tamanho do passo de adaptação.

3.4.3 Critério Baseado em Correntropia para Equalização Cega

Na Seção 3.3, foi apresentada a definição da medida de correntropia, bem como da função de auto-correntropia, a qual nos dá uma caracterização de um processo

aleatório comparando a relação entre suas amostras em instantes de tempo distintos, tal qual a função de autocorrelação, mas com a diferença de implicitamente incorporar informações dos momentos de ordem superior a dois.

Com base nestas características, (SANTAMARÍA *et al.*, 2006) propuseram um critério alternativo para equalização não-supervisionada baseado em correntropia para fontes pré-codificadas. Em suma, a ideia proposta é realizar o casamento entre as funções de autocorrentropia da fonte e do sinal equalizado. Tendo como hipótese que o perfil de correntropia da fonte é conhecido, busca-se o equalizador que consiga gerar em sua saída um sinal cuja função de auto-correntropia seja a mais parecida possível com a da fonte.

Em termos matemáticos, (SANTAMARÍA *et al.*, 2006) propuseram a minimização da seguinte função custo:

$$J_{corr}(\mathbf{w}) = \sum_{m=1}^B (V_S(m) - V_Y(m))^2, \quad (3.29)$$

onde V_S é a correntropia da fonte, V_Y é a correntropia da saída do equalizador e B é o número de atrasos que estão sendo considerados. O critério propõe, portanto, a minimização do erro quadrático total entre os valores de correntropia (V_S) referentes à fonte e os valores da correntropia observada na saída do equalizador (V_Y), considerando um número máximo de atrasos B . Neste caso, a estrutura temporal da fonte pré-codificada é capturada pelos valores da função de correntropia $V_S(m)$ para diferentes atrasos m , a qual também contém informações estatísticas de ordem elevada, e pode ser comparada com aquela associada ao sinal equalizado, através da função de correntropia $V_Y(m)$. Dessa maneira, a estrutura temporal da fonte é efetivamente incorporada e considerada no projeto do equalizador, e certas informações estatísticas dos sinais envolvidos são implicitamente forçadas a serem iguais devido às propriedades da correntropia, remetendo-nos às ideias dos teoremas BGR e SW.

Apesar dos atrativos apresentados por este critério, o uso efetivo da correntropia para a equalização cega requer a escolha de alguns parâmetros que são: a largura do *kernel* σ , o número de atrasos considerados B e o número de amostras L empregadas na estimação da correntropia da saída do equalizador. Uma vez que o pré-codificador é assumido como sendo conhecido, é usual considerar que os valores de $V_S(m)$ podem ser analiticamente computados, enquanto que os valores de $V_Y(m)$ são substituídos pelos seus valores estimados $\hat{V}_Y(m)$, expressos pela Equação (3.13). Sob uma perspectiva teórica, é possível também considerar o caso em que $V_Y(m)$ é obtida analiticamente, o que permite um entendimento mais profundo do comportamento do critério e da influência dos seus parâmetros ajustáveis no desempenho dos filtros equalizadores projetados. Esta possibilidade será explorada neste trabalho por meio da derivação da expressão exata da correntropia $V_Y(m)$ em alguns cenários.

Algoritmo de Gradiente Descendente para o Critério Baseado em Correntropia para Equalização Cega

O objetivo aqui é minimizar o erro quadrático entre os valores de correntropia V_S e os valores da correntropia \hat{V}_Y , considerando um número de atrasos B . Derivando a Equação (3.29) com respeito a um coeficiente w_k do filtro, obtém-se a direção do ajuste a ser realizado (SANTAMARÍA *et al.*, 2006):

$$\frac{\partial \hat{V}_Y(m)}{\partial w_k} = -\frac{1}{\sigma^2(L_j - m)} \sum_{i=n-L_j-m+1}^n G_\sigma(y(i) - y(i-m)) \times (y(i) - y(i-m))(x_k(i) - x_k(i-m)). \quad (3.30)$$

Os coeficientes do filtro são atualizados na direção definida pelo gradiente a cada nova amostra conforme a seguir (SANTAMARÍA *et al.*, 2006):

$$w_k(n+1) = w_k(n) + \mu \sum_{m=1}^B (V_S(m) - V_Y(m)) \frac{\partial \hat{V}_Y[m]}{\partial w_k}, \quad (3.31)$$

onde $k = 1, \dots, K$ e μ define o tamanho do passo de adaptação.

3.5 Conclusão

Neste capítulo, foram apresentados alguns conceitos e fundamentos referentes ao aprendizado baseado na teoria da informação, bem como os principais critérios de ITL que são empregados tanto na vertente supervisionada quanto na vertente não-supervisionada do problema de equalização de canais. Mais detalhadamente, apresentamos a medida denominada correntropia para a caracterização de um processo aleatório, a qual possui propriedades importantes que tornam possível explorar de forma simultânea informações estatísticas de ordem superior dos sinais de interesse e sua estrutura temporal. Em seguida, descrevemos o critério de casamento de correntropia, proposto por Santamaria *et al.* (2006), que visa trazer à fruição as características vantajosas da correntropia durante o projeto do filtro equalizador. Este critério, no âmbito de fontes pré-codificadas, representa o principal foco deste trabalho. Para isto, faremos o desenvolvimento das expressões analíticas exatas do critério em (3.29) considerando alguns tipos de fontes (e.g., BPSK e Gaussiana) com um estágio de pré-codificação. As derivações teóricas para o critério de casamento de correntropia serão apresentadas em detalhes no próximo capítulo.

Parte II

Contribuições

4 Correntropia Analítica

Neste capítulo, vamos derivar as expressões exatas da correntropia (isto é, sem utilizar o estimador) relacionada ao sinal observado na saída do equalizador ($V_Y(m)$), assim como as expressões da correntropia da fonte considerando alguns tipos específicos de sinal.

O cenário a ser abordado é exatamente equivalente ao descrito na Seção 2.2. Para facilitar o acompanhamento das deduções, repetimos, na Figura 7, o diagrama de blocos do problema de equalização cega com fonte pré-codificada, mostrado anteriormente na Figura 4.

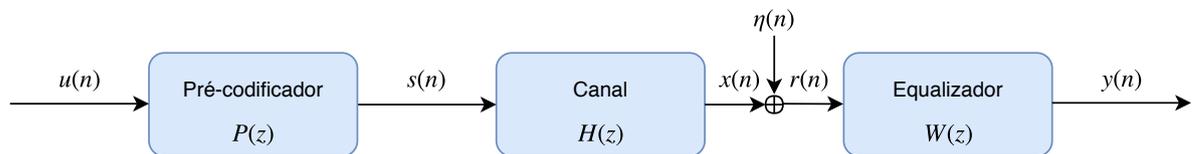


Figura 7 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas.

A fonte $u(n)$ corresponde a um processo aleatório i.i.d. e passa por um estágio de pré-codificação, caracterizado pela função de transferência $P(z)$, a qual introduz dependências temporais entre as amostras $s(n)$. O canal é modelado pela função de transferência $H(z)$, enquanto o equalizador é um sistema do tipo FIR, com coeficientes dados pelo vetor $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_{K-1}]^T$ e função de transferência $W(z) = w_0 + \dots + w_{K-1}z^{-K+1}$. O sinal transmitido também é distorcido pela presença de um ruído aditivo branco e gaussiano (AWGN), de média nula e variância σ_η^2 .

Serão considerados três casos: o primeiro caso a ser tratado envolve uma fonte com distribuição Gaussiana, i.e., quando $u(n) \sim N(0, \sigma_u^2)$. Aqui, vamos reproduzir a dedução analítica originalmente desenvolvida em (FANTINATO, 2017).

No segundo caso abordado, a fonte está associada à modulação digital BPSK e, momentaneamente, não consideramos a presença de ruído no canal. A derivação teórica da correntropia para este cenário é uma das contribuições deste trabalho (OLIVEIRA *et al.*, 2017).

Finalmente, o terceiro caso abordado traz um modelo de fonte novamente contínuo, inspirado no trabalho (BOCCATO *et al.*, 2016), o qual pode ser visto como uma extensão do caso BPSK. A derivação teórica para este tipo de fonte também é feita pela primeira vez neste trabalho de mestrado.

4.1 Fonte Gaussiana

Considerando a fonte como um sinal gaussiano branco $u(n)$, sabe-se que uma combinação linear de $u(n)$ também produz um sinal gaussiano $s'(n)$, com variância:

$$\sigma_{S'}^2 = E[s'(n)^2] = \mathbf{p}'^T E[\mathbf{u}(n)\mathbf{u}(n)^T] \mathbf{p}' = \mathbf{p}'^T \sigma_U^2 \mathbf{p}', \quad (4.1)$$

onde \mathbf{p}' é o vetor do pré-codificador modificado, que tem comprimento $M_{p'} + 1$, com função de transferência $P'(z) = P(z)(1 - 1z^{-m})$ e $\mathbf{u}(n)$ é o vetor do sinal na entrada do pré-codificador. O sistema $P'(z)$, além de aplicar a codificação da fonte, também considera a diferença entre amostras do sinal defasadas por m unidades de tempo, como pode ser visto na Equação (3.13). Dessa forma, a expressão da correntropia analítica se torna (FANTINATO, 2017):

$$\begin{aligned} V_S(m) &= \int f_{S'}(\mathbf{v}) G_{\sigma^2}(\mathbf{v}) d\mathbf{v} \\ &= \int G_{\sigma_{S'}^2}(\mathbf{v}) G_{\sigma^2}(\mathbf{v}) d\mathbf{v} = G_{\sigma_{S'}^2 + \sigma^2}(0). \end{aligned} \quad (4.2)$$

Assumindo um cenário sem ruído, a computação analítica da correntropia associada com a saída de um equalizador pode ser facilmente obtida pela definição do sistema combinado $Q'(z) = P(z)H(z)W(z)(1 - 1z^{-m})$. Dessa maneira, dispondo de $Q'(z)$ em mãos, a correntropia $V_Y(m)$ pode ser determinada de maneira exata com o auxílio das Equações (4.1) e (4.2) (FANTINATO, 2017).

4.2 Fonte BPSK (sem ruído)

Quando a fonte é discreta, a correntropia pode ser analiticamente computada a partir de uma PMF associada a um sinal filtrado de acordo com a mesma estratégia utilizada no contexto da fonte Gaussiana. Considere que, na saída do pré-codificador $P(z)$, exista um sistema com função de transferência $X(z) = 1 - 1z^{-m}$, o que nos leva a ter um sistema combinado $P'(z) = P(z)X(z) = P(z)(1 - 1z^{-m})$. O processo de filtragem antes do canal se torna:

$$s'(n) = \mathbf{p}'^T \mathbf{u}(n), \quad (4.3)$$

onde \mathbf{p}' é o vetor do pré-codificador modificado, que tem comprimento $M_{p'} + 1$, com função de transferência $P'(z) = P(z)(1 - 1z^{-m})$ e $\mathbf{u}(n)$ é o vetor do sinal na entrada do pré-codificador. Se a PMF associada a $s'(n)$ é

$$p_{S'}(v) = \sum_{i \in \mathcal{A}_{S'}} P(v = a_{S'}(i)) \delta(v - a_{S'}(i)), \quad (4.4)$$

onde $\mathcal{A}_{S'}$ é o alfabeto de todas as possíveis ocorrências de $s'(n)$ que são associadas aos estados S' , $a_{S'}(i)$ é o i -ésimo símbolo $\in \mathcal{A}_{S'}$ e $P(v = a_{S'}(i))$ é a probabilidade de $v = a_{S'}(i)$,

então a correntropia da fonte para o atraso m pode ser simplesmente calculada como

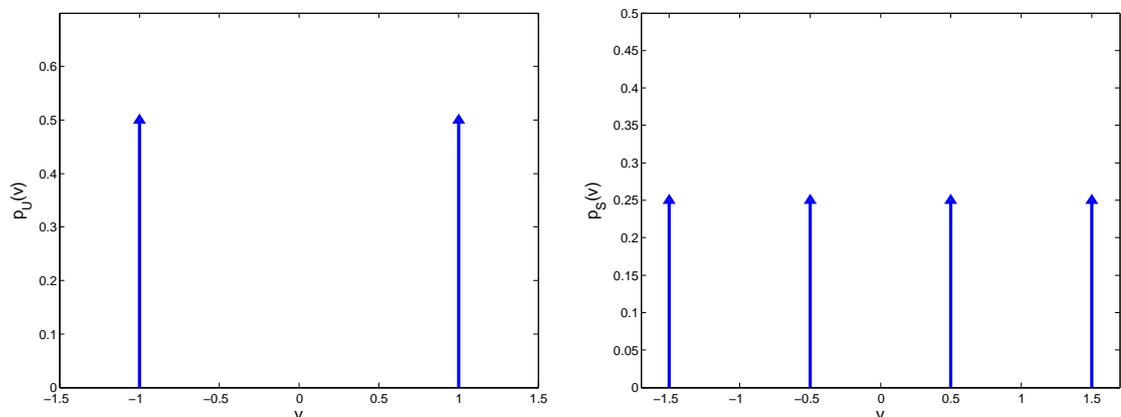
$$V_S(m) = \sum_{i \in \mathcal{A}_{S'}} p_{S'}(a_{S'}(i)) G_\sigma(a_{S'}(i)), \quad (4.5)$$

que é resultado da aplicação da Equação (4.4) em (3.12). Considera-se, como exemplo, o caso em que $u(n)$ é um sinal com modulação BPSK. Se o vetor \mathbf{p}' tem comprimento $M_{p'}+1$, existem então $2^{M_{p'}+1}$ possíveis ocorrências para $\mathbf{u}(n)$ na Equação (4.3), dadas todas as possíveis permutações dos símbolos $\{+1, -1\}$ em um vetor de comprimento $M_{p'}+1$. Neste caso, o i -ésimo estado de S' , $a_{S'}(i)$, é simplesmente a saída dada pela Equação (4.3) para a i -ésima possível ocorrência de $\mathbf{u}(n)$, com probabilidade $P(v = a_{S'}(i)) = 1/2^{M_{p'}+1}$, salvo quando existem estados coincidentes para S' , cuja probabilidade será dada pela soma de ocorrências de $\mathbf{u}(n)$ associadas ao mesmo estado.

De uma maneira interessante, assumindo um cenário sem ruído, o cálculo analítico da correntropia associada com a saída do equalizador pode ser facilmente obtida pela definição do sistema combinado $G'(z) = P(z)H(z)W(z)(1 - 1z^{-m})$, cujo sinal resultante é dado por $y'(n) = \mathbf{g}'^T \mathbf{u}(n)$. A PMF resultante, denotada por $p_{Y'}(v)$, pode ser escrita de uma maneira muito similar à Equação (4.4), com $\mathcal{A}_{Y'}$ sendo o alfabeto de todas as possíveis ocorrências de Y' . Dessa maneira, dispondo de $p_{Y'}(v)$ em mãos, a correntropia $V_Y(m)$ pode ser determinada de maneira exata similarmente à Equação (4.5).

- Exemplo:

Seja $P(z) = 1 + 0.5z^{-1}$ o pré-codificador. A PMF da fonte $u(n)$ é ilustrada na Figura 8a, comprovando ser esta uma fonte do tipo BPSK. Na saída do pré-codificador, o sinal $s(n)$ tem a PMF mostrada na Figura 8b. Por fim, a PMF de $s'(n) = s(n) - s(n-1)$ pode ser observada na Figura 9. Com base em $p_{S'}(v)$, podemos obter a correntropia $V_S(m)$ de forma exata por meio da Equação (4.5).



(a) PMF de $u(n)$.

(b) PMF de $s(n)$ - sinal pré-codificado.

Figura 8 – Distribuições - $u(n)$ e $s(n)$.

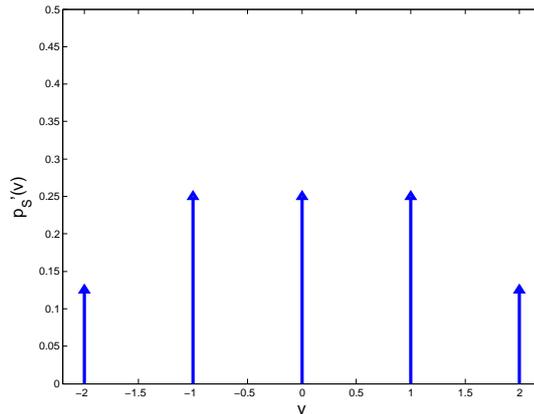


Figura 9 – PMF de $s'(n) = s(n) - s(n-1)$.

4.3 Fonte Contínua (com ruído)

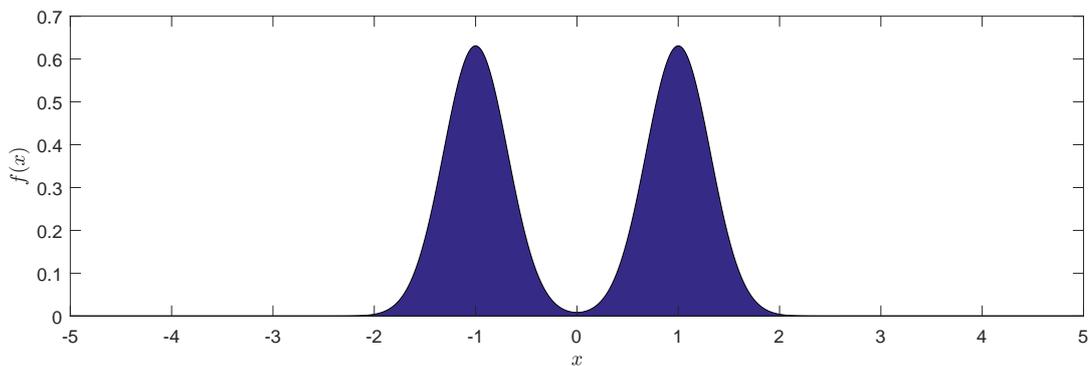


Figura 10 – Formato da PDF da fonte contínua para $\sigma_g^2 = 0,1$.

O terceiro caso abordado traz um modelo de fonte novamente contínuo, inspirado no trabalho (BOCCATO *et al.*, 2016). O sinal da fonte corresponde a um processo aleatório i.i.d., cujas amostras resultam da adição de um ruído Gaussiano $g(n)$ com média nula e variância σ_g^2 , representado por $g(n)$, aos símbolos $u'(n)$ provenientes de um alfabeto binário com símbolos $\{-1, +1\}$ (modulação BPSK). Dessa maneira, temos que:

$$u(n) \sim \frac{1}{2}\mathcal{N}(+1, \sigma_g^2) + \frac{1}{2}\mathcal{N}(-1, \sigma_g^2), \quad (4.6)$$

onde $\mathcal{N}(\mu, \sigma^2)$ representa uma distribuição Gaussiana com média μ e variância σ^2 . Este modelo relativamente incomum para a fonte é particularmente interessante por dois motivos: (i) a modulação BPSK pode ser vista como um caso particular que é atingido quando $\sigma_g^2 = 0$, e (ii) ele se mostrou útil para a análise das propriedades dos critérios para equalização supervisionada baseados em entropia e correntropia, como mostrado em (BOCCATO *et al.*, 2016). Esse modelo de PDF mais complexo, pode potencialmente conduzir a ganhos por meio do uso do critério baseado em correntropia, dada a sua capacidade de explorar informação estatística de ordem elevada. A seguir, serão apresentadas

as formulações teóricas da correntropia da saída do equalizador (V_Y) e da correntropia da fonte (V_S) para a fonte Contínua.

4.3.1 Correntropia da Saída do Equalizador

Dada uma diferença de tempo m entre duas observações, devemos montar um sinal $e_m(n) = y(n) - y(n - m)$ e então olhar a sua PDF para obter o valor da correntropia, definido como:

$$V(e_m) = E [G_\sigma(e_m)] = \int_{-\infty}^{\infty} G_\sigma(e_m) f_{e_m}(e_m = e) de. \quad (4.7)$$

O pré-codificador $P(z)$, que tem comprimento R , e o canal $H(z)$, de comprimento D , tem resposta combinada $Q(z) = P(z)H(z)$, cujo comprimento é $Q = R + D - 1$. Por sua vez, o ruído é matematicamente modelado como um processo estocástico estacionário i.i.d. com distribuição Gaussiana de média nula e variância σ_η^2 . Assume-se que o equalizador é um filtro do tipo FIR de comprimento K , sendo seu sinal de entrada $\mathbf{r}(n)$, que é dado por

$$\mathbf{r}(n) = \mathbf{Q}\mathbf{u}'(n) + \mathbf{Q}\mathbf{g}(n) + \boldsymbol{\eta}(n) = \mathbf{Q}\mathbf{u}(n) + \boldsymbol{\eta}(n), \quad (4.8)$$

onde \mathbf{Q} representa a matriz de convolução do sistema combinado pré-codificador/canal ($Q(z)$), definida por:

$$\mathbf{Q} = \begin{bmatrix} q_0 & q_1 & \dots & q_{Q-1} & 0 & \dots & 0 & \dots & 0 \\ 0 & q_0 & \dots & q_{Q-2} & q_{Q-1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & \dots & \dots & 0 & q_0 & \dots & q_{Q-1} \end{bmatrix}, \mathbf{Q} \in \mathbb{R}^{K \times (K+Q-1)}. \quad (4.9)$$

Temos então que $\mathbf{u}'(n)$ é um vetor intermediário do sinal não-codificado, que contém uma possível sequência de símbolos $\{-1, +1\}$, dado por:

$$\mathbf{u}'(n) = [u'(n) \dots u'(n - K - Q + 2)]^T, \quad (4.10)$$

$\mathbf{g}(n)$ é outro vetor intermediário do sinal não-codificado, que contém uma sequência produzida por um sinal Gaussiano i.i.d. com $G \sim \mathcal{N}(0, \sigma_g^2)$, dado por:

$$\mathbf{g}(n) = [g(n) \dots g(n - K - Q + 2)]^T, \quad (4.11)$$

$\mathbf{u}(n)$ é o vetor efetivo da entrada do pré-codificador dado por:

$$\mathbf{u}(n) = \mathbf{u}'(n) + \mathbf{g}(n) = [u(n) \dots u(n - K - Q + 2)]^T, \quad (4.12)$$

e $\boldsymbol{\eta}(n)$ denota o vetor ruído no instante n , dado por:

$$\boldsymbol{\eta}(n) = [\eta(n) \dots \eta(n - K + 1)]^T. \quad (4.13)$$

O sinal produzido na saída do equalizador é definido, então, pela expressão:

$$y(n) = \mathbf{w}^T \mathbf{r}(n) = \mathbf{w}^T \mathbf{Q} \mathbf{u}'(n) + \mathbf{w}^T \mathbf{Q} \mathbf{g}(n) + \mathbf{w}^T \boldsymbol{\eta}(n). \quad (4.14)$$

O sinal montado $e_m(n)$, que será empregado no cálculo da correntropia, é dado por:

$$\begin{aligned} e_m(n) &= y(n) - y(n-m) \\ &= \mathbf{w}^T \mathbf{Q} \{\mathbf{u}'(n) - \mathbf{u}'(n-m)\} + \mathbf{w}^T \mathbf{Q} \{\mathbf{g}(n) - \mathbf{g}(n-m)\} \\ &\quad + \mathbf{w}^T \{\boldsymbol{\eta}(n) - \boldsymbol{\eta}(n-m)\}. \end{aligned} \quad (4.15)$$

Sendo assim, dado um valor de m , devemos analisar $e_m(n)$ e calcular sua PDF associada para obter o respectivo valor da correntropia. Podemos definir o vetor $\mathbf{u}'_a(n)$ como sendo um vetor resultante da concatenação dos vetores $\mathbf{u}'(n)$ e $\mathbf{u}'(n-m)$. Deste modo, o vetor $\mathbf{u}'_a(n)$ é dado por:

$$\mathbf{u}'_a(n) = [u'(n) \dots u'(n-K-Q+2) \dots u'(n-m-K-Q+2)]^T. \quad (4.16)$$

Note que ocorre certo grau de interseção, isto é, há elementos em comum, entre $\mathbf{u}'(n)$ e $\mathbf{u}'(n-m)$, ou, analogamente, entre $\mathbf{g}(n)$ e $\mathbf{g}(n-m)$, e entre $\boldsymbol{\eta}(n)$ e $\boldsymbol{\eta}(n-m)$ para $m \leq K + Q - 2$.

Com o intuito de facilitar a montagem do sinal $e_m(n)$, é proposta a matriz \mathbf{Q}_m^y , de modo que a operação $\mathbf{Q}\{\mathbf{u}'(n) - \mathbf{u}'(n-m)\}$ se reduza a $\mathbf{Q}_m^y \mathbf{u}'_a(n)$. A matriz \mathbf{Q}_m^y é dada por:

$$\mathbf{Q}_m^y = [[\mathbf{Q} \quad \mathbf{0}] - [\mathbf{0} \quad \mathbf{Q}]], \quad (4.17)$$

onde $\mathbf{0}$ é uma matriz de zeros com dimensões ($K \times m$). A expressão de $e_m(n)$ pode ser então simplificada em:

$$e_m(n) = \mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_a(n) + \mathbf{w}^T \mathbf{Q} \{\mathbf{g}(n) - \mathbf{g}(n-m)\} + \mathbf{w}^T \{\boldsymbol{\eta}(n) - \boldsymbol{\eta}(n-m)\}. \quad (4.18)$$

O segundo termo da Equação (4.18), referente ao sinal gaussiano, também pode ser reescrito usando a mesma abordagem: explorando a matriz \mathbf{Q}_m^y e definindo o vetor $\mathbf{g}_a(n) = [g(n) \dots g(n-m-K-Q+2)]^T$, temos que:

$$e_m(n) = \mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_a(n) + \mathbf{w}^T \mathbf{Q}_m^y \mathbf{g}_a(n) + \mathbf{w}^T \{\boldsymbol{\eta}(n) - \boldsymbol{\eta}(n-m)\}. \quad (4.19)$$

Vamos definir, então, o sinal $g_m(n) = \mathbf{w}^T \mathbf{Q}_m^y \mathbf{g}_a(n)$. No fundo, $g_m(n)$ é dado por uma combinação linear de VAs Gaussianas e independentes. Logo, $g_m(n) \sim \mathcal{N}(0, \sigma_{g_m}^2 = \sigma_g^2 [\mathbf{w}^T \mathbf{Q}_m^y \mathbf{Q}_m^{yT} \mathbf{w}])$.

Seguindo um raciocínio parecido com aquele que nos levou a construir a matriz \mathbf{Q}_m^y , o termo $\mathbf{w}^T \{\boldsymbol{\eta}(n) - \boldsymbol{\eta}(n-m)\}$ pode também ser reescrito na forma $\mathbf{w}^T \mathbf{I}_m^y \boldsymbol{\eta}_a(n)$, onde definimos a matriz \mathbf{I}_m^y como:

$$\mathbf{I}_m^y = [[\mathbf{I} \quad \mathbf{0}] - [\mathbf{0} \quad \mathbf{I}]] \quad (4.20)$$

em que o vetor $\boldsymbol{\eta}_a(n) = [\eta(n) \dots \eta(n-m-K+1)]^T$ e \mathbf{I} é uma matriz identidade com dimensões ($K \times K$). O terceiro termo da Equação (4.19), referente ao ruído, será denotado por $\eta_m(n) = \mathbf{w}^T \mathbf{I}_m^y \boldsymbol{\eta}_a(n)$. Como $\eta_m(n)$ resulta de uma combinação linear de VAs Gaussianas e independentes, $\eta_m(n) \sim \mathcal{N}(0, \sigma_{\eta_m}^2 = \sigma_\eta^2 [\mathbf{w}^T \mathbf{I}_m^y \mathbf{I}_m^y \mathbf{w}])$.

Com os três termos de (4.15) colocados na forma simplificada, o sinal $e_m(n)$ pode ser escrito como:

$$e_m(n) = \mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_a(n) + g_m(n) + \eta_m(n) \quad (4.21)$$

onde $g_m(n)$ e $\eta_m(n)$ são variáveis aleatórias Gaussianas e independentes. Logo, dada uma realização do vetor $\mathbf{u}'_a(n)$, contendo uma possível sequência de $(K+Q+m-1)$ símbolos BPSK, a PDF condicional de $e_m(n)$ é dada por:

$$e_m(n) | \mathbf{u}'_a(n) \sim \mathcal{N}(\mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_a(n), \sigma_{g_m}^2 + \sigma_{\eta_m}^2). \quad (4.22)$$

Sabendo que todas as possíveis sequências $\mathbf{u}'_a(n)$ são equiprováveis e explorando a lei da probabilidade total, a PDF de $e_m(n)$ é dada por:

$$f_{e_m(n)}(e) = \frac{1}{2^{K+Q+m-1}} \sum_{i=0}^{2^{K+Q+m-1}-1} G_{\sqrt{\sigma_{g_m}^2 + \sigma_{\eta_m}^2}}(e - \mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_{a(i)}(n)). \quad (4.23)$$

Sendo assim, a correntropia na saída do equalizador se torna:

$$\begin{aligned} V_Y(e_m(n)) &= \int_{-\infty}^{\infty} G_{\sqrt{\sigma_k^2}}(e) \left[\frac{1}{2^{K+Q+m-1}} \sum_{i=0}^{2^{K+Q+m-1}-1} G_{\sqrt{\sigma_{g_m}^2 + \sigma_{\eta_m}^2}}(e - \mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_{a(i)}(n)) \right] de \\ &= \frac{1}{2^{K+Q+m-1}} \sum_{i=0}^{2^{K+Q+m-1}-1} G_{\sqrt{\sigma_k^2 + \sigma_{g_m}^2 + \sigma_{\eta_m}^2}}(\mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_{a(i)}(n)). \end{aligned} \quad (4.24)$$

4.3.2 Correntropia da Fonte

A correntropia da fonte ($V_S(m)$) pode ser analiticamente computada a partir da PDF associada com um sinal $e_{S_m}(n)$ montado de acordo com a metodologia proposta a seguir. O pré-codificador $P(z)$, que é representado por um filtro FIR e tem comprimento R , é responsável por embutir a dependência temporal e produzir o sinal $s(n)$. Deste modo, o sinal $s(n)$ na saída do pré-codificador $P(z)$ é dado por:

$$s(n) = \mathbf{p}^T \mathbf{u}_s(n), \quad (4.25)$$

em que \mathbf{p} é dado por $\mathbf{p} = [p_0 \dots p_{R-1}]^T$ e $\mathbf{u}_s(n)$ representa o valor do vetor efetivo da entrada do pré-codificador no instante n que é equivalente ao vetor $\mathbf{u}_s(n) = [u(n) \dots u(n-R+1)]^T$. O sinal montado $e_{S_m}(n)$, que será empregado no cálculo da correntropia da fonte, é dado por:

$$\begin{aligned} e_{S_m}(n) &= s(n) - s(n-m) \\ &= ([\mathbf{p} \quad \mathbf{0}]^T - [\mathbf{0} \quad \mathbf{p}]^T) \mathbf{u}_p(n), \end{aligned} \quad (4.26)$$

em que $\mathbf{0}$ é um vetor de zeros com dimensão m e $\mathbf{u}_p(n) = [u(n) \dots u(n - m - R + 1)]^T$. Sendo assim, dado um valor de m , devemos montar o sinal $e_{Sm}(n)$ e olhar sua PDF para obter o valor da correntropia na saída do pré-codificador. Definindo o vetor $\mathbf{p}_0 = [\mathbf{p} \ \mathbf{0}] - [\mathbf{0} \ \mathbf{p}]$, a expressão de $e_{Sm}(n)$ pode ser simplificada em:

$$e_{Sm}(n) = \mathbf{p}_0^T \mathbf{u}_p(n) = \mathbf{p}_0^T [\mathbf{u}'_p(n) + \mathbf{g}_p(n)], \quad (4.27)$$

em que $\mathbf{u}'_p(n)$ é um vetor contendo uma possível sequência de símbolos $\{-1, +1\}$, dado por $\mathbf{u}'_p(n) = [u'(n) \dots u'(n - m - R + 1)]^T$ e $\mathbf{g}_p(n)$ é o vetor equivalente a $\mathbf{g}_p(n) = [g(n) \dots g(n - m - R + 1)]^T$. Considerando $g_{pm}(n) = \mathbf{p}_0^T \mathbf{g}_p(n)$, temos que $g_{pm}(n) \sim \mathcal{N}(\mu = 0, \sigma_{pm}^2 = \mathbf{p}_0^T \sigma_g^2 \mathbf{p}_0)$. Logo, dada uma realização de $\mathbf{u}'_p(n)$, contendo uma possível sequência de $(m - R + 1)$ símbolos BPSK, a PDF condicional de $e_{Sm}(n)$ é dada por:

$$e_{Sm}(n) | \mathbf{u}'_p(n) \sim \mathcal{N}(\mathbf{p}_0^T \mathbf{u}'_p(n), \sigma_{pm}^2). \quad (4.28)$$

Sabendo que todas as possíveis sequências $\mathbf{u}'_p(n)$ são equiprováveis e explorando a lei da probabilidade total, a PDF de $e_{Sm}(n)$ é dada por:

$$f_{e_{Sm}(n)}(e) = \frac{1}{2^{R+m}} \sum_{i=0}^{2^{R+m}-1} G_{\sqrt{\sigma_{pm}^2}}(e - \mathbf{p}_0^T \mathbf{u}'_p(n)). \quad (4.29)$$

Desta forma, a expressão da correntropia da fonte se torna:

$$\begin{aligned} V_S(e_{Sm}(n)) &= \int_{-\infty}^{\infty} G_{\sqrt{\sigma_k^2}}(e) \left[\frac{1}{2^{R+m}} \sum_{i=0}^{2^{R+m}-1} G_{\sqrt{\sigma_{pm}^2}}(e - \mathbf{p}_0^T \mathbf{u}'_p(n)) \right] de \\ &= \frac{1}{2^{R+m}} \sum_{i=0}^{2^{R+m}-1} G_{\sqrt{\sigma_k^2 + \sigma_{pm}^2}}(\mathbf{p}_0^T \mathbf{u}'_p(n)). \end{aligned} \quad (4.30)$$

As Equações (4.24) e (4.30) nos serão particularmente úteis para avaliar o comportamento teórico da correntropia.

4.4 Conclusão

Tendo como motivação principal a busca de um melhor entendimento do critério baseado em correntropia para equalização não-supervisionada, foi apresentado o desenvolvimento teórico da correntropia para diferentes tipos de fontes, que diferem quanto à estrutura estatística: Fonte Gaussiana, Fonte BPSK e Fonte Contínua. Primeiramente, é reproduzida a dedução analítica originalmente desenvolvida em (FANTINATO, 2017) para fontes Gaussianas, que será objeto de avaliação por este trabalho. No segundo caso abordado, é apresentada a derivação teórica da correntropia para o cenário em que a fonte está associada à modulação digital BPSK, que, inicialmente, não considera a presença de ruído no canal. A derivação teórica da correntropia para este cenário é uma das contribuições que foram desenvolvidas neste trabalho (OLIVEIRA *et al.*, 2017). Finalmente, o

terceiro caso abordado traz um modelo de fonte novamente contínuo, inspirado no trabalho (BOCCATO *et al.*, 2016), que também pode ser visto como uma generalização do caso BPSK. A derivação teórica para este tipo de fonte também foi apresentada.

Partindo do estudo das formulações teóricas desenvolvidas para a correntropia analítica - Equações (4.2), (4.5) e (4.24), é possível observar que as funções custo analíticas possuem uma quantidade bem maior de parâmetros ajustáveis em comparação com a versão estimada, dada na Equação (3.13). O único parâmetro que ambas as versões compartilham é a largura do *kernel* (σ), que está ligada diretamente à propriedade da correntropia de ponderar os momentos estatísticos da PDF dos sinais envolvidos de maneira diferente. Recai sobre três únicos parâmetros ajustáveis – a largura do *kernel* (σ), o número de atrasos (B) e o número de amostras empregadas (L) – a responsabilidade de ajustar a versão estimada do critério às condições do cenário avaliado: PDF da fonte, canal empregado e variância do ruído. Neste sentido, a largura do kernel, na versão estimada do critério, é responsável por desempenhar outros ajustes, além do ajuste das propriedades intrínsecas da correntropia. O número de atrasos (B) é responsável por definir os número de atrasos para os valores de correntropia que o critério busca igualar efetivamente, ou seja, ele está diretamente relacionado à estrutura temporal da fonte. O número de amostras (L) empregado pela versão estimada do critério está diretamente ligado à qualidade da estimação da correntropia e constitui um fator crítico quando se pensa em aplicações reais que, frequentemente, dispõem de um número limitado de amostras.

5 Experimentos Computacionais e Resultados

Neste capítulo, vamos investigar, por meio de simulações computacionais, o comportamento do critério de casamento de correntropia, tanto em sua versão estimada, cuja expressão é dada em (3.13), quanto em sua versão teórica, derivada no Capítulo 4 para diferentes tipos de sinal.

Inicialmente, vamos comparar as duas versões do critério tendo em vista as características das superfícies de cada função custo. Para que esta visualização seja possível, restringiremos o número de coeficientes do equalizador em $K = 2$. Nesta análise comparativa, também vamos avaliar o efeito que a largura do *kernel* (σ) exerce sobre as superfícies da função custo, bem como sobre a própria solução (mínimo global).

Outra questão importante com a qual vamos lidar refere-se ao desempenho das soluções fornecidas pelo critério de correntropia na equalização do canal. As soluções obtidas para as versões teórica e estimada do critério de correntropia serão avaliadas em termos da medida de interferência inter-simbólica residual (ISI), cuja definição é dada a seguir:

$$ISI_{dB} = 10 \log_{10} \frac{\left(\sum_{i=0}^{M_c} |c_i|^2\right) - \max_j |c_j|^2}{\max_j |c_j|^2}, \quad (5.1)$$

onde $\mathbf{c} = [c_0 \ c_1 \ \dots \ c_{M_c}]^T$ representa a resposta ao impulso combinada canal-equalizador (que possui função de transferência $C(z) = H(z)W(z)$). Quanto melhor for a solução para o equalizador (\mathbf{w}), mais parecida com um impulso será a resposta combinada canal-equalizador. Assim, menor será a diferença entre a soma dos módulos dos coeficientes ao quadrado ($\sum_{i=0}^{M_c} |c_i|^2$) e o máximo coeficiente ao quadrado ($\max_j |c_j|^2$). Logo, quanto menor for o valor da ISI, mais próximo o equalizador chegou da condição de cancelamento do canal. É interessante ressaltar que a ISI não é influenciada por fatores de escala associados ao vetor \mathbf{w} .

Em todos os experimentos envolvendo a versão estimada do critério de correntropia (\hat{J}_{corr}), a análise de desempenho baseada na ISI será feita a partir da média dos resultados obtidos em N_E simulações independentes. O pré-codificador foi escolhido de maneira arbitrária e fixado para os primeiros resultados. Por fim, será realizada uma análise detalhada sobre o seu efeito. Além disso, a fim de estabelecer uma referência de desempenho, vamos mostrar também os valores de ISI obtidos pela solução de Wiener no contexto de uma fonte pré-codificada. O Apêndice A traz a dedução completa da solução ótima de mínimo erro quadrático médio para os coeficientes do equalizador quando a fonte possui uma estrutura de dependência gerada por um pré-codificador $P(z)$.

Conforme destacado no Capítulo 4, o critério baseado em correntropia para equalização cega, em sua forma estimada, possui três parâmetros importantes: (1) a largura do *kernel* (σ); (2) o número de atrasos (B), e (3) o número de amostras do sinal de interesse (no caso, da saída do equalizador) para a estimação da PDF. O impacto destes parâmetros também será discutido durante os experimentos apresentados neste capítulo. No caso dos parâmetros σ e B , esta discussão englobará também a versão teórica do critério.

O projeto efetivo do equalizador explorando o critério de casamento de correntropia passa pela otimização da função custo em (3.29), seja na forma estimada, seja na forma analítica. Contudo, a função custo em questão possui uma dependência não-linear e não-quadrática com respeito aos parâmetros do equalizador, de modo que a superfície de otimização pode apresentar múltiplos ótimos locais e outros tipos de comportamentos (e.g., pontos de sela) que dificultam a busca pelo ótimo global.

Diante destas características e, tendo em vista que desejamos avaliar o critério com base na solução ideal que ele consegue fornecer, o uso de um método de busca mais robusto se mostra pertinente. Neste contexto, metaheurísticas populacionais emergem como opções particularmente interessantes, pois se mostram capazes de balancear uma exploração adequada do espaço de soluções candidatas com um refinamento local em torno de soluções mais promissoras. Embora não se possa garantir que este tipo de estratégia encontra, de fato, o ótimo global, tais métodos têm um potencial notável de realizar a busca pelo ótimo, de escapar de soluções localmente ótimas e de, até mesmo, encontrar múltiplas soluções ótimas em uma única execução.

Por estas razões, a busca pelos ótimos globais de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ será realizada neste trabalho com o auxílio da meta-heurística denominada evolução diferencial (PRICE *et al.*, 2005).

5.1 Evolução Diferencial

A busca pelos ótimos globais das funções custo é realizada com o auxílio da metaheurística denominada Evolução Diferencial (ED), que consiste em uma técnica eficiente para explorar o espaço de busca e evitar a convergência para ótimos locais (PRICE *et al.*, 2005). A sua principal característica é o fato de as soluções candidatas serem adaptadas por meio de mecanismos que exploram a informação sobre o espaço de busca que está presente na população atual, ao invés de empregar operadores convencionais baseados em perturbações aleatórias (para mais detalhes, sugere-se a referência (PRICE *et al.*, 2005)). Os parâmetros do algoritmo ED são o tamanho da população N_P , o tamanho do passo F , a taxa de *crossover* CR e o número máximo de iterações I_T , que serão ajustados de acordo com cada cenário avaliado. Adicionalmente, com o intuito de assegurar a con-

vergência adequada do algoritmo ED, foram realizados alguns testes por busca em grade que confirmaram a eficácia do algoritmo ED para lidar com o problema de equalização considerado.

No método de Evolução Diferencial, novos vetores de parâmetros são gerados via mutação através da adição do vetor diferença entre dois indivíduos a um terceiro indivíduo. Em termos matemáticos, para cada indivíduo (*target vector*), denotado por $\mathbf{x}_{i,G}$, $i = 1, \dots, N_p$, onde N_p é o tamanho da população e G o índice da geração atual, um novo vetor será gerado segundo a expressão:

$$\mathbf{v}_{i,G+1} = \mathbf{x}_{r_1} + F(\mathbf{x}_{r_3} - \mathbf{x}_{r_2}), \quad (5.2)$$

onde $i, r_1, r_2, r_3 \in \{1, 2, \dots, N_p\}$, são índices mutuamente distintos e F é uma constante real que denota o passo a ser dado na direção do vetor diferença.

Após esta operação, é feita a recombinação entre $\mathbf{x}_{i,G}$, i.e., os indivíduos originalmente presentes na população no início da geração G , e os respectivos vetores mutados $\mathbf{v}_{i,G+1}$, produzindo, assim, os chamados *trial vectors*, representados por $\mathbf{u}_{i,G+1} = [u_{1i,G+1} \ u_{2i,G+1} \ \dots \ u_{Di,G+1}]^T$. A estratégia de recombinação utilizada é muito semelhante ao famoso esquema de *crossover* uniforme (BÄCK *et al.*, 2000): cada elemento de $\mathbf{u}_{i,G+1}$ é uma cópia do elemento correspondente de um dos pais, i.e., de $\mathbf{x}_{i,G}$ ou de $\mathbf{v}_{i,G}$; a escolha de qual pai transfere o atributo para $\mathbf{u}_{i,G+1}$ é probabilística, sendo esta probabilidade um parâmetro do algoritmo, denotado por CR .

Por fim, a seleção é feita com base na comparação de *fitness* entre o *target vector* $\mathbf{x}_{i,G}$ e o *trial vector* associado, $\mathbf{u}_{i,G+1}$: aquele que possuir o melhor valor de *fitness* será mantido na população.

A fim de proporcionar um melhor entendimento do funcionamento do método de evolução diferencial, o seu pseudocódigo é apresentado a seguir.

Pseudocódigo 5.1 Evolução Diferencial

```

1: função  $\mathbf{x} = \text{ED}(N_p, CR, F, range, f)$ 
2:    $\mathbf{x} \leftarrow \text{random}(range, N_p)$ 
3:    $fit_x \leftarrow f(\mathbf{x})$ 
4:   enquanto critério de parada não for satisfeito faça
5:     para  $i \leftarrow 1$  até  $N_p$  faça
6:        $\mathbf{v}_{i,G+1} \leftarrow \text{mutação}(\mathbf{x}_{i,G}, F)$ 
7:        $\mathbf{u}_{i,G+1} \leftarrow \text{crossover}(\mathbf{x}_{i,G}, \mathbf{v}_{i,G+1}, CR)$ 
8:     fim para
9:      $fit_u \leftarrow f(\mathbf{u})$ 
10:    para  $i \leftarrow 1$  até  $N_p$  faça
11:      se  $fit_u(i) > fit_x(i)$  então
12:        devolve  $\mathbf{x}_{i,G+1} \leftarrow \mathbf{u}_{i,G+1}$ 
13:      senão
14:        devolve  $\mathbf{x}_{i,G+1} \leftarrow \mathbf{x}_{i,G}$ 
15:      fim se
16:    fim para
17:  fim enquanto
18: fim função

```

Definido o método de busca, investigaremos a seguir, através de simulações computacionais, os três casos de fontes previamente considerados.

5.2 Fonte BPSK (sem ruído)

Nesta seção, vamos investigar, por meio de simulações computacionais, o comportamento do critério de casamento de correntropia para a fonte BPSK, tanto em sua versão estimada quanto em sua versão teórica, no cenário em que a fonte é BPSK. O pré-codificador utilizado no cenário analisado é fixo e tem função de transferência $P(z) = 1 - 0,5z^{-1} + 0,3z^{-2}$.

A expressão do critério de casamento de correntropia em sua versão estimada é

$$\hat{J}_{corr}(\mathbf{w}) = \sum_{m=1}^B \left(V_S(m) - \hat{V}_Y(m) \right)^2, \quad (5.3)$$

onde B representa o número de atrasos considerados pelo critério e o cálculo da correntropia na saída do equalizador (\hat{V}_Y) é realizado de maneira estimada, cuja expressão é dada em (3.13). A expressão da correntropia estimada é reproduzida abaixo:

$$\hat{V}_Y(m) = \frac{1}{L - m + 1} \sum_{n=m}^L G_\sigma(y(n) - y(n-m)), \quad (5.4)$$

onde L é o tamanho da janela ou o número de amostras e $G_\sigma(x)$ é o *kernel* Gaussiano com largura de *kernel* σ .

A expressão do critério de casamento de correntropia na sua versão teórica é

$$J_{corr}(\mathbf{w}) = \sum_{m=1}^B (V_S(m) - V_Y(m))^2, \quad (5.5)$$

onde o cálculo da correntropia na saída do equalizador (V_Y) é realizado de maneira analítica, conforme a Expressão em (4.5), reproduzida abaixo:

$$V_Y(m) = \sum_{i \in \mathcal{A}_{Y'}} p_{Y'}(a_{Y'}(i)) G_{\sigma^2}(a_{Y'}(i)). \quad (5.6)$$

onde $\mathcal{A}_{Y'}$ é o alfabeto de todas as possíveis ocorrências de Y' , $a_{Y'}(i)$ é o i -ésimo símbolo $\in \mathcal{A}_{Y'}$ e $P(v = a_{Y'}(i))$ é a probabilidade de $v = a_{Y'}(i)$.

Tanto para a versão estimada quanto para a versão analítica do critério de casamento de correntropia, o cálculo da correntropia da fonte (V_S) é realizado de maneira analítica, sendo calculada da seguinte forma:

$$V_S(m) = \sum_{i \in \mathcal{A}_{S'}} p_{S'}(a_{S'}(i)) G_{\sigma^2}(a_{S'}(i)). \quad (5.7)$$

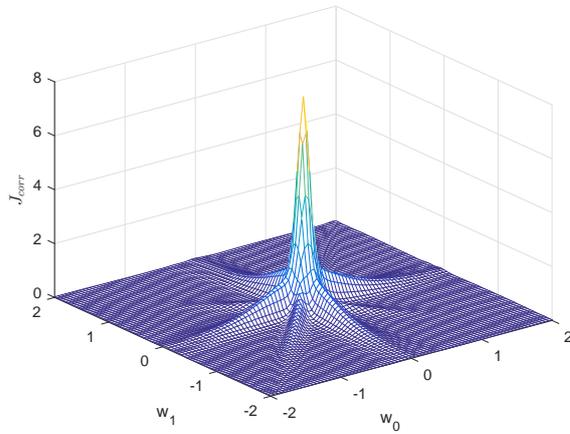
onde $\mathcal{A}_{S'}$ é o alfabeto de todas as possíveis ocorrências de S' , $a_{S'}(i)$ é o i -ésimo símbolo $\in \mathcal{A}_{S'}$ e $P(v = a_{S'}(i))$ é a probabilidade de $v = a_{S'}(i)$, isso para um sinal filtrado e proveniente de uma modulação BPSK $\{-1, +1\}$.

5.2.1 Canal de fase mínima

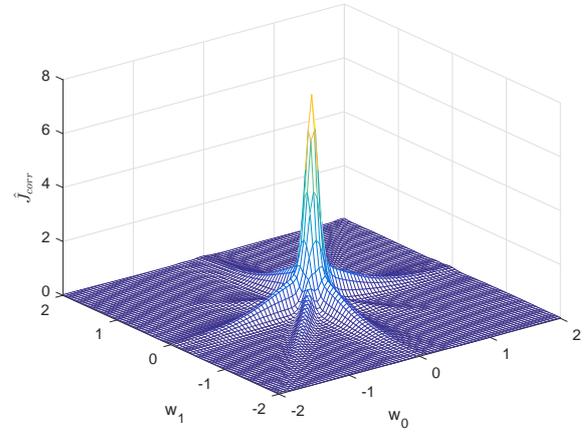
No cenário considerado, o canal é um sistema de fase mínima com função de transferência dada por $H(z) = 1 + 0,5z^{-1}$. Um equalizador com dois coeficientes $\mathbf{w} = [w_0 \ w_1]^T$ é adotado, cujos pesos irão variar no intervalo de -2 a 2 para se obter os contornos das funções custo $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$. Os parâmetros utilizados no algoritmo ED foram $N_P = 100$, $F = 0,5$, $CR = 0,9$ e $I_T = 100$ iterações.

Primeiramente, analisou-se o efeito do tamanho do *kernel* σ nas superfícies. Fixou-se o número de atrasos considerados no critério em $B = 2$ para ambas as funções custo - analítica e estimada - e utilizou-se $L = 5000$ para a versão estimada. Assim, variando os coeficientes do equalizador, foram obtidas as superfícies de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,175$ e $\sigma = 0,9$, como mostrado nas Figuras 11 e 12, respectivamente. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. Foram obtidos também as curvas de nível das superfícies $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para estes valores de σ , como mostrado nas Figuras 13 e 14, respectivamente. Adicionalmente, também mostramos nas Figuras 15a e 15b as soluções encontradas pelo algoritmo ED, que são representadas por asteriscos (*) para o custo analítico e por pontos (·) para a função custo estimada, considerando $N_E = 100$ experimentos independentes. Por fim, são apresentados nas Figuras 16 e 17 os histogramas dos

coeficientes do equalizador para a versão estimada do critério quando $\sigma = 0,175$ e $\sigma = 0,9$, respectivamente.

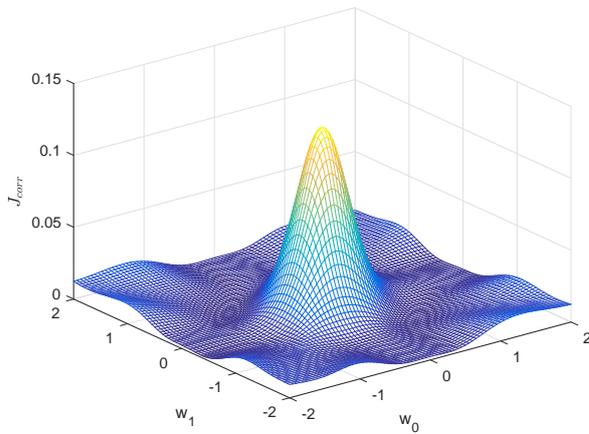


(a) Superfície de $J_{corr}(\mathbf{w})$

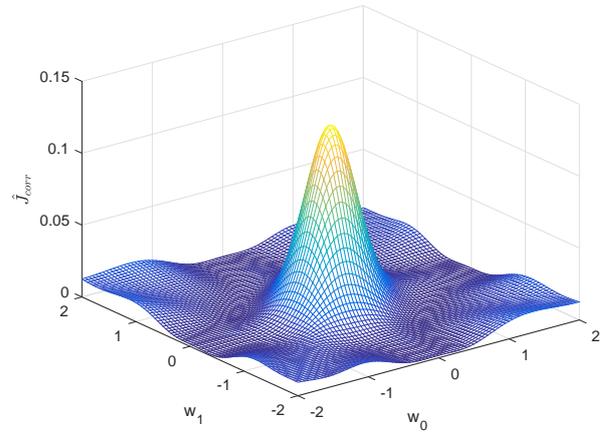


(b) Superfície de $\hat{J}_{corr}(\mathbf{w})$

Figura 11 – Superfícies das funções custo analítica e estimada para $\sigma = 0,175$.



(a) Superfície de $J_{corr}(\mathbf{w})$



(b) Superfície de $\hat{J}_{corr}(\mathbf{w})$

Figura 12 – Superfícies das funções custo analítica e estimada para $\sigma = 0,9$.

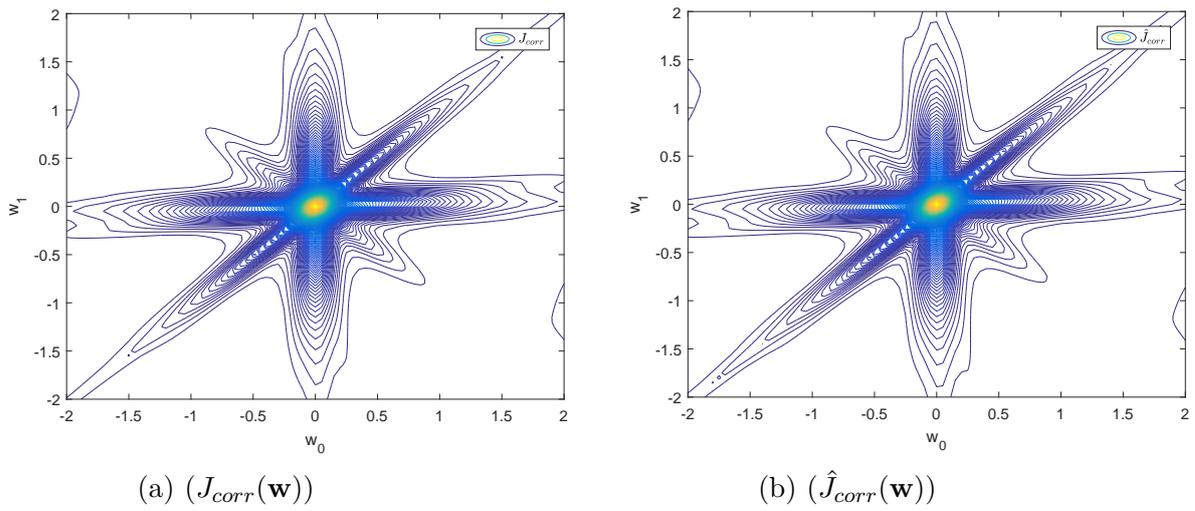


Figura 13 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,175$.

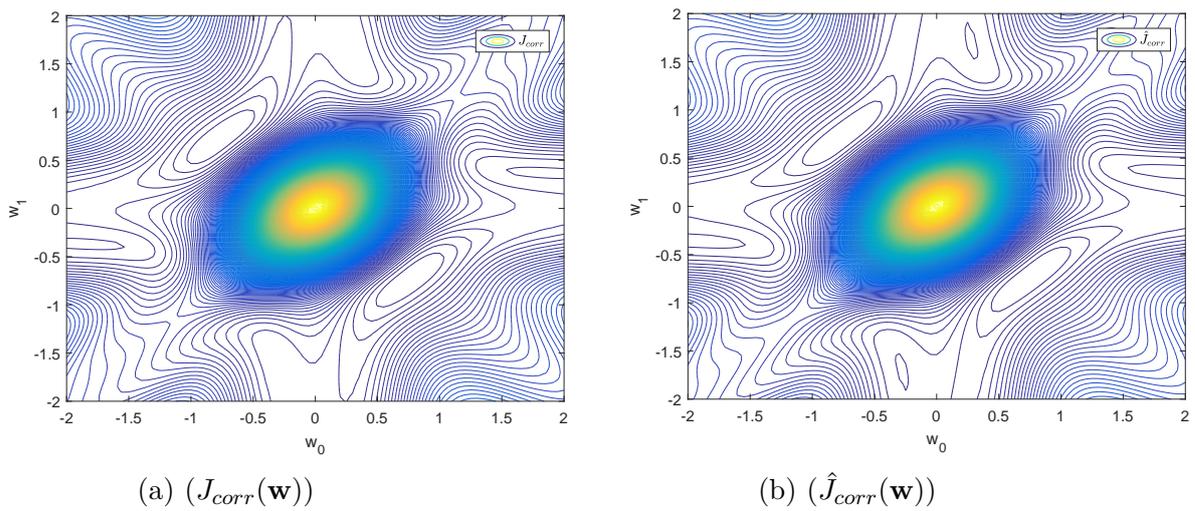


Figura 14 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,9$.

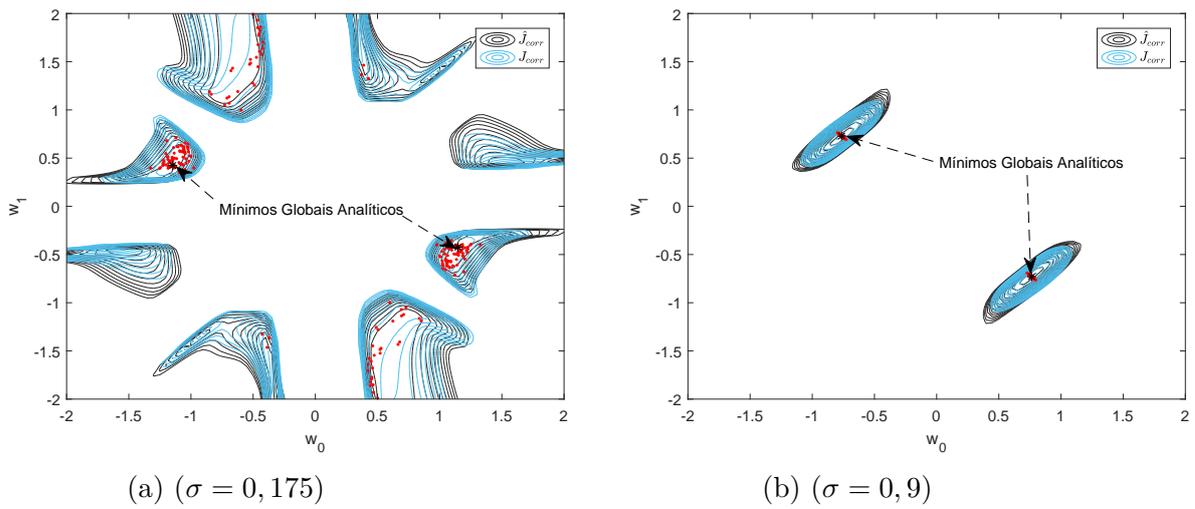


Figura 15 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$.

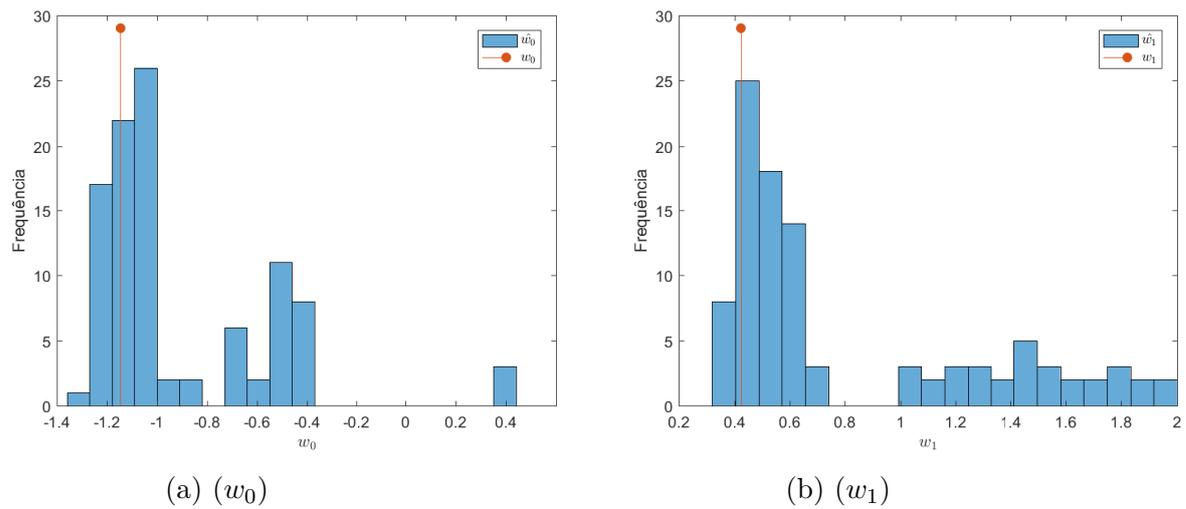


Figura 16 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,175$.

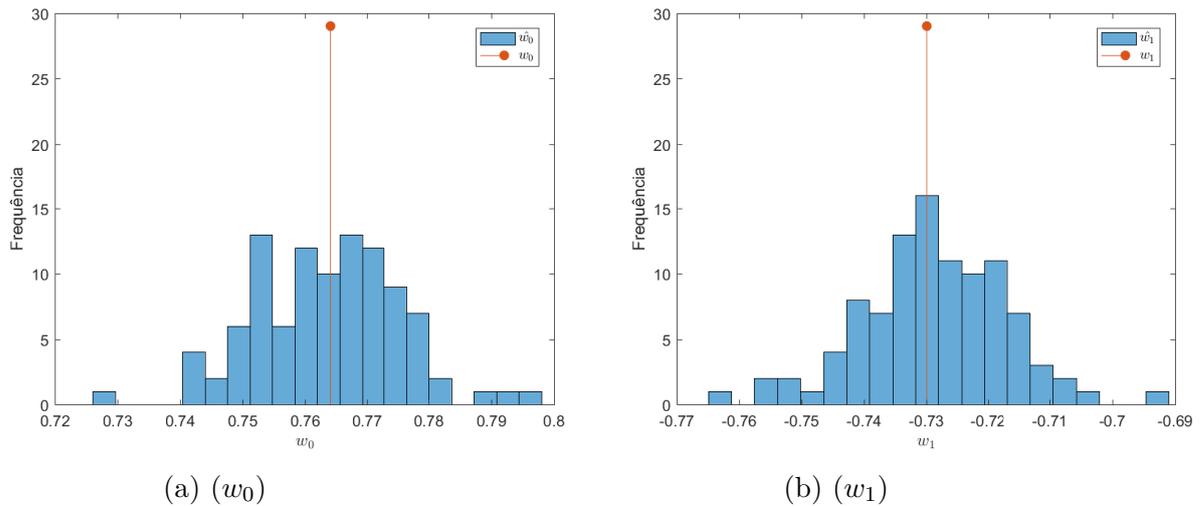


Figura 17 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,9$.

Tabela 1 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,175$	$\sigma = 0,9$
Estimada	-8,9518	-3,6213
Análítica	-12,9129	-3,6109

É possível observar que, em ambos os casos, as curvas de nível das superfícies das funções custo analítica e estimada são, de certo modo, similares. Porém as mudanças no tamanho do *kernel* produzem uma diferença significativa no formato das superfícies. Os pontos em vermelho exibidos na Figura 15 correspondem à solução ótima para a função custo estimada quando $\sigma = 0,175$ para cada uma das simulações independentes e seu conjunto de amostras. É possível perceber que eles acabam se afastando consideravelmente do ponto ótimo da função custo analítica e isso sugere que, para o presente cenário, valores pequenos para o tamanho do *kernel* podem resultar em erros de estimação elevados.

Em contrapartida, quando o tamanho do kernel é elevado para $\sigma = 0,9$, as soluções ótimas obtidas para \hat{J}_{corr} ficam mais concentradas na vizinhança dos mínimos globais analíticos, como podemos perceber na Figura 15b. Os histogramas mostrados na Figura 17 também confirmam esta constatação. É possível observar a mudança do intervalo de incursão dos coeficientes do equalizador nos histogramas das Figuras 16 e 17.

Entretanto, o aumento da largura do *kernel* causa uma degradação no desempenho do equalizador, como podemos ver na Tabela 1, que apresenta os valores de ISI alcançados pelas soluções associadas com as funções custo analítica e estimada. Dessa maneira, mesmo que \hat{J}_{corr} forneça soluções que sejam mais parecidas com a analítica para $\sigma = 0,9$, o equalizador não é o melhor que poderia ser obtido com o emprego deste tipo

de critério, visto que um desempenho superior pode ser alcançado utilizando um *kernel* menor com J_{corr} .

Prosseguindo com o mesmo cenário, agora é analisada a influência do tamanho do *kernel* no desempenho em ISI associado a cada versão do critério baseado em correntropia. Variamos σ de 0,025 até 1 em passos de 0,025, considerando $B = 2$ e um total de $L = 5000$ amostras para o custo estimado. Para cada valor de σ percorrido, é mostrado na Figura 18 o valor da ISI associado com a melhor solução fornecida pela função custo analítica, que foi obtida por meio do algoritmo ED, assim como a ISI média para as soluções da função custo estimada. A área sombreada representa o intervalo de ISI considerando os valores máximo e mínimo associados ao conjunto de todas as soluções encontradas para \hat{J}_{corr} . Para efeitos de comparação, também mostramos o desempenho em ISI para a solução de Wiener para um sistema sem pré-codificador, considerando o melhor atraso de equalização, neste caso $\alpha = 0$ (HAYKIN, 2013). O emprego da solução de Wiener para um sistema sem pré-codificador ocorre porque ela representa o limitante de desempenho alcançável para cada uma das versões do critério baseado em correntropia. O seu emprego ficará melhor justificado na Seção 5.5.

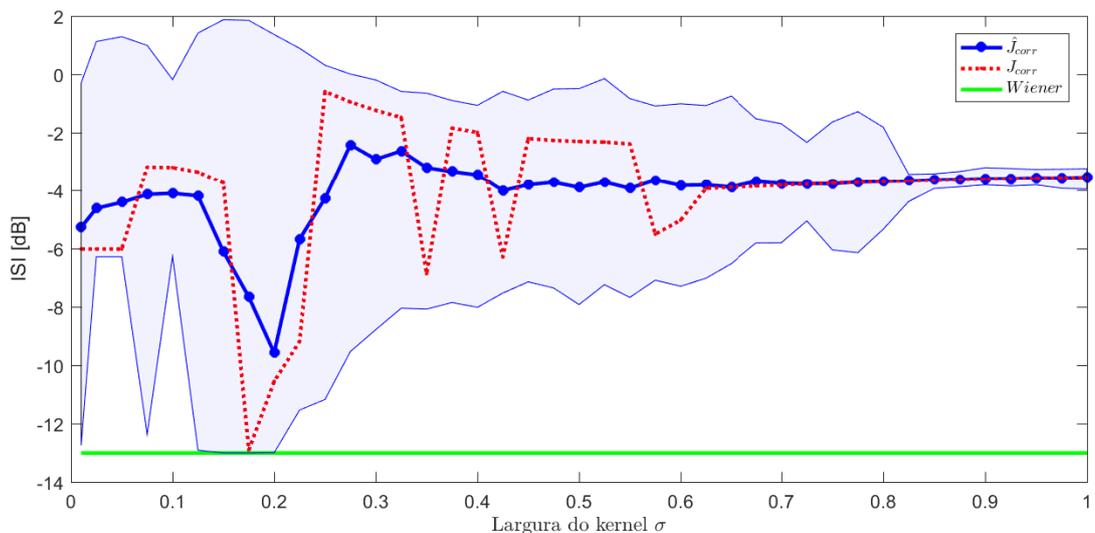


Figura 18 – ISI como uma função da largura do *kernel* σ .

É possível verificar na Figura 18 que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ podem diferir significativamente daqueles de $\hat{J}_{corr}(\mathbf{w})$ para $\sigma \leq 0,6$. Para $\sigma > 0,6$, os desempenhos em ISI são similares, o que sugere que as soluções são próximas para os casos estimado e analítico, como já fora verificado na Figura 15b. Entretanto, os valores de ISI associados a estas soluções são elevados se comparados à mínima ISI atingida pelo critério em sua versão analítica, o que ocorreu na vizinhança de $\sigma = 0,2$. O melhor caso para a versão estimada do critério de correntropia também se deu na vizinhança de $\sigma = 0,2$. Entretanto, como mostrado na Figura 15a, as soluções associadas a

$\hat{J}_{corr}(\mathbf{w})$ podem variar de maneira significativa, assim como pode também ser observado na Figura 18 pela área sombreada.

De uma maneira particular, o fato da ISI associada com $\hat{J}_{corr}(\mathbf{w})$ possuir maior variabilidade para valores pequenos de σ sugere que a estimação pode não ser precisa o suficiente para o número de amostras L escolhido. Sendo assim, avaliou-se o efeito de L no critério estimado variando ele de 1000 para 40000 em passos de 1000. Durante essa análise, foram mantidos fixos os valores $B = 2$ e $\sigma = 0,175$ para $J_{corr}(\mathbf{w})$ e para $\hat{J}_{corr}(\mathbf{w})$. Os valores de desempenho em ISI são mostrados na Figura 19 na qual a área sombreada representa o desvio padrão das soluções para $\hat{J}_{corr}(\mathbf{w})$ e a ISI obtida para a solução de Wiener é também indicada. É possível observar que, à medida que L aumenta, o desempenho médio em termos de ISI obtida com $\hat{J}_{corr}(\mathbf{w})$ se aproxima do desempenho correspondente à $J_{corr}(\mathbf{w})$. Entretanto, a grande quantidade de amostras necessárias para se obter um desempenho em ISI para $\hat{J}_{corr}(\mathbf{w})$ mais próximo do desempenho em ISI para $J_{corr}(\mathbf{w})$ pode se tornar um fator de preocupação, dependendo do tipo de aplicação. Por outro lado, o emprego de tamanhos de *kernel* σ maiores pode requerer um número menor de amostras para uma estimação adequada. Contudo, ele pode também conduzir a desempenhos inferiores, como mostrado na Figura 18.

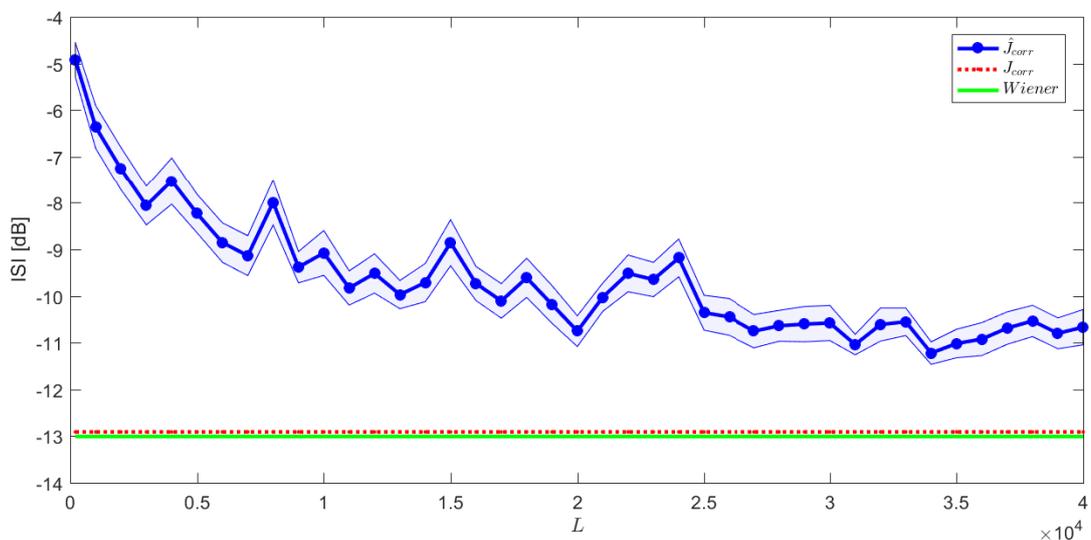


Figura 19 – ISI como uma função do número de amostras L .

Finalmente, foi verificado o efeito do número de atrasos B para $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$, quando $\sigma = 0,175$ e $L = 20000$ amostras. A Figura 20 mostra o perfil de correntropia da fonte - composta pelos valores de correntropia que o critério explorado busca casar - para os atrasos de tempo m de 1 até 10, juntamente com o perfil de correntropia na saída do equalizador para ambas as versões analítica e estimada (para uma média de 100 simulações de Monte Carlo), considerando $B = 2$ e $B = 10$. Os valores de ISI alcançados pelas soluções associadas com as funções custo analítica e estimada para ambos os atrasos

considerados pelo critério são apresentados na Tabela 2.

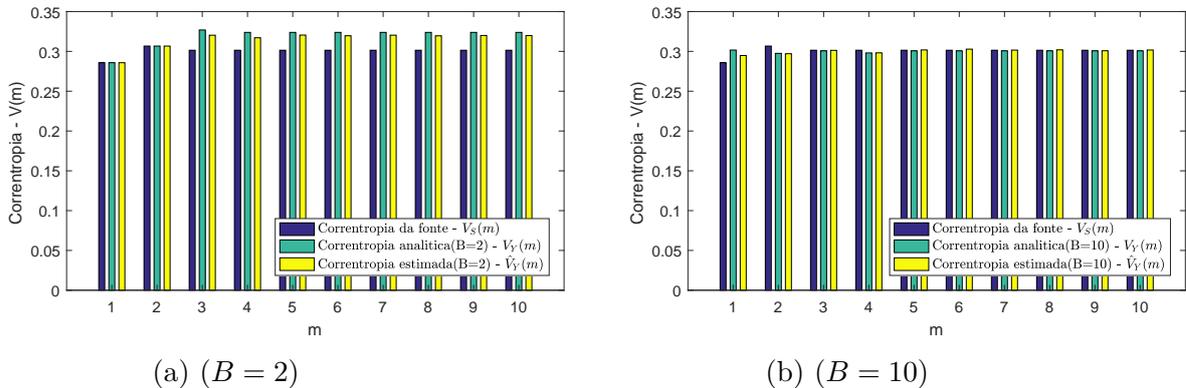


Figura 20 – Perfil de correntropia como uma função do atraso m .

Tabela 2 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Atrasos	Estimada	Análítica
$B = 2$	-9,6852	-10,5571
$B = 10$	-2,4519	-13,0102

Como pode ser observado na Figura 20a, tanto os valores da correntropia analítica (V_Y) quanto os valores da correntropia estimada (\hat{V}_Y), ambos na saída do equalizador, casam com os valores da correntropia na fonte (V_S) para $m = 1$ e $m = 2$, os quais são os atrasos que o critério explorado busca efetivamente igualar. Entretanto, para os atrasos m subsequentes, os valores de V_Y e \hat{V}_Y não casam perfeitamente com V_S e as soluções da função custo estimada ($\hat{J}_{corr}(\mathbf{w})$) apresentam um desempenho inferior por causa de erros de estimação, como mostrado na Tabela 2. Quando B aumenta para 10, vide Figura 20b, os perfis de correntropia não se casam perfeitamente para o caso analítico e nem para o caso estimado, mas, mesmo assim, a solução associada à função custo analítica ($J_{corr}(\mathbf{w})$) aponta para um patamar menor em termos de ISI, enquanto a versão estimada da correntropia perde bastante em desempenho, muito provavelmente devido ao acúmulo de erros de estimação ao longo dos vários atrasos considerados (B). Esse resultado mostra que empregar um número maior de atrasos no critério de casamento para $\hat{J}_{corr}(\mathbf{w})$ pode levar a desempenhos inferiores em termos de ISI.

5.2.2 Canal de fase máxima

No segundo cenário abordado, o canal é um sistema de fase máxima com função de transferência dada por $H(z) = 1 + 1,5z^{-1}$. Um equalizador com dois coeficientes $\mathbf{w} = [w_0 \ w_1]^T$ é adotado e os parâmetros do algoritmo ED foram $N_P = 100$, $F = 0,5$, $CR = 0,9$ e $I_T = 100$ iterações. Primeiramente, analisamos o efeito do tamanho do *kernel*

σ nas superfícies. Fixamos o número de atrasos considerados no critério em $B = 2$ para ambas as funções custo - analítica e estimada - e empregamos $L = 5000$ para a versão estimada. Assim, variando os coeficientes do equalizador, foram obtidas as superfícies de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,225$ e $\sigma = 0,95$, como mostrado nas Figuras 21 e 22, respectivamente. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. Foram obtidas também as curvas de nível de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,225$ e $\sigma = 0,95$, como mostrado nas Figuras 23 e 24, respectivamente. Adicionalmente, também mostramos nas Figuras 25a e 25b, as soluções encontradas pelo algoritmo ED, que são representadas por asteriscos (*) para o custo analítico e por pontos (·) para a função custo estimada, considerando $N_E = 100$ experimentos independentes. Finalmente, são apresentados os histogramas dos coeficientes do equalizador para a versão estimada do critério quando $\sigma = 0,225$ e $\sigma = 0,95$, nas Figuras 26 e 27, respectivamente.

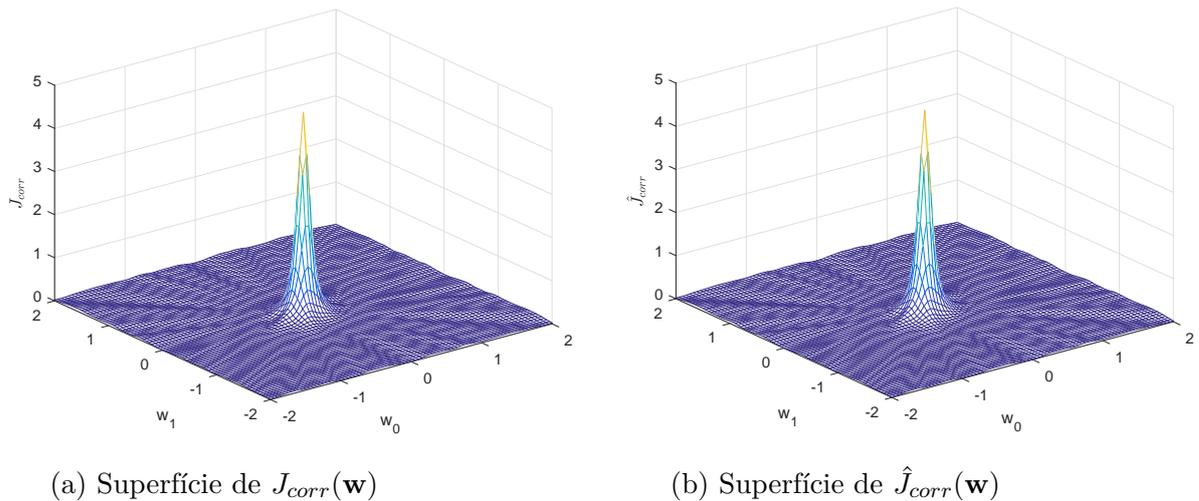
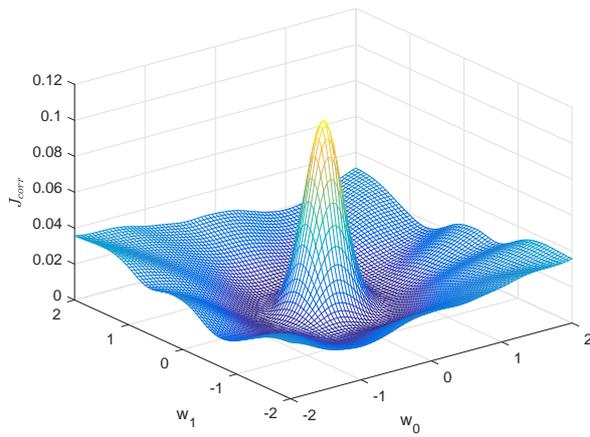
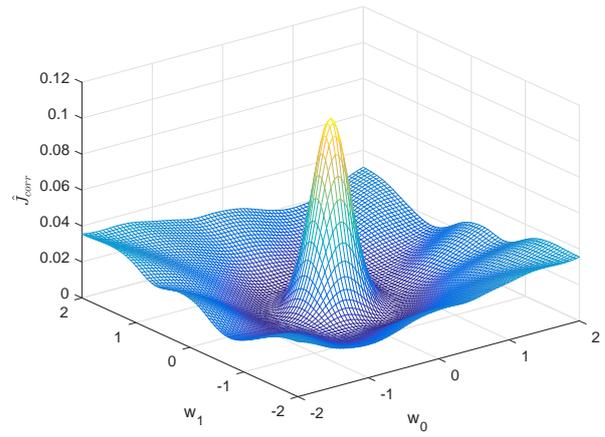
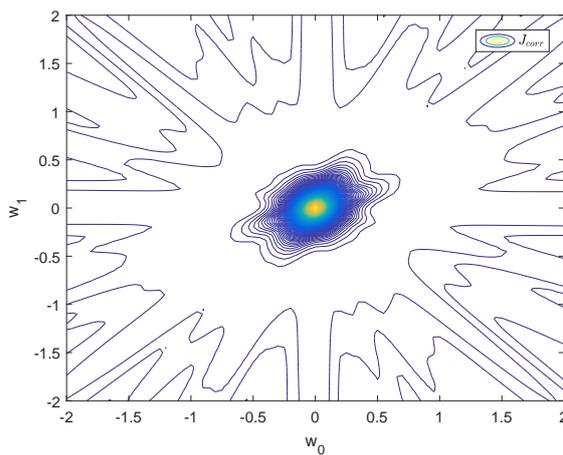
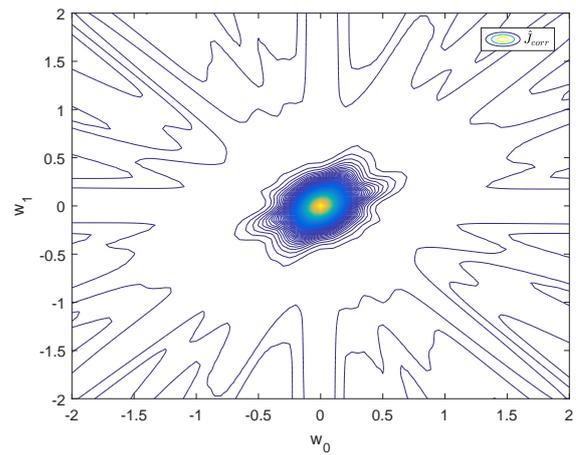


Figura 21 – Superfícies das funções custo analítica e estimada para $\sigma = 0,225$.

(a) Superfície de $J_{corr}(\mathbf{w})$ (b) Superfície de $\hat{J}_{corr}(\mathbf{w})$ Figura 22 – Superfícies das funções custo analítica e estimada para $\sigma = 0,95$.(a) ($J_{corr}(\mathbf{w})$)(b) ($\hat{J}_{corr}(\mathbf{w})$)Figura 23 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,225$.

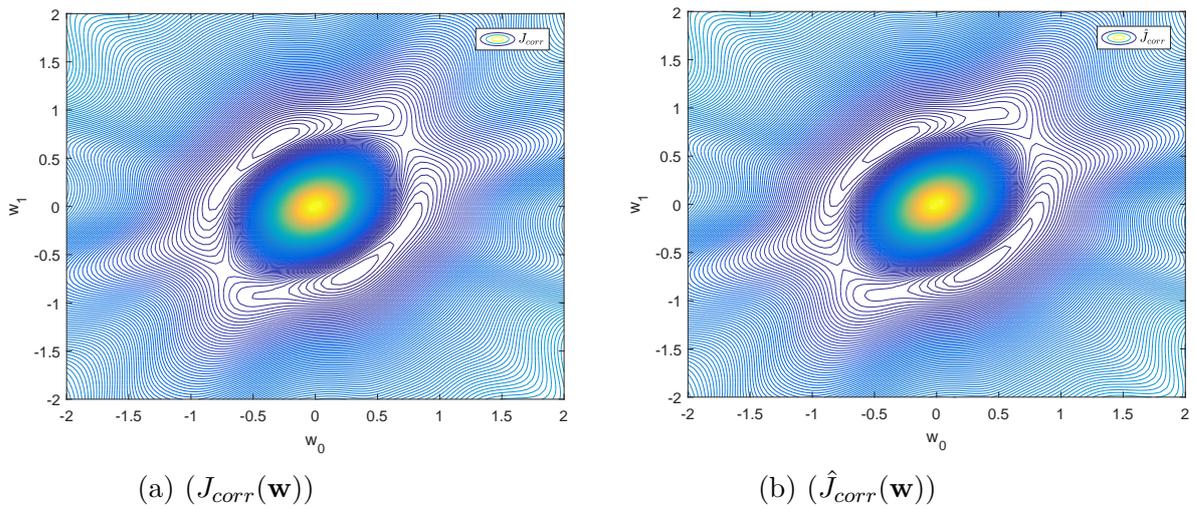


Figura 24 – Curvas de nível da superfície das funções custo analítica e estimada para $\sigma = 0,9$.

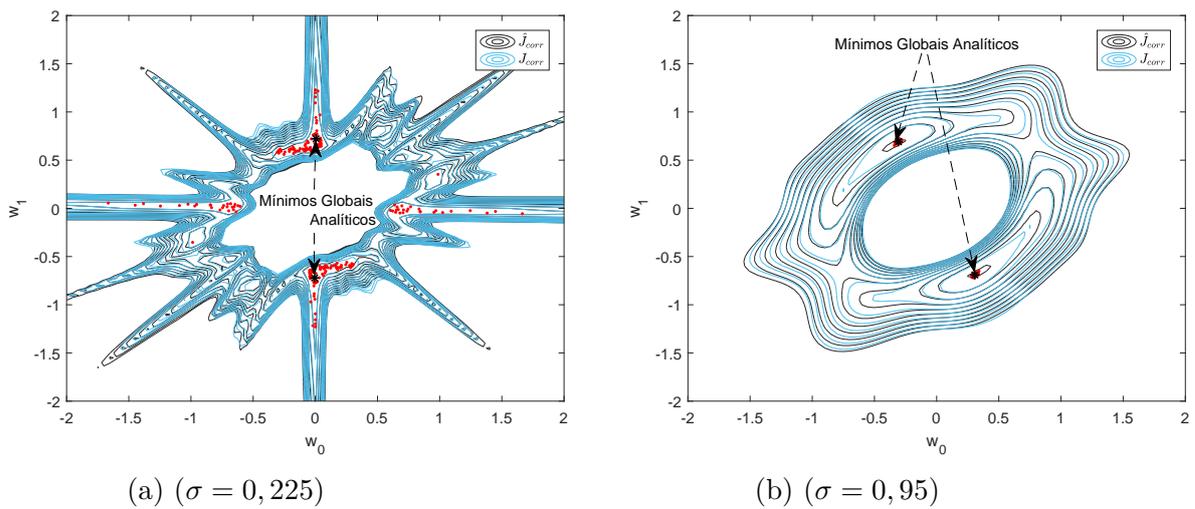


Figura 25 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$.

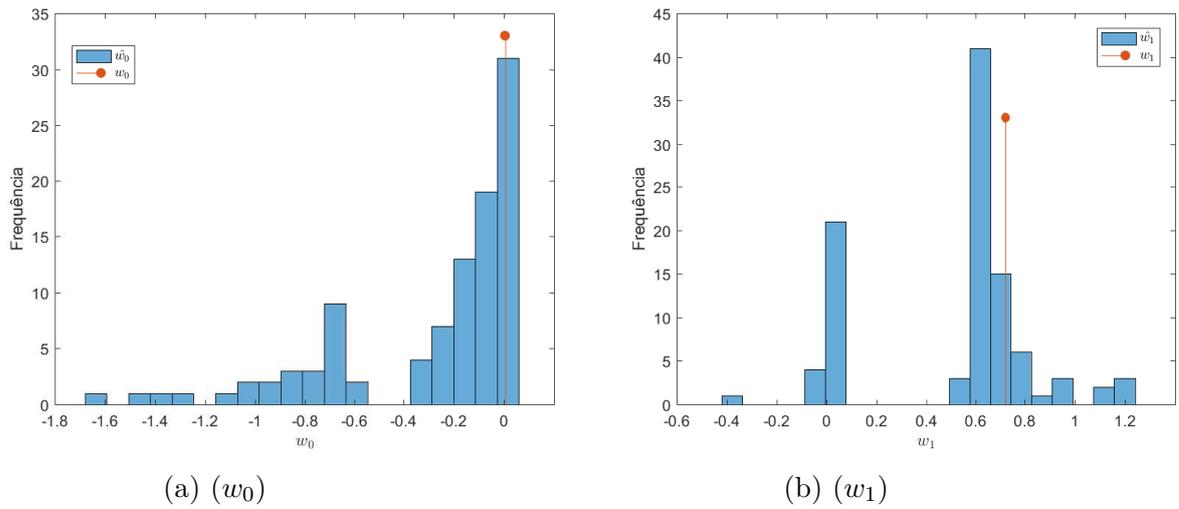


Figura 26 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,225$.

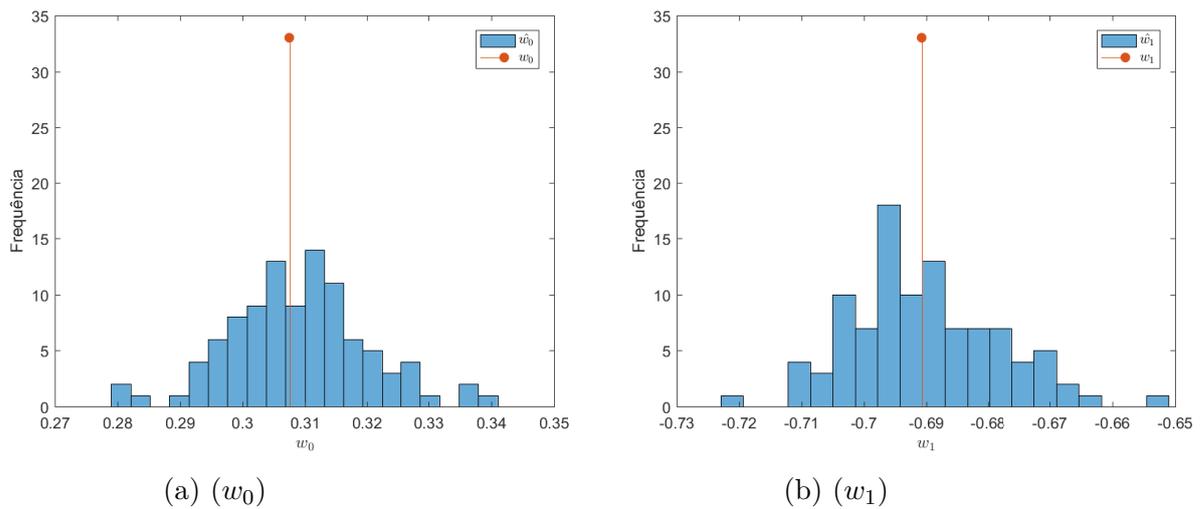


Figura 27 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,95$.

Tabela 3 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,225$	$\sigma = 0,95$
Estimada	-4,6901	-8,6105
Análítica	-3,4197	-8,6286

Analisando as Figuras 21 a 27, percebemos um comportamento semelhante àquele observado no cenário anterior: quando σ é pequeno, as soluções que minimizam \hat{J}_{corr} podem se afastar de forma considerável da solução ótima analítica; quando, porém, aumentamos o valor de σ , a distribuição das soluções de \hat{J}_{corr} fica mais concentrada em

uma vizinhança pequena em torno dos mínimos globais analíticos, de modo que as respectivas soluções são bastante parecidas. Interessantemente, ao contrário do que ocorreu no cenário anterior, podemos ver na Tabela 3 que o desempenho dos equalizadores melhora conforme a largura do *kernel* aumenta. Chamamos a atenção do leitor para a mudança do intervalo de incursão dos coeficientes do equalizador nos histogramas das Figuras 26 e 27.

Prosseguindo com este cenário, agora é analisada a influência do tamanho do *kernel* no desempenho em ISI associado a cada versão do critério baseado em correntropia. Variou-se σ de 0,025 para 1 em passos de 0,025, considerando $B = 2$ e um total de $L = 5000$ amostras para o custo estimado. Para cada valor de σ , é mostrado na Figura 28 o valor da ISI associado com a melhor solução fornecida pela função custo analítica, que foi obtida por meio do algoritmo ED, similar ao que foi mostrado na Figura 18. Para efeitos de comparação, também é mostrado o desempenho em ISI para a solução de Wiener para um sistema sem pré-codificador, considerando o melhor atraso de equalização (neste caso, $\alpha = 2$) (HAYKIN, 2013).

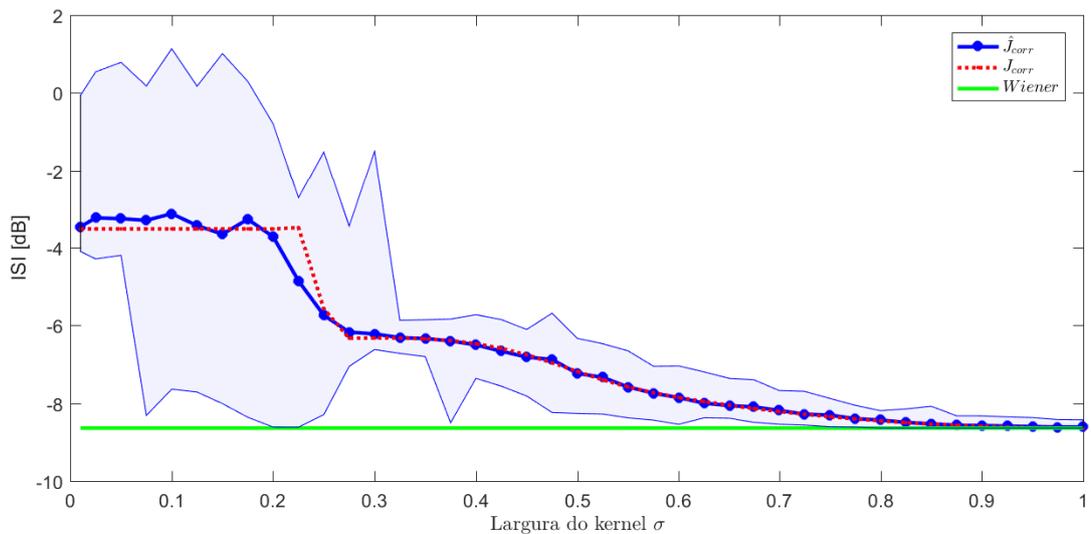


Figura 28 – ISI como uma função da largura do *kernel* σ .

É possível verificar na Figura 28 que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ se aproximam significativamente daqueles de $\hat{J}_{corr}(\mathbf{w})$ para o intervalo de valores de σ percorridos. Esse resultado sugere, para este cenário, que as soluções obtidas são muito próximas para os casos estimado e analítico. Vale mencionar também que o melhor desempenho em ISI é obtido para valores acima de $\sigma = 0,875$, para ambos os casos, que também alcançam o patamar de desempenho em ISI associado à solução de Wiener.

5.3 Fonte Gaussiana

Nesta seção, vamos investigar, por meio de simulações computacionais, o comportamento do critério de casamento de correntropia no contexto de uma fonte Gaussiana, tanto em sua versão estimada, quanto em sua versão teórica. O pré-codificador utilizado no cenário analisado é o mesmo que fora empregado na Seção 5.2, cuja função de transferência é $P(z) = 1 - 0,5z^{-1} + 0,3z^{-2}$.

Neste cenário, a expressão da correntropia analítica da saída do equalizador, denotada por $V_Y(m)$, é dada pela expressão em (4.2), que é reproduzida abaixo:

$$\begin{aligned} V_Y(m) &= \int f_{Y'}(\mathbf{v})G_{\sigma^2}(\mathbf{v})d\mathbf{v} \\ &= \int G_{\sigma_{Y'}^2}(\mathbf{v})G_{\sigma^2}(\mathbf{v})d\mathbf{v} = G_{\sigma_{Y'}^2 + \sigma^2}(0). \end{aligned} \quad (5.8)$$

onde $\sigma_{Y'}^2$ é a variância do sinal gaussiano $y'(n)$ produzido pelo efeito do pré-codificador modificado com função de transferência $Q'(z) = P(z)H(z)W(z)(1 - 1z^{-m})$, isso para um sinal gaussiano branco filtrado.

Tanto para a versão estimada quanto para a versão analítica do critério de casamento de correntropia, o cálculo da correntropia na saída da fonte (V_S) é realizado de maneira analítica, cuja expressão também é reproduzida abaixo:

$$\begin{aligned} V_S(m) &= \int f_{S'}(\mathbf{v})G_{\sigma^2}(\mathbf{v})d\mathbf{v} \\ &= \int G_{\sigma_{S'}^2}(\mathbf{v})G_{\sigma^2}(\mathbf{v})d\mathbf{v} = G_{\sigma_{S'}^2 + \sigma^2}(0). \end{aligned} \quad (5.9)$$

onde $\sigma_{S'}^2$ é a variância do sinal gaussiano $s'(n)$ produzido pelo efeito do pré-codificador modificado com função de transferência $Q'(z) = P(z)(1 - 1z^{-m})$.

5.3.1 Canal de fase mínima

No cenário considerado, o canal é um sistema de fase mínima com função de transferência dada por $H(z) = 1 + 0,5z^{-1}$. Um equalizador com dois coeficientes $\mathbf{w} = [w_0 \ w_1]^T$ é adotado, cujos pesos irão variar no intervalo de -2 a 2 para se obter os contornos das funções custo $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$. Os parâmetros utilizados no algoritmo ED foram $N_P = 100$, $F = 0,5$, $CR = 0,9$ e $I_T = 100$ iterações.

Primeiramente, analisamos o efeito do tamanho do *kernel* σ nas superfícies. Considerando $B = 2$ e $L = 5000$, foram obtidas as superfícies das funções custo $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,05$ e $\sigma = 2,2$, as quais são exibidas nas Figuras 29 e 30, respectivamente. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. Por sua vez, as Figuras 31 e 32 trazem as correspondentes curvas de nível de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$. Adicionalmente, também mostramos nas

Figuras 33a e 33b, as soluções encontradas pelo algoritmo ED, que são representadas por asteriscos (*) para o custo analítico e por pontos (.) para a função custo estimada, considerando $N_E = 100$ experimentos independentes. Por fim, apresentamos nas Figuras 34 e 35 os histogramas dos coeficientes w_0 e w_1 do equalizador obtidos nos N_E experimentos independentes envolvendo a função custo estimada, tanto para $\sigma = 0,05$ quanto para $\sigma = 2,2$.

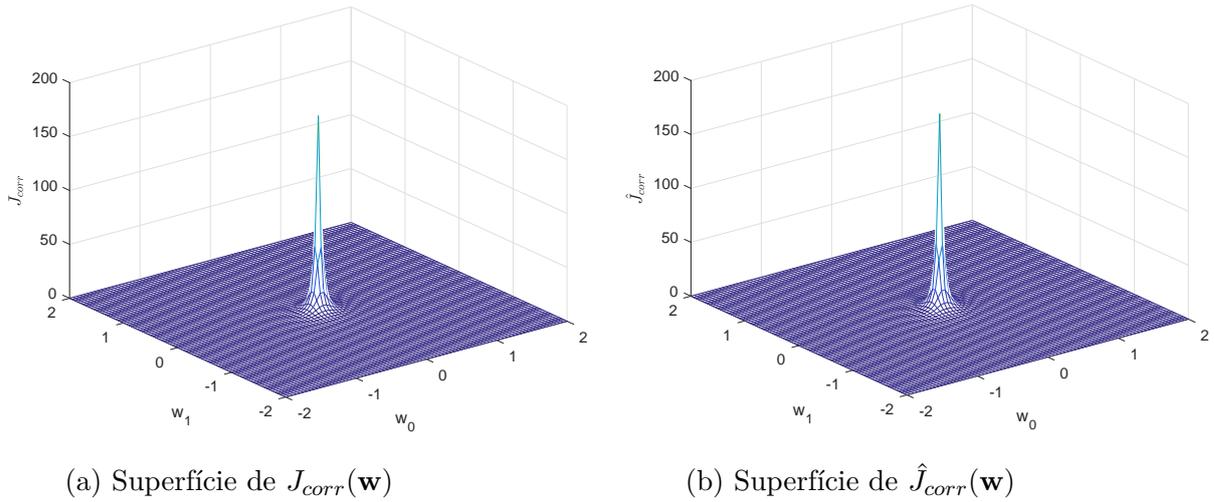


Figura 29 – Superfícies das funções custo analítica e estimada para $\sigma = 0,05$.

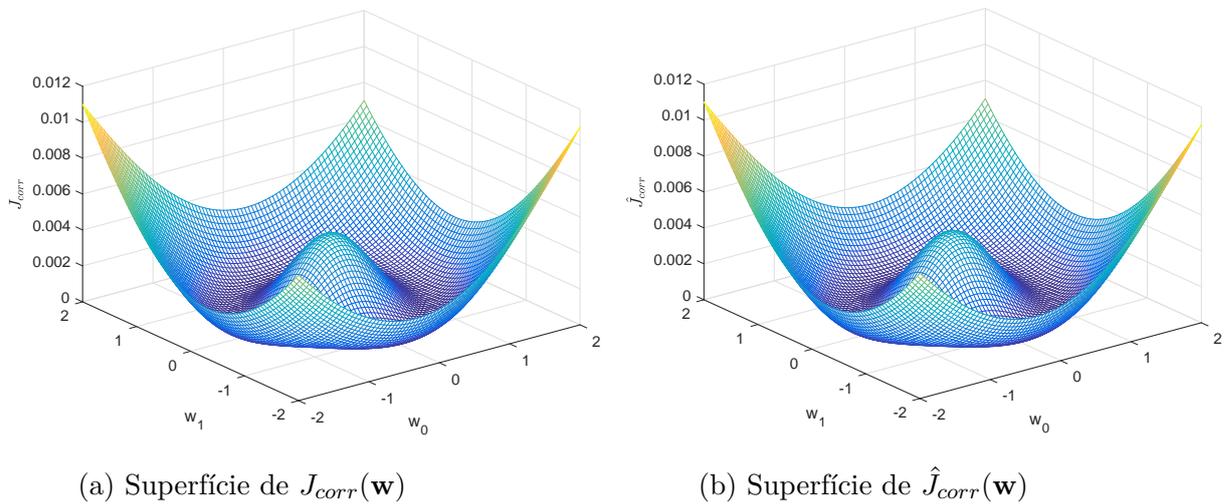


Figura 30 – Superfícies das funções custo analítica e estimada para $\sigma = 2,2$.

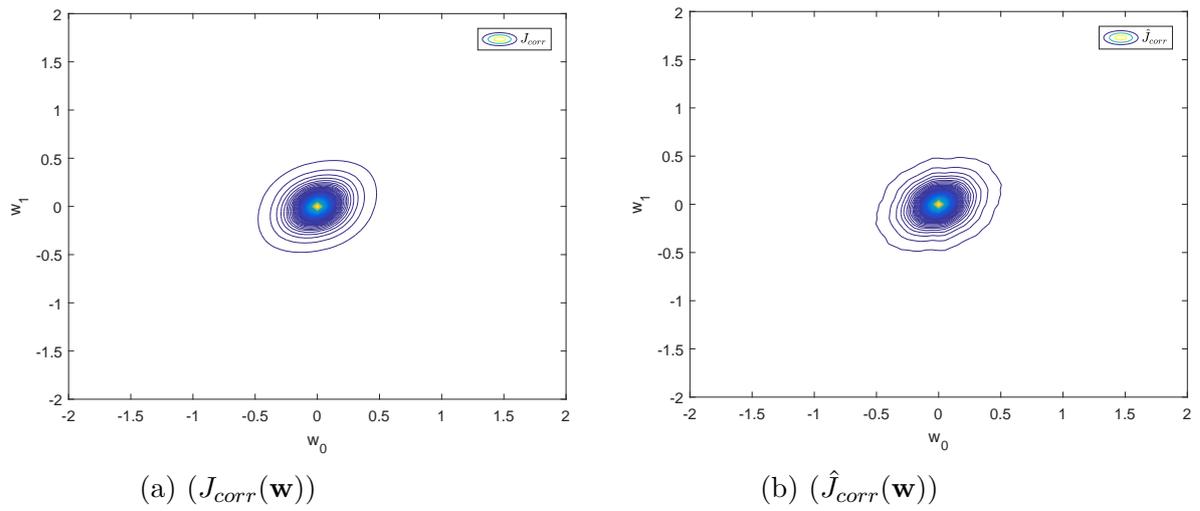


Figura 31 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,05$.

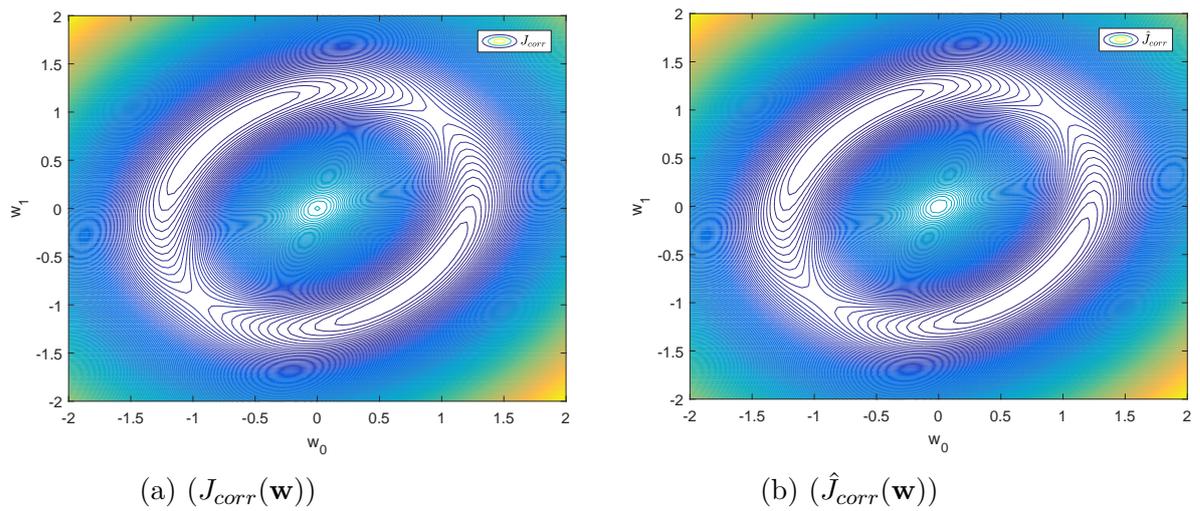


Figura 32 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 2,2$.

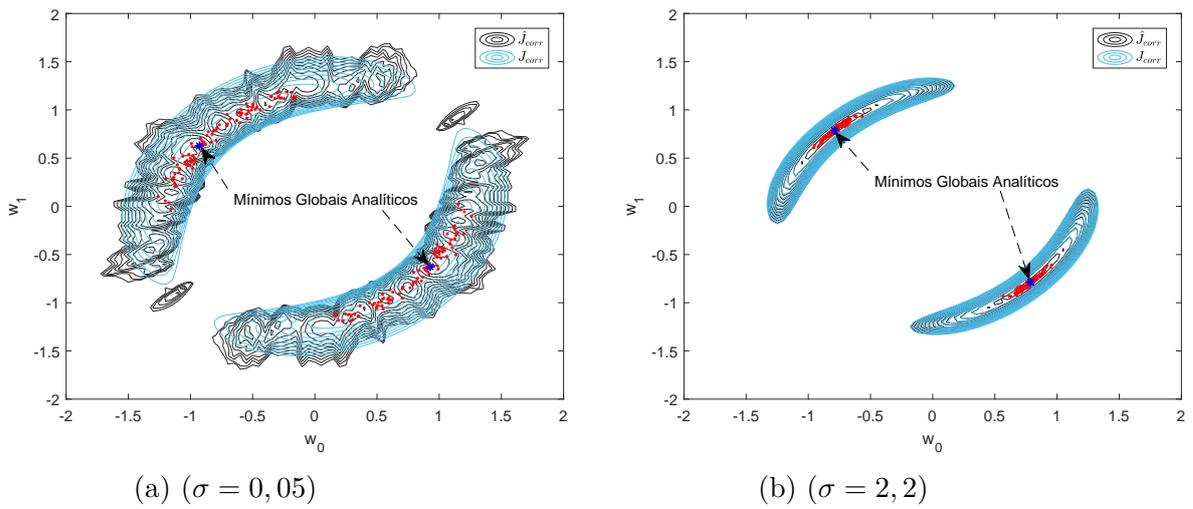


Figura 33 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$.

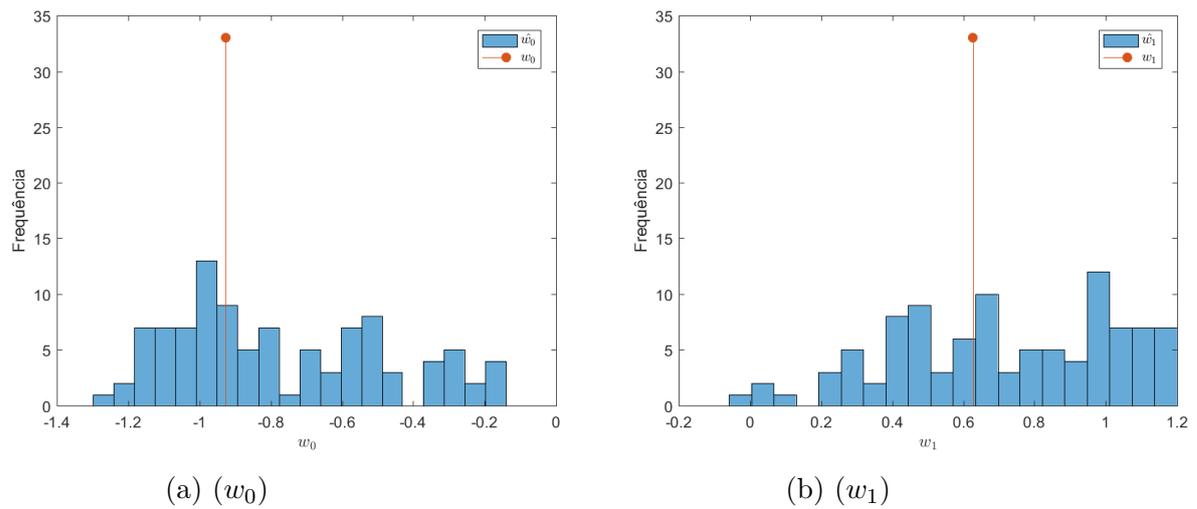


Figura 34 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,05$.

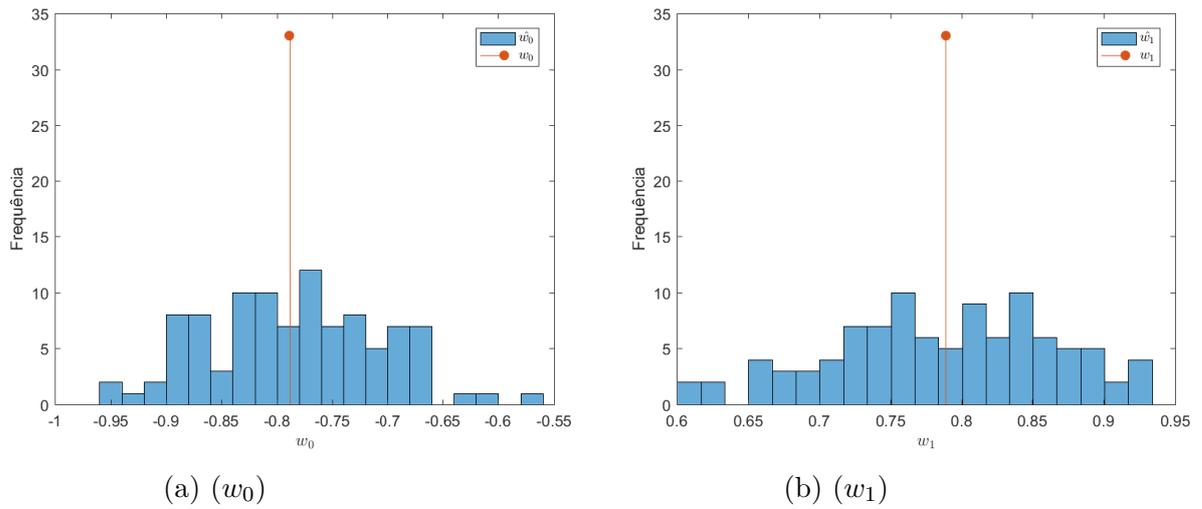


Figura 35 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 2, 2$.

Tabela 4 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,05$	$\sigma = 2, 2$
Estimada	-6,0121	-3,0379
Analítica	-8,3983	-3.0103

Como verificado anteriormente para a fonte BPSK, é possível observar que, em ambos os casos, as curvas de nível das superfícies das funções custo analítica e estimada são, de certo modo, similares, porém as mudanças no tamanho do *kernel* produzem uma diferença significativa em seus formatos. Além disso, o afastamento entre os coeficientes ótimos do equalizador para a versão estimada e para a versão analítica é reduzido conforme a largura do kernel aumenta. Porém, aumentar σ faz com que o desempenho alcançado pelo equalizador, em termos de ISI, sofra uma degradação, conforme mostra a Tabela 4.

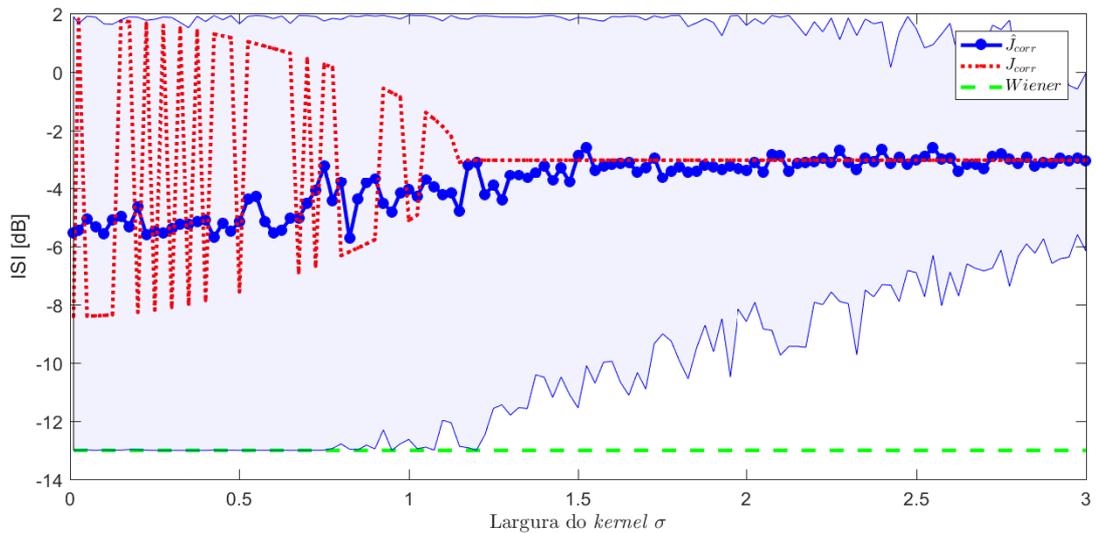


Figura 36 – ISI como uma função da largura do *kernel* σ .

À semelhança do que foi feito na Seção 5.2, analisamos também a influência do tamanho do *kernel* sobre o desempenho, medido pela ISI, associado a cada versão do critério baseado em correntropia. Variou-se σ de 0,025 até 3 em passos de 0,025, considerando $B = 3$ e um total de $L = 10000$ amostras para o custo estimado. De modo semelhante ao que foi verificado para a fonte BPSK, é possível observar, na Figura 36, que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ podem diferir de forma expressiva daqueles de $\hat{J}_{corr}(\mathbf{w})$ para $\sigma \leq 1,5$. Vale mencionar ainda que o desempenho em termos de ISI para $J_{corr}(\mathbf{w})$ apresenta uma oscilação significativa para $\sigma < 1$, podendo ir do melhor para o pior desempenho em um pequeno intervalo de σ . Para efeitos de comparação, também mostramos o desempenho em ISI para a solução de Wiener para um sistema sem pré-codificador, considerando o melhor atraso de equalização (neste caso, $\alpha = 0$) (HAYKIN, 2013).

5.3.2 Canal de fase máxima

O segundo caso estudado envolve o caso de fase máxima considerado na Seção 5.2.2, cuja função de transferência é $H(z) = 1 + 1,5z^{-1}$. Um equalizador com dois coeficientes $\mathbf{w} = [w_0 \ w_1]^T$ é adotado e os parâmetros do algoritmo ED foram $N_P = 100$, $F = 0,5$, $CR = 0,9$ e $I_T = 100$ iterações.

As Figuras 37 e 38 mostram as superfícies das funções custo analítica e estimada do critério de correntropia considerando $B = 3$, $L = 5000$, e os valores de $\sigma = 0,25$ e $\sigma = 1,525$. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. As curvas de nível de J_{corr} e \hat{J}_{corr} para estes dois casos são apresentadas nas Figuras 39 e 40, respectivamente. Além disso, também mostramos nas Figuras 41a

e 41b, as soluções encontradas pelo algoritmo ED, que são representadas por asteriscos (*) para o custo analítico e por pontos (·) para a função custo estimada, considerando $N_E = 100$ experimentos independentes. Por fim, são apresentados os histogramas dos coeficientes do equalizador para a versão estimada do critério para $\sigma = 0,25$ e $\sigma = 1,525$, nas Figuras 42 e 43, respectivamente.

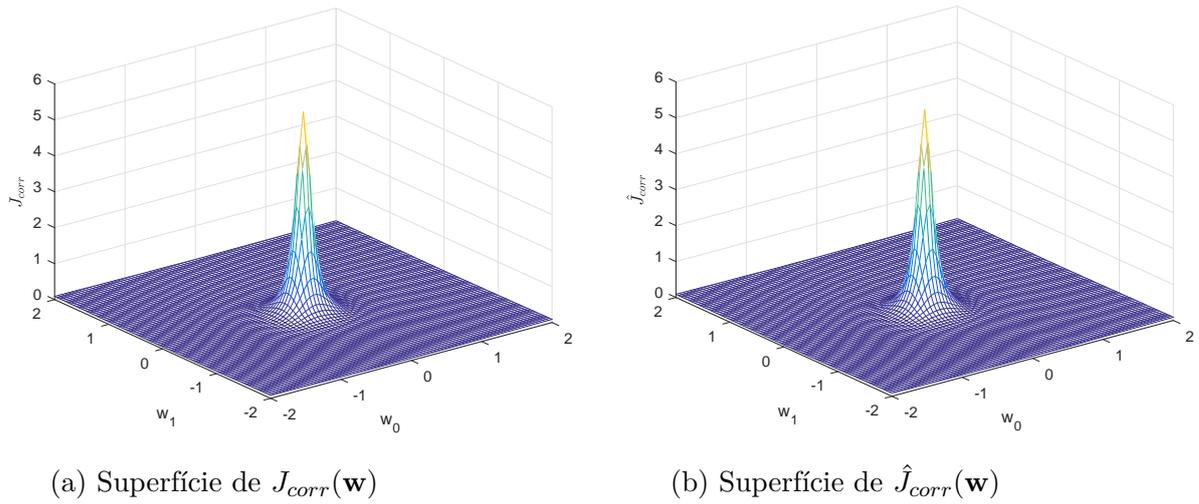


Figura 37 – Superfícies das funções custo analítica e estimada para $\sigma = 0,25$.

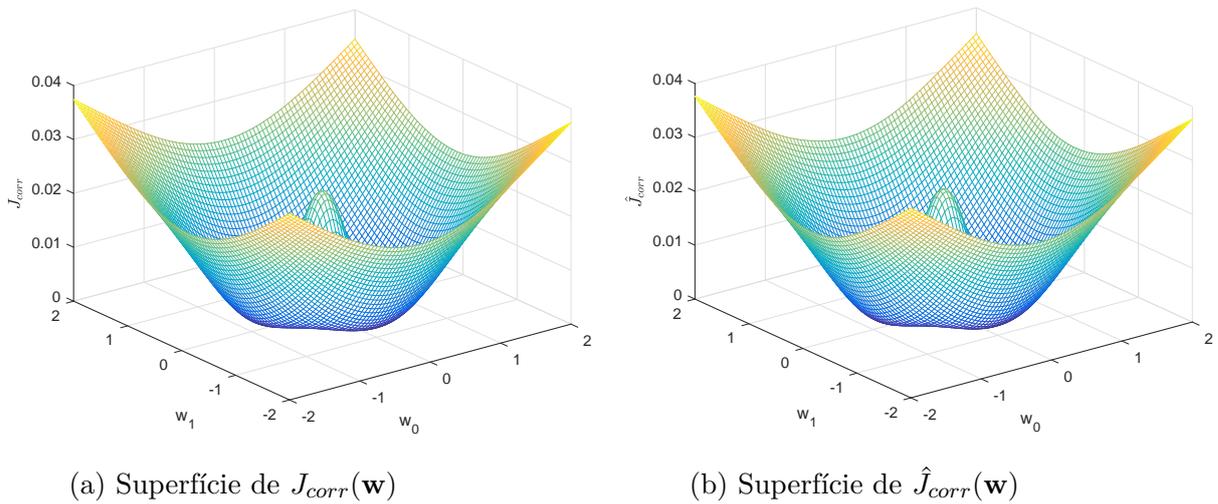


Figura 38 – Superfícies das funções custo analítica e estimada para $\sigma = 1,525$.

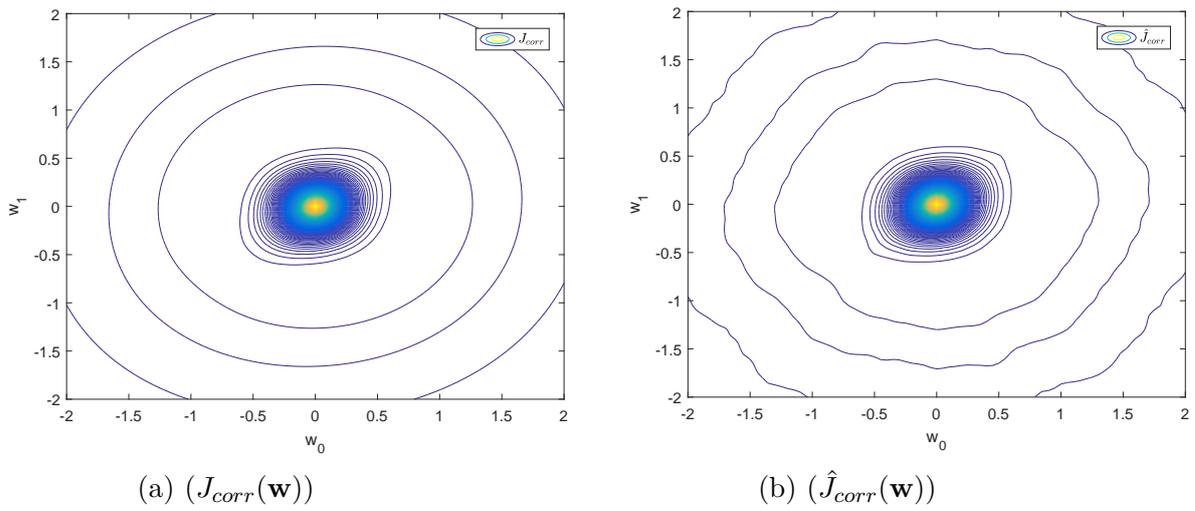


Figura 39 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,25$.

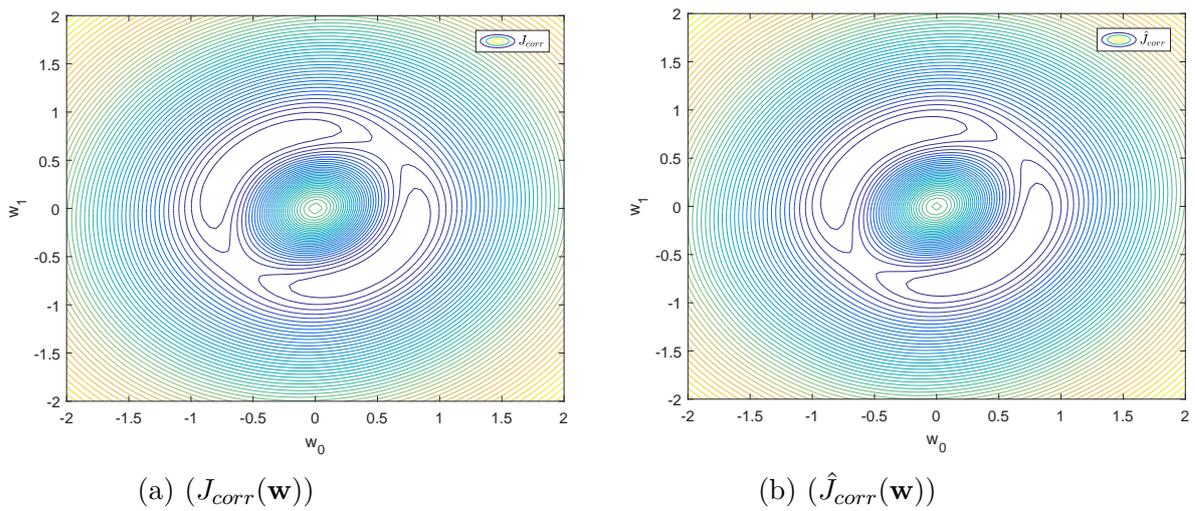


Figura 40 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 1,525$.

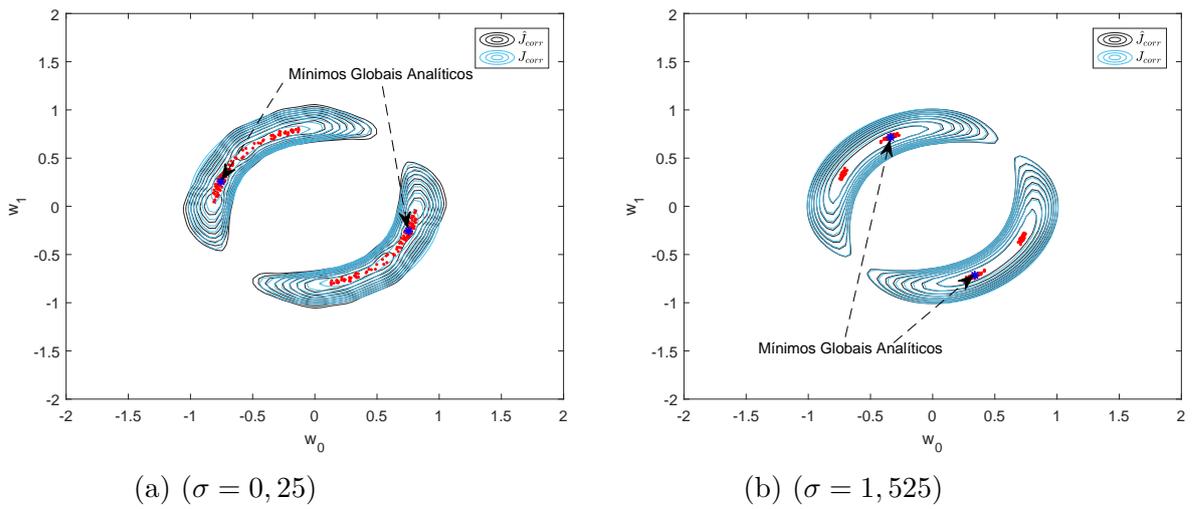


Figura 41 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$.

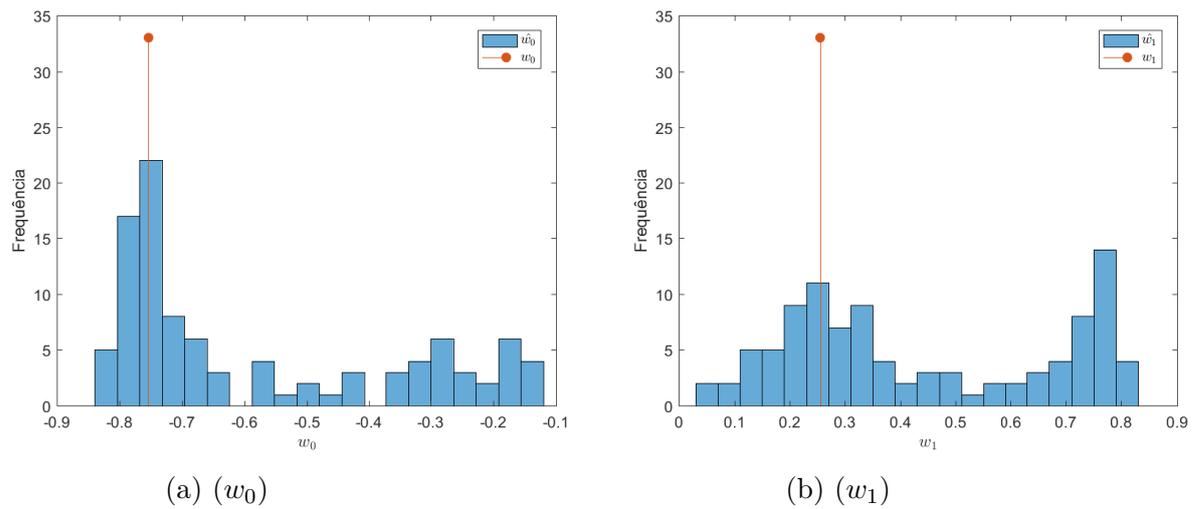


Figura 42 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,25$.

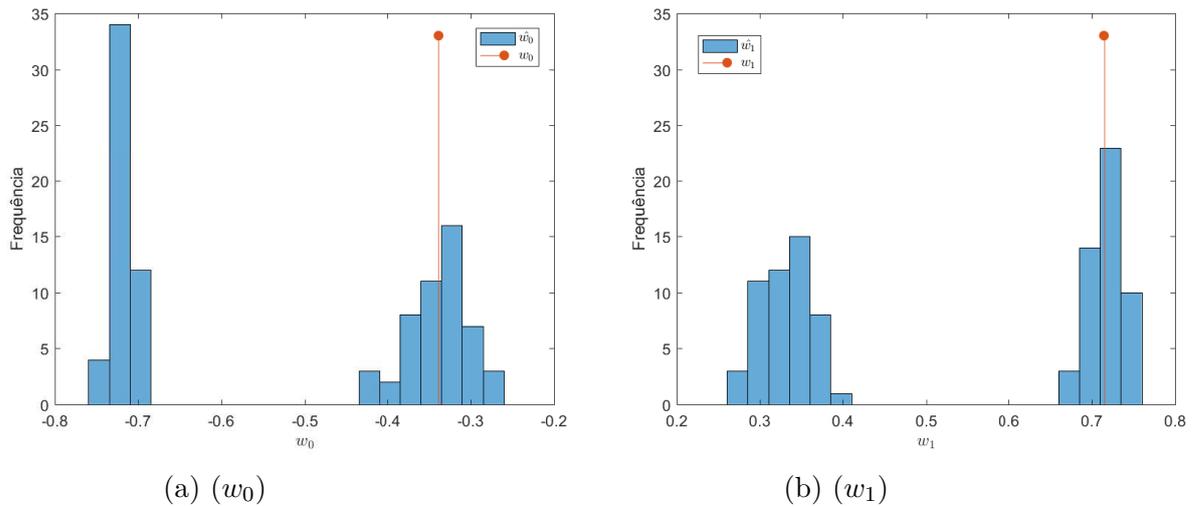


Figura 43 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 1,525$.

Tabela 5 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,25$	$\sigma = 1,525$
Estimada	-2,7560	-3,5807
Análítica	-0,2948	-8,6344

De modo geral, percebemos um comportamento das superfícies e das soluções ótimas muito semelhante àquele observado na Seção 5.2.2, com fonte BPSK e o mesmo canal. Em suma, é possível notar que as superfícies e as curvas de nível de J_{corr} e de \hat{J}_{corr} são relativamente parecidas para ambos os valores de σ . Porém, para $\sigma = 0,25$, as soluções ótimas de \hat{J}_{corr} acabam se espalhando para regiões um pouco mais afastadas em relação à solução ótima analítica (i.e., da função custo J_{corr}). Isto também é percebido quando verificamos o histograma da Figura 42.

Quando a largura do kernel é alterada para $\sigma = 1,525$, as soluções ótimas de \hat{J}_{corr} ficam um pouco mais concentradas em torno da solução analítica. Todavia, em várias realizações, também foram obtidas soluções ótimas em torno de outro ponto. Este segundo agrupamento também é percebido nos histogramas da Figura 43, algo que não havia ocorrido no caso da fonte BPSK, para este mesmo canal, quando um valor maior de σ foi utilizado.

Com relação ao desempenho dos equalizadores, a ISI associada às soluções estimadas e analíticas é reduzida (e, portanto, o equalizador se tornou mais bem-sucedido na tarefa de cancelar o canal) quando σ cresce, o que também foi observado para uma fonte BPSK na Seção 5.2.2. Seguindo o mesmo procedimento descrito na Seção 5.3.1, foram obtidas as curvas da ISI média associada às soluções ótimas da função custo estimada,

considerando $L = 10000$, e da ISI analítica, em função da largura do kernel. Estas curvas são apresentadas na Figura 44, juntamente com o desempenho em termos de ISI para a solução de Wiener do sistema sem pré-codificador, considerando o melhor atraso (neste caso, $\alpha = 2$).

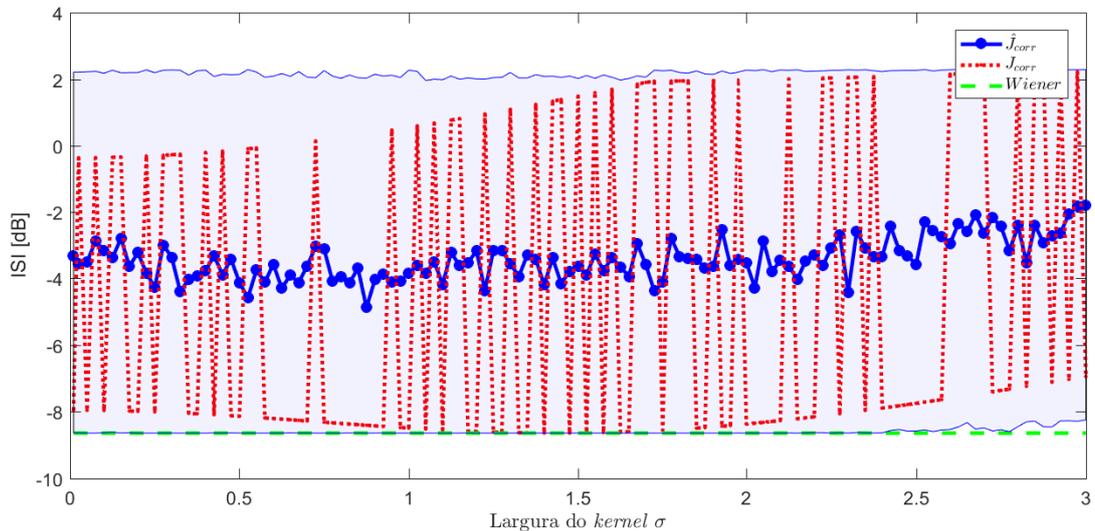


Figura 44 – ISI como uma função da largura do $kernel \sigma$.

É possível constatar na Figura 44 que os valores de ISI associados a $J_{corr}(w)$ diferem significativamente daqueles de $\hat{J}_{corr}(\mathbf{w})$ para todo o intervalo de valores de σ considerado. Vale mencionar ainda que o desempenho em termos de ISI para $J_{corr}(\mathbf{w})$ apresenta uma oscilação significativa, podendo ir do melhor para o pior desempenho em um pequeno intervalo de σ . O melhor desempenho em ISI é obtido para valores na vizinhança de $\sigma = 1,5$ para ambos os casos, sendo que $J_{corr}(\mathbf{w})$ foi capaz de alcançar o patamar de desempenho em ISI associado à solução de Wiener.

5.4 Fonte Contínua (com ruído)

Nesta seção, faremos a análise do critério baseado em correntropia considerando o modelo de fonte contínua apresentado na Seção 4.3, o qual representa uma generalização do caso BPSK. Seguiremos, aqui, os mesmos passos empregados nas Seções 5.2 e 5.3 para avaliar o critério, tanto na forma estimada quanto na forma analítica. Os dois casos tratados serão, novamente, aqueles já considerados nas análises anteriores: (1) canal de fase mínima e (2) canal de fase máxima. Mais uma vez, o pré-codificador utilizado possui função de transferência $P(z) = 1 - 0,5z^{-1} + 0,3z^{-2}$.

As expressões da correntropia analítica da saída do equalizador e da própria fonte são dadas pelas Equações (4.24) e (4.30), respectivamente. Para facilitar, reprodu-

zimos aqui ambas as expressões:

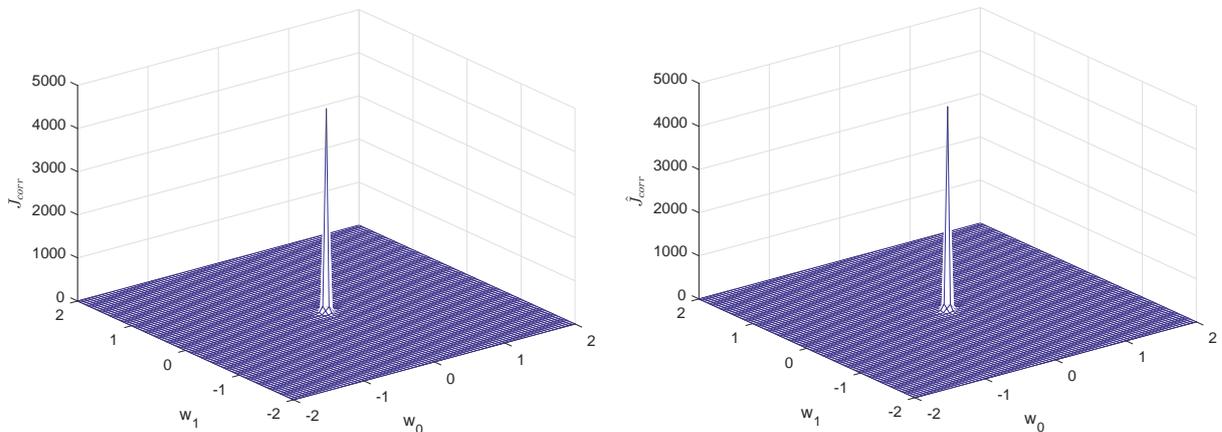
$$V_Y(e_m(n)) = \frac{1}{2^{K+Q+m-1}} \sum_{i=0}^{2^{K+Q+m-1}-1} G \sqrt{\sigma_k^2 + \sigma_{g_m}^2 + \sigma_{\eta_m}^2} (\mathbf{w}^T \mathbf{Q}_m^y \mathbf{u}'_{a(i)}(n)), \quad (5.10)$$

$$V_S(e_{S_m}(n)) = \frac{1}{2^{R+m}} \sum_{i=0}^{2^{R+m}-1} G \sqrt{\sigma_k^2 + \sigma_{p_m}^2} (\mathbf{p}_0^T \mathbf{u}'_p(n)). \quad (5.11)$$

Um fator importante que surge, porém, neste cenário é o ruído de canal. Agora, o sinal que chega ao equalizador também contém ruído aditivo branco e Gaussiano.

5.4.1 Canal de fase mínima

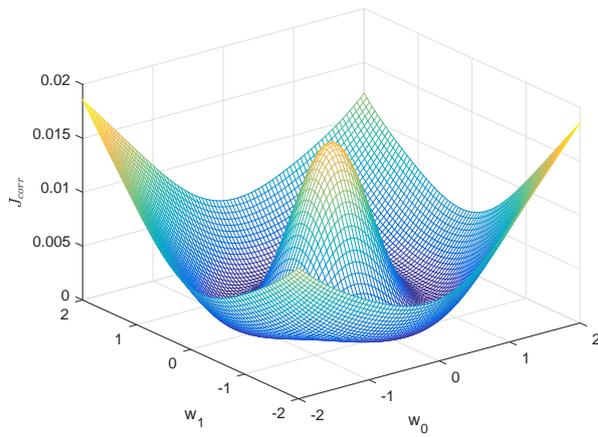
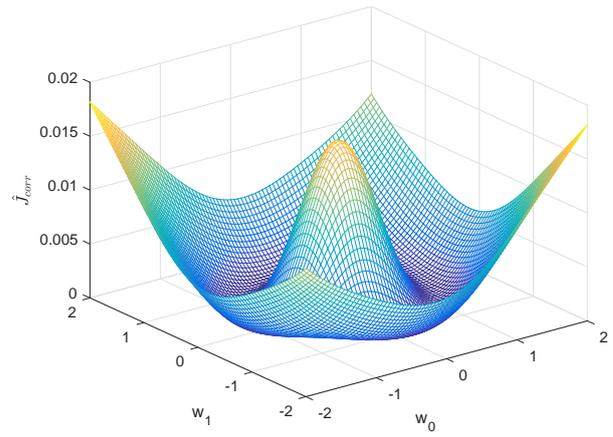
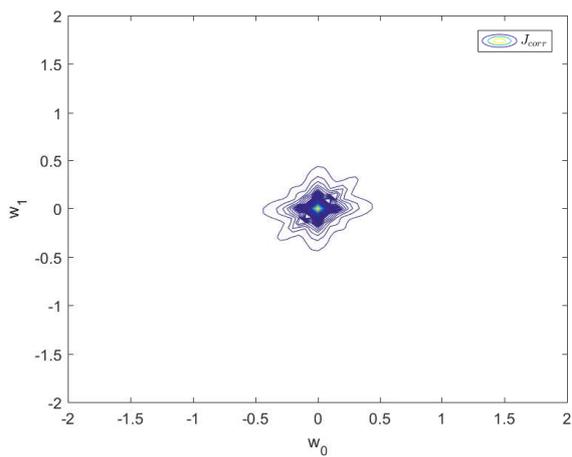
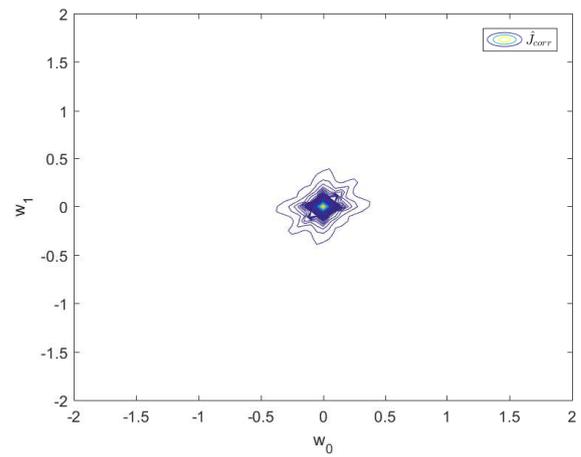
Primeiramente, foi analisado o efeito do tamanho do *kernel* σ nas superfícies. Fixou-se o número de atrasos considerados no critério em $B = 3$, a variância da fonte contínua $\sigma_g^2 = 0,1$ para ambas as funções custo - analítica e estimada - e adotou-se $L = 10000$ para a versão estimada. Assim, variando os coeficientes do equalizador, foram obtidas as superfícies de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,01$ e $\sigma = 1,75$, como mostrado nas Figuras 45 e 46, respectivamente, e para uma relação sinal-ruído (SNR, do inglês *signal-to-noise ratio*) de 30 dB. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. Foram obtidas também as curvas de nível da superfície de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os dois valores de σ considerados, como mostrado nas Figuras 47 e 48, respectivamente. Adicionalmente, também mostramos nas Figuras 49a e 49b, as soluções encontradas pelo algoritmo ED, que são representadas por asteriscos (*) para o custo analítico e por pontos (.) para a função custo estimada, considerando $N_E = 100$ experimentos independentes. Por fim, são apresentados os histogramas dos coeficientes do equalizador para a versão estimada do critério para $\sigma = 0,01$ e $\sigma = 1,75$, nas Figuras 50 e 51, respectivamente.



(a) Superfície de $J_{corr}(\mathbf{w})$

(b) Superfície de $\hat{J}_{corr}(\mathbf{w})$

Figura 45 – Superfícies das funções custo analítica e estimada para $\sigma = 0,01$.

(a) Superfície de $J_{corr}(\mathbf{w})$ (b) Superfície de $\hat{J}_{corr}(\mathbf{w})$ Figura 46 – Superfícies das funções custo analítica e estimada para $\sigma = 1,75$.(a) ($J_{corr}(\mathbf{w})$)(b) ($\hat{J}_{corr}(\mathbf{w})$)Figura 47 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,01$.

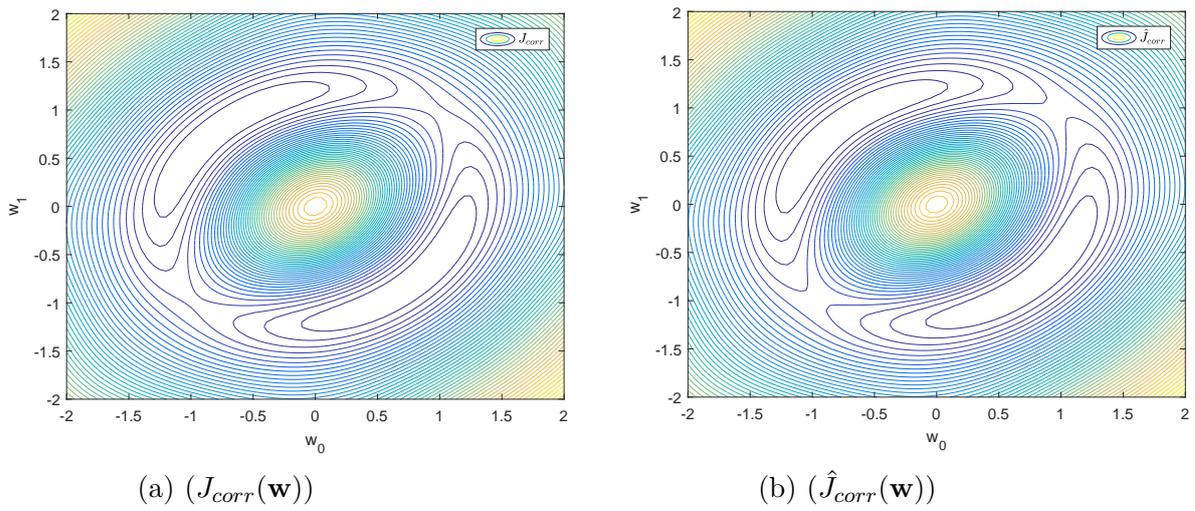


Figura 48 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 1,75$.

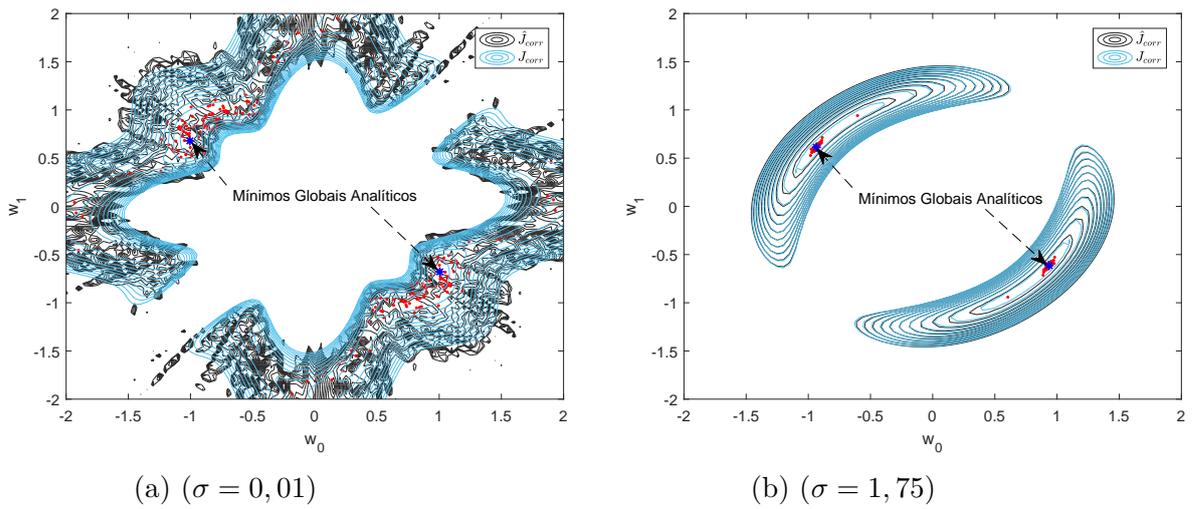


Figura 49 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$ (SNR = 30 dB).

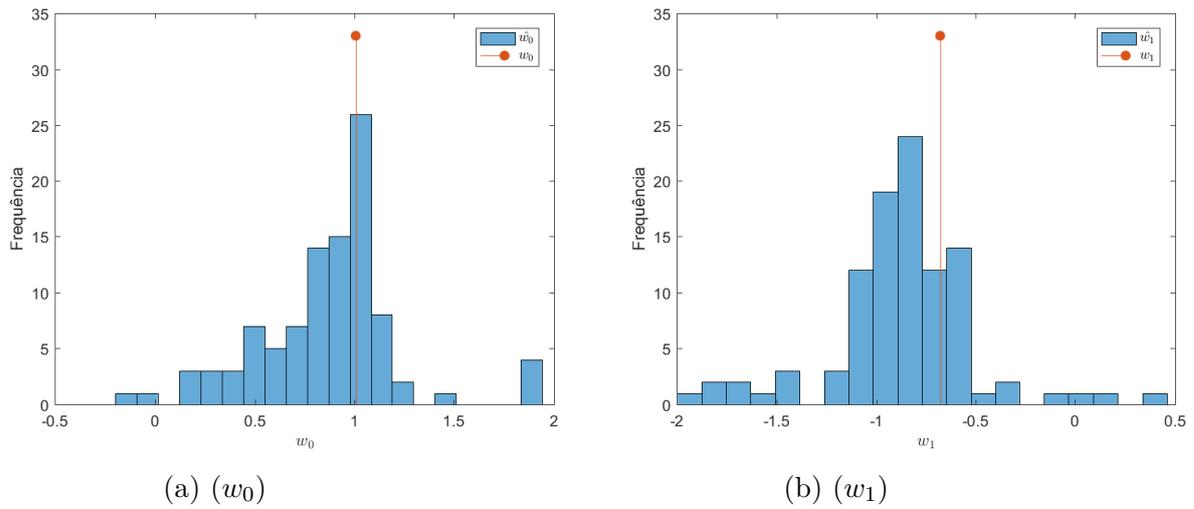


Figura 50 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0,01$.

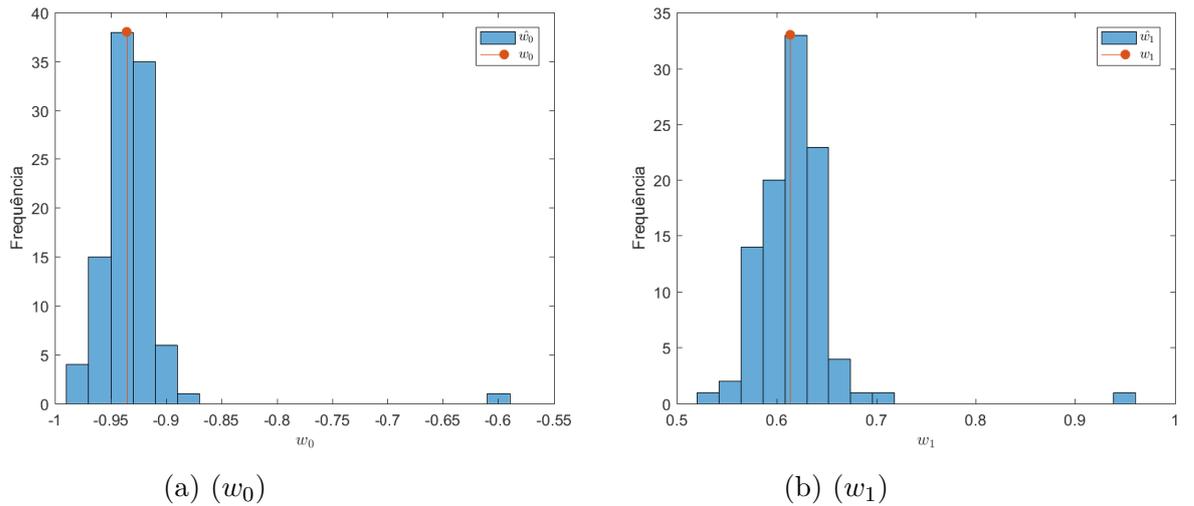


Figura 51 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 1,75$.

Tabela 6 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,01$	$\sigma = 1,75$
Estimada	-4,8611	-8,6482
Analítica	-8,3882	-8,7789

Analisando as Figuras 45 a 51, percebemos um comportamento semelhante àquele observado no caso BPSK: quando σ é pequeno, as soluções que minimizam \hat{J}_{corr} podem se afastar de forma considerável da solução ótima analítica; quando, porém, aumentamos o valor de σ , a distribuição das soluções de \hat{J}_{corr} fica mais concentrada em uma

vizinhança pequena em torno dos mínimos globais analíticos, de modo que as respectivas soluções são bastante parecidas. Interessantemente, ao contrário do que ocorreu no caso BPSK, podemos ver na Tabela 6 que o desempenho dos equalizadores melhora conforme a largura do *kernel* aumenta.

Prosseguindo com o mesmo cenário, agora é analisada a influência do tamanho do *kernel* no desempenho, em termos de ISI, para cada versão do critério baseado em correntropia. Variou-se σ de 0,05 para 2 em passos de 0,05, considerando $B = 3$, um total de $L = 10000$ amostras e $N_E = 100$ experimentos independentes para o custo estimado para SNRs de 12 dB e de 30 dB, como mostrado nas Figuras 52 e 53, respectivamente. Para efeitos de comparação, também mostramos o desempenho em ISI para a solução de Wiener considerando o melhor atraso de equalização (neste caso, $\alpha = 0$). É importante observar que, ao contrário do que foi observado para a fonte BPSK, o melhor desempenho alcançado em termos de ISI não está associado a um valor muito pequeno de σ .

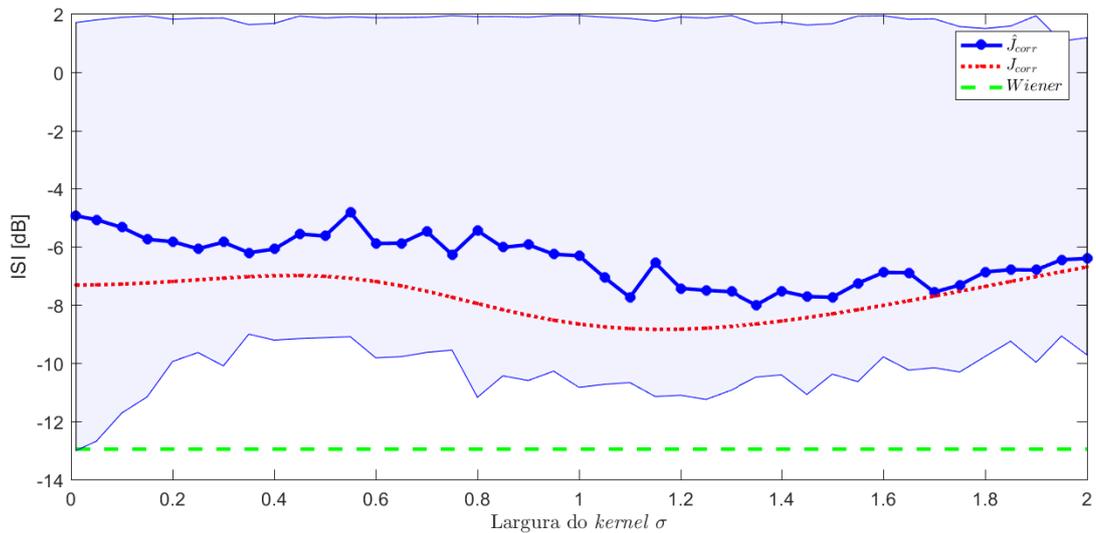


Figura 52 – ISI como uma função da largura do *kernel* σ (SNR = 12 dB).

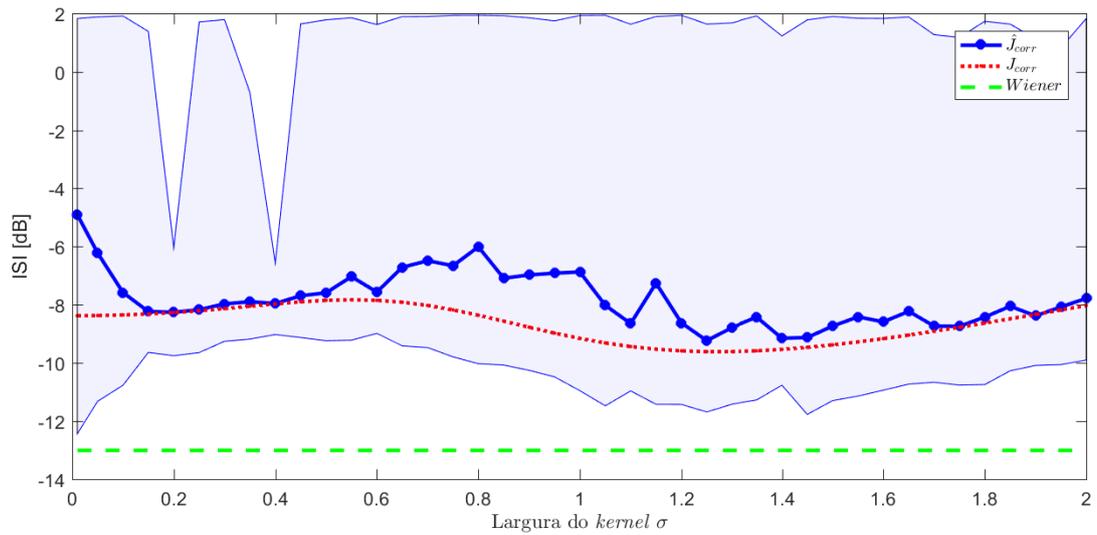


Figura 53 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB).

Como uma forma de explorar melhor o efeito do ruído, agora é analisada a influência da SNR no desempenho em ISI associado a cada versão do critério baseado em correntropia. Variou-se a SNR de 0 dB para 40 dB em passos de 2 dB, considerando $\sigma = 1,2$, $\sigma_g^2 = 0,1$, $B = 3$, um total de $L = 10000$ amostras e $N_E = 100$ experimentos independentes para o custo estimado, como mostrado na Figura 54. É importante observar que o desempenho em termos de ISI melhora na medida em que a SNR aumenta e que a versão estimada do critério apresentou um desempenho pior. Note que a ambas as versões do critério tendem a um patamar constante de desempenho em termos de ISI para SNR maior.

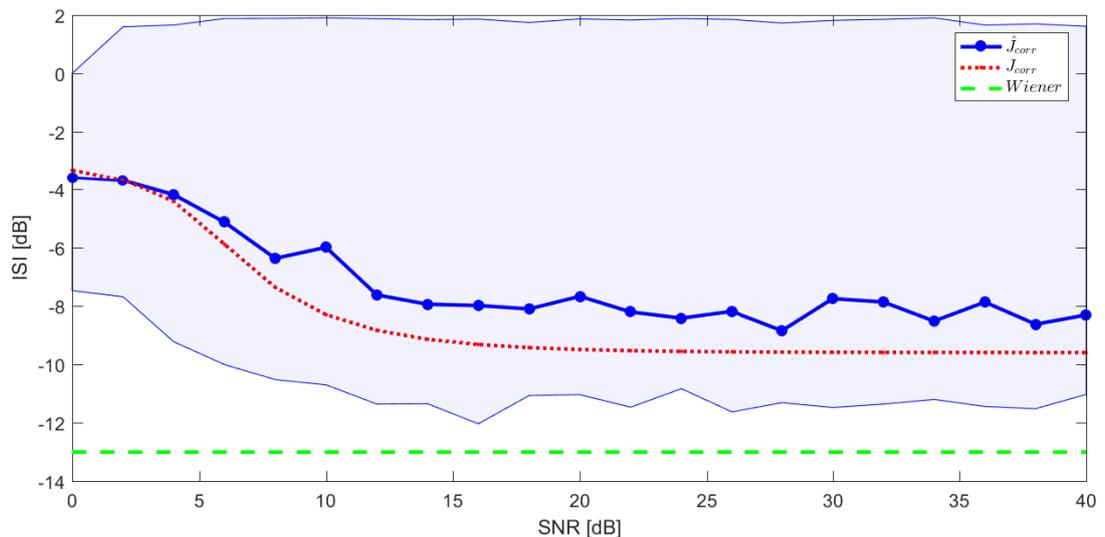


Figura 54 – ISI como uma função da largura da SNR (dB) ($\sigma = 1,2$).

Com o intuito de avaliar o impacto do modelo da fonte, especialmente em

comparação ao caso BPSK, agora é analisada a influência da variância da fonte contínua σ_g^2 no desempenho em ISI associado a cada versão do critério baseado em correntropia. Variou-se σ de 0,05 para 2 em passos de 0,05, considerando agora a variância da fonte contínua $\sigma_g^2 = 0,5$, $B = 3$, um total de $L = 10000$ amostras e $N_E = 100$ experimentos independentes para o custo estimado para SNR de 30 dB, como mostrado na Figura 55. É importante observar que, ao contrário do que foi observado para $\sigma_g^2 = 0,1$, o desempenho em termos de ISI piora conforme a largura do *kernel* σ aumenta e que a versão estimada do critério apresentou um desempenho pior. Esse resultado pode ser explicado pela mudança no formato da PDF da fonte contínua que requer um ajuste diferenciado da correntropia por parte da largura do kernel σ .

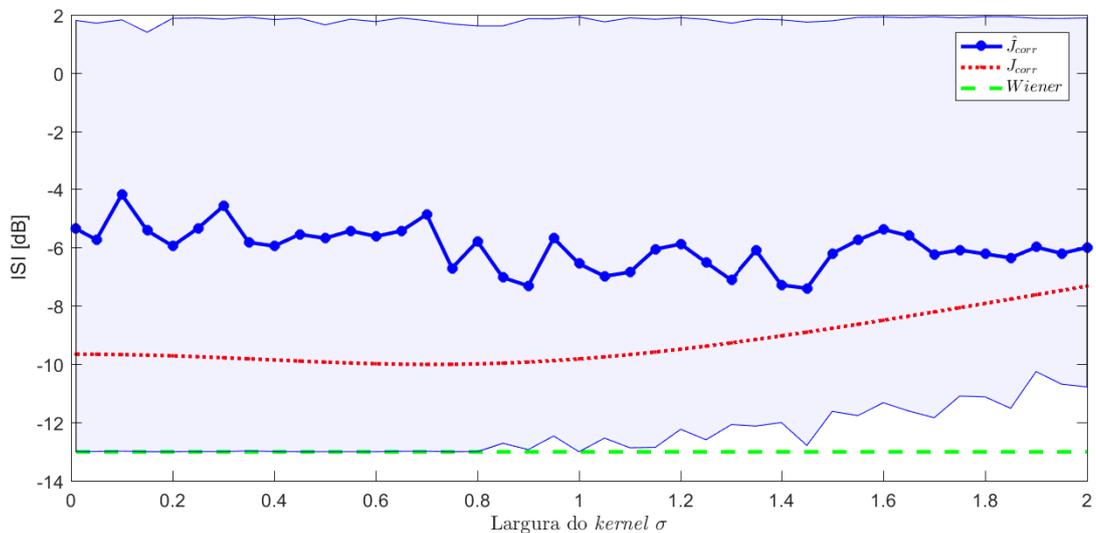


Figura 55 – ISI como uma função da largura do *kernel* σ para $\sigma_g^2 = 0,5$ (SNR = 30 dB).

5.4.2 Canal de fase máxima

Considerando, agora, o canal de fase máxima, com $H(z) = 1 + 1,5z^{-1}$, vamos analisar em primeiro lugar o efeito do tamanho do *kernel* (σ) sobre as superfícies de J_{corr} e \hat{J}_{corr} . Fixamos o número de atrasos considerados no critério em $B = 2$ para ambas as funções custo e utilizamos $L = 5000$ para a versão estimada. Assim, variando os coeficientes do equalizador, foram obtidas as superfícies de $J_{corr}(\mathbf{w})$ e $\hat{J}_{corr}(\mathbf{w})$ para os tamanhos de *kernel* de $\sigma = 0,1$ e $\sigma = 2$, como mostrado nas Figuras 56 e 57, respectivamente, e para uma SNR de 30 dB. Estes valores de σ foram escolhidos por representarem a maior divergência e a maior proximidade entre os desempenhos em termos de ISI para as versões estimada e analítica do critério. As Figuras 58 e 59 mostram as respectivas curvas de nível, enquanto as Figuras 60a e 60b apresentam as soluções encontradas pelo algoritmo ED para a função custo estimada, considerando $N_E = 100$ experimentos independentes, juntamente com as soluções ótimas para a função custo analítica. Finalmente, os histo-

gramas de cada coeficiente do equalizador referentes à versão estimada do critério são mostrados nas Figuras 61 e 62 para $\sigma = 0,1$ e $\sigma = 2$, respectivamente.

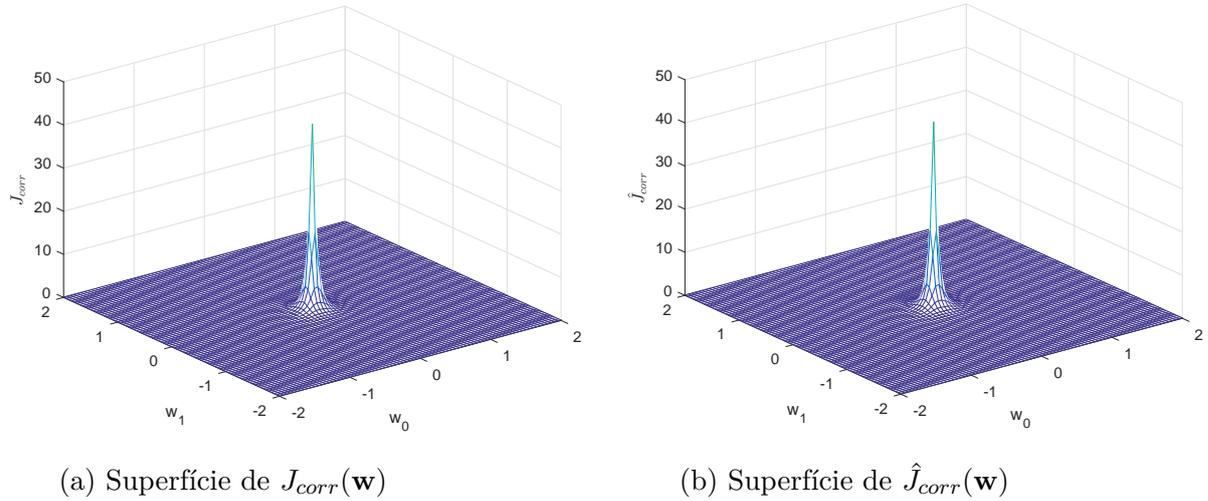


Figura 56 – Superfícies das funções custo analítica e estimada para $\sigma = 0,1$.

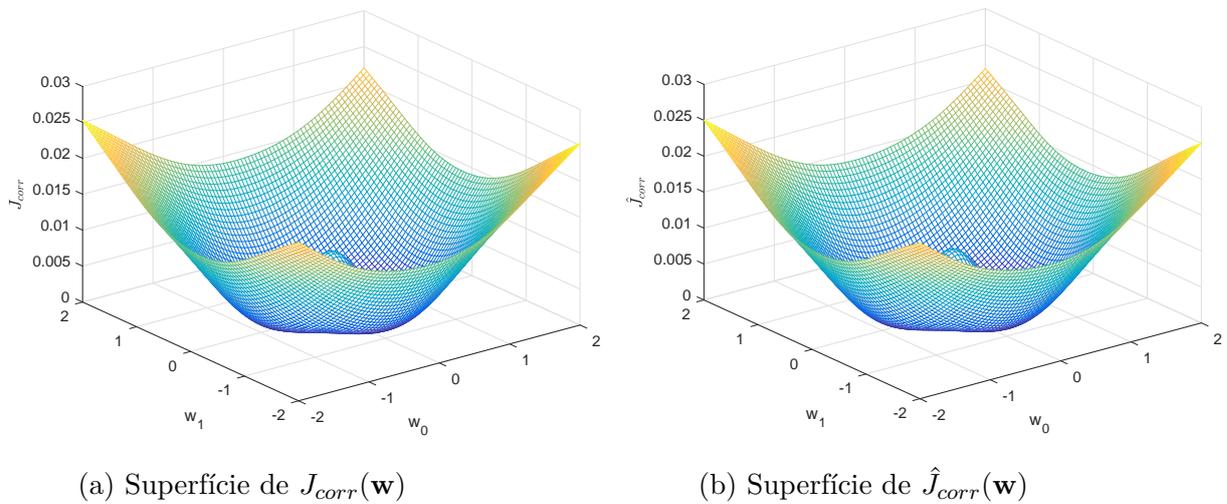


Figura 57 – Superfícies das funções custo analítica e estimada para $\sigma = 2$.

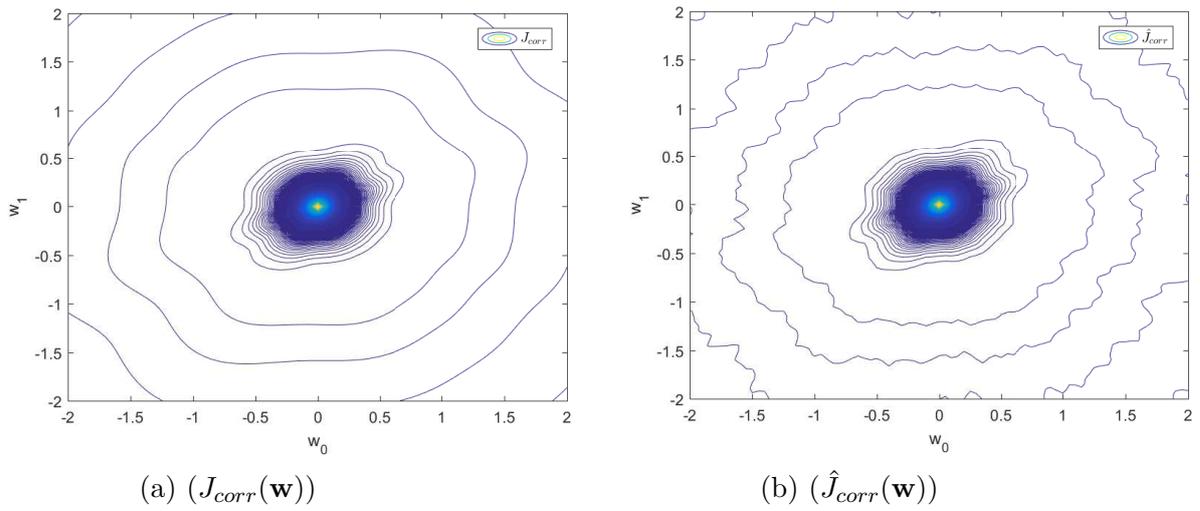


Figura 58 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 0,1$.

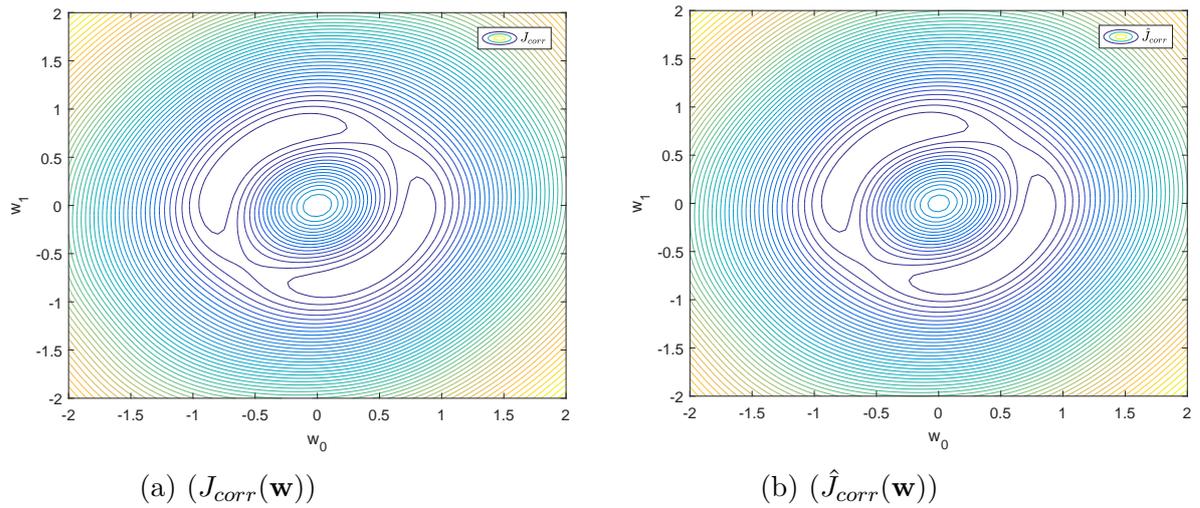


Figura 59 – Curvas de nível das superfícies das funções custo analítica e estimada para $\sigma = 2$.

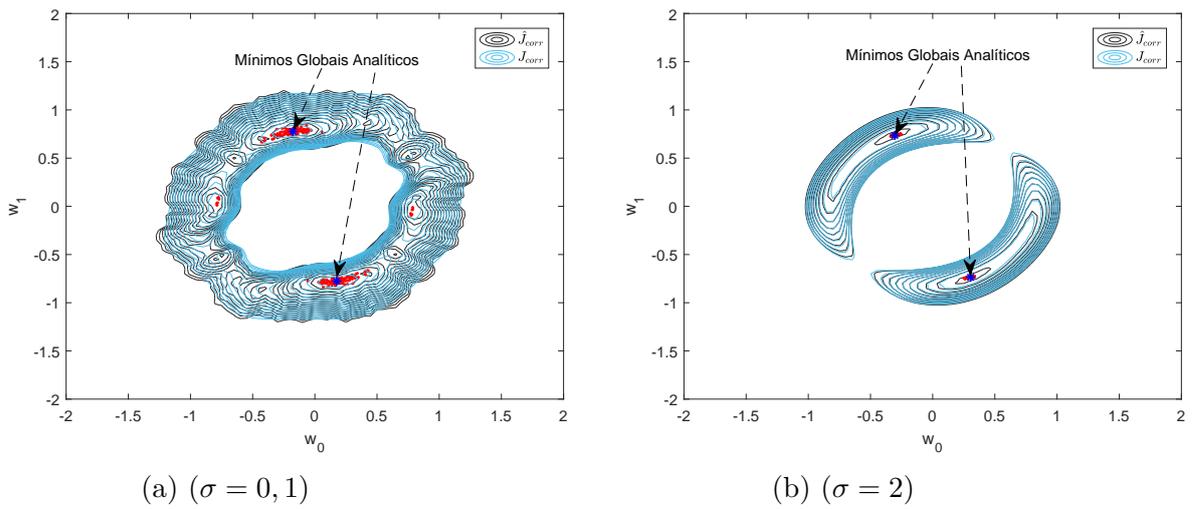


Figura 60 – Soluções obtidas pelo algoritmo ED durante a minimização de $\hat{J}_{corr}(\mathbf{w})$. Ao fundo, aparecem as curvas de nível de $J_{corr}(\mathbf{w})$ e de $\hat{J}_{corr}(\mathbf{w})$ (SNR = 30 dB).

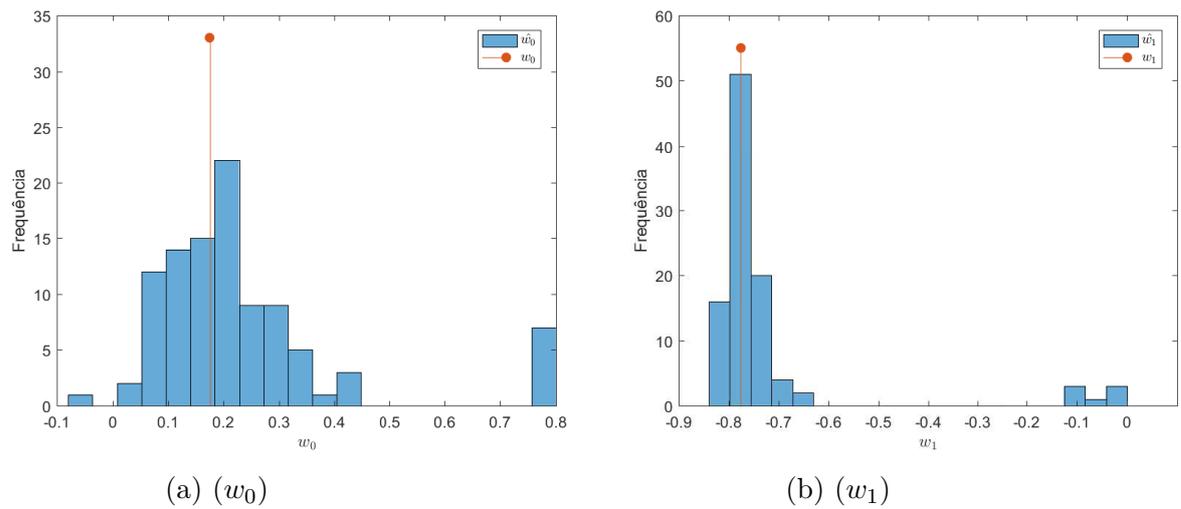


Figura 61 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 0, 1$.

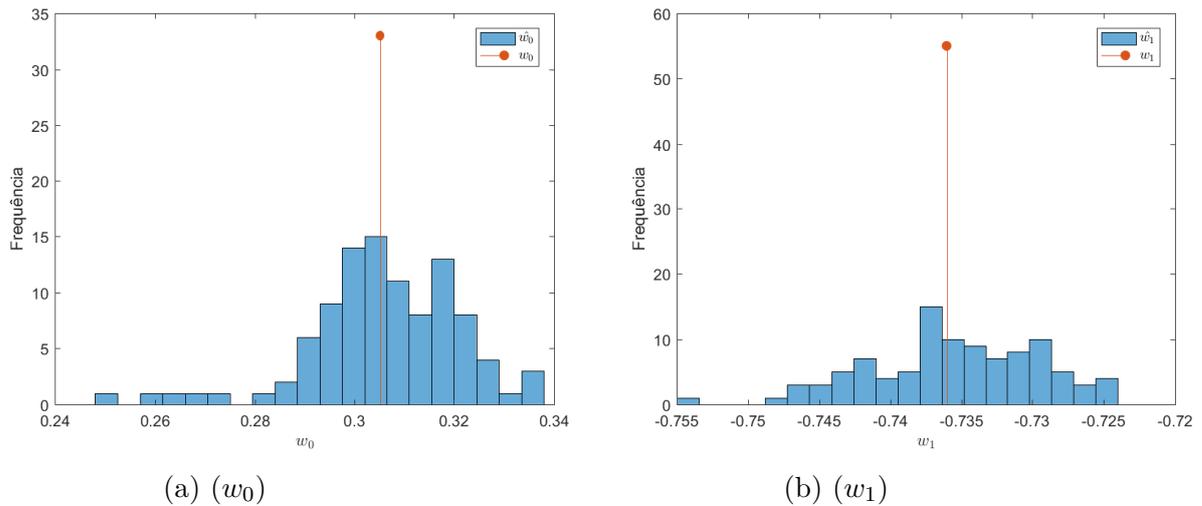


Figura 62 – Histogramas dos coeficientes do equalizador para a versão estimada para $\sigma = 2$.

Tabela 7 – ISI (em dB) associada com as soluções ótimas de J_{corr} e \hat{J}_{corr} .

Função Custo	$\sigma = 0,1$	$\sigma = 2$
Estimada	-6,4163	-8,5243
Analítica	-6,6406	-8,5404

Observando as Figuras 56 a 62, podemos perceber que: (i) as superfícies das funções custo estimada e analítica são, até certo ponto, parecidas, e sofrem modificações expressivas quando a largura do *kernel* é alterada; (ii) assim como ocorreu no caso das fontes BPSK e Gaussiana, as soluções ótimas de \hat{J}_{corr} ficam bem próximas à solução analítica quando σ assume um valor mais elevado; e (iii) se a largura do *kernel* é pequena, o ótimo global de \hat{J}_{corr} pode se afastar bastante do ótimo da função custo analítica.

A influência do tamanho do *kernel* no desempenho em ISI associado a cada versão do critério baseado em correntropia pode ser melhor apreciada com base nas Figuras 63 e 64, que foram geradas variando-se σ de 0,05 para 2 em passos de 0,05, considerando $B = 3$, $L = 10000$ amostras, $N_E = 100$ experimentos independentes envolvendo a função custo estimada, e os valores de SNR de 12 dB e de 30 dB, respectivamente. A área sombreada representa o intervalo de ISI considerando os valores de máximo e de mínimo associados ao conjunto de todas as soluções encontradas para \hat{J}_{corr} . Para efeitos de comparação, também mostramos o desempenho em ISI para a solução de Wiener considerando o melhor atraso de equalização (neste caso, $\alpha = 2$) (HAYKIN, 2013).

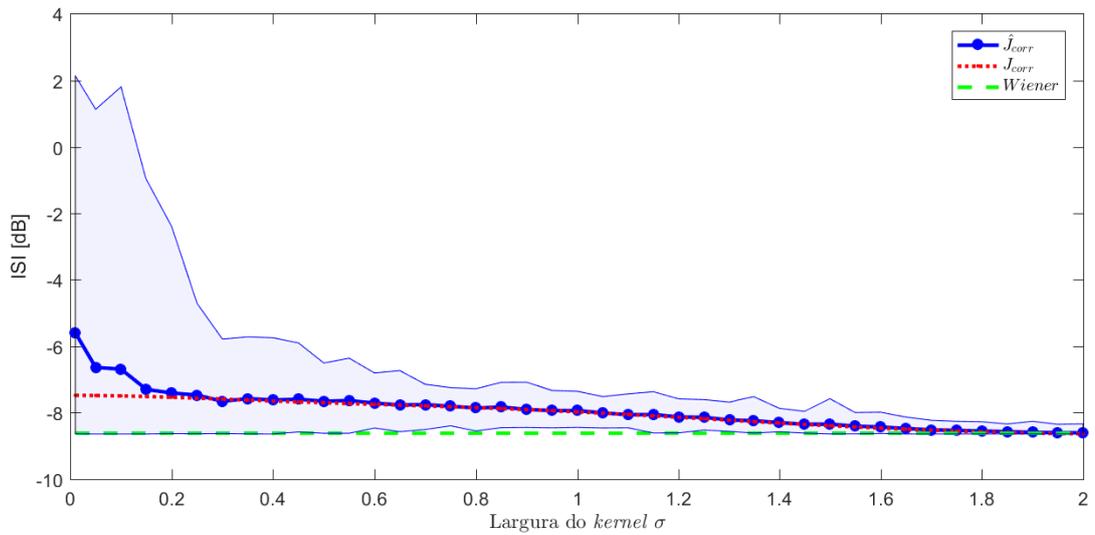


Figura 63 – ISI como uma função da largura do *kernel* σ (SNR = 12 dB).

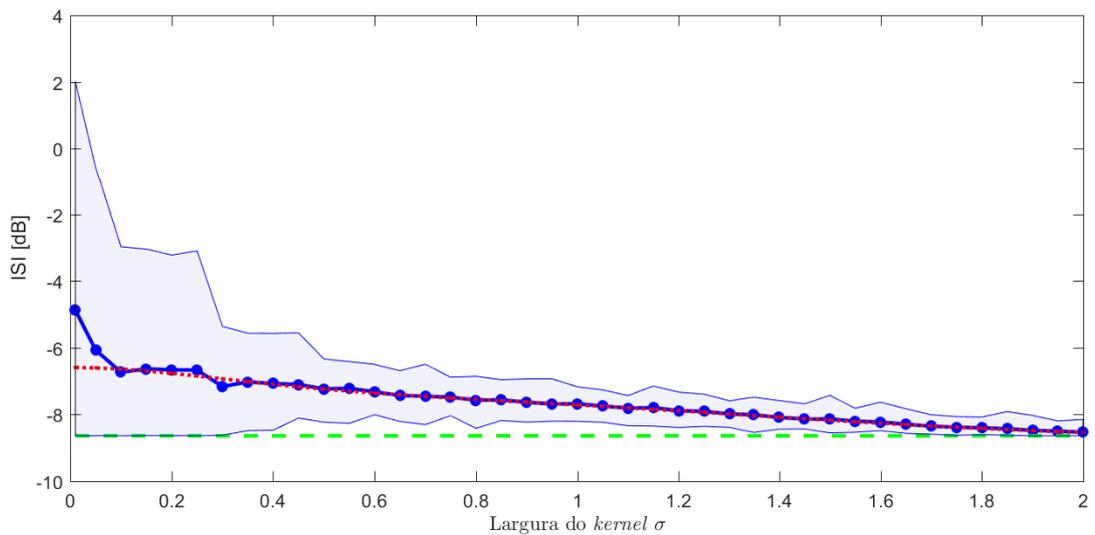


Figura 64 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB).

É possível verificar, nas Figuras 63 e 64, que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ se aproximam significativamente daqueles de $\hat{J}_{corr}(\mathbf{w})$ para o intervalo de valores de σ considerados. Esse resultado sugere que as soluções obtidas são próximas para os casos estimado e analítico. Vale mencionar também que, para as SNR consideradas de 12 dB e 30 dB, o melhor desempenho em ISI é obtido para valores acima de $\sigma = 1,8$, que, inclusive, alcançam o patamar de desempenho em ISI associado à solução de Wiener. É importante ressaltar também que, neste caso, o critério baseado em correntropia se mostrou robusto para lidar com uma SNR mais baixa, uma vez que as curvas de desempenho em termos de ISI são muito parecidas para ambos os valores de SNR considerados.

Curiosamente, comparando as Figuras 63 e 64, é possível notar que para valores pequenos de σ , o desempenho em ISI é um pouco melhor quando a SNR é de 12 dB. Para tentar compreender um pouco melhor a razão disto, analisamos os perfis de correntropia gerados na saída dos respectivos equalizadores (construídos de modo a maximizar o casamento de correntropia teórica para dois atrasos, com $\sigma = 0,05$) e percebemos que: (1) para a SNR de 30 dB, o casamento acaba sendo um pouco melhor, levando a um valor um pouco inferior de $J_{corr}(\mathbf{w})$; (2) para a SNR de 12 dB, o perfil de correntropia é um pouco mais distante daquele associado à fonte; porém, a ISI correspondente é, de fato, um pouco menor.

Com o intuito de avaliar o impacto do ajuste da fonte, agora é analisada a influência da variância da fonte contínua σ_g^2 no desempenho em ISI associado a cada versão do critério baseado em correntropia. Variou-se σ de 0,05 para 2 em passos de 0,05, considerando agora a variância da fonte contínua $\sigma_g^2 = 0,5$, $B = 3$, um total de $L = 10000$ amostras e $N_E = 100$ experimentos independentes para o custo estimado para SNR de 30 dB, como mostrado na Figura 65. É possível verificar que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ se aproximam significativamente daqueles de $\hat{J}_{corr}(\mathbf{w})$ para o intervalo de valores de σ considerados, de maneira muito similar ao que foi observado para $\sigma_g^2 = 0,1$, com exceção do intervalo de valores $\sigma < 0,45$.

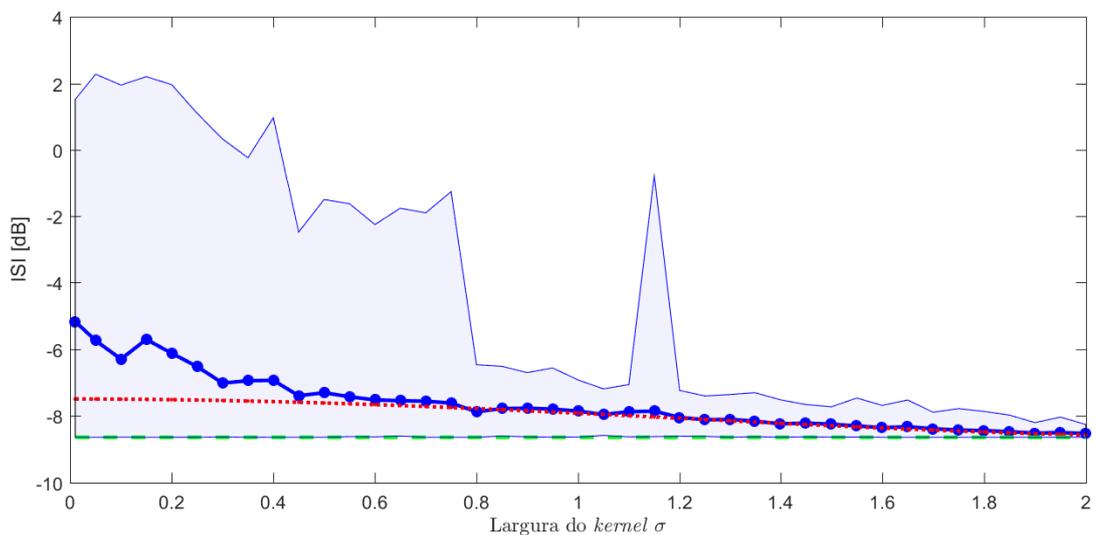


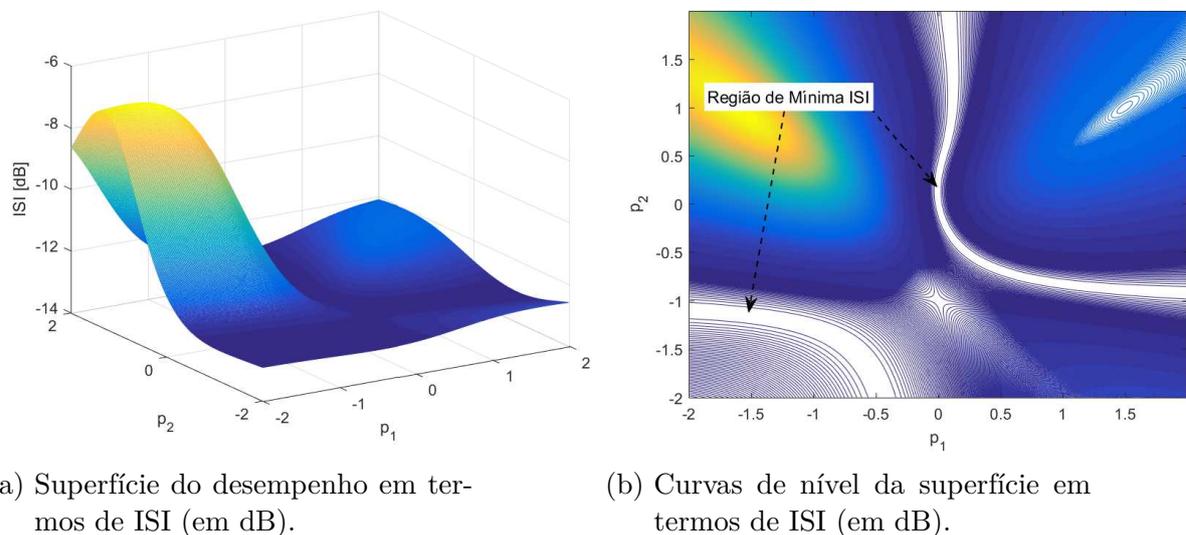
Figura 65 – ISI como uma função da largura do *kernel* σ para $\sigma_g^2 = 0,5$ (SNR = 30 dB).

5.5 Análise do Efeito do Pré-Codificador

Com o intuito de investigar o efeito do pré-codificador no desempenho alcançado pelos filtros equalizadores, analisaremos a sua influência em termos do desempenho em ISI e do erro quadrático médio ($J_{Wiener}(\mathbf{w})$) associados aos seus correspondentes filtros equalizadores projetados por meio da solução de Wiener para o sistema com pré-

codificador (apresentada no Anexo A deste trabalho). No cenário considerado, o canal é um sistema de fase mínima com função de transferência dada por $H(z) = 1 + 0,5z^{-1}$. Um pré-codificador com três coeficientes $\mathbf{p} = [1 \ p_1 \ p_2]^T$ é adotado, cujos pesos p_1 e p_2 irão variar do intervalo de -2 a 2 para se obter as curvas de nível da superfície do desempenho em termos de ISI e da função custo $J_{Wiener}(\mathbf{w})$ associados às soluções de Wiener obtidas para o sistema considerado.

Primeiramente, analisamos o efeito do pré-codificador no desempenho alcançado pelos filtros equalizadores em termos da ISI. Assim, variando os coeficientes p_1 e p_2 do pré-codificador em passos de $0,011$, obtivemos a superfície do desempenho em termos de ISI (em dB) e o seu contorno, como mostrado nas Figuras 66a e 66b, respectivamente.



(a) Superfície do desempenho em termos de ISI (em dB).

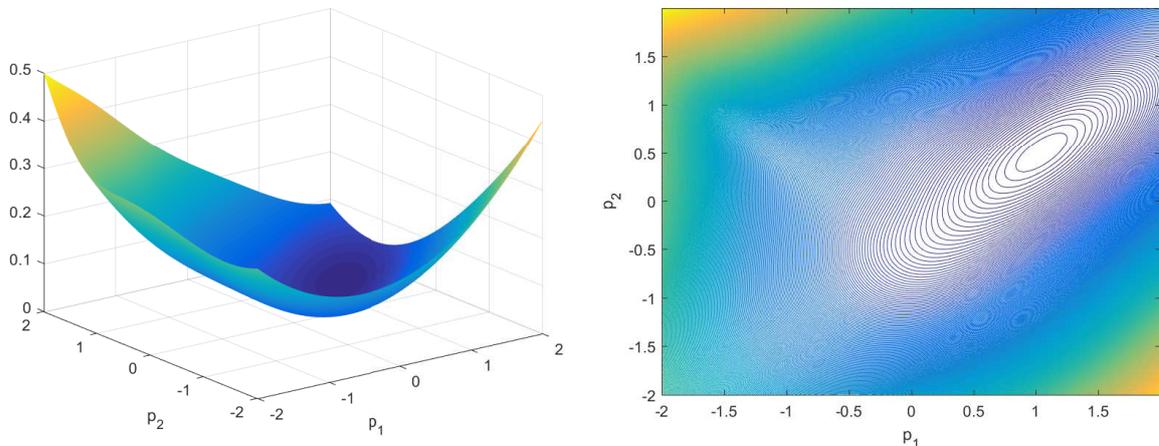
(b) Curvas de nível da superfície em termos de ISI (em dB).

Figura 66 – Desempenho em termos de ISI (em dB).

É possível verificar, nas Figuras 66a e 66b, que os valores de desempenho em ISI associados aos filtros equalizadores, que foram projetados por meio da Solução de Wiener para o sistema com pré-codificador, formam o que podemos chamar de uma região ótima, considerado um nível de desempenho, em termos de ISI, inferior a -13 dB. Esse resultado sugere que a escolha adequada do pré-codificador pode levar a resultados melhores, em termos de ISI, para os equalizadores projetados. Entretanto, a análise do desempenho em termos de ISI apresenta o problema de não ser invariável a um fator de escala das soluções encontradas. Com o intuito de se obter uma descrição mais completa do efeito do pré-codificador, surge uma nova possibilidade de análise direta em termos de $J_{Wiener}(\mathbf{w})$, que será explorada a seguir.

Prosseguindo com este cenário, agora é analisada a influência do pré-codificador em termos do erro quadrático médio associado aos seus correspondentes filtros equalizadores projetados por meio da Solução de Wiener. Assim, variando os coeficientes p_1 e p_2 do pré-codificador em passos de $0,011$, obtivemos a superfície e o contorno da função custo

$J_{Wiener}(\mathbf{w})$, como mostrado nas Figuras 67a e 67b, respectivamente.



(a) Superfície do desempenho em termos de $J_{Wiener}(\mathbf{w})$.

(b) Contorno da superfície em termos de $J_{Wiener}(\mathbf{w})$.

Figura 67 – Desempenho em termos de $J_{Wiener}(\mathbf{w})$.

É possível verificar, nas Figuras 67a e 67b, que existe um valor ótimo para os coeficientes p_1 e p_2 do pré-codificador em termos de $J_{Wiener}(\mathbf{w})$. No caso, o pré-codificador ideal é representado pela função de transferência dada por $P(z) = 1 + 1,0522z^{-1} + 0,5081z^{-2}$. Vale mencionar também que a solução ótima em termos de $J_{Wiener}(\mathbf{w})$, para um intervalo de -2 a 2 , está fora da região de desempenho ótimo para o pré-codificador em termos de ISI.

Com o intuito de compreender melhor o efeito do pré-codificador, considera-se agora o caso limite onde o pré-codificador é considerado fixo e igual ao que foi encontrado anteriormente para o melhor desempenho em termos de $J_{Wiener}(\mathbf{w})$. Para este caso, avaliam-se agora os desempenhos em termos de ISI e $J_{Wiener}(\mathbf{w})$ para as soluções (encontradas pelo ED) de cada versão do critério baseado em correntropia. Fixou-se o número de atrasos considerados no critério em $B = 4$ para uma fonte BPSK (sem ruído) para ambas as funções custo - analítica e estimada para $\sigma = 1$ - e utilizando $L = 10000$ para a versão estimada. Adicionalmente, para efeito de comparação também são apresentados os resultados para a solução de Wiener para um sistema com pré-codificador e sem pré-codificador, considerando o melhor atraso de equalização.

Tabela 8 – Desempenhos em ISI (em dB) e $J_{Wiener}(\mathbf{w})$ associados com as soluções ótimas de J_{corr} e \hat{J}_{corr} para $\sigma = 1$.

Desempenho	ISI (em dB)	$J_{Wiener}(\mathbf{w})$
Wiener (sem pré-codificador)	-13,0103	0,0476
Wiener (com pré-codificador)	-12,0904	0,0228
J_{corr}	-12,8959	0,0358
\hat{J}_{corr}	-12,8595	0,0404

Os valores de ISI e de $J_{Wiener}(\mathbf{w})$ alcançados pelas soluções associadas com as funções custo analítica e estimada para ambos os atrasos considerados pelo critério, bem como com os resultados para a solução de Wiener para um sistema com pré-codificador e sem pré-codificador, são apresentados na Tabela 8. É possível observar, na Tabela 8, que os critérios baseados em correntropia, tanto na sua versão analítica quanto na estimada, são capazes de fornecer soluções que conseguem alcançar um desempenho, em termos de ISI, superior ao alcançado pela solução de Wiener com pré-codificador. Foi possível observar também que a solução de Wiener com pré-codificador é responsável por alcançar o menor valor em termos de MSE ($J_{Wiener}(\mathbf{w})$), entretanto, isso não leva a um melhor desempenho em termos de ISI. Esse resultado demonstra a robustez do critério baseado em correntropia frente à abordagem supervisionada da solução de Wiener para um sistema com pré-codificador, dado que a correntropia é uma medida capaz de explorar a informação estatística e a dependência temporal, incorporando assim a presença do pré-codificador.

Ao olhar a solução de Wiener para o sistema com pré-codificador de uma maneira isolada, é possível entender que o pré-codificador atua como um aumento do número de coeficientes do filtro equalizador empregado para cancelar o efeito do canal. Portanto, o sistema com pré-codificador está relacionado com um equalizador capaz de minimizar o valor de MSE. Entretanto, a solução de Wiener para um sistema com pré-codificador não implica em uma melhora de desempenho em termos de ISI da resposta combinada canal/equalizador. É importante ressaltar também que Wiener não explora estatísticas de ordem superior, ao contrário do critério baseado em correntropia. O efeito do pré-codificador merece uma investigação mais extensiva de modo que os seus resultados possam ser interpretados de uma maneira mais segura.

5.5.1 Pré-codificador escolhido com base no desempenho em ISI

Com o intuito de compreender melhor o efeito do pré-codificador, considera-se agora o caso onde o pré-codificador é considerado fixo e escolhido com base no nível de desempenho, em termos de ISI, associado aos filtros equalizadores que foram projetados por meio da Solução de Wiener para o sistema com pré-codificador. Foi analisada a influência da largura do *kernel* no desempenho em ISI de cada uma das versões do critério baseado em correntropia para a fonte contínua. Fixou-se o número de atrasos considerados no critério em $B = 3$, a variância da fonte contínua em $\sigma_g^2 = 0,1$ para ambas as funções custo - analítica e estimada - e empregou-se $L = 10000$ para a versão estimada do critério, dada uma SNR de 30 dB. Variou-se σ de 0,05 para 2 em passos de 0,05 e considerou-se $N_E = 100$ experimentos independentes para o custo estimado.

O pré-codificador escolhido tem função de transferência $P(z) = 1 + 0.4z^{-1} - 0.6z^{-2}$ com uma ISI igual a $-13,0092$ dB para a sua solução de Wiener, uma vez que a solução de Wiener (sem pré-codificador) apresenta uma ISI igual a $-13,0103$ dB.

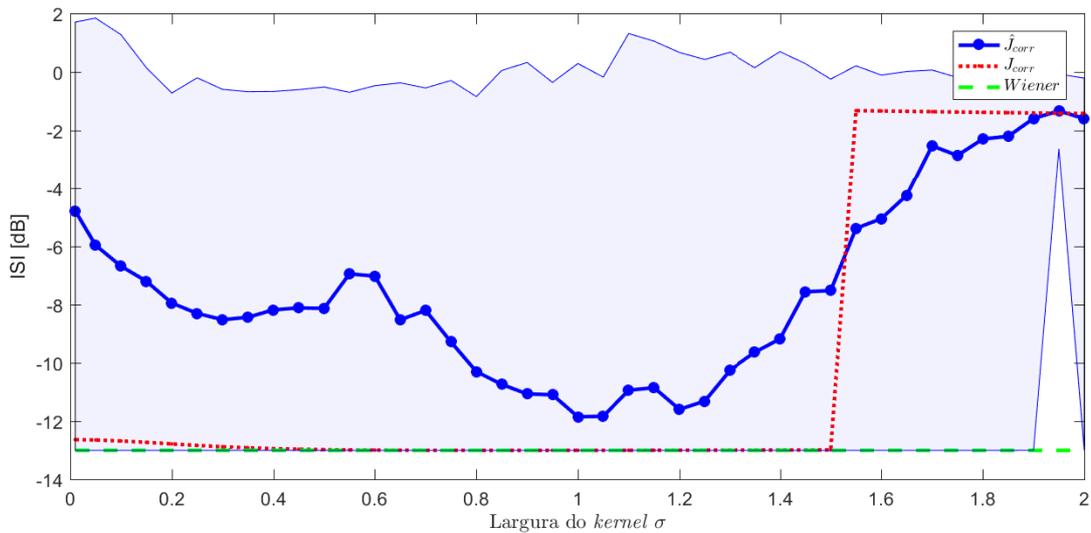


Figura 68 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB) para o pré-codificador com função de transferência $P(z) = 1 + 0.4z^{-1} - 0.6z^{-2}$.

É possível verificar, na Figura 68, que, tanto $J_{corr}(\mathbf{w})$ quanto $\hat{J}_{corr}(\mathbf{w})$ foram capazes de alcançar soluções com níveis de desempenho em ISI melhores em comparação com os resultados que foram obtidos para o pré-codificador que foi empregado nos cenários anteriores, com função de transferência $P(z) = 1 - 0,5z^{-1} + 0.3z^{-2}$. É possível observar também que os valores de desempenho em ISI associados com $J_{corr}(\mathbf{w})$ para a fonte contínua foram capazes de alcançar o mesmo patamar de desempenho da solução de Wiener para um grande intervalo de valores de σ , tanto para valores pequenos ($\sigma \leq 0,2$) quanto para valores maiores ($0,4 \leq \sigma \leq 1,35$), fato que não acontecia quando o pré-codificador possuía a função de transferência $P(z) = 1 - 0,5z^{-1} + 0,3z^{-2}$. Vale mencionar ainda que o melhor desempenho, em termos de ISI, para a versão estimada do critério, é alcançado na vizinhança de $\sigma = 1$. Esses resultados mostram que a escolha do pré-codificador pode impactar de maneira significativa no desempenho alcançável pelo critério baseado em correntropia e mostra também o potencial deste critério em extrair informação estatística e temporal do sistema em questão e empregá-la de maneira efetiva no contexto de equalização de canais.

5.6 Emprego de Outras Normas no Casamento da Correntropia

Até este ponto, o trabalho foi dedicado à análise do critério de casamento de correntropia, cuja definição foi dada em (SANTAMARÍA *et al.*, 2006) (vide Equação (3.29)). Este critério traduz a ideia de tornar a correntropia na saída do equalizador o mais parecida possível com a correntropia da fonte em termos da minimização do erro quadrático entre estas duas medidas considerando um número máximo (B) de atrasos.

Esta forma de expressar a noção de casamento de correntropia constitui, porém, apenas uma possibilidade dentre várias alternativas. Neste contexto, podemos pensar em empregar outras normas, diferentes da Euclidiana, para calcular o desvio entre os perfis de correntropia. Em outras palavras, podemos trabalhar com outras definições para o critério de casamento de correntropia que não explorem a norma Euclidiana.

Esta perspectiva será explorada, ainda que de maneira preliminar, no restante deste trabalho. Em particular, propomos estudar dois critérios ligados à noção de casamento de correntropia, mas baseados em normas L_p , com $p \neq 2$.

De uma maneira direta, o critério baseado em correntropia, na forma proposta por Santamaria et al. (2006), está intuitivamente relacionado à noção geométrica de comprimento, ou seja, ele propõe a minimização da distância quadrática entre o perfil de correntropia observado na saída do equalizador com o perfil de correntropia produzido pelo efeito do pré-codificador no sinal gerado pela fonte. Neste sentido, os perfis de correntropia podem ser vistos como vetores de comprimento igual ao número de atrasos B .

Dado um espaço vetorial X^B sobre um conjunto de números reais \mathfrak{R} , uma função $\|\cdot\| : X^B \rightarrow \mathfrak{R}$ é chamada de norma se, para quaisquer vetores $\mathbf{x}, \mathbf{y} \in X^B$ e todo $\alpha \in \mathfrak{R}$ (GONIN, 2017):

1. $\|\mathbf{x}\| = 0 \Rightarrow \mathbf{x} = \mathbf{0}$. Caso essa condição não seja atendida, a função poderá no máximo ser uma seminorma.
2. $\|\alpha\mathbf{x}\| = |\alpha| \|\mathbf{x}\|$.
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (desigualdade triangular). Se o espaço vetorial X^B tem uma norma, ele é chamado de espaço normado e passa a ser denotado por $(X^B, \|\cdot\|)$. Um espaço normado completo na métrica induzida pela norma é chamado espaço de Banach.

Pode-se dizer que a norma induz uma métrica d em X , onde a distância entre dois vetores quaisquer \mathbf{x} e \mathbf{y} é dada por: $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$. A partir daí, considerando espaços de dimensão finita, podem ser definidas as chamadas normas canônicas que são também conhecidas como normas L_p :

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^B |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty. \quad (5.12)$$

O caso particular em que $p = 2$ é correspondente à norma Euclidiana. Em alguns casos, devido à natureza do problema, o emprego de outras normas pode se tornar interessante em termos de desempenho e redução do custo computacional.

Sendo assim, outras possibilidades para o casamento de correntropia surgem quando consideramos normas L_p diferentes da Euclidiana. Aqui, vamos abordar dois casos específicos, envolvendo as normas L_1 e L_∞ .

A norma L_1 é definida como sendo a soma dos módulos das componentes, ou seja:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^B |x_i|. \quad (5.13)$$

Logo, ela dá origem ao seguinte critério de casamento de correntropia baseado na norma L_1 .

$$J_{corr(L_1)}(\mathbf{w}) = \sum_{m=1}^B |V_S(m) - V_Y(m)|. \quad (5.14)$$

A norma L_∞ é definida como sendo o módulo da maior componente:

$$\|\mathbf{x}\|_\infty = \max_{i=1}^B |x_i|. \quad (5.15)$$

Com isto, podemos elaborar o seguinte critério de casamento de correntropia baseado na norma L_∞ .

$$J_{corr(L_\infty)}(\mathbf{w}) = \max_{m=1}^B |V_S(m) - V_Y(m)|. \quad (5.16)$$

A norma L_2 procura atenuar o efeito de erros pequenos, enquanto os erros maiores apresentam uma penalização maior no seu cálculo, por conta do seu efeito quadrático. A norma L_1 , por outro lado, procura penalizar a contribuição dos erros menores com um maior peso, enquanto que os erros maiores são penalizados com um menor peso, quando comparada à norma L_2 . A norma L_∞ por sua vez considera apenas o maior erro na sua computação, não levando em conta se os demais erros são pequenos ou também muito elevados. Em (RICE; WHITE, 1964), é observado que a efetividade das normas L_p para estimação depende essencialmente da distribuição do erro. No fundo o emprego dos critérios baseados no casamento de correntropia buscam obter uma melhor estimativa da correntropia na saída do equalizador e, para isso, buscam minimizar uma medida de erro no casamento dos atrasos considerados.

Agora analisamos a influência das normas L_1 , L_2 e L_∞ no desempenho em ISI associado a cada versão do critério que é baseada no casamento entre as correntropias da fonte e do sinal equalizado. Fixou-se o número de atrasos considerados no critério em $B = 3$, a variância da fonte contínua $\sigma_g^2 = 0,1$ para ambas as funções custo - analítica e estimada - e empregou-se $L = 10000$ para as versões estimadas do critério. Variou-se σ de 0,01 para 3 em passos de 0,01 e considerou-se $N_E = 100$ experimentos independentes para o custo estimado para uma SNR de 30 dB. Os parâmetros utilizados no algoritmo ED foram $N_P = 250$, $F = 0,5$, $CR = 0,9$ e $I_T = 150$ iterações.

Para cada valor de σ , é mostrado nas Figuras 69, 70 e 71, o valor da ISI associado com a melhor solução fornecida pela função custo analítica, que foi obtida por

meio do algoritmo ED, assim como a ISI média para as soluções da função custo estimada para cada uma das variantes do critério baseado em correntropia. A área sombreada representa o intervalo de ISI considerando os valores de máximo e de mínimo associados ao conjunto de todas as soluções encontradas para \hat{J}_{corr} . Para efeitos de comparação, também mostramos o desempenho em ISI para a solução de Wiener considerando o melhor atraso de equalização (HAYKIN, 2013).

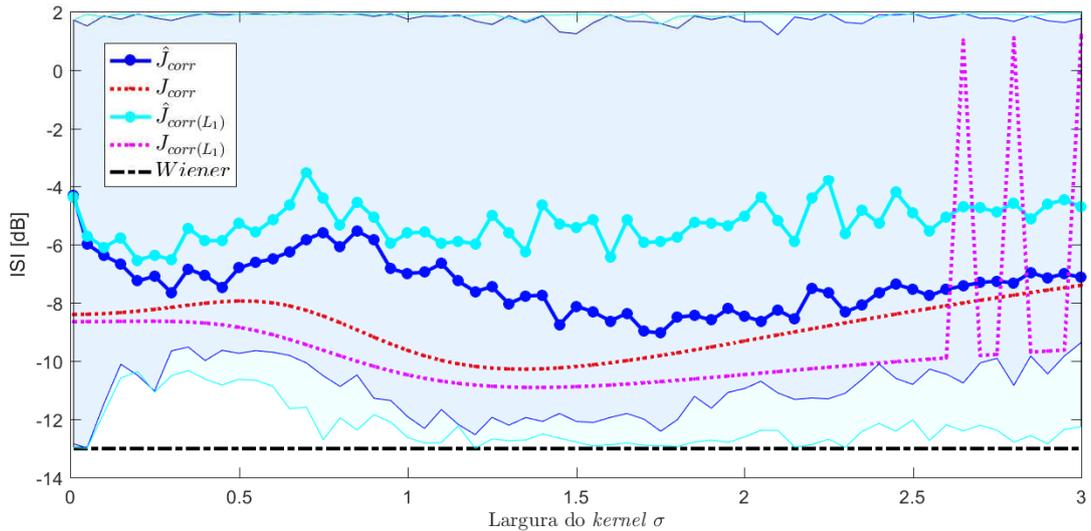


Figura 69 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB) para as normas L_2 e L_1 .

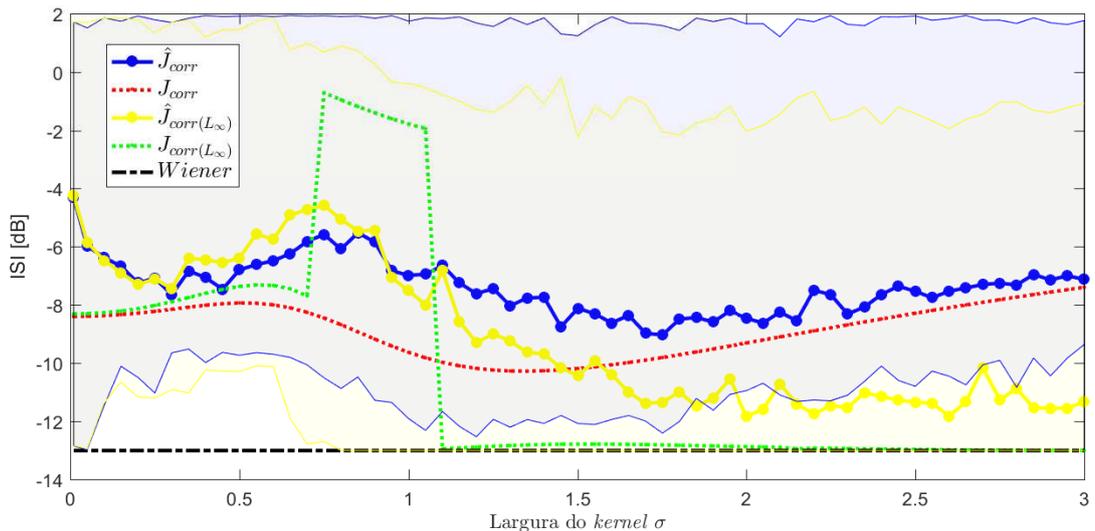


Figura 70 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB) para as normas L_2 e L_∞ .

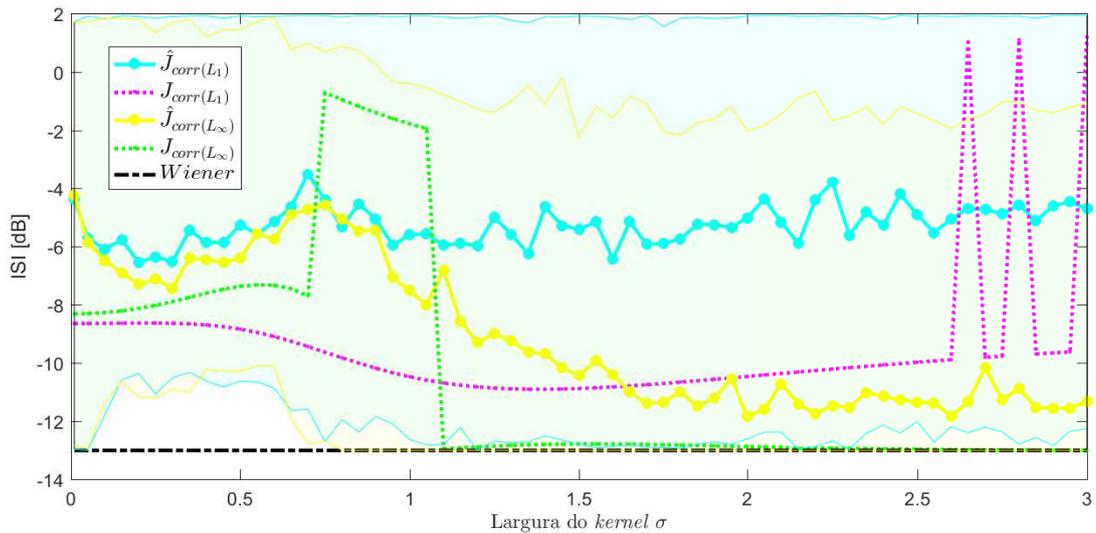


Figura 71 – ISI como uma função da largura do *kernel* σ (SNR = 30 dB) para as normas L_1 e L_∞ .

É possível verificar, nas Figuras 69, 70 e 71, que os valores de desempenho em ISI associados com as normas L_1 e L_∞ não diferem muito para valores de σ inferiores a 0,4 para as versões estimadas e analíticas do critério baseado em correntropia. Entretanto, para valores de σ maiores do que 1, foi possível observar que o emprego da norma L_∞ resultou em um desempenho significativamente melhor quando comparado com os desempenhos das normas L_1 e L_2 , para este mesmo cenário. Esse desempenho notável da norma L_∞ foi capaz inclusive de alcançar o patamar de desempenho da solução de Wiener para um grande intervalo de valores de σ na sua versão analítica. A versão estimada do critério para a norma L_∞ foi também capaz de obter níveis de ISI consideravelmente melhores do que as versões do critério que empregam as normas L_1 e L_2 .

Intuitivamente, este resultado mostra o potencial de se explorar outras formas de casamento entre o perfil de correntropia observado na saída do equalizador com o perfil de correntropia produzido pelo efeito do pré-codificador no sinal gerado pela fonte. Esse resultado reforça ainda mais a ideia de que o emprego de um número maior de atrasos para o casamento de correntropia no critério não necessariamente implica em uma melhoria de desempenho em termos de ISI. Isso pode ser explicado pelo fato que as normas L_1 e L_2 ponderam igualmente todos os atrasos considerados para o casamento de correntropia e que o critério baseado na norma L_∞ tenta minimizar apenas o atraso que apresenta a maior diferença, ignorando os demais atrasos. Uma nova perspectiva levantada seria a aplicação de normas ponderadas que fossem capazes de privilegiar o peso de alguns atrasos específicos considerados pelo critério, ou seja, explorando a possível relação entre normas L_p e a norma L_2 .

Conclusão

É importante destacar as principais contribuições deste trabalho na análise do critério baseado em correntropia para fontes pré-codificadas. Para este propósito, foram derivadas as expressões da correntropia analítica para fonte BPSK (sem ruído) e para fonte contínua (com ruído). Foi também derivada a solução de Wiener para o sistema com pré-codificador, com o objetivo de realizar o ajuste do pré-codificador $P(z)$. Por fim, foi realizada uma investigação dos principais parâmetros no critério baseado em correntropia, considerando tanto a sua versão estimada como a sua versão analítica.

No Capítulo 2, foram introduzidos os principais conceitos e os desafios referentes ao problema de equalização de canais de comunicação, quando o sinal transmitido é composto por amostras independentes, tanto no contexto supervisionado quanto no caso cego. Em particular, são apresentados também os resultados fundamentais de Benveniste-Goursat-Ruget (BGR) e de Shalvi-Weinstein (SW) que sustentam as condições estatísticas necessárias para uma equalização cega. Por fim, é apresentado o problema de equalização de canais para fontes pré-codificadas (i.e., em que o sinal da fonte apresenta dependência temporal), que ocupa um papel central neste trabalho.

O Capítulo 3 apresentou os conceitos e fundamentos referentes ao aprendizado baseado na Teoria da Informação, bem como os seus principais critérios que são empregados tanto na vertente supervisionada quanto na não-supervisionada do problema de equalização de canais. Neste Capítulo também é introduzida a medida de correntropia, cujas propriedades são exploradas pelo critério baseado em sua aplicação, que é objeto direto de estudo e avaliação por este trabalho.

Tendo como motivação a busca de um melhor entendimento do critério baseado em correntropia para equalização não-supervisionada, foi realizado o desenvolvimento teórico da correntropia para diferentes tipos de fontes. Nesse sentido, foram apresentadas no Capítulo 4 as derivações teóricas da correntropia analítica para os três tipos de fontes abordados neste trabalho: Fonte Gaussiana, Fonte BPSK e Fonte Contínua. Primeiramente, é reproduzida a dedução analítica originalmente desenvolvida em (FANTINATO, 2017) para fontes Gaussianas, que também foi objeto de avaliação por este trabalho e que serviu como base para as deduções que foram desenvolvidas. No segundo caso abordado, é apresentada a derivação teórica da correntropia para o cenário em que a fonte está associada à modulação digital BPSK e, que momentaneamente, não considera a presença de ruído no canal. A derivação teórica da correntropia para este cenário é uma das contribuições que foram desenvolvidas por este trabalho (OLIVEIRA *et al.*, 2017). Finalmente, o terceiro caso abordado traz um modelo de fonte novamente contínuo, inspirado no tra-

balho (BOCCATO *et al.*, 2016), que pode ser visto como uma extensão do caso BPSK. A derivação teórica para este tipo de fonte também foi apresentada de forma inédita neste trabalho de mestrado.

Por fim, o Capítulo 5 apresenta uma análise detalhada do critério baseado em correntropia para equalização cega, dedicando especial atenção ao efeito dos seus parâmetros ajustáveis, bem como à análise comparativa entre as suas versões analítica e estimada. As análises seguintes – em termos da superfície de erro e desempenho de equalização – serviram como suporte para entender melhor o critério baseado em correntropia e a influência de cada um dos seus parâmetros no desempenho atingível. Foram também propostos novos critérios de equalização inspirados na noção de casamento de correntropia, mas fazendo uso de normas L_p do vetor de erro entre os perfis de correntropia distintas da Euclidiana. Os testes preliminares realizados foram promissores, apontando ganhos de desempenho, especialmente com a versão do critério de correntropia baseada na norma L_∞ .

É possível concluir, portanto, que o estudo realizado revelou contribuições importantes para o problema de equalização não-supervisionada para fontes pré-codificadas. Nesse contexto, foram providos elementos que auxiliaram para uma melhor compreensão e esclarecimento sobre o critério baseado em correntropia.

Perspectivas para Trabalhos Futuros

Das abordagens seguidas neste trabalho surgiram algumas interessantes possibilidades a serem exploradas em trabalhos futuros:

- Análise do critério baseado em correntropia para um conjunto mais diversificado de cenários, incluindo aqueles com canais de fase mista e/ou não-lineares. Insere-se também nesse estudo o comportamento das funções custo, que visará esclarecer, por exemplo, a formação de platôs nas superfícies a partir de variações do canal.
- Estudo de alguma técnica para projeto de pré-codificador que seja capaz de embutir tal dependência temporal de modo a maximizar o desempenho alcançado pelo critério baseado em correntropia. Também seria interessante, olhando a propriedade da correntropia de explorar estatísticas de ordem elevada, empregar alguma estrutura de pré-codificador que seja capaz de produzir um sinal com dependência não linear, ou seja, embutir alguma estrutura temporal que seja mais eficiente para o uso da correntropia. O emprego da correntropia se mostrou promissor para lidar com sinais que apresentam outros tipos de estrutura estatística e temporal, como, por exemplo, em aplicações que se mostraram promissoras para sinais provenientes de

interface cérebro-computador (BCI, do inglês *Brain-Computer Interface*) (URIBE *et al.*, 2017).

- Explorar outras formulações da noção de casamento de correntropia entre o sinal do pré-codificador e o sinal do equalizador, como, por exemplo, pelo emprego de normas ponderadas ou mesmo aprofundando o estudo inicial que foi feito por este trabalho para as normas L_1 e L_∞ .
- Avaliar o comportamento dos critérios baseados em casamento de correntropia para estruturas mais complexas, como filtros IIR e redes neurais.

Referências

- BÄCK, T.; FOGEL, D. B.; MICHALEWICZ, Z. *Evolutionary computation 1: Basic algorithms and operators*. [S.l.]: CRC press, 2000. v. 1. Citado na página 61.
- BARRY, J. R.; LEE, E. A.; MESSERSCHMITT, D. G. *Digital communication*. [S.l.]: Springer, 2003. Citado na página 23.
- BOCCATO, L.; FANTINATO, D. G.; SILVA, D. G.; FERRARI, R.; NEVES, A.; ATTUX, R. Analysis of itl criteria in the context of fir channel equalization. *Journal of Communication and Information Systems*, v. 31, n. 1, 2016. Citado 4 vezes nas páginas 50, 53, 58 e 109.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 36.
- COVER, T. M.; THOMAS, J. A. *Elements of information theory*. [S.l.]: John Wiley & Sons, 2006. Citado na página 35.
- EMMERT-STREIB, F.; DEHMER, M. *Information theory and statistical learning*. [S.l.]: Springer, 2009. Citado 2 vezes nas páginas 39 e 40.
- ERDOGMUS, D.; PRINCIPE, J. C. An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems. *IEEE Transactions on Signal Processing*, IEEE, v. 50, n. 7, p. 1780–1786, 2002. Citado na página 45.
- FANTINATO, D. G. New methods for adaptive equalization based on information theoretic learning. *Universidade Estadual de Campinas - UNICAMP*, 2017. Citado 4 vezes nas páginas 50, 51, 57 e 108.
- FANTINATO, D. G.; NEVES, A.; ATTUX, R. Analysis of a novel density matching criterion within the itl framework for blind channel equalization. *Circuits, Systems, and Signal Processing*, Springer, p. 1–29, 2017. Citado 3 vezes nas páginas 17, 33 e 41.
- FERRARI, R.; SUYAMA, R.; LOPES, R. R.; ATTUX, R. R.; ROMANO, J. M. An optimal mmse fuzzy predictor for siso and mimo blind equalization. In: *Proceedings of the IAPR Workshop on Cognitive Information Processing (CIP), Santorini, Greece*. [S.l.: s.n.], 2008. Citado na página 33.
- GODARD, D. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE transactions on communications*, IEEE, v. 28, n. 11, p. 1867–1875, 1980. Citado na página 31.
- GONIN, R. *Nonlinear Lp-norm estimation*. [S.l.]: Routledge, 2017. Citado na página 104.
- HAYKIN, S. *Adaptive filter theory*. [S.l.]: Pearson Higher Ed, 2013. Citado 12 vezes nas páginas 17, 23, 27, 28, 30, 31, 35, 68, 75, 81, 97 e 106.
- LIN, S.; COSTELLO, D. J. *Error control coding*. [S.l.]: Pearson Education India, 2001. Citado na página 33.

- LIU, W.; POKHAREL, P. P.; PRÍNCIPE, J. C. Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Transactions on Signal Processing*, IEEE, v. 55, n. 11, p. 5286–5298, 2007. Citado 2 vezes nas páginas 41 e 46.
- OLIVEIRA, V. A.; FANTINATO, D. G.; FERRARI, R.; ATTUX, R.; BOCCATO, L. Analysis of the correntropy-based criterion for blind equalization with precoded sources. *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBrT2017*, 2017. Citado 3 vezes nas páginas 50, 57 e 108.
- PARZEN, E. On estimation of a probability density function and mode. *The annals of mathematical statistics*, JSTOR, v. 33, n. 3, p. 1065–1076, 1962. Citado na página 35.
- PRICE, K.; STORN, R. M.; LAMPINEN, J. A. *Differential evolution: a practical approach to global optimization*. [S.l.]: Springer, 2005. Citado na página 60.
- PRINCIPE, J. C. *Information theoretic learning: Renyi's entropy and kernel perspectives*. [S.l.]: Springer, 2010. Citado 9 vezes nas páginas 18, 24, 36, 38, 39, 40, 41, 45 e 46.
- RÉNYI, A. *On measures of entropy and information*. [S.l.], 1961. Citado na página 35.
- RICE, J. R.; WHITE, J. S. Norms for smoothing and estimation. *SIAM review*, SIAM, v. 6, n. 3, p. 243–256, 1964. Citado na página 105.
- ROBBINS, H.; MONRO, S. A stochastic approximation method. *The annals of mathematical statistics*, JSTOR, p. 400–407, 1951. Citado na página 45.
- ROMANO, J. M. T.; ATTUX, R.; CAVALCANTE, C. C.; SUYAMA, R. *Unsupervised signal processing: channel equalization and source separation*. [S.l.]: CRC Press, 2011. Citado 9 vezes nas páginas 17, 24, 27, 29, 30, 31, 32, 33 e 35.
- SANTAMARÍA, I.; POKHAREL, P. P.; PRINCIPE, J. C. Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing*, IEEE, v. 54, n. 6, p. 2187–2197, 2006. Citado 7 vezes nas páginas 18, 23, 36, 42, 47, 48 e 103.
- SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal*, Wiley Online Library, v. 27, n. 3, p. 379–423, 1948. Citado na página 35.
- SILVERMAN, B. W. *Density estimation for statistics and data analysis*. [S.l.]: CRC press, 1986. v. 26. Citado 2 vezes nas páginas 39 e 40.
- STEINWART, I.; CHRISTMANN, A. *Support vector machines*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 36.
- SUYAMA, R. *et al.* Sobre o critério do módulo constante para equalização não-supervisionada e suas relações com a teoria de wiener. [sn], 2003. Citado na página 32.
- TREICHLER, J.; AGEE, B. A new approach to multipath correction of constant modulus signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, IEEE, v. 31, n. 2, p. 459–472, 1983. Citado na página 32.

URIBE, L. F. S.; FILHO, C. A. S.; OLIVATTO, V. B.; OLIVEIRA, V.; SORIANO, D. C.; BOCCATO, L.; CASTELLANO, G.; ATTUX, R. A comparative analysis of correlation and correntropy in graph-based brain computer interfaces. *XXXV Simpósio Brasileiro de Telecomunicações, São Pedro, Brasil*, 2017. Citado na página 110.

WIDROW, B.; HOFF, M. E. Adaptive switching circuits. In: MIT PRESS. *Neurocomputing: foundations of research*. [S.l.], 1988. p. 123–134. Citado na página 45.

Anexos

ANEXO A – Solução de Wiener para o Sistema com Pré-codificador

O problema de equalização de canais de comunicação para fontes pré-codificadas pode ser representado pelo diagrama da Figura 72.



Figura 72 – Diagrama de blocos apresentando os elementos envolvidos no problema de equalização de canais de comunicação com fontes pré-codificadas.

Dado um pré-codificador $P(z)$, o equalizador deve tentar recuperar na saída uma estimativa - com algum atraso (α) - do sinal $s(n)$. Definindo o vetor do sinal na saída do pré-codificador $\mathbf{s}(n) = [s(n) \dots s(n - K - D + 2)]^T$, o vetor de ruído $\boldsymbol{\eta}(n) = [\eta(n) \dots \eta(n - K - D + 2)]^T$ e a matriz de convolução do canal

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \dots & h_{D-1} & 0 & \dots & 0 & \dots & 0 \\ 0 & h_0 & \dots & h_{D-2} & h_{D-1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & \dots & \dots & 0 & h_0 & \dots & h_{D-1} \end{bmatrix}, \mathbf{H} \in \mathbb{R}^{K \times (K+D-1)},$$

podemos escrever que:

$$\mathbf{r}(n) = \mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}(n), \quad (\text{A.1})$$

onde:

$$\begin{bmatrix} r(n) \\ r(n-1) \\ \vdots \\ r(n-K+1) \end{bmatrix} = \begin{bmatrix} h_0 s(n) & + & \dots & + & h_{D-1} s(n-D+1) \\ h_0 s(n-1) & + & \dots & + & h_{D-1} s(n-D) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_0 s(n-K+1) & + & \dots & + & h_{D-1} s(n-K-D+2) \end{bmatrix} + \begin{bmatrix} \eta(n) \\ \eta(n-1) \\ \vdots \\ \eta(n-K+1) \end{bmatrix}$$

Seja o vetor do sinal não-codificado $\mathbf{u}(n) = [u(n) \dots u(n - K - D - R + 3)]^T$ e a matriz de convolução do pré-codificador

$$\mathbf{P} = \begin{bmatrix} p_0 & p_1 & \dots & p_{R-1} & 0 & \dots & 0 & \dots & 0 \\ 0 & p_0 & \dots & p_{R-2} & p_{R-1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & \dots & \dots & 0 & p_0 & \dots & p_{R-1} \end{bmatrix}, \mathbf{P} \in \mathbb{R}^{(K+D-1) \times (K+D+R-2)}.$$

Agora, pode-se representar o vetor $s(n)$ como uma função do sinal não-codificado $u(n)$:

$$\begin{aligned} \mathbf{s}(n) &= [s(n) \ s(n-1) \ \dots \ s(n-K-D+2)]^T \\ &= \begin{bmatrix} p_0 u(n) & + & \dots & + & p_{R-1} u(n-R+1) \\ p_0 u(n-1) & + & \dots & + & p_{R-1} u(n-R) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_0 u(n-K-D+2) & + & \dots & + & p_{R-1} u(n-K-D-R+3) \end{bmatrix} \end{aligned}$$

Dessa forma, temos que $\mathbf{s}(n) = \mathbf{P}\mathbf{u}(n)$ e podemos escrever:

$$\mathbf{r}(n) = \mathbf{H}\mathbf{P}\mathbf{u}(n) + \boldsymbol{\eta}(n). \quad (\text{A.2})$$

A solução ótima para os coeficientes do vetor $\mathbf{w} = [w_0 \dots w_{K-1}]^T$ do filtro equalizador no sentido de minimizar o erro quadrático será dada por:

$$\mathbf{w}^* = \mathbf{R}_x^{-1} \mathbf{p}_{xs}, \quad (\text{A.3})$$

onde $\mathbf{R}_x = E\{\mathbf{r}(n)\mathbf{r}(n)^T\}$ é a matriz de autocorrelação e $\mathbf{p}_{xs} = \{\mathbf{r}(n)s(n-d)\}$ é o vetor de correlação cruzada. Desenvolvendo a expressão da matriz de autocorrelação \mathbf{R}_x , obtemos:

$$\begin{aligned} \mathbf{R}_x &= E\{[\mathbf{H}\mathbf{P}\mathbf{u}(n) + \boldsymbol{\eta}(n)][\mathbf{H}\mathbf{P}\mathbf{u}(n) + \boldsymbol{\eta}(n)]^T\} \\ &= \mathbf{H}\mathbf{P}\mathbf{P}^T\mathbf{H}^T E\{\mathbf{u}(n)\mathbf{u}(n)^T\} + E\{\boldsymbol{\eta}(n)\boldsymbol{\eta}(n)^T\} \\ &= \mathbf{H}\mathbf{P}\mathbf{P}^T\mathbf{H}^T + \sigma_\eta^2 \mathbf{I}, \end{aligned} \quad (\text{A.4})$$

onde \mathbf{I} representa uma matriz identidade com dimensões $(K+D-1 \times K+D-1)$. Desenvolvendo a expressão do vetor de correlação cruzada \mathbf{p}_{xs} , obtemos:

$$\begin{aligned} \mathbf{p}_{xs} &= E\{\mathbf{r}(n)s(n-d)\} \\ &= E\{[\mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}(n)]s(n-d)\} \\ &= \mathbf{H}E\{\mathbf{s}(n)s(n-d)\}. \end{aligned} \quad (\text{A.5})$$

Se considerarmos $\mathbf{R}_s = E\{\mathbf{s}(n)\mathbf{s}(n)^T\} = \mathbf{P}\mathbf{P}^T$ e desenvolvermos a sua expressão, temos:

$$\mathbf{R}_s = \begin{bmatrix} E\{s(n)s(n)\} & \dots & E\{s(n)s(n-K-D+2)\} \\ E\{s(n-1)s(n)\} & \dots & E\{s(n-1)s(n-K-D+2)\} \\ \vdots & \ddots & \vdots \\ E\{s(n-K-D+2)s(n)\} & \dots & E\{s(n-K-D+2)s(n-K-D+2)\} \end{bmatrix}.$$

Como o atraso de equalização pode variar de $d_{min} = 0$ a $d_{max} = K + D - 2$, o vetor que desejamos - $E\{\mathbf{s}(n)s(n-d)\}$ - corresponde à $(d+1)$ -ésima coluna da matriz $\mathbf{R}_s = \mathbf{P}\mathbf{P}^T$. Para selecionar esta coluna em particular, basta multiplicar \mathbf{R}_s por um vetor coluna \mathbf{z}_f formado por zeros e cujo $(d+1)$ -ésimo elemento é igual a 1. Este vetor é dado por: $\mathbf{z}_f = [0 \dots 1 \dots 0]^T$.

Com estas manipulações, o vetor de correlação cruzada se torna:

$$\mathbf{p}_{xd} = \mathbf{H}\mathbf{P}\mathbf{P}^T\mathbf{z}_f. \quad (\text{A.6})$$

Finalmente, substituindo as expressões de \mathbf{R}_x e de \mathbf{p}_{xd} , dadas pelas Equações (A.4) e (A.6), respectivamente, na definição em (A.3), a solução de Wiener para o sistema com pré-codificador se torna:

$$\mathbf{w}^* = (\mathbf{H}\mathbf{P}\mathbf{P}^T\mathbf{H}^T + \sigma_\eta^2\mathbf{I})^{-1}(\mathbf{H}\mathbf{P}\mathbf{P}^T\mathbf{z}_f). \quad (\text{A.7})$$

É importante salientar que caso o pré-codificador seja representado pelo vetor $\mathbf{p} = [1 \ 0 \ \dots \ 0]^T$, o resultado se torna idêntico ao fornecido pela solução clássica de Wiener.