

**An Immune-Inspired, Dependence-Based
Approach to Blind Inversion
of Wiener Systems**

Stephanie Milena Alvarez Fernandez

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Brasília, Março de 2016

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

An Immune-Inspired, Dependence-Based
Approach to Blind Inversion
of Wiener Systems

Stephanie Milena Alvarez Fernandez

ORIENTADOR: Daniel Guerreiro e Silva

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA
ELÉTRICA

PUBLICAÇÃO: PPGEE.DM - 621/2016
BRASÍLIA/DF: MARÇO - 2016

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia
Mestrado em Engenharia Elétrica

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**An Immune-Inspired, Dependence-Based
Approach to Blind Inversion
of Wiener Systems**

Stephanie Milena Alvarez Fernandez

*Dissertação submetida como requisito parcial de obtenção
de grau de Mestre em Engenharia Elétrica*

Banca Examinadora

Prof. Dr. Daniel Guerreiro e Silva, UnB
Orientador

Prof. Dr. Jugurta Rosa Montalvão Filho, UFS
Examinador externo

Prof. Dr. Francisco Assis Nascimento, UnB
Examinador interno

Brasília, Março de 2016

FICHA CATALOGRÁFICA

FERNANDEZ, STEPHANIE MILENA ALVAREZ

An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems [Distrito Federal] 2016.

vii, 60p., 297 mm (ENE/FT/UnB, Mestre, 2016) Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

1. Sistemas de Wiener

2. Processamento de Sinais

3. Aprendizado Baseado na Teoria da Informação

4. Correntropia

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

FERNANDEZ, S. M. A., (2016). An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems, Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGEE.DM-621/2016, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 60p.

CESSÃO DE DIREITOS

AUTOR: Stephanie Milena Alvarez Fernandez

TÍTULO: An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems

GRAU: Mestre

ANO: 2016

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Stephanie Milena Alvarez Fernandez

SCLN 410 Bloco A Apto 113

70.865-510 Brasília – DF – Brasil.

Dedicatória

Aos meus pais:

Ovani e Yadire, suas vidas são exemplo para mim. O amor que vocês me dão é minha força e alegria.

Aos meus irmãos:

Harol e Giovanni, obrigada por cuidarem sempre de mim.

Stephanie Milena Alvarez Fernandez

Agradecimentos

A Deus, por me iluminar e guiar em todos os momentos da minha vida.

Devo muito ao meu orientador Professor Daniel Guerreiro e Silva pela oportunidade de trabalhar sobre sua excelente tutela, pela paciência ao longo de longas discussões, pelo grande apoio e incentivo e, acima de tudo, por ter me ensinado a possuir grande respeito e admiração pelo campo de processamento de sinais. Trabalhar ao seu lado tem sido uma experiência incrivelmente recompensadora e espero que essa parceria dure por um longo tempo.

Ao Professor Jugurta Rosa Montalvão, ao Professor Romis Ribeiro de Faissol Attux e ao Denis Gustavo Fantinato pela confiança e oportunidade de contribuir neste trabalho, pela disponibilidade e acompanhamento e por todas as correções e sugestões que me permitiram buscar sempre padrões mais altos no meu trabalho. Ao Professor Ugo Silva Dias e ao Professor Francisco Nascimento, aos membros da comissão do Programa de Pós-Graduação do ENE e aos membros da secretaria do ENE, pela constante atuação e colaboração durante o desenvolvimento do meu mestrado.

À minha família e amigos que, direta ou indiretamente, contribuíram para a realização de mais uma etapa significativa em minha vida. Um agradecimento especial à família Marques Marinho pelo apoio manifestado em cada momento durante minha estância no Brasil e pelo acolhimento tão caloroso, em particular à Marcelita por esses bolos cheios de amor.

Ao apoio financeiro da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), o qual foi de vital importância para a realização deste trabalho.

Aos colegas do GPDS, do LATITUDE e do LASP, em especial, ao Helard Becerra Martinez, ao Lucas Soares de Brito e ao Ricardo Kehrlé Miranda pela amizade e companheirismo, pela colaboração e por proporcionarem momentos divertidos nos momentos difíceis.

Finalmente, expresso meu grande amor e agradecimento ao Marco, “meu amor, meu companheiro e tudo”, obrigada pelo seu amor, pelo cuidado e pela paciência. Você sabe que sem o seu apoio eu não tivesse conseguido fazer possível este trabalho.

Sua coragem me inspira.

Sua força me sustém.

Seu amor me salva.

Stephanie Milena Alvarez Fernandez

An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems

Autor: Stephanie Milena Alvarez Fernandez

Orientador: Daniel Guerreiro e Silva

Programa de Pós-graduação em Engenharia Elétrica

Brasília, Março de 2016

Nas últimas décadas, o estudo de métodos para a inversão cega de sistemas de Wiener tem recebido uma atenção significativa, especialmente em áreas como a biologia, química, sociologia e na indústria. Um grande número de métodos tem sido desenvolvidos com diferentes abordagens e análises teóricas do problema, que incluem algoritmos de gradiente para minimizar a taxa de informação mútua do sinal extraído, algoritmos genéticos para executar a tarefa de procurar os parâmetros ótimos assim como algoritmos imuno-inspirados. Estes métodos têm como requisito fundamental que o sinal de entrada seja originalmente i.i.d., além de algumas outras condições de suavidade. Cenários de aplicação que cumprem com este requisito podem ser difíceis de ocorrer, na prática; por isso, considerar fontes não-independentes tem se tornado uma importante abordagem. Neste trabalho, propõem-se dois métodos baseados nas funções de autocorrelação e autocorrentropia para explorar a estrutura do tempo de um determinado sinal, com a finalidade de promover a inversão cega dos sistemas de Wiener usando sistemas Hammerstein. Filtros lineares com e sem realimentação são considerados e um algoritmo imuno-inspirado é usado para permitir a otimização de parâmetros sem a necessidade de manipular analiticamente a função custo, ao mesmo tempo que se aumenta a probabilidade de convergência global. Os resultados experimentais indicam que ambas as funções proporcionam meios eficazes para a inversão do sistema e também ilustram o efeito de realimentação linear sobre o desempenho global do sistema.

Palavras Chave: Sistemas de Wiener, Processamento de Sinais, Aprendizado Baseado na Teoria da Informação, Correntropia.

An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems**Author: Stephanie Milena Alvarez Fernandez****Supervisor: Daniel Guerreiro e Silva****Programa de Pós-graduação em Engenharia Elétrica****Brasília, March of 2016**

In the last decades, the study of blind inversion of Wiener systems has received significant attention, in a special manner in areas such as biology, chemistry, sociology, psychology and industry. A large number of methods have been developed with different approaches and theoretical analysis of the problem, which include a gradient algorithm to minimize the mutual information rate of the extracted signal, genetic algorithms to perform the task of searching for the optimal parameters as well as immune-inspired algorithms. These methods have the particular requirement that the input signal must be i.i.d. and, besides some smoothness conditions. This requirement may be hard to be present in real-world problems, hence, considering non-independent sources have become an interesting approach. In this work, we propose two methods based on the autocorrelation and autocorentropy functions for representing the time structure of a given signal, in order to cope with the unsupervised inversion of Wiener systems by Hammerstein systems. Linear filters with and without feedback are considered and an immune-inspired algorithm is used to allow parameter optimization without the need for explicitly manipulating the cost function, with the additional benefit of increasing the probability of global convergence. The experimental results indicate that both functions provide effective means for system inversion and also illustrate the effect of linear feedback on the overall system performance.

Keywords: Wiener systems, Signal Processing, Information Theoretic Learning, Correntropy.

CONTENTS

1	INTRODUCTION	5
2	INFORMATION THEORY AND ADAPTIVE FILTERING	9
2.1	ENTROPY	10
2.2	MUTUAL INFORMATION	13
2.3	ADAPTIVE FILTERING	15
2.4	INFORMATION THEORETIC LEARNING	18
2.4.1	CORRENTROPY FUNCTION	20
2.5	SUMMARY	22
3	BLIND INVERSION OF WIENER SYSTEMS	23
3.1	RELATED WORK	24
3.1.1	TALEB, SOLÉ-CASALS AND JUTTEN METHODOLOGY FOR BLIND INVERSION OF WIENER SYSTEMS	25
3.1.2	SILVA, <i>et al.</i> IMMUNE INSPIRED METHODOLOGY FOR BLIND INVERSION OF WIENER SYSTEMS	26
3.2	PROPOSAL	27
3.2.1	DEPENDENCE MEASURES	28
3.2.2	THE CLONALG ALGORITHM	29
3.3	SUMMARY	31
4	RESULTS	32
4.1	PRELIMINARY EXPERIMENTS	32
4.2	EXPERIMENTAL ANALYSIS	37
4.2.1	CONTINUOUS CASE	38
4.2.2	DISCRETE CASE	44
4.3	SUMMARY	52
5	CONCLUSION	53
	REFERENCES	55

LIST OF FIGURES

1	Diagrama do sistema de Wiener e do sistema Hammerstein.....	2
1.1	Diagram of Wiener and Hammerstein system.....	6
2.1	Binary entropy function $H_b(p)$ as a function of the probability of p	10
2.2	Venn diagram: Relationship among entropies and mutual information.....	14
2.3	Block diagram of the general adaptive system identification problem.	16
2.4	Block diagram of the general adaptive prediction problem.	16
2.5	Block diagram of the adaptive interference canceling problem.....	17
2.6	Block diagram of the general adaptive inverse modeling problem.....	17
2.7	Correntropy of a zero-mean Gaussian process.....	21
3.1	Block diagram of the proposal to the Wiener system inversion.....	28
4.1	Block diagram of a system with linear precoding.....	32
4.2	Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with Uniform source.	34
4.3	Example of $s(n)$ and $y(n)$, with uniform source in a simple Wiener-Hammerstein configuration.	34
4.4	Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with Laplace source.	35
4.5	Example of $s(n)$ and $y(n)$, with Laplace source in a simple Wiener-Hammerstein configuration.	35
4.6	Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with binary source.....	36
4.7	Example of $s(n)$ and $y(n)$, with binary source in a simple Wiener-Hammerstein configuration.	36
4.8	Kernel performance for the continuous case with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v)\sqrt[3]{ v }$	38
4.9	Kernel performance for the continuous case with parameters $M = 1$, and $Q = 1$ minimum phase system $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and nonlinear function $f[v] = \tanh(3v)$	39

4.10	Kernel performance for the discrete case in the first scenario with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v)\sqrt[3]{ v }$	45
4.11	Kernel performance for the discrete case in the third scenario with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and nonlinear distortion $f[v] = \tanh(3v)$	45
4.12	Experiment with the duobinary $\{+1, -1\}$ signal with the autocorrentropy criterion. The Figure of the left correspond to the signal after passing through the linear channel $H(z) = 1 + 0.5z^{-1}$ and the other one through $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$	50
4.13	Filtered duobinary $\{+1, -1\}$ signal after passing through the nonlinear distortion $f[v] = \tanh(3v)$. The Figure of the left correspond to the signal after passing through the linear channel $H(z) = 1 + 0.5z^{-1}$ and the other one through $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$	50
4.14	Example of algorithm results with duobinary signal source $\{+1, -1\}$, $M = 2$, and $Q = 2$ for third scenario.	51

LIST OF TABLES

1	Comparação de alguns métodos de inversão cega para sistemas não-lineares recentemente disponíveis	3
1.1	Comparison of recently available blind inversion methods for nonlinear systems.	7
3.1	Main characteristics of the present proposal in the context of blind inversion of Wiener systems.....	31
4.1	FIR and IIR filters obtained by the criterion $J_{cor}(\cdot)$ for inverting the channel $H(z) = 1 + 0.5z^{-1}$	33
4.2	PERFORMANCE RESULTS FOR THE FIRST SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	40
4.3	PERFORMANCE RESULTS FOR THE SECOND SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	41
4.4	PERFORMANCE RESULTS FOR THE THIRD SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	42
4.5	PERFORMANCE RESULTS FOR THE FOURTH SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	43
4.6	PERFORMANCE RESULTS FOR THE FIRST SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	46
4.7	PERFORMANCE RESULTS FOR THE SECOND SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	47
4.8	PERFORMANCE RESULTS FOR THE THIRD SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	48
4.9	PERFORMANCE RESULTS FOR THE FOURTH SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$	49

LIST OF SYMBOLS

κ_σ	Gaussian kernel
α	Order of Rényi's entropy
$E[\cdot]$	Mathematical expectation operator
$J(\cdot)$	Cost function
'	First derivative
$f[\cdot]$	Nonlinear function
$g[\cdot]$	Nonlinear function
$\hat{\cdot}$	Estimated value
m	Lag being considered
P	Largest considered lag
N	Sample length
σ	Window width parameter
$*$	Convolution
θ	Equalizer parameters

Sumário

Modelagem inversa adaptativa é uma tarefa importante, tendo em conta o seu vasto horizonte de aplicações práticas na área de processamento de sinais. Isto é devido à necessidade de compreender, analisar, prever e controlar sistemas reais, que tem crescido rapidamente com o avanço tecnológico e industrial [1]. Ao longo dos anos, esta questão recebeu mais atenção devido aos crescentes avanços tecnológicos em diversas áreas do conhecimento humano, inclusive no contexto de estruturas não-lineares, que têm sido eficazes em diversas áreas de aplicações onde a modelagem linear falha: por exemplo, tecnologia de microondas e RF [2, 3], processos químicos [4, 5] e biologia [6, 7]. Sistemas não-lineares também podem ser usados no modelo de controle preditivo [8].

Tradicionalmente, métodos de identificação de sistemas não-lineares assumem que o sinal de referência está disponível. No entanto, é uma situação relativamente comum, no mundo real, não se possuir acesso à entrada do sistema. A inversão cega de sistemas não-lineares torna-se, então, uma ferramenta necessária. O problema de inversão cega (intimamente relacionada com a identificação cega e equalização) de sistemas não-lineares tem uma longa história, com muitos resultados e aplicações teóricas, especialmente em telecomunicações [9].

Os reconhecidos sistemas de Wiener e de Hammerstein são dois tipos de modelos não-lineares usados em muitos domínios, devido à sua simplicidade e significado físico, onde o comportamento em estado estacionário do sistema é determinado totalmente pela não-linearidade estática, enquanto o comportamento dinâmico do sistema é determinada por ambos, a não-linearidade e os componentes do modelo linear [10].

Neste trabalho, consideramos o uso de sistemas Hammerstein para a modelagem inversa de sistemas de Wiener. Em termos mais específicos, um sistema de Wiener consiste de um subsistema linear invariante no tempo, do inglês *Linear Time-Invariant* (LTI) seguido por uma distorção sem memória e não-linear, como ilustrado na Figura 1. Um sistema Hammerstein é apenas um sistema de Wiener estruturalmente invertido, isto é, um bloco estático não-linear é seguido por um bloco de resposta linear. As não-linearidades em sistemas de Wiener e de Hammerstein podem ser contínua e/ou descontínua, contanto que inversíveis. Uma vantagem da distinção em blocos lineares e não-lineares é que a estabilidade do sistema é determinada apenas pelas partes lineares do modelo, que podem ser facilmente verificadas.

Apesar de sua simplicidade estrutural, esta modelagem tem encontrado aplicação na indústria, sociologia e psicologia; é empregada com sucesso para o estudo de fenômenos importantes na biologia e na química [4, 5, 6, 7]. Com efeito, a principal motivação para a utilização de sistemas de

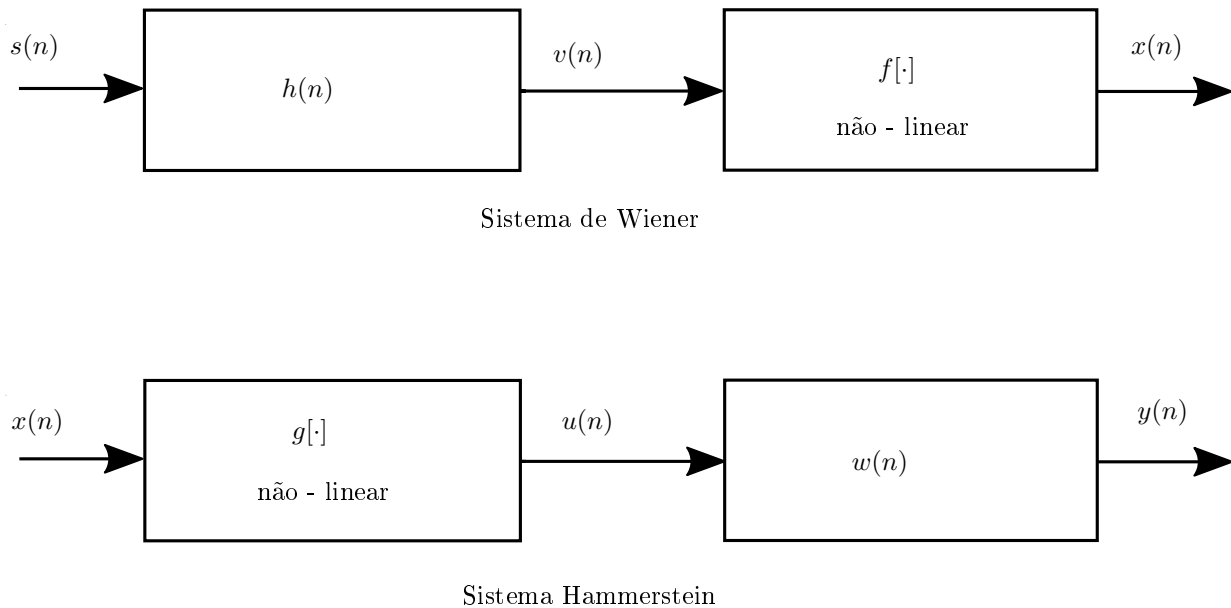


Figure 1: Diagrama do sistema de Wiener e do sistema Hammerstein

Wiener é a manipulação de dinâmicas lineares e não-lineares através de um tratamento matemático simples. A Tabela 1 apresenta uma visão geral e comparação de alguns métodos de inversão cega mais recentes que encontram-se disponíveis para sistemas não-lineares.

Taleb *et al.* [11] e Silva *et al.* [18] propõem métodos para a inversão cega de sistemas de Wiener usando sistemas Hammerstein, nestes trabalhos os autores consideram que $h(n)$, $f[\cdot]$ e $s(n)$ são desconhecidos, mas assume-se que $s(n)$ é composto por amostras independentes e identicamente distribuídas, i.i.d.. Neste contexto, o problema de inversão consiste em encontrar $w(n)$ e $g[\cdot]$ de tal modo que as estatísticas do sinal de saída $y(n)$ do sistema Hammerstein sejam as mais próximas daquelas conhecidas do sinal original $s(n)$, apesar da ambiguidade da escala e atraso do tempo.

No entanto, se as amostras de $s(n)$ são dependentes, as abordagens mencionadas anteriormente não são capazes de obter a solução, uma vez que consideram critérios de independência máxima para estimar os parâmetros ótimos do sistema Hammerstein. Neste caso, é plausível adotar alguma medida de dependência como um novo critério para a tarefa de inversão. Além disso, assumindo amostras dependentes estatisticamente, torna-se possível considerar outros tipos de sinais, por exemplo, sinais codificados.

No contexto do Aprendizado Baseado na Teoria da Informação, do inglês *Information Theoretic Learning* (ITL) [19, 20, 21], uma nova função de correlação generalizada, chamada correntropia, foi recentemente introduzida. Juntamente com a função de autocorrelação, a função de correntropia compartilha o fato de considerar a estrutura de tempo do processo aleatório [22]. No entanto, a correntropia não está limitada a momentos de segunda ordem. Com isto em mente, este trabalho pretende apresentar uma comparação entre estes critérios onde será possível formar uma visão mais clara das suas potencialidades e também de algumas peculiaridades do problema em si, propondo uma versão modificada do método original proposto em [18], considerando critérios baseados na função da autocorrelação e na correntropia, estudando em mais detalhes os aspectos teóricos de

Modelo (abordagem)	Métodos
Wiener-Hammerstein (Taleb <i>et al.</i> , 2001 [11]). Informação Mútua	Entrada: não Gaussiana i.i.d. Não-linearidade e filtro inversível Quase-não paramétrico
Wiener-Hammerstein (Solé Casals <i>et al.</i> , 2002 [12]). Informação Mútua	Entrada: não Gaussiana i.i.d. Paramétrico Redes neuronais artificiais ou polinômios
Wiener-Hammerstein (Babaie-Zadeh <i>et al.</i> , 2003 [13]). Informação Mútua	Entrada: não Gaussiana i.i.d. Não-linearidade e filtro inversível Minimização pelo gradiente
Wiener-Hammerstein (Zhang and Chan, 2004 [14] e Solé Casals <i>et al.</i> 2005 [15]). Informação Mútua	Entrada: não Gaussiana i.i.d. Técnicas de Gaussianização Paramétrico
Wiener-Hammerstein (Rojas <i>et al.</i> , 2007 [16]). Curtose	Entrada: não Gaussiana i.i.d. Paramétrico polinomial Algoritmo genético
Wiener-Hammerstein (Solé-Casals and Caiafa, 2013 [17]). Informação Mútua	Entrada: não Gaussiana i.i.d. Filtro desconhecido e inversível Paramétrico Algoritmo de implementação acelerada
Wiener-Hammerstein (Silva <i>et al.</i> , 2015 [18]). Informação Mútua	Entrada: não Gaussiana i.i.d. Sistema Hammerstein com estrutura FIR ou IIR Algoritmo de otimização imuno-inspirado

Table 1: Comparação de alguns métodos de inversão cega para sistemas não-lineares recentemente disponíveis .

cada critério, potencialidades e novos algoritmos para busca dos parâmetros ótimos.

A primeira parte desta dissertação apresenta os aspectos fundamentais e históricos da Teoria da Informação, uma área do conhecimento fundada pelo Claude E. Shannon, a qual fornece uma base teórica para atividades como observação, medida, compressão e armazenamento de informação e, que tem atraído o esforço de pesquisadores em diferentes áreas do conhecimento humano, tais como economia, física, comunicações e processamento de sinais, entre outras. Então, são introduzidas as idéias principais de Aprendizado de Máquina Baseado na Teoria da Informação, ao mesmo tempo que são apresentadas ferramentas e técnicas de processamento adaptativo de sinais, as quais têm sido usadas no contexto de identificação, predição, cancelamento de ruído e equalização.

Em seguida, é apresentado o problema fundamental tratado nesta dissertação, além de mencionar as características do problema de inversão cega de sistemas de Wiener usando sistemas Hammerstein, são revisados alguns conceitos teóricos necessários à compreensão dos procedimentos propostos e dos resultados obtidos e, com base nessa teoria, são discutidas algumas das recentes metodologias propostas até então, as quais empregam a abordagem de minimização da informação mútua como critério para um processo de busca, bem como a consideração de sinais de entradas compostos com amostras estatisticamente independentes.

Dadas as características do problema e, focando nas oportunidades de contribuições fundamentadas nesta dissertação, a nova proposta inclui (i) a utilização de um critério baseado na correntropia, bem como um critério baseado na função de autocorrelação, que exploram o perfil temporal de sinais estatisticamente dependentes; (ii) a utilização de filtro linear com realimentação no sistema Hammerstein para melhorar a capacidade de inversão; e (iii) um algoritmo imuno-inspirado com risco reduzido de convergência local, responsável pela busca dos valores ótimos dos parâmetros do sistema inverso.

Finalmente, um conjunto de simulações numéricas é apresentado. Neste sentido, um experimento preliminar de caráter qualitativo foi realizado a fim de avaliar o critério baseado na função da correntropia, os resultados obtidos neste estudo prévio permitiram então a realização dos experimentos de análise da nova proposta no contexto do problema Wiener-Hammerstein. Na análise experimental da proposta, as simulações consideram sinais de entrada com amostras estatisticamente dependentes, que sofrem distorção por um canal de fase mínima e uma função não-linear. Os resultados dos experimentos mostram que ambos os métodos estatísticos são viáveis nos cenários avaliados, e também revelam o potencial da utilização de filtros com realimentação no sistema Hammerstein.

Chapter 1

Introduction

Adaptive inverse modeling is an important task, in view of its vast horizon of practical applications in the signal processing area. This is due to the necessity to understand, analyze, predict and control real systems, which has grown quickly with the technological and industrial advance [1]. Over the years, this issue received further attention due to the increasing technological advances in several areas of human knowledge, which led to nonlinear model structures that have been effective in several application areas, where linear modelling has failed: e.g. microwave and RF technology [2, 3], chemical processes [4, 5], biology [6, 7] and predictive control [8].

Traditionally, nonlinear systems identification methods assume that a reference signal is available. However, in a real-world situation, one may have no access to the system input, hence, the blind inversion of nonlinear systems becomes a necessary tool. Blind inversion (closely related to blind identification and equalization) of nonlinear systems has a long development, with many theoretical results and applications, especially in telecommunications [9].

The well-known Wiener and Hammerstein systems are nonlinear models that are employed within many domains, due to their simplicity and physical meaning, where the system steady-state behavior is determined completely by the static nonlinearities, while the system dynamic behavior is determined by both the nonlinearity and the linear model components. Formally speaking, a Wiener system, depicted in Figure 1.1, consists of a linear time-invariant (LTI) filter subsystem, with impulse response $h(n)$, followed by a memoryless, nonlinear distortion $f[\cdot]$. A Hammerstein system (Figure 1.1, bottom part) is just a Wiener system structurally reversed, i.e. a nonlinear static block $g[\cdot]$ is followed by a linear dynamic block with impulse response $w(n)$. The nonlinearities in Wiener and Hammerstein systems can be continuous or discontinuous, as long as they are invertible. An advantage of the distinction into nonlinear and linear blocks is that the system stability is determined solely by the linear parts of the model, which can be easily checked.

In this work, we consider the inverse modeling of a Wiener system. Despite its simplicity, it has been applied in many areas, such as industry, sociology and psychology; moreover, it is successfully employed as a model for important phenomena in biology and chemistry [4, 5, 6, 7]. Indeed, a key motivation for the use of Wiener systems is the handling of linear and non-linear dynamics within a simple mathematical treatment.

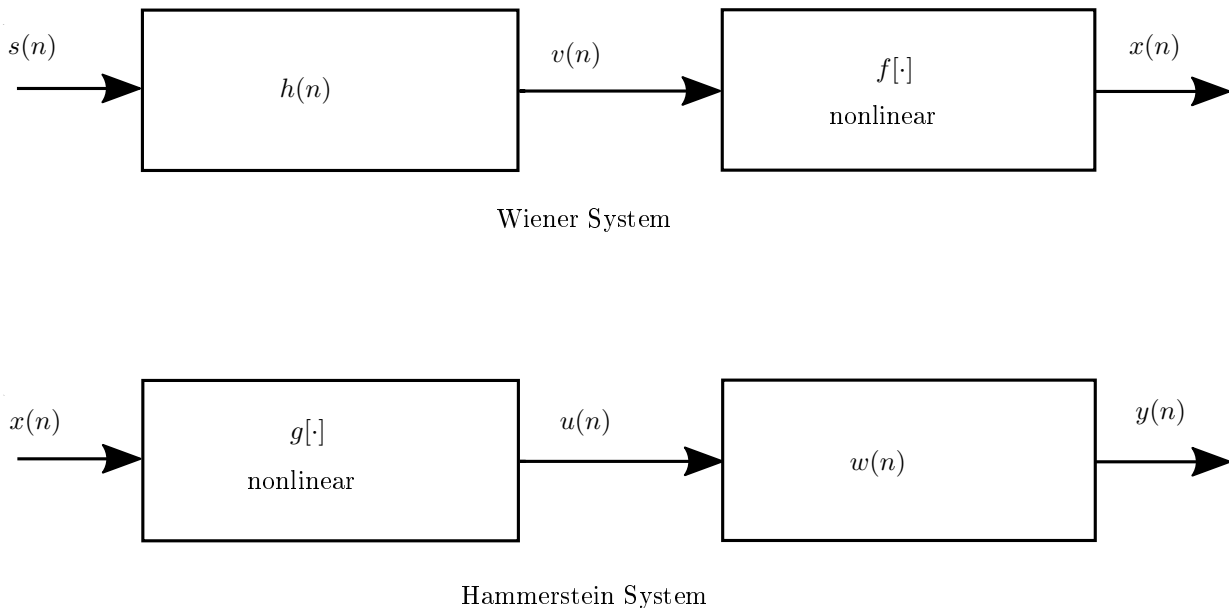


Figure 1.1: Diagram of Wiener and Hammerstein system

Various methods have been developed for the task of Wiener system inversion, such as techniques based on linear optimization, nonparametric regression, and nonlinear optimization with different nonlinear models such as polynomials, artificial neural networks and orthogonal functions [23, 24, 25]. Table 1.1 gives an overview and comparison of the most recently available blind inversion methods for nonlinear systems. Every approach illustrates the kind of source that is employed as the input signal $s(n)$ to the system, the model characteristics for the nonlinearity $g[\cdot]$ and the LTI filter $w(n)$, the criterion chosen to adjust the parameters of the Hammerstein system as well as the search method for estimating the optimal values of these parameters.

Taleb *et al.* [11] and Silva *et al.* [18] proposed blind methods to invert Wiener systems through Hammerstein models, assuming that $h(n)$ and $f[\cdot]$ are unknown, as well as $s(n)$, but with the restriction that its samples are i.i.d.. In this context, the inversion problem consists of finding $w(n)$ and $g[\cdot]$ such that statistics of the output signal $y(n)$ of the Hammerstein system be as close as possible to the known statistics of the original signal $s(n)$, despite a scale and time delay ambiguity.

However, if the samples of $s(n)$ are *dependent*, the previously mentioned approaches are not capable of obtaining the solution, since they consider maximal independence criterion to estimate the Hammerstein system optimal parameters. In this case, it is plausible to adopt some dependence measure as a new criterion for the inversion task. Additionally, by assuming statistically dependent samples, it becomes possible to consider different types of signals, for instance, encoded signals.

Blind inversion of Wiener systems is a challenging problem in signal processing and it has potential applications in various real-world applications. It is important to reinforce that, in the literature, there exists a very wide variety of methods and studies in the context of blind inversion where the input of the Wiener system is assumed originally i.i.d., but in spite of efforts like [11, 12, 13, 15, 16, 17, 18], the inversion task for dependent sources still demands a dedicated

Model (approach)	Methods
Wiener-Hammerstein (Taleb <i>et al.</i> , 2001 [11]). Mutual information	Input: non-Gaussian i.i.d. Invertible nonlinearity and filter Quasi-nonparametric
Wiener-Hammerstein (Solé-Casals <i>et al.</i> , 2002 [12]). Mutual information	Input: non-Gaussian i.i.d. Artificial neural networks or polynomials Parametric
Wiener-Hammerstein (Babaie-Zadeh <i>et al.</i> , 2003 [13]). Mutual information	Input: non-Gaussian i.i.d. Invertible nonlinearity and filter Minimization projection
Wiener-Hammerstein (Zhang and Chan, 2004 [14] and Solé Casals <i>et al.</i> 2005 [15]). Mutual information	Input: non-Gaussian i.i.d. Gaussianization techniques Parametric
Wiener-Hammerstein (Rojas <i>et al.</i> , 2007 [16]). Kurtosis	Input: non-Gaussian i.i.d. Polynomial parametric Genetic algorithm
Wiener-Hammerstein (Solé-Casals and Caiafa, 2013 [17]). Mutual information	Input: non-Gaussian i.i.d. Unknown and invertible filter Parametric Accelerated implementation algorithm
Wiener-Hammerstein (Silva <i>et al.</i> , 2015 [18]). Mutual information	Input: non-Gaussian i.i.d. Hammerstein system with FIR or IIR structure Immune-inspired optimization algorithms

Table 1.1: Comparison of recently available blind inversion methods for nonlinear systems.

study. It should be mentioned that dependent sources are practically important in view of the potential application of different types of codes before signal transmission.

In recent times, Information Theoretic Learning (ITL) has gained attention in the signal processing area [19, 20, 21] and a new generalized correlation function, called correntropy, has been introduced. Correntropy is a positive definite function which yields a generalized similarity measure between random variables (or between time samples of a stochastic process) and it involves higher-order statistics of input signals, therefore it can be a promising candidate for a diverse set of applications in machine learning and signal processing.

Based on the promising properties of correntropy and considering, as well, a traditional dependence measure as the autocorrelation function, in this work, we propose a modified version of the original method proposed in [18] to the blind inversion of Wiener systems. The proposal includes (i) the use of a correntropy-based criterion as well as an autocorrelation-based function that explore the temporal profile of the signal of interest; (ii) the use of linear filter with and without feedback in the Hammerstein system to improve inversion capability; and (iii) an immune-inspired search algorithm with a relatively reduced risk of local convergence.

The outline of this work is as follows:

- In Chapter 2, the concepts of Information Theoretic Learning are introduced, starting from primordial aspects of Information Theory, to then reach its connection with the adaptive signal processing and machine learning theories.
- Supported by the concepts of the previous chapter, Chapter 3 gives a detailed description of the blind inversion of Wiener systems problem, as well as the strategy of employing a Hammerstein system for this task. It is introduced basics concepts of Artificial Immune Systems, which are necessary for the parameters optimization step. Furthermore, two strategies are introduced, in order to quantify the temporal structure of the involved signal and, consequently, to represent the inversion criterion: the autocorrelation and autocorrentropy functions.
- In Chapter 4 the performance of the procedure described in Chapter 3 is evaluated in various scenarios. In addition to the numerical experiments of the proposal, the chapter presents an empirical analysis of the correntropy-based criterion in a qualitative experiment.
- Finally, in Chapter 5, the final remarks about the work and future perspectives are drawn.

Chapter 2

Information Theory and Adaptive Filtering

This chapter initiates the theoretical basis of this dissertation, by presenting fundamental aspects of Information Theory and the definition of its main measures, as well as the Renyi's quadratic entropy and a new generalized correlation function. Notwithstanding, we link those concepts with the main ideas of adaptive filtering and machine learning, leading to the tools that will be able to tackle, in this case, the blind inversion of Wiener systems.

All communication schemes lie in between the two limits on the compressibility of data and the capacity of a channel. Information Theory can offer means to achieve these theoretical limits, more specifically, Information Theory deals with the measurement of information and the representation of it (for example, coding) and the capacity of communication systems to transmit and process information [26]. It was initiated by communication scientists who were studying the statistical structure of electrical communication equipment and was principally founded by Claude E. Shannon in 1948 [27]. A key step in Shannon's work was his realization that, in order to have a theory, communication signals must be treated in isolation from the meaning of the messages that they transmit. Shannon produced a formula that showed how the bandwidth of a channel and its signal-to-noise ratio affected its capacity to carry signals. In doing so, he was able to suggest strategies for maximizing the capacity of a given channel and showed the limits of what was possible with a given technology. This was a great utility to engineers, who could focus consequently on individual cases and understand the specific trade-offs involved. The principles and applications of the Information Theory have deep connections with probability theory, statistics, electrical engineering, signal processing and it has attracted an amount of research effort and provided several insights into many research fields, not only communication and signal processing in electrical engineering, but also physics, computer science, economics, biology, etc [28].

A fundamental concept in Information Theory is that the amount of information contained in a message is a well-defined and measurable mathematical value. We talk about entropy, which does not refer to quantify the amount of data, but the probability of a message within a set of possible messages be received. In most practical applications, the choice is between messages that have

different probabilities of being sent. The term entropy has been borrowed from thermodynamics to designate the average amount of information in these messages.

2.1 Entropy

The concept of information may be too large to be captured by a single definition, however, for any probability distribution, it is possible to define a so-called entropy amount that has many properties which are in accordance with the intuitive sense of what an information measure should be. Shannon's entropy is the primary concept in Information Theory studies and indicates the average degree of uncertainty associated with a given discrete random variable [29]:

$$H(X) = - \sum_{x \in \mathcal{X}} p_X(x) \log_{\alpha}[p_X(x)], \quad (2.1)$$

where $p_X(x)$ is the probability mass function (PMF) and \mathcal{X} is the set of possible values taken by the random variable X .

Eq. 2.1 allows us to speak of the information content or the entropy of a random variable. Note that entropy is a function of the distribution of the random variable, therefore, it does not depend on the values assumed by it, only on their probabilities.

Consider a random variable that takes on only two values, one with probability p and the other with probability $(1 - p)$. Entropy is a concave function of this distribution, and equals 0 if $p = 0$ or $p = 1$:

$$H(X) = -p \log_{\alpha}(p) - (1 - p) \log_{\alpha}(1 - p) \doteq H_b(p), \quad (2.2)$$

where $H_b(p)$ is the binary entropy function. The function $H_b(p)$ is shown in Figure 2.1.

If the \log_{α} in Eq. 2.1 is taken to be 2, then entropy is expressed in *bits*. If it is taken to be the natural log, then entropy is expressed in *nats*. Commonly, entropy is expressed in bits, and unless otherwise noted, we will assume a logarithm with base 2.

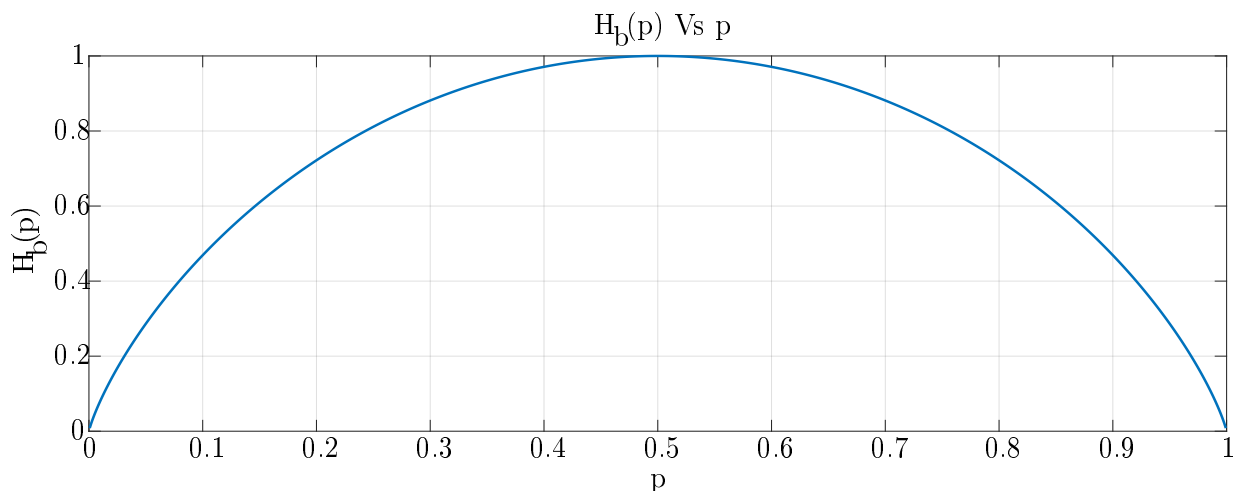


Figure 2.1: Binary entropy function $H_b(p)$ as a function of the probability of p .

The extent of this definition for the continuous case is called differential entropy, and its calculation for a random variable X is:

$$h(X) = - \int_{\mathcal{X}} f_X(x) \ln[f_X(x)] dx, \quad (2.3)$$

where $f_X(x)$ is the probability density function (PDF) of X .

In the following, we introduce another relevant definitions and properties related to $H(X)$, which will be useful for future considerations:

Property 1: $H(X) = 0$ if and only if there is a single event with non-zero probability (i.e., unitary). This means that the entropy is zero when there is no uncertainty about the outcome of a random experiment.

Property 2: Joint entropy is the entropy of a joint probability distribution. The joint entropy of two random variables, X and Y , is defined by

$$H(X; Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{XY}(x, y), \quad (2.4)$$

where $p_{XY}(x, y)$ is the joint PMF of X and Y . This definition gives in turn a significant inequality:

$$H(X; Y) \leq H(X) + H(Y), \quad (2.5)$$

whose equality occurs if and only if X and Y are statistically independent, i.e., when

$$p_{XY}(x, y) = p_X(x)p_Y(y). \quad (2.6)$$

The joint entropy can be written as the sum

$$H(X, Y) = H(X) + H(Y|X), \quad (2.7)$$

where $H(Y|X)$ is the conditional entropy of Y given X . We, of course, also have

$$H(X, Y) = H(Y) + H(X|Y). \quad (2.8)$$

Property 3: The conditional entropy of Y given X is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p_X(x) H(Y|X = x) \\ &= \sum_{x \in \mathcal{X}} p_X(x) \sum_{y \in \mathcal{Y}} p_{Y|X=x}(y) \log p_{Y|X=x}(y) \cdot \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{Y|X=x}(y|x) \end{aligned} \quad (2.9)$$

The joint entropy and conditional entropy are related by equations 2.7 and 2.8. It should seem intuitive that the joint entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other. Thus, if we have three random variables X, Y, Z , the

conditionalizing of the joint distribution of any two of them, upon the third, is also expressed by the Chain Rule:

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z). \quad (2.10)$$

A consequence of the Chain Rule for entropy is that if we have many different random variables X_1, X_2, \dots, X_n , then the sum of all their individual entropies is an upper bound on their joint entropy:

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i). \quad (2.11)$$

Their joint entropy only reaches this upper bound if all of the random variables are independent.

Property 4: From Eqs. 2.5 and 2.7, it is possible to obtain

$$H(Y) \geq H(Y|X). \quad (2.12)$$

which reveals that the uncertainty of a random variable is never increased by knowledge of another variable.

Property 5: The relative entropy or *Kullback – Leibler* distance between two probability mass functions $p_X(x)$ and $q_X(x)$ with the same support \mathcal{X} is defined by

$$D(p \parallel q) = \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}. \quad (2.13)$$

The properties given by the Eqs. 2.5 - 2.13 are directly extensible to the continuous case, just consider integrals instead of sums and PDFs instead of PMFs. Moreover, there are other important properties in Shannon's work related to the continuous case:

Property 6: The differential entropy can be negative.

Property 7: Under the restriction that a continuous random variable is limited to a finite volume of space, the probability density function with maximum entropy is uniform.

Property 8: If the covariance matrix of a continuous random variable is fixed *a priori*, the probability density function with maximum entropy is Gaussian.

The rest of Shannon's work represents an exceptional effort that established a number of key results to sources and continuous channels. Thanks to Shannon's work, Information Theory became a new science topic, and a slew of studies have been developed in order to understand, consolidate and expand the initial contributions made by Shannon.

Additionally, extensions of Shannon's original work have resulted in many alternative measures of information or entropy. For instance, Rényi was able to extend Shannon's entropy to the Rényi entropy of α order [30]. It is expressed, in the discrete case as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} p_X^\alpha(x), \quad (2.14)$$

and for continuous variables

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \int_{\mathcal{X}} f_X^\alpha(x) dx. \quad (2.15)$$

The α parameter of Rényi's entropy allows several uncertainty measurements for the same distribution. Considering the continuous case, two scenarios are important to mention in this context: first, when $\lim_{\alpha \rightarrow 1} h_\alpha(X) = h(X)$, which is Shannon's entropy and finally, when $\alpha = 2$, which is the so-called Rényi's quadratic entropy

$$\begin{aligned} h_2(X) &= -\log \int_{\mathcal{X}} f_X^2(x) dx \\ &= -\log E[f_X(x)], \end{aligned} \quad (2.16)$$

where $E[\cdot]$ denotes the statistical expectation.

The quadratic entropy plays an important role in Information Theoretic Learning because it originates a family of estimators that has interesting features (from the machine learning perspective) such as being non-parametric, continuous and computationally simple to calculate.

2.2 Mutual Information

As we discussed in Section 2.1, Information theory is also capable of dealing with a pair or a collection of random variables. It is able to quantify the amount of information that one variable conveys about the other. Equivalently, the mutual information measures the average reduction in uncertainty about X that results from learning about Y . Mutual information has risen in recent years as an important measure of statistical dependence, mainly in the context of unsupervised learning methods [31, 32]. It is defined as

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}. \quad (2.17)$$

Note that if the two random variables X and Y are independent, then the numerator inside the logarithm equals to the denominator

$$p_{XY}(x, y) = p_X(x)p_Y(y), \quad (2.18)$$

the log term vanishes, and mutual information equals to zero $I(X; Y) = 0$, i.e. no information about X is gained once Y is received.

Mutual information is a nonnegative measure, i.e. $I(X; Y) \geq 0$. In the event that the two random variables are perfectly correlated, then their mutual information is the entropy of either one alone. Another way to say this is: $I(X; X) = H(X)$, the mutual information of a random variable with itself is just its entropy.

These properties are reflected in three equivalent definitions for the mutual information between X and Y :

$$I(X; Y) = H(X) - H(X|Y), \quad (2.19)$$

$$I(X; Y) = H(Y) - H(Y|X) = I(Y; X), \quad (2.20)$$

i.e., it is symmetric and able to detect nonlinear relationships between variables. Eq. 2.20 has made mutual information a very popular criterion for feature selection [33, 34, 35]. And the last definition is given by

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (2.21)$$

The relationship between these measurements are expressed in Figure 2.2, recall that $H(X|Y)$ is the conditional entropy defined in Eq. 2.9, which measures the average uncertainty that remains about X when Y is known.

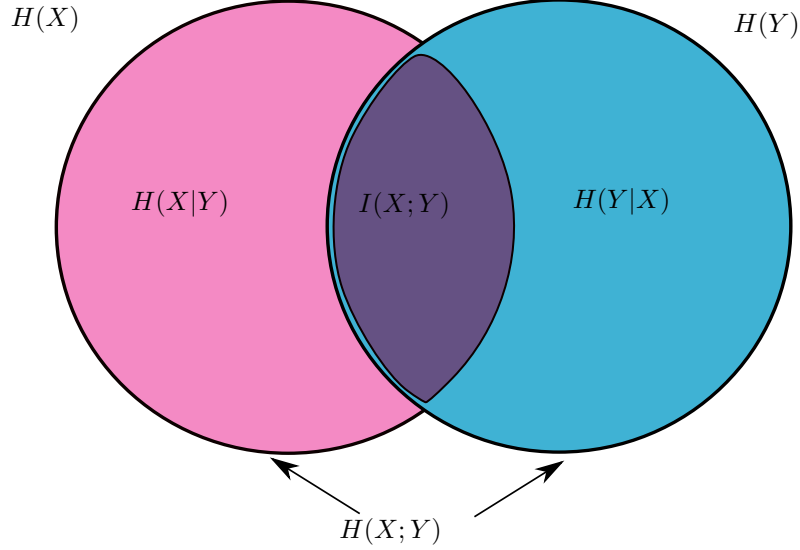


Figure 2.2: Venn diagram: Relationship among entropies and mutual information.

In a sense, the mutual information $I(X;Y)$ is the *intersection* between $H(X)$ and $H(Y)$, since it represents their statistical dependence. In the Venn diagram, the portion of $H(X)$ that does not lie within $I(X;Y)$ is just $H(X|Y)$. The portion of $H(Y)$ that does not lie within $I(X;Y)$ is just $H(Y|X)$.

Similarly to entropy, mutual information can also be extended to the continuous case:

$$I(X;Y) = \int_{\mathcal{Y}} \int_{\mathcal{X}} f_{XY}(x,y) \ln \left[\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} \right] dx dy, \quad (2.22)$$

where $f_{XY}(x,y)$ is the joint probability density function and $f_X(x)$, $f_Y(y)$ are the marginal density functions.

Furthermore, for a random vector $\mathbf{Z} \in \mathbb{R}^m$, it is defined as

$$\begin{aligned} I(\mathbf{Z}) &= I(Z_1; \dots; Z_m) \\ &= \int_{\mathbb{R}^m} f_{\mathbf{Z}}(\mathbf{u}) \log \left(\frac{f_{\mathbf{Z}}(\mathbf{u})}{\prod_{i=1}^m f_{Z_i}(u_i)} \right) d\mathbf{u} \end{aligned} \quad (2.23)$$

and in terms of differential entropy

$$I(\mathbf{Z}) = \sum_{i=1}^m h(Z_i) - h(\mathbf{Z}), \quad (2.24)$$

where $h(Z_i)$ and $h(\mathbf{Z})$ denote, respectively, the differential entropy of Z_i and the joint differential entropy of \mathbf{Z} .

Finally, the notion of entropy and mutual information can also be formalized for a stochastic process, i.e., consider the real-valued, discrete-time process $\mathcal{Z} = \{z(n)\}$, one can define the differential entropy rate, as

$$H(\mathcal{Z}) = \lim_{N \rightarrow \infty} \frac{h(z(-N); \dots; z(N))}{2N + 1} \quad (2.25)$$

when the limit exists, which is true when \mathcal{Z} is stationary [29]. Consequently, from Eq. 2.24 we can also obtain the mutual information rate of the stationary process \mathcal{Z} , given by

$$\begin{aligned} I(\mathcal{Z}) &= \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \left[\sum_{n=-N}^N h(z(n)) - h(z(-N); \dots; z(N)) \right] \\ &= h(z(\tau)) - h(\mathcal{Z}), \end{aligned}$$

where $\tau \in \{\dots, -2, -1, 0, 1, 2, \dots\}$ is arbitrary due to the stationarity assumption.

2.3 Adaptive Filtering

Adaptive filtering involves the changing of filter parameters (coefficients) over time, to adapt to changing signal characteristics. As the signal into the filter continues, the adaptive filter coefficients adjust themselves to achieve the desired result, such as identifying an unknown filter or canceling noise in the input signal. Techniques of adaptive filters are related to the design of filters using statistical methods that take into account statistical information present in the variable to be filtered. The goal in the design process may be to minimize the mean square error of a certain signal, usually defined as the difference between a given reference signal and the filter output. The solution to this problem optimizes the design of such filters, and can be addressed by the Wiener method, which results in the known Wiener filter, an optimal choice in the sense of minimizing the mean square error between a reference and the output of the designed filter, when the input has known and invariant statistical values over time [36, 37]. Adaptive filters that are based on Wiener's criterion make use of learning algorithms to converge to the optimal solution that minimizes the mean square error, having little or no explicit information about the input signals probability distribution.

The development of the most common techniques in adaptive filtering aim to obtain a model that can identify an unknown system. However, these same techniques can be applied to solve various problems. In particular, one can identify four different topologies that allow the adaptive filter to perform various applications, including:

- Identification: In this case, adaptive filtering techniques identify an unknown system making that the input of the system and the adaptive filter, $x(n)$, are the same. The error signal, $e(n) = y(n) - d(n)$, where $y(n)$ is the filter output signal and $d(n)$ is the output signal of the system, is used by the corresponding iterative algorithm to minimize the cost function. The

basic layout used in identification applications is shown in Figure 2.3. It is important to note that successful adaptive systems identification start with choosing correctly the adaptive filter structure. When the system response is oscillatory in nature or has an asymptotic output, an infinite impulse response adaptive filter updated might be used. When the system response is short, an finite impulse response transversal filter is usually preferred for stability reasons [37].

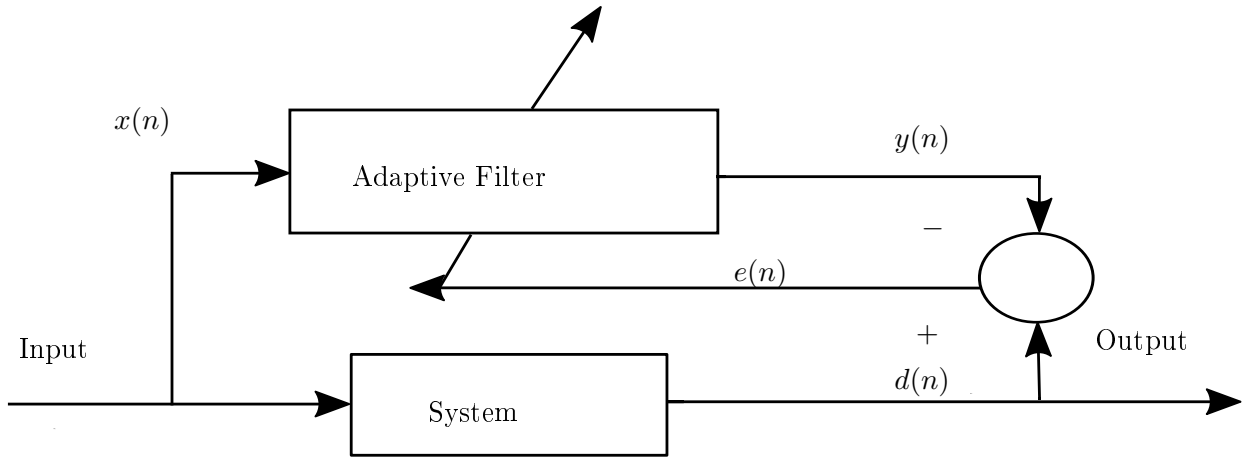


Figure 2.3: Block diagram of the general adaptive system identification problem.

- Prediction: In this case, the adaptive filter is used to predict the current value of a random signal, providing an input to the filter, $x(n)$, composed of previous samples of the mentioned signal, as shown in Figure 2.4. The iterative algorithm responsible for adjusting the adaptive filter coefficients must minimize a cost function where the error $e(n)$, is the difference between the random signal $d(n)$, and the filter output, $y(n)$. Depending on whether the system output is the output of the adaptive filter or the error signal $e(n)$, it is obtained the predicted random signal or the error respectively, which in this case yields the prediction error filter [37].

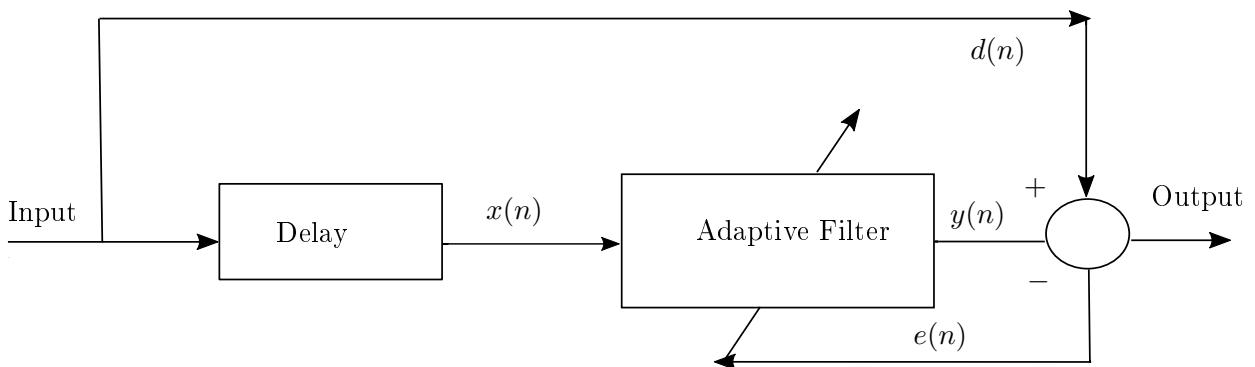


Figure 2.4: Block diagram of the general adaptive prediction problem.

- Interference cancellation: In this case, adaptive filtering is intended to remove certain interference present in the input signal $d(n)$ to the system. To achieve this result, reference $x(n)$

is often used as input to the adaptive filter, comprising a signal correlated with the interference. In Figure 2.5 a basic outline of the configuration of this type of application is shown. A frequently mentioned application of adaptive interference canceling is cleaning power-line interference from weak sensor signals. This is essential in applications such as recording electrocardiograms (ECGs) [38], weak vibration measurements, audio frequency measurements using microphones, and many other applications that employ sensors to collect input data [39]. The interference can be occasionally reduced by proper grounding and using shielded cables, but can not be completely eliminated.

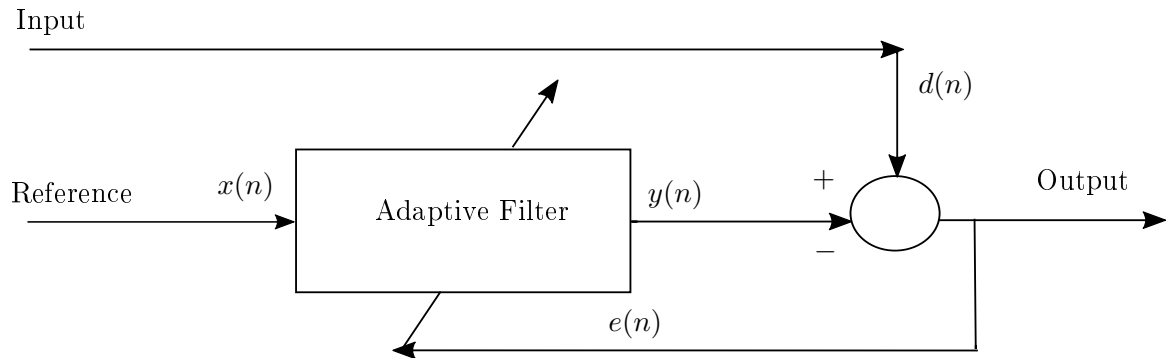


Figure 2.5: Block diagram of the adaptive interference canceling problem.

- Inverse modeling: In this case, an adaptive filtering is used to dynamically perform the task of counterbalancing the effects of a system. The inverse modeling using adaptive filtering is shown in Figure 2.6. Inverse modeling has found many practical applications in control systems and communication systems. The most widely used application of this technique is channel equalization, where the physical system is a communication channel and the adaptive filter is referred to as an adaptive channel equalizer filter. Channel equalizers are usually implemented as adaptive transversal FIR filters. Adaptive inverse modeling is precisely the kind of problem under study in this dissertation, but with the restriction of not having access to the reference signal $d(n)$.

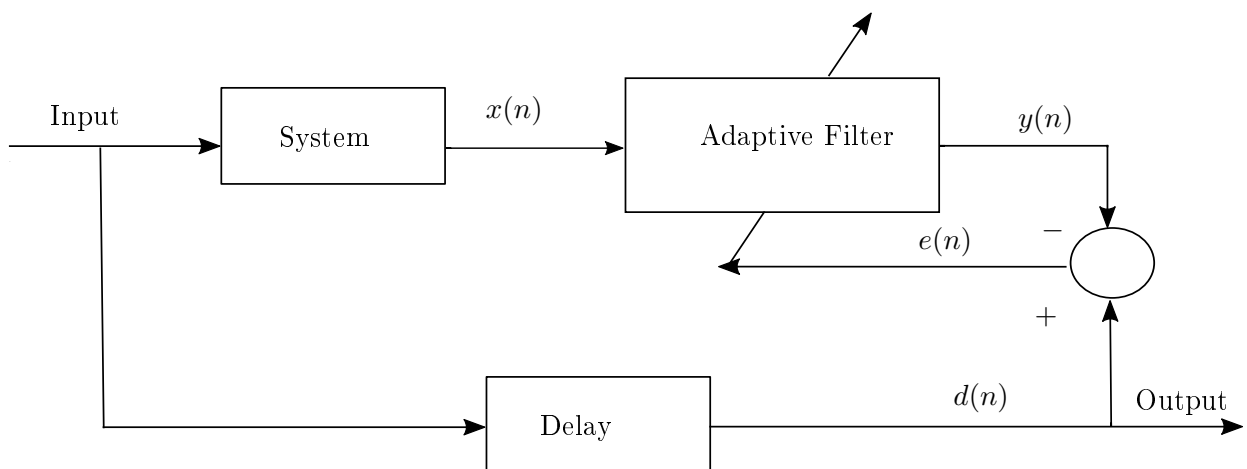


Figure 2.6: Block diagram of the general adaptive inverse modeling problem.

2.4 Information Theoretic Learning

In recent years Information Theory has had an increasing impact on the important issue of extracting information directly from data, i.e., learning from examples. The learning-from-examples scenario starts with a data set that carries information about a real-system, and the objective is to capture the information in the parameters of a learning machine. Machine learning has been extraordinarily successful in providing tools and practical algorithms for extracting information from massive data sets. In this context, the implicit problem usually involves an adaptation process, where the parameters of the learning system are adjustable in a way that implementation improves through repeated presentation of exemplars to the system.

The performance measure that is adopted will determine the type of information which can be extracted from the data. Through the work of Wiener [40], as already introduced in Section 2.3, it was established the possibility of using adaptive filtering structures under a probabilistic perspective. Two premises were fundamental in the consolidation of this approach: the use of linear structures for adaptation and a criterion for setting parameters based on second-order statistics such as the mean square error (MSE), variance and correlation.

Since early researches were concentrated on linear adaptive systems, the adoption of such second-order statistics optimality measures resulted in quadratic performance surfaces, for which the analytical expression of the optimal solution could be easily obtained [41]. Thus, the mainstay of adaptive systems has been second-order statistics criterion. However, for some machine learning problems, second-order statistics may be not sufficient to extract the structure of data. Examples of such problems are typically found on unsupervised tasks, i.e. when there is no reference signal, e.g. blind source separation, blind deconvolution and equalization, clustering, subspace projections [12, 13, 42, 43]. These challenging scenarios make necessary the use of cost functions which can capture higher order statistical properties of the data. Such cost functions can be provided by Information Theory, since, as we could see in Sections 2.1 and 2.2, information-theoretic measures are capable of capturing all the data statistics, since they are nonlinear functions of probability densities.

Through the inspiring work of Principe *et al.*, [19], which was the first to formalize the notion of Information Theoretic Learning (ITL), the interest has risen in the use of criteria derived from Information Theory and that would allow to overcome the limitations of second-order statistics. This progress can be achieved with continuous data, as well as with discrete data. Figure 2.6 from Section 2.3 could illustrates a typical ITL system. In information-theoretic machine learning, the output is given by $y(n) = f[\mathbf{W}, x(n)]$, where $x(n)$ is the signal presented to the system, the function $f[\mathbf{W}]$ represents a possibly non-linear data transformation, which depends on the parameters given by \mathbf{W} , $e(n)$ is the error signal (or only the output $y(n)$, if there is no desired response), which is provided as input to the criterion for adjusting the filter parameters. The goal may be to train the system to perform a specific task, according to an information-theoretic criteria.

Specifically, Principe *et al.*, [19] argued that one should make as few assumptions as possible about the structure of the probability density functions in question. The Parzen window method

of PDF estimation is fundamental in all efforts to create algorithms to manipulate entropy. The principal approach of designing practical information-theoretic criteria is by using Rényi's quadratic entropy. Quadratic entropy can be easily integrated with the Parzen window estimator.

The Parzen window estimator is a technique based on the use of kernel functions to approximate the PDF $f_X(x)$ of a vector of continuous random variables X . The problem can be described as follow: let $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ be a set of N i.i.d. m -dimensional observations drawn from an unknown PDF $f_X(\mathbf{x})$. It is assumed that there is an adequate approximation $\hat{f}_X(\mathbf{x})$ given by [44]:

$$\hat{f}_X(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \kappa(\mathbf{x} - \mathbf{x}_i), \quad (2.26)$$

where $\kappa(\cdot)$ is called the Parzen window or kernel. Kernel functions are a special class of functions that meets the required properties for the role of window function and, therefore, are the most used in the method. More specifically, the circular Gaussian kernel function is one of the most widely adopted, with the expression:

$$\kappa_\sigma(\mathbf{x}, \mathbf{x}_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right), \quad (2.27)$$

where σ is the window width parameter. The Parzen window method using a Gaussian kernel is given by

$$\hat{f}_X(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \kappa_\sigma(\mathbf{x} - \mathbf{x}_i). \quad (2.28)$$

With a method to estimate the probability function that describes the data, one can apply its expression to Shannon's and Rényi's differential entropy. When Shannon's entropy is used along with this PDF estimation strategy, an algorithm to estimate entropy becomes complicated as Viola [45] realized. Fortunately, Rényi's quadratic entropy can be easily integrated with the Parzen window estimator, hence providing a means to estimate the entropy directly from the data set. As we defined in Section 2.1, Rényi's entropy quadratic is given by Eq. 2.16. Since the logarithm is a monotonic function, the quantify of interest is $V(X) = \int_X f^2(\mathbf{x})d\mathbf{x}$, which is called the *information potential* [19]. We have then

$$\begin{aligned} \hat{V}(X) &= \int_X \frac{1}{N} \sum_{j=1}^N \kappa_\sigma(\mathbf{x} - \mathbf{x}_j) \frac{1}{N} \sum_{k=1}^N \kappa_\sigma(\mathbf{x} - \mathbf{x}_k) \\ &= \frac{1}{N^2} \int_X \left[\sum_{j=1}^N \sum_{k=1}^N \kappa_\sigma(\mathbf{x} - \mathbf{x}_j) \kappa_\sigma(\mathbf{x} - \mathbf{x}_k) \right] d\mathbf{x} \\ &= \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \int_X \kappa_\sigma(\mathbf{x} - \mathbf{x}_j) \kappa_\sigma(\mathbf{x} - \mathbf{x}_k) d\mathbf{x} \\ &= \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \kappa_{2\sigma}(\mathbf{x}_j - \mathbf{x}_k) \end{aligned} \quad (2.29)$$

Rényi's quadratic entropy estimator is the starting point for various ITL criteria, such as the MEE (minimum entropy error) supervised criterion and also for new information measures, such as correntropy, which we discuss in the following.

2.4.1 Correntropy Function

Correntropy or, more specifically, the autocorrentropy function was first introduced by Santamaria *et al.* [46], who suggested an initial application to blind deconvolution. Its name stresses the connection to correlation, but also indicates the fact that its mean value across time or dimensions is associated with entropy, more precisely to the information potential, which is the argument of the log in Renyi's quadratic entropy estimated with Parzen windows as was derived in Eq. 2.29. This relation to quadratic entropy shows that correntropy contains information beyond second-order moments.

Correntropy generalizes the autocorrelation function to nonlinear spaces: if $\{\mathbf{x}_i, i \in N\}$ is a stochastic process within an index set N , then the correntropy function $V(i_1, i_2)$ is defined as

$$V(i_1, i_2) = E[\kappa_\sigma(\mathbf{x}_{i_1} - \mathbf{x}_{i_2})], \quad (2.30)$$

where $E[\cdot]$ denotes the statistical expectation and $\kappa_\sigma(\cdot)$ is the Gaussian kernel, given in Eq. 2.27.

Using a Taylor series expansion for the Gaussian kernel, Eq. 2.30 can be rewritten as

$$V(i_1, i_2) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E \|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|^{2n} \quad (2.31)$$

which involves all the even-order moments of the random variable $\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|$. More specifically, the term corresponding to $n = 1$ in Eq. 2.31 is proportional to

$$E[\|\mathbf{x}_{i_1}\|^2] + E[\|\mathbf{x}_{i_2}\|^2] - 2E[\langle \mathbf{x}_{i_1} - \mathbf{x}_{i_2} \rangle] = \sigma_{\mathbf{x}_{i_1}}^2 + \sigma_{\mathbf{x}_{i_2}}^2 - 2C_x(i_1, i_2) \quad (2.32)$$

where $C_x(i_1, i_2)$ is the covariance function of the random process; this shows that the information provided by the conventional covariance function (the autocorrelation for zero mean processes) is included within the correntropy.

For a discrete-time strictly stationary stochastic processes, correntropy can be easily estimated through the sample mean:

$$\hat{V}(m) = \frac{1}{N - m + 1} \sum_{n=m}^N \kappa_\sigma(\mathbf{x}_n - \mathbf{x}_{n-m}), \quad (2.33)$$

where N is the size of the data window used to estimate the correntropy, m is the lag being considered and \mathbf{x}_n is the data samples available $\{\mathbf{x}_n, n = 0, 1, \dots, N - 1\}$.

From the definition of correntropy, the kernel size σ is a crucial factor, since it influences in the nature of the performance surface, like the presence of local optima and rate of convergence [47].

Correntropy has interesting properties that make it useful for machine learning and signal processing. Several properties of correntropy and their proofs are presented in [46, 48]. Here we present, without proof, only the ones that are relevant to this dissertation.

Property 1: $V(m)$ is a symmetric function $V(-m) = V(m)$.

Property 2: $V(m)$ reaches its maximum at the origin, i.e., $V(m) \leq V(0), \forall m$.

Property 3: $V(m) \geq 0$ and $V(0) = 1/\sqrt{2\pi}\sigma$.

Property 4: Let $\{x_n \in \mathcal{R}, n \in N\}$ be a discrete-time wide-sense stationary, zero-mean Gaussian process with autocorrelation function $R[m] = E[x_n x_{n-m}]$. The correntropy function for this process is illustrated in Figure 2.7 and is given by

$$V(m) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}, & m = 0 \\ \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma^2[m])}}, & m \neq 0 \end{cases}, \quad (2.34)$$

where σ is the kernel size and $\sigma^2[m] = 2(R[0] - R[m])$. Since the correntropy function is defined as $V(m) = E[\kappa_\sigma(x_n - x_{n-m})]$, where x_n is a zero-mean Gaussian random process, for $m \neq 0$, $z_m = x_n - x_{n-m}$ is also a zero-mean Gaussian random variable with variance $\sigma^2[m] = 2(R[0] - R[m])$. Therefore

$$V(m) = \int_{-\infty}^{\infty} k_\sigma(z_m) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z_m^2}{2\sigma^2[m]}\right) dz_m. \quad (2.35)$$

Since we are considering a Gaussian kernel with variance σ^2 , Eq. 2.35 is the convolution of two zero-mean Gaussians of variances σ^2 and $\sigma^2[m]$ evaluated at the origin; this yields Eq. 2.34.

Property 4 illustrates that correntropy function conveys information about the time structure of the process.

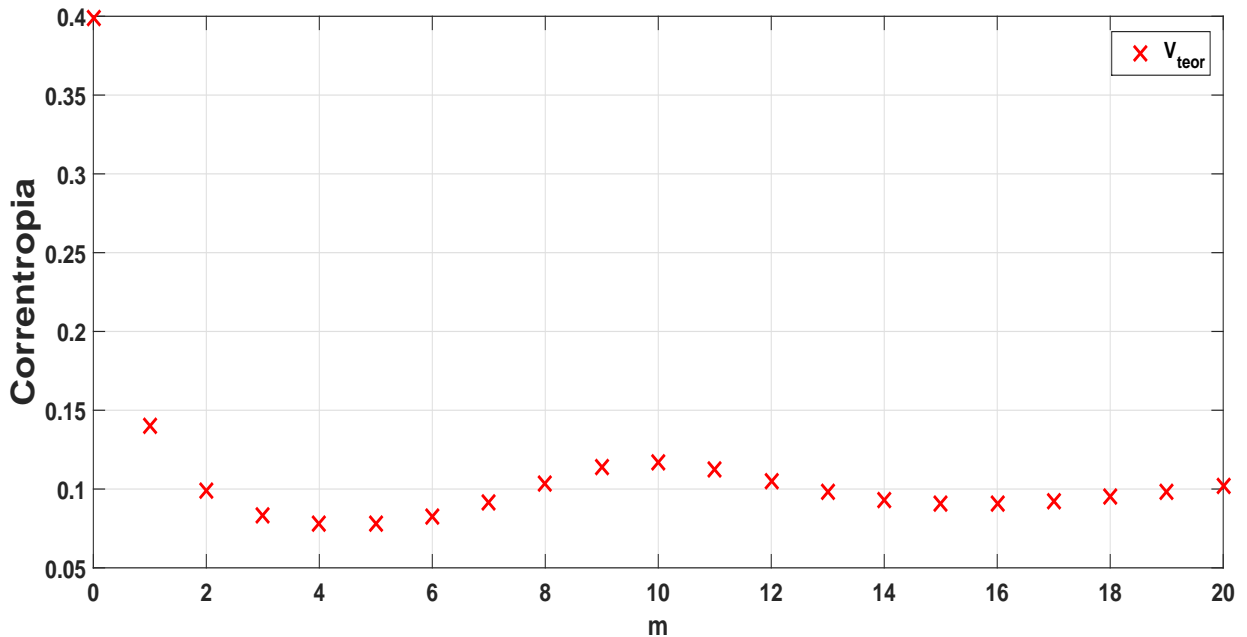


Figure 2.7: Correntropy of a zero-mean Gaussian process.

2.5 Summary

This chapter presented the fundamental and theoretical aspects of Information Theory as well as the main ideas of adaptive filtering, proposing a series of techniques and tools in the context of learning machine. Finally, it is presented the main aspects and the origins of the recent introduced area called Information Theoretic Learning, the mentioned concepts and criteria can be used for a very large number of different applications. In the next chapter is shown the appliance of them in the context of blind inversion of Wiener systems.

Chapter 3

Blind Inversion of Wiener Systems

As we discussed in Chapter 1, the study of nonlinear systems identification/inversion has received much attention in recent decades, since the demand to understand, analyze, predict, and control real systems have grown rapidly with technological and industrial advances. Identification and/or the inversion of such systems is a well-established field with a number of approaches and algorithms. The key task of system identification is to find out a best suitable mathematical model. Models can be useful for gaining a better understanding of the system and to predict or simulate a system's behavior. After selecting the model structure and possible identification strategies, based on the inputs and outputs, the parameters of the model will be determined by an optimization process with minimization (or maximization) of a criterion to solve the data-fitting problem. A number of successful methods are available identify/invert nonlinear systems, in time or in frequency domain, using iterative or non iterative schemes. Actually, all physical systems are nonlinear to an extent. A system is called nonlinear if the input-output steady state relation is nonlinear. Because nonlinear models are able to describe the system behavior in a much larger operating region than corresponding linear models, it is reasonable and necessary to characterize or predict the behavior of real nonlinear processes directly using nonlinear models to improve inversion/identification performance over the whole operating range [10].

Over the years, this issue received further attention and has become a demanding problem in signal processing, with practicable applications in various real-world applications. While all the blind techniques for inversion of a nonlinear model like the Wiener system require the source signal to be independent, non-independent sources do exist in many applications. Inherent dependence character exist in some types of sources such as images [34, 49], or multiple-input multiple-output wireless relay systems [50, 51], to name a few.

In order to describe adequately the nonlinear behavior of the system over the entire range of operating conditions, a nonlinear block-oriented model is often used and the system is generally subdivided into linear dynamic sub-systems and nonlinear static sub-systems. In this work, as already mentioned in Chapter 1, the inverse modeling of a particular and appealing model, the Wiener system, is considered. A Wiener system (remember Figure 1.1 from Chapter 1), is composed of a linear time-invariant (LTI) sub-system $h(n)$ followed by a memoryless, nonlinear distortion

$f[\cdot]$. Despite its simplicity, it can be applied within many contexts [1, 5, 6, 7], by providing a simple mathematical treatment for handling nonlinear dynamics. The following describes the main methods that deal, in an unsupervised manner, with this model, plus a critical analysis of the main points to improve, and the chapter is concluded by addressing our proposal based on the autocorrelation and autocorentropy functions for representing the time structure of a given signal, in the context of new criteria for blind inversion of Wiener systems by Hammerstein systems.

3.1 Related Work

Many methods based on linear and nonlinear optimization with different nonlinear models such as nonlinear control, model predictive control, neural networks, fuzzy control, polynomials and combinations of them have been extensively developed to solve Wiener system inversion, thereby achieving desired performance. To date, these approaches have attracted much attention in industrial applications [23, 24, 25]. Specially, one straightforward structure to invert a Wiener system is the well-known Hammerstein system (see Figure 1.1 from Chapter 1), which consists as mentioned before, of a nonlinear memoryless element followed by a linear dynamical sub-system.

As briefly introduced in Chapter 1, Taleb *et al.* [11] and Silva *et al.* [18] proposed blind methods to invert Wiener systems through Hammerstein models, assuming that $h(n)$ and $f[\cdot]$ are unknown, as well as $s(n)$, but with the restriction that the samples are i.i.d. In this context, the inversion problem consists of finding $w(n)$ and $g[\cdot]$ such that the output $y(n)$ of the Hammerstein system be as close as possible to the original signal $s(n)$, despite a scale and time delay ambiguity.

Assume that $\mathcal{S} = \{s(n)\}$ is a discrete-time stochastic process and that both the LTI sub-system with impulse response $h(n)$ and the nonlinear mapping $f[\cdot]$ are invertible. We can represent the output of the Wiener system as:

$$\begin{aligned} x(n) &= f[v(n)], \\ &= f[h(n) * s(n)] \end{aligned} \tag{3.1}$$

where $s(n) * h(n)$ is the convolution between $s(n)$ and $h(n)$. The result is submitted to the nonlinear distortion $f[\cdot]$.

As already mentioned, we consider the Hammerstein system defined in Figure 1.1, to invert the Wiener system. Likewise to the idea underlying Eq. 3.1, we mathematically obtain the model output as

$$y(n) = w(n) * u(n) \tag{3.2}$$

where $u(n) = g[x(n)]$. In the following, the main strategies to cope with the solution of Eq. 3.2 are discussed.

3.1.1 Taleb, Solé-Casals and Jutten Methodology for Blind Inversion of Wiener Systems

Given the i.i.d. hypothesis of the source signal, Taleb *et al.* developed in their work a theoretical analysis of the Wiener system inversion problem that led to the proposal of a gradient algorithm to minimize the mutual information rate of the output signal, hence assuming that the independence recovery (with the known ambiguities), as well. An overview of their results can be studied in [11, 52]. In the context of the Wiener-Hammerstein structure illustrated in Figure 1.1, as \mathcal{S} is stationary and $h(n)$, $w(n)$ are LTI systems, and following the notation introduced in Section 2.2¹:

$$I(\mathcal{Y}) = h_d(y(\tau)) - h_d(\mathcal{Y}), \quad (3.3)$$

where τ is an arbitrary instant. Moreover, Taleb *et al.* establish that, since $w(n)$ is invertible, the criterion represented in Eq. 3.3 can be expressed in terms of the parameters of each sub-system and of $\mathcal{X} = \{x(n)\}$, which yields

$$\hat{I}(\mathcal{Y}) = h_d(y(\tau)) - \frac{1}{2\pi} \int_0^{2\pi} \log |W(\theta)| d\theta - E[\log g'(x(\tau))] - h_d(\mathcal{X}), \quad (3.4)$$

where $W(\theta)$ is the discrete time Fourier transform of $w(n)$ and $E[\cdot]$ is the expected value of a random variable. The above criterion uses all higher order statistics by the means of the entropy function. The minimization of 3.4 is done by a gradient technique with respect to the coefficients of a non-casual, finite impulse response (FIR) structure that represents the linear stage of the inverse system. The input-output relation this model is given by

$$y(n) = \sum_{k=-M_a}^{M_a} a_k u(n-k), \quad (3.5)$$

where $2M_a + 1$ is the number of adjustable coefficients a_k .

The authors of [11] proposed a *quasi* nonparametric algorithm to model the nonlinear mapping $g[\cdot]$. As mentioned before, the algorithm proposed is based on gradient descent and, therefore, vulnerable to risk of inopportune convergence to suboptimal solutions. Afterwards, an alternative with a parametric model was introduced in a following work, by Solé-Casals *et al.* [12], adopting artificial neural networks and polynomial functions. These approaches have shown promising results in applications of seismic data and even in the inversion problem for correlated signals, as long as the correlation filter is known by the user. But it is important to emphasize that the inversion procedure, in both cases, is based on the minimization of the mutual information rate of the inverse system output.

The work of Taleb *et al.* has been further developed by, among others, Babaie-Zadeh [13], Zhang and Chan [14], Solé-Casals [15] and Solé-Casals and Caiafa [17]. The approaches followed by these authors are based on similar ideas and some alternative extensions were pursued, which include the use of the score function difference [13], the use of Gaussianization techniques in order

¹Hereinafter is employed the notation $h_d(\cdot)$ for entropy to avoid confusion with the impulse response of the Wiener sub-system $h(n)$.

to reduce the effects of nonlinearity estimation [14, 15] and the use of *spline* interpolation functions to decrease the cost estimation of score functions [17].

It is appropriate to lay emphasis on the idea that the essence of the algorithm proposed in [11], as well of its subsequent extensions is some method of gradient-based search as well as a minimum mutual information criterion.

Also, as an alternative to gradient-based algorithms, Rojas *et al.* [16] employ a genetic algorithm (GA), in order to perform the search task for the optimal values of the parameters. The nonlinear part of the Hammerstein system may be approximated by n -th order odd polynomials. Similarly to the methodology proposed by Taleb *et al.*, the linear part is an anti-causal FIR filter and the criterion chosen is the maximization of the Hammerstein system output *kurtosis*.

These works, despite their differences, still employ the premise of i.i.d..

3.1.2 Silva, *et al.* Immune Inspired Methodology for Blind Inversion of Wiener Systems

Section 3.1.1 discussed that the core of the algorithm proposed by Taleb *et al.* is a method of gradient-based search, which means that the obtained solutions may be suboptimal. Silva *et al.* [18], based on these statements and the limitations of other previous works [11, 12, 13, 16], proposed a new approach to the blind inversion of Wiener systems. They proposed a solution that combines immune-inspired search with information-theoretic criteria. The algorithm proposed a new framework bringing together [52]:

- A novel estimation strategy to derive the mutual information-based cost function: the first criterion is the sum of pairwise mutual information [13], which estimates the value of mutual information between each pair of signals, by employing the mutual information estimator based on adaptive partitioning of the observation space

$$\hat{J}_I(g, w) = \sum_{m=1}^D \hat{I}_A(y(n); y(n-m)), \quad (3.6)$$

where D is the number of lags that are considered and \hat{I}_A is the mutual information estimated according to Darbellay *et al.* in [42]. The second criterion is based on the definition of mutual information rate

$$\hat{J}_R(g, w) = \hat{h}_o(y(\tau)) - \frac{1}{K} \sum_{k=0}^{K-1} \log |W_K(k)| - \frac{1}{N} \sum_{i=1}^N \log g'(x(i)). \quad (3.7)$$

The first term of the equation $\hat{h}_o(y(\tau))$ is the order-statistics based entropy estimator, proposed by Pham [53] in the context of blind source separation, the second term is obtained by employing the K -point discrete Fourier transform (DFT) of $w(n)$, denoted by $W_K(k)$ and finally the last term is the sample mean estimator over the N samples of the signal $x(n)$. The derivative of $g(x)$ is analytically evaluated according to the parametric model chosen to represent the nonlinearity.

- Infinite impulse response (IIR) filters in the Hammerstein system to improve inversion capability: the Hammerstein system has a linear part that can be composed of either FIR or IIR filters, thus increasing the potential of effective inversion under a limited number of free parameters, whose input-output relation is defined by

$$y(n) = \sum_{k=0}^M a_k u(n-k) + \sum_{k=1}^Q b_k y(n-k), \quad (3.8)$$

where $M + 1$ and Q are the number of adjustable coefficients a_k , b_k of the model.

- An immune-inspired optimization algorithm: the CLONALG algorithm is adopted to perform the adaptation of the parameters of the related model.

The authors presented experimental results which indicates that the proposed framework is efficient in performing the inversion task, being more accurate when the second criterion of estimating the mutual information rate of the output signal is employed. Moreover, the adoption of an IIR filter improved the solutions for the cases where the perfect inversion of the original linear sub-system was not possible.

3.2 Proposal

As seen in Section 3.1, the main blind inversion approaches assume that $h(n)$, $f[\cdot]$ and $s(n)$ are unknown, but the latter is assumed to be composed of i.i.d. samples. Based on the assumption that the signals are i.i.d., it is conjectured in the work of Taleb *et al.* 2001 [11] that an independence maximization criterion can retrieve the original signal. Then, the algorithm proposed in [11] is based on gradient descent search, and as a result, is inherently vulnerable to risk of premature convergence to a local minimum. On the other hand, we also found that the technique developed by Rojas *et al.* [16], although it is more robust in terms of the potential for local convergence consider that the original source $s(n)$ is a non-Gaussian i.i.d. process, which makes exploring a limited set of signals. Additionally, Silva *et al.* [18] proposed a solution combining immune inspired algorithms and information theory that yield a robust scheme. Nevertheless, the three methods limit the input of the Wiener system to i.i.d. signals.

Consequently, if the samples of $s(n)$ are dependent, the previously mentioned approaches are not capable of obtaining an appropriate solution. In this case, it is plausible to adopt dependence measures as new criteria for the inversion task, since there are real-world situations - as frequently in sensor and measurement applications [54] - where the system input is dependent. Besides, as we discussed in Chapter 1, by assuming statistically dependent samples, it becomes possible to consider a new set of signals, for instance, encoded signals.

Hence, this work develops a new framework for blind inversion that consider dependence measures as cost function, at the same time that maintains a search strategy by means of population metaheuristics. Figure 3.1 illustrates an overview of the proposal, with each of its characteristic aspects highlighted by the numbered circles. In the following, we detail these aspects and in Section 3.2.2 the CLONALG method is described.

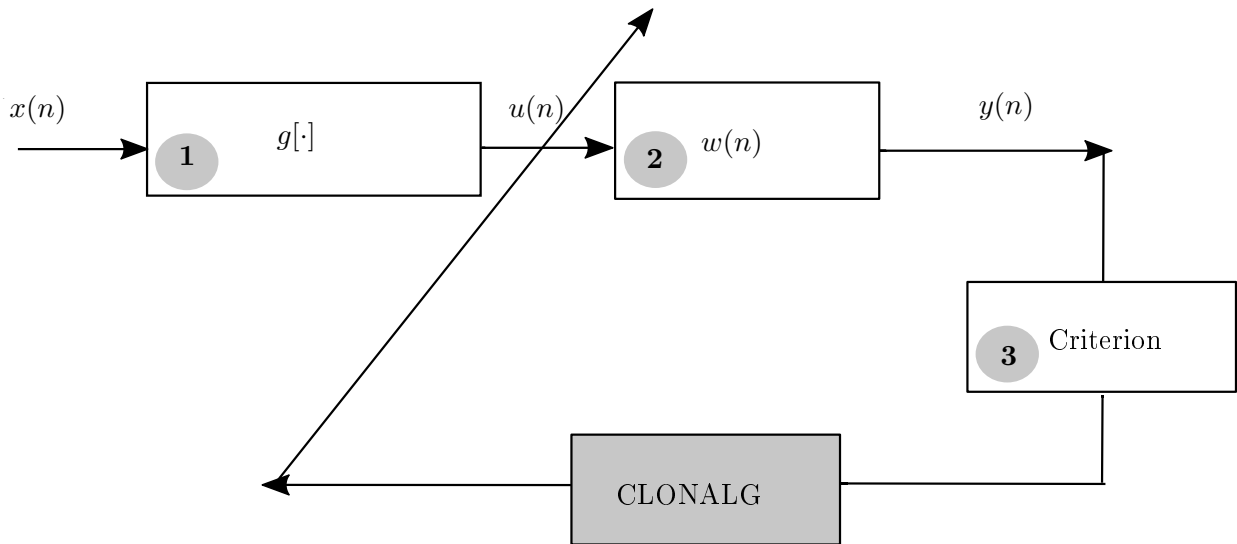


Figure 3.1: Block diagram of the proposal to the Wiener system inversion.

1. The nonlinear function $g[\cdot]$ is defined as a polynomial of k -order (composed entirely of odd degree monomials) with strictly positive coefficients:

$$g[x] = c_1x^1 + c_2x^3 + \dots + c_kx^{2k-1}, \quad c_k \geq 0, \quad \forall k.$$

These restrictions make that the polynomial is monotonically increasing (and therefore invertible), because its first derivative is positive for any value.

2. Despite the adoption by some authors of a FIR model for the linear sub-system [11], we employ, in consonance with [18], a more powerful linear structure which is represented by an IIR filter with finite number of coefficients ($M + 1$ delay taps FIR and Q feedback loops), whose transfer function has been previously described in Eq. 3.8, which theoretically improves the ability to find good solutions within a limited number of coefficients.
3. The optimization criterion is chosen between the two dependence measures described in Section 3.2.1.

Finally, the identification of Hammerstein systems via CLONALG algorithm with real encoding is the search procedure responsible to seek the parameters of $g[\cdot]$ and $w(n)$ that minimize the cost function $J(\cdot)$, assessed by one of the two criteria described in the following section. Due to the difficulties of gradient-based methods to avoid local convergence and stability issues with IIR filters, and based on the successful results, in this context, of the previous proposal of Silva *et al.* [18], this work maintains CLONALG as the optimization method.

3.2.1 Dependence Measures

The straightforward dependence measure is the autocorrelation function, easily defined for a wide sense stationary stochastic process $\{x_n\}$ as

$$R_x(m) = E[x_n x_{n-m}]. \quad (3.9)$$

In the context of the blind deconvolution of signals composed of statistically dependent samples, Fantinato *et al.* [22] employ the autocorrelation function as a criterion to explore the temporal profile of the signal of interest. In this sense, we similarly use an autocorrelation-based criterion, defined as:

$$J_R(\theta) = \sum_{m=1}^P (R_y(m) - R_s(m))^2, \quad (3.10)$$

where $R_s(m)$ and $R_y(m)$ are the autocorrelation functions of the signals $s(n)$ and $y(n)$, respectively, θ are the equalizer parameters and P is the largest considered lag.

Recalling the ideas seen in Chapter 1, this work considers another attractive dependence measure: the correntropy function. This function, as we discussed before in Chapter 2, is an appealing generalization, since it is capable of taking into account the time structure of signals as well as the use of richer statistical information. Currently, one can find several applications of correntropy in different domains, e.g. nonlinear regression, equalization, blind source separation, independence tests [55], etc.. In the context of blind deconvolution, the authors of [46] proposed a correntropy-based criterion to be minimized:

$$J_{cor}(\theta) = \sum_{m=1}^P (V_s(m) - V_y(m))^2. \quad (3.11)$$

Where θ are the equalizer parameters and P is the number of lags. Note that the lag $m = 0$ is not considered since it is always equal to $\kappa(0) = 1/(\sqrt{2\pi}\sigma)$. This criterion tries to match the correntropy $V_s(m)$ associated with the source $s(n)$ to the correntropy $V_y(m)$ of the equalizer output $y(n)$.

In the context of this work, where $s(n)$ is the input to the Wiener system and $y(n)$ is the output of the Hammerstein system, we propose, analogously to the linear blind deconvolution problem, to employ Eqs. 3.11 and 3.10 as criteria to estimate the inverse model. The idea is that the temporal dependence signature of the original signal provides sufficient information to estimate the Hammerstein system and, consequently, to obtain an estimate for $s(n)$. It is important to note that with these criteria, the user does not know the samples $s(n)$, only the autocorrelation or autocorrentropy of $s(n)$.

3.2.2 The CLONALG Algorithm

Artificial Immune Systems (AIS) are algorithms and systems that use the human immune system as inspiration. The human immune system is robust, error tolerant and extremely adaptive. Such properties are highly desirable for the development of novel computer systems. The field of AIS encompasses a spectrum of algorithms, where each one implements different mechanisms of the immune system. All AIS algorithms mimic the behavior and properties of immunological cells, specifically B-cells (a particular type of lymphocyte, white blood cell), T-cells (a type of white blood cell that play a central role in cell-mediated immunity) and dendritic cells (DCs), but the resultant algorithms exhibit different levels of complexity and can perform a wide range of tasks.

The B-cell model seemed mature for exploitation, given the similarities with local search and optimization techniques. In the case of clonal selection principle, this was initially based on works carried in the 1970 by Burnett [56], where affinity metrics were first characterized mathematically, this work served as inspiration for CLONALG [57], a popular AIS algorithm involving an abstract version of the cloning and hypermutation process. All clonal selection-based algorithms essentially center around a repeated cycle of *match*, *clone*, *mutate* and *replace*, and numerous parameters can be tuned, including the cloning rate, the initial number of antibodies, and the mutation rate for the clones.

Standard genetic algorithms and other bio-inspired proposals and immune-inspired algorithms such as CLONALG have generated high-quality solutions to complex problems in signal processing (see, for example, the results of Dias *et al.* [43], Wada *et al.* [58] and Romano *et al.* [59]) and CLONALG particularly has the intrinsic ability of balancing the exploitation of the best solutions with the exploration of the search space, which can be very important to increase the probability of finding the global optimum or a good solution.

The CLONALG algorithm, described in Algorithm 3.1, is inspired in the Clonal Selection principle [56], and is characterized by a population of antibodies, Ab , whose affinity (or fitness) with respect to the antigen Ag is represented by the objective function. The algorithm initializes with an Ab pool of fixed size $N_{initial}$, in which every Ab_i represents an element from the parameter space, possibly an optimal solution. First, the fitness function evaluates $f^{Ag}(Ab_i)$ for all Ab members, then, it proceeds by selecting a subset of n antibodies that have the highest affinities, which are subsequently cloned (nC clones per antibody). The set of clones C are then conducted to an affinity maturation process, where the number of modifications is inversely proportional to their parent's affinity. In the sequence, the clones are compared to their parent in order to select the one with the highest affinity. The main loop is concluded with a random generation of b new antibodies that will replace the lowest affinity Ab in the current population. The process repeats itself until a number of iterations $maxIT$ is executed. After that, the solution is the best individual of Ab (higher affinity).

Algoritmo 3.1 Pseudo-code of CLONALG algorithm for optimization

Require: $[Ab] = \text{clonalg}(N_{initial}, nC, b, \text{range})$

Ensure: $Ab = \text{random}(N_{initial}, \text{range})$

- 1: **while** iteration $\geq maxIT$ **do**
 - 2: Solve $fit = \text{affinity}(Ab)$
 - 3: $C = \text{clone}(Ab, nC)$
 - 4: $C^* = \text{mutate}(C, fit)$
 - 5: $Fit' = \text{affinity}(C^*)$
 - 6: $P = \text{select}(C^*, Fit')$
 - 7: $Ab = \text{replace}(P, \text{random}(b, \text{range}))$
 - 8: **end while**
-

3.3 Summary

In this chapter a new framework was presented. The approach addressed the blind inversion of the Wiener systems with a new dependence-based criteria that explore the temporal structure of a given signal and an immune-inspired algorithm to allow parameter optimization with a significant probability of global convergence. Also, the proposed method employs IIR structure in the Hammerstein system and a set of non-independent sources is considered as the input of the Wiener system. Table 3.1 revisits the main characteristics of the present proposal.

Model (approach)	Method
Wiener-Hammerstein Dependence-based	Input: non i.i.d. Hammerstein system with FIR or IIR structure Immune-inspired optimization algorithm Parametric polynomial

Table 3.1: Main characteristics of the present proposal in the context of blind inversion of Wiener systems.

In the following, the experimental simulations of this new framework covering the inversion task are presented.

Chapter 4

Results

This chapter presents the experimental results of the proposed framework. It begins with a qualitative study of the correntropy-based criterion to determine empirically whether this criterion has the minimum requirements for the recovery of dependent signals. Then, it is presented a more thorough experimental analysis, in order to assess the proposed algorithm behavior.

4.1 Preliminary Experiments

After the presentations of the concepts about the problem of blind inversion of Wiener systems, the existing solution strategies and the proposal that we introduced in this work, it may be interesting to note, in a qualitative experiment, if the correntropy-based criterion of the system output works as a contrast.

The implementation in Matlab consists of performing an exhaustive search within the deconvolution filter parameters in order to build the mapping of the cost function stated by Eq. 3.11, with respect to the parameters of a simplified Hammerstein system.

In this preliminary experiment we have N samples of some correlated source signal $s(n)$, the source distribution changes among uniform, Laplace and a binary signal $\{+1, -1\}$. We consider a baud rate sampled baseband representation of a linearly precoded digital communication system illustrated in Figure 4.1 [46]. An i.i.d. source is linearly precoded at the transmitter to form a sequence of correlated symbols $s(n)$, which mimes the idea of some error coding channel. Through the present work we employ the linear precoder given by $P(z) = 1 + z^{-1}$, which is the filter used in a duobinary modulation.

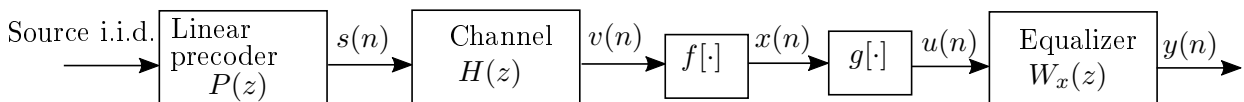


Figure 4.1: Block diagram of a system with linear precoding.

For the three cases of distributions, $N = 10000$ samples, with zero mean and unit variance

are submitted to the system, which after going through the precoder $P(z)$, suffers distortion of a minimum phase channel $H(z) = 1 + 0.5z^{-1}$ and the nonlinearity $f[v] = \text{sign}(v)\sqrt[3]{|v|}$. The number of lags used in the $J_{cor}(\cdot)$ cost function for equalization is $P = 10$ and the kernel size is defined as $\sigma = 0.4$. We have removed the zero lag for reasons already explained in Section 3.2.1. Since it is necessary the knowledge of the autocorrentropy function of $s(n)$, which is a cumbersome task in the continuous case, it is estimated from 2000 samples of $s(n)$, which are distinct from the (unknown to the algorithm) samples to be submitted to the Wiener system. For the binary signal $\{+1, -1\}$, the autocorrentropy function is analytically given by [46]

$$V_s(m) = \begin{cases} \kappa(0), & m = 0 \\ \frac{1}{2}\kappa(0) + \frac{1}{2}\kappa(2), & m = \pm 1 \\ \frac{3}{8}\kappa(0) + \frac{1}{2}\kappa(2) + \frac{1}{8}\kappa(4), & |m| > 1 \end{cases} \quad (4.1)$$

In the first preliminary scenario, the Hammerstein system is modeled by a linear filter defined by the transfer function

$$W_1(z) = \frac{1}{1 + \theta z^{-1}}, \quad (4.2)$$

and nonlinear distortion

$$g[x] = x^3 + \phi x. \quad (4.3)$$

On the second preliminary scenario, the Hammerstein system is modeled by a linear filter defined by the transfer function

$$W_2(z) = 1 + \theta z^{-1}, \quad (4.4)$$

and the same nonlinear distortion defined in Eq. 4.3.

The expected result in both scenarios is that the output signal $y(n)$ is the most similar possible to $s(n-d)$. Table 4.1 shows the MSE between $y(n)$ and $s(n-d)$ for the three distributions as well as the solution obtained for each filter model.

	Uniform	Laplace	Binary
$W_1(z)$	$\frac{1}{1+0.5500z^{-1}}$	$\frac{1}{1+0.5000z^{-1}}$	$\frac{1}{1+0.5000z^{-1}}$
MSE	2.1044^{-04}	1.4116^{-31}	4.5958^{-32}
$W_2(z)$	$1 - 0.5000z^{-1}$	$1 - 0.4500z^{-1}$	$1 - 0.3500z^{-1}$
MSE	0.0737	0.0816	0.0537

Table 4.1: FIR and IIR filters obtained by the criterion $J_{cor}(\cdot)$ for inverting the channel $H(z) = 1 + 0.5z^{-1}$

Figures 4.2 and 4.3 illustrate an example of the results obtained for the uniform distribution, with the linear system $W_1(z)$, analyzing the solution presented in Table 4.1 as well as the mapping of the cost function, we can note that the solution may be considered perfect since the MSE value is small.

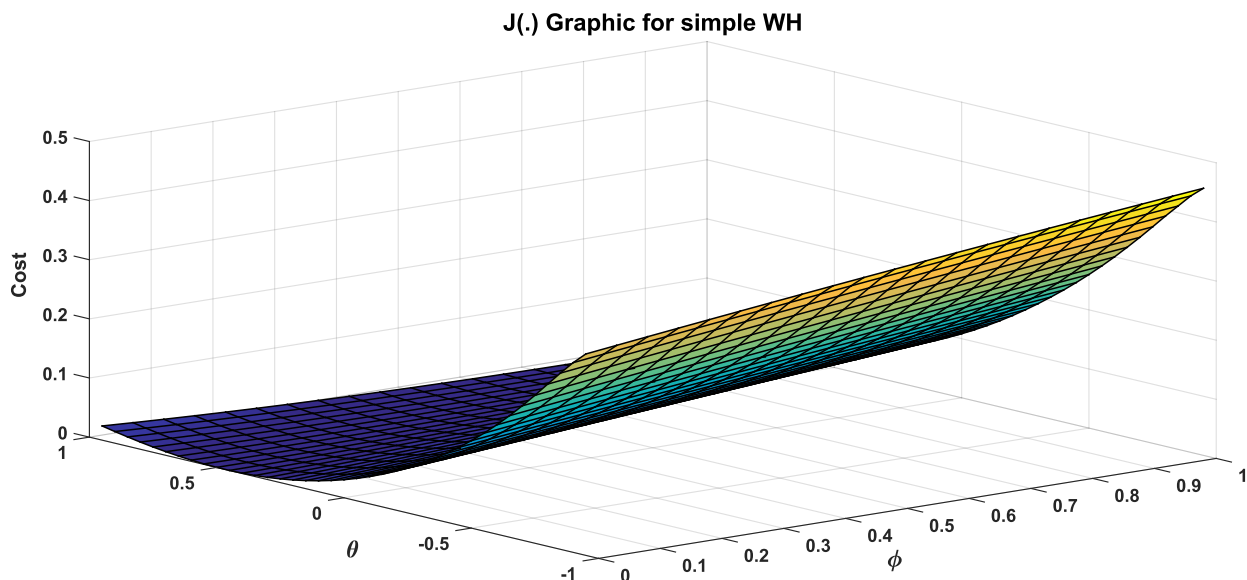


Figure 4.2: Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with Uniform source.

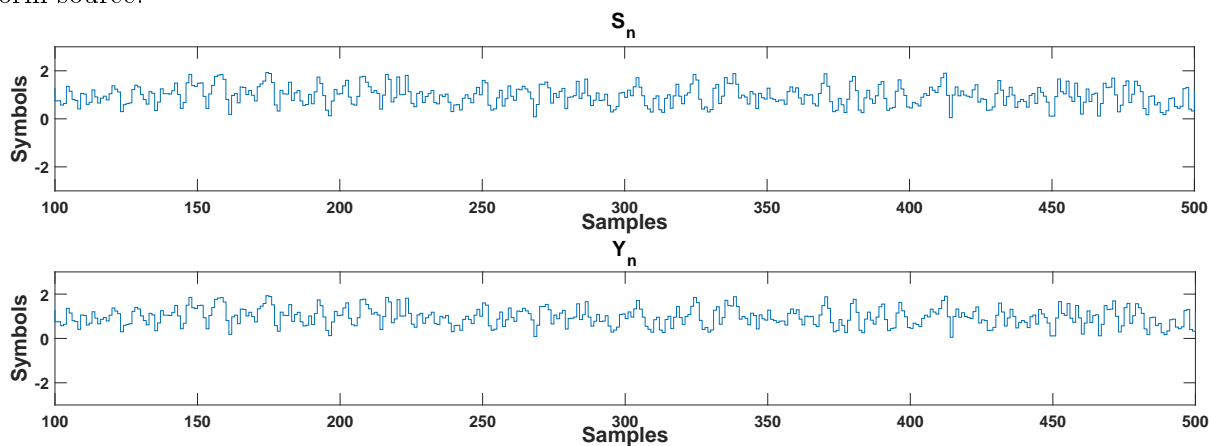


Figure 4.3: Example of $s(n)$ and $y(n)$, with uniform source in a simple Wiener-Hammerstein configuration.

Figures 4.4 and 4.5 illustrate an example of the results obtained in this case for the Laplace distribution, with the linear system $W_1(z)$, note that Figure 4.4 is quite different from what was seen for the uniform scenario, unlike the valley format that we saw in the previous case. Also, the solution provided in this case presented a perfect inversion channel. Moreover, the output signal $y(n)$ is very similar to respect the input signal $s(n)$.

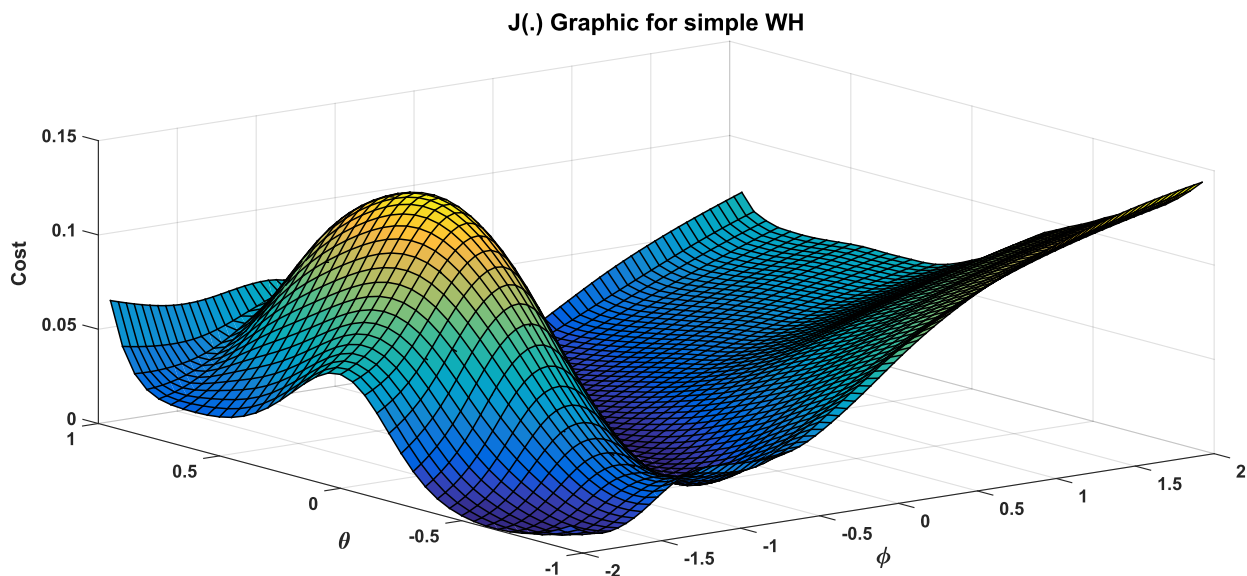


Figure 4.4: Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with Laplace source.

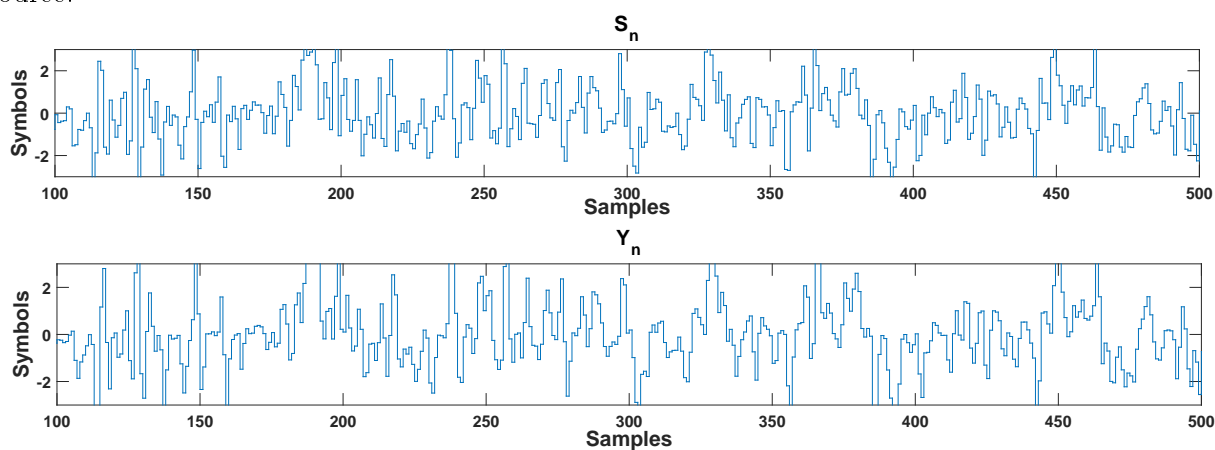


Figure 4.5: Example of $s(n)$ and $y(n)$, with Laplace source in a simple Wiener-Hammerstein configuration.

Finally, Figure 4.6 shows that when the source is binary, the criterion acquire a perfect solution, the regions of optimal points are easier to identify in this case, which are equivalent to the solution presented in Table 4.1. This can be better understood if we look at the signals $s(n)$ and $y(n)$ of the solution which is shown in Figure 4.7.

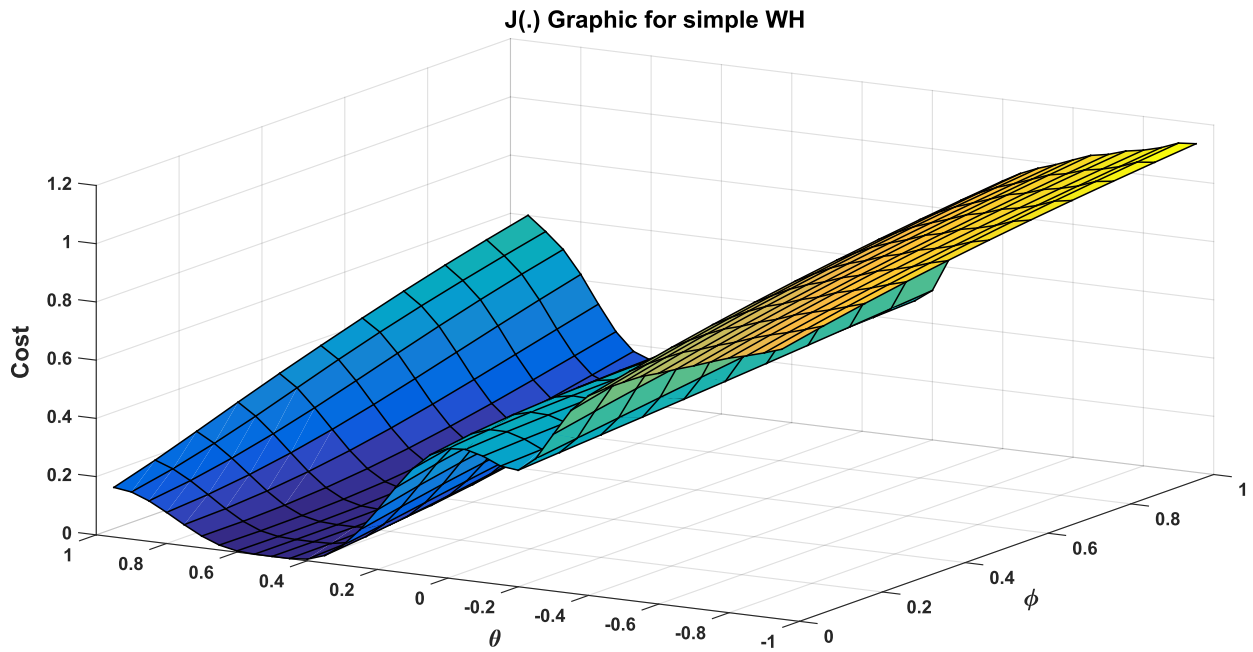


Figure 4.6: Example of $J_{cor}(\cdot)$ surface for a simple Wiener-Hammerstein configuration with binary source.

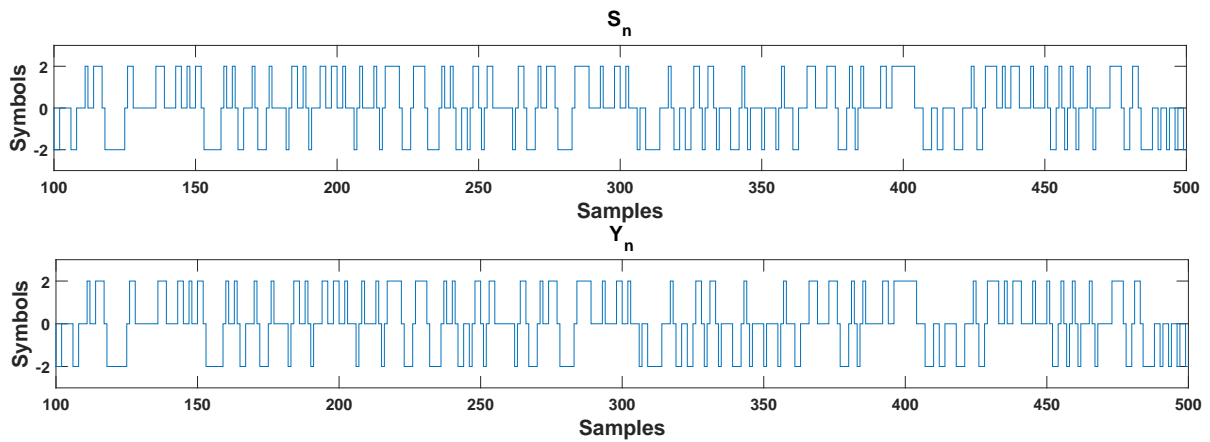


Figure 4.7: Example of $s(n)$ and $y(n)$, with binary source in a simple Wiener-Hammerstein configuration.

With the preliminary experiment one can consider that the autocorrentropy-based criterion has the potential to be tested now in full condition of blind inversion of Wiener systems, in the context of our proposal presented in Section 3.2. The obtained results in this qualitative experimental, with the autocorrentropy-based criterion suggested in this proposal, confirm that the criterion in question presents advantages in non-independent sources. In the follow sections, the autocorrentropy as well as the autocorrelation criterion will be analyze in more complex scenarios.

4.2 Experimental Analysis

This section tests the proposal performance in two sets of experiments. The first experiment comprises input signals continuously distributed, in Section 4.2.1, and the second experiment considers input signals discretely distributed, in Section 4.2.2. The first experiment analyzes the algorithm behavior with two different distributions, the original sequence $s(n)$ is either uniform or laplacian. For the second experiment, tests were made where the source distribution changes among the Alternate Mark Inversion (AMI) [60] and an i.i.d. signal with samples drawn from the alphabet $\{+1, -1\}$. All signals, except AMI, are submitted to the linear precoder $P(z)$ used in the preliminary experiments, in order to generate dependent samples. They were then submitted to different situations with different settings of the deconvolution filter $w(n)$, considering FIR filters and also IIR structures with different values of M and Q . The polynomial part vary the amount of coefficients from $k = 3$ and $k = 5$.

To eliminate the indeterminacies, at every fitness evaluation, the output of the nonlinear stage $u(n)$, $n = 1, 2, \dots, N$ is centered and normalized as well as the output of the linear stage $y(n)$. For example, the mean and variance of $u(n)$ are calculated as follow

$$\hat{\mu}_u = \frac{1}{N} \sum_{n=1}^N u(n) \quad (4.5)$$

$$\hat{\sigma}_u = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (u(n) - \hat{\mu}_u)^2} \quad (4.6)$$

and $u(n)$ is centered and normalized by

$$u(n) \leftarrow \frac{u(n) - \hat{\mu}_u}{\hat{\sigma}_u}. \quad (4.7)$$

The same normalization scheme is applied for $y(n)$.

For all scenarios, $N = 2000$ samples of $s(n)$ are considered, where the resulting signal $x(n)$ is provided to the algorithm. The number of lags used in the $J_{cor}(\cdot)$ as well as in the $J_R(\cdot)$ cost function for equalization is $P = 10$, the kernel size is discussed in the following Section and we have removed the zero lag for reasons previously clarified in Section 3.2.1.

The CLONALG parameters were adjusted to the values suggested in [18], i.e. 300 iterations, 50 individuals, and 10% of new individuals inserted per iteration. Other CLONALG parameters were defined with the aid of a preliminary cross-validation routine, which comprised 10 independent trials of the algorithm, with the correntropy cost function, for each possible configuration: the clone number parameter $\beta \in \{0.1, 0.2, 0.3\}$ and the mutation rate $\rho \in \{2, 3, 4, \dots, 8\}$.

In order to evaluate the average behavior, we compute the mean values of SNR over a set of 10 independent algorithm executions for each experiment, calculated between the output signal and the original signal for the optimal equalization delay. The SNR can be measured with the output signal to noise ratio

$$\frac{\sigma_s^2}{\sigma_n^2} = E[y^2(n)]/E[(s(n) - y(n))^2], \quad (4.8)$$

where σ_n^2 is the error power and σ_s^2 is the estimated signal power.

4.2.1 Continuous Case

First, we consider the input signal to be either an uniform or laplacian i.i.d. sequence that is submitted to the linear precoder $P(z)$. Thereby, an i.i.d. source is linearly precoded to form a sequence of correlated symbols, hence we have dependent samples $s(n)$ [46]. As we discussed in Section 4.1, the autocorrentropy as well as the autocorrelation are estimated from 500 samples of $s(n)$, which are distinct from the (unknown to the algorithm) samples to be submitted to the Wiener system. Then, we analyze the performance of the algorithm in a series of scenarios varying the parameters of the Wiener system (the linear $H(z)$ and nonlinear distortion $f[\cdot]$) as well as the number of coefficients of the Hammerstein linear sub-system $W(z)$.

Before starting the experimental analysis of our framework we must empirically define an appropriate value for the kernel size of the autocorrentropy criterion, thus, we test the performance of the algorithm with respect to this key parameter for both distributions. We choose 5 kernel sizes: 0.1, 0.4, 1, 2.2, and 4.6. For each kernel size the Wiener system is modeled by a minimum phase system with coefficients $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v)\sqrt[3]{|v|}$. The polynomial model is set to $k = 3$ and the linear sub-system parameters are set to $M = 1$ and $Q = 1$. The results are presented in Figure 4.8 where the best kernel size is defined as $\sigma = 0.4$. However, note that when the kernel size was $\sigma = 2.2$ the Laplace distribution perform a similar course.

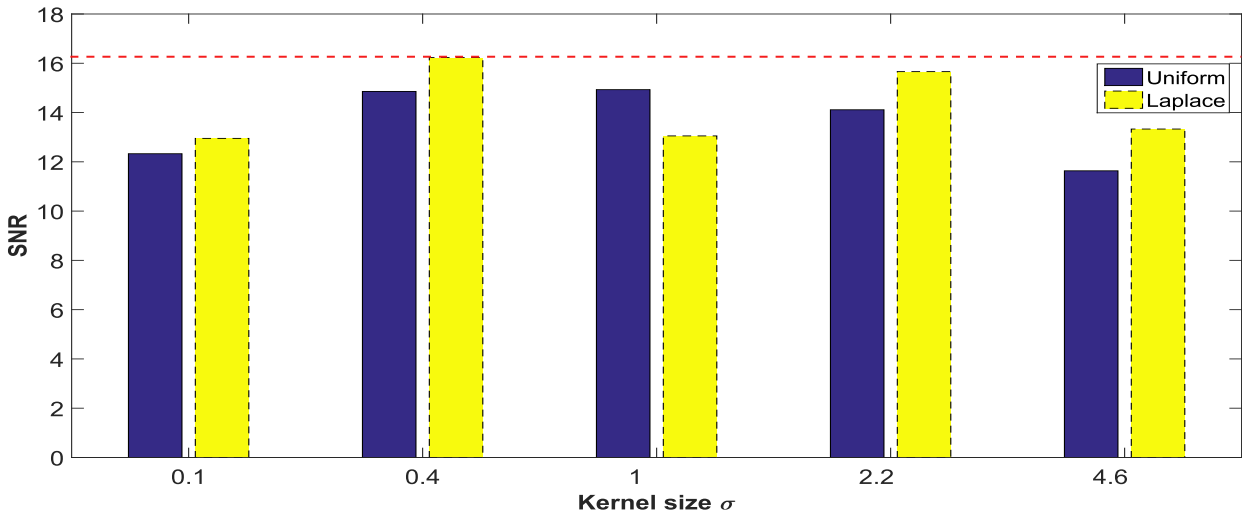


Figure 4.8: Kernel performance for the continuous case with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v)\sqrt[3]{|v|}$.

In order to clear doubts, we repeat the experiment, nonetheless by this time we increase the number of coefficients of the Wiener system i.e. $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and we modify the cubic root function to represent a harder nonlinear distortion, $f[v] = \tanh(3v)$, consequently we increment the flexibility

of the polynomial model of the Hammerstein system by setting $k = 5$. Figure 4.9 shows that in this more complex test the kernel size $\sigma = 0.4$ presented the best performance with the uniform distribution. However, the kernel size $\sigma = 2.2$ presented again a similar result, the red line point out that the small kernel won narrowly. These evidences were enough to keep the kernel size as $\sigma = 0.4$ in the following scenarios.

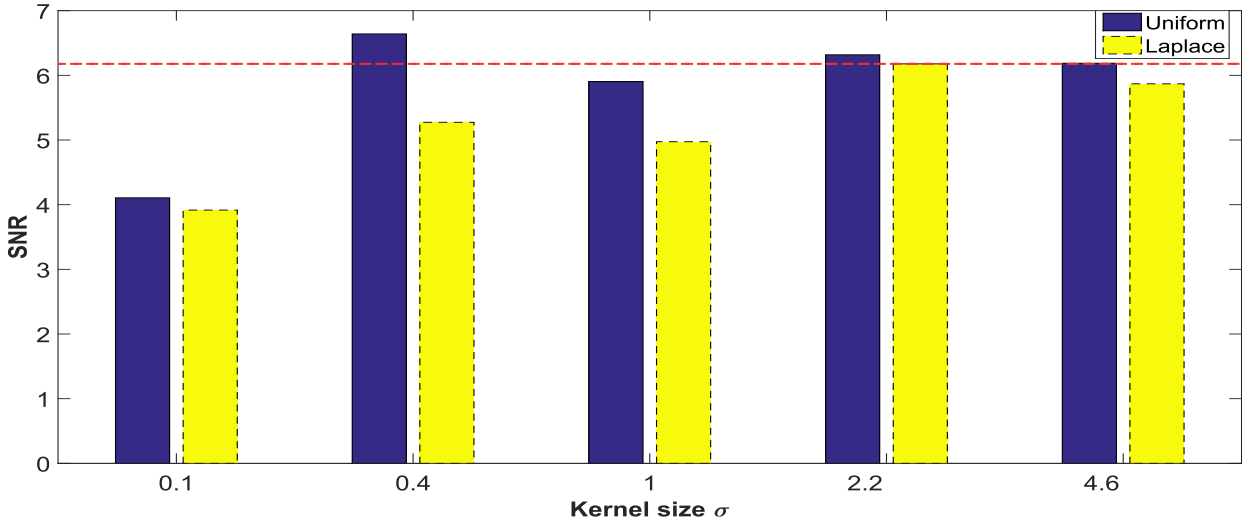


Figure 4.9: Kernel performance for the continuous case with parameters $M = 1$, and $Q = 1$ minimum phase system $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and nonlinear function $f[v] = \tanh(3v)$.

After asserting the kernel size $\sigma = 0.4$ of the autocorretropy-based algorithm, we continue to test the algorithm. In the first scenario, the Wiener system is modeled by a minimum phase system with coefficients $H(z) = 1 + 0.5z^{-1}$ and the nonlinear distortion is a cubic root $f[v] = \text{sign}(v) \sqrt[3]{|v|}$. The polynomial model is fit to $k = 3$. Table 4.2 shows the mean and standard deviation of SNR for both distributions. The highlighted values correspond to the best results for each criterion. This case was interesting because when $s(n)$ was uniform, the best configuration for both criteria was obtained with the IIR filter that correspond to the perfect inverse of $h(n)$, i.e. $M = 1$, $Q = 1$. For the Laplace distribution, the performance of the correntropy-based algorithm was inferior with the IIR configuration, the SNR values obtained with the FIR configuration (the first column), are actually satisfactory. In general, the autocorrelation criterion keeps a behavior among both distributions which can be considered good.

TABLE 4.2: PERFORMANCE RESULTS FOR THE FIRST SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Uniform			
1	N/A	14.8568 ±3.9057	10.5581±6.4042
		14.6424 ±3.5231	12.1894 ±4.8719
2	6.9127 ±4.6778	6.5298 ±4.4481	6.5864 ±2.3673
	9.6851 ±3.4061	9.9486 ±5.2356	8.0404 ±4.8041
3	4.2240 ±2.1388	7.8640 ±3.6413	7.0788 ±3.8619
	7.1719 ±3.3396	7.2597 ±4.1099	6.1361 ±3.6377
4	5.2360 ±3.4400	5.7305 ±3.0323	4.3977 ±2.8700
	8.1795 ±3.6980	5.1590 ±2.3961	9.0349 ±5.1689
Laplace			
1	N/A	4.8299 ±0.5993	3.6216 ±0.7221
		4.9733 ±0.6191	4.6350 ±0.8511
2	12.7572 ±4.4355	8.6372 ±3.4385	6.2717 ±5.0701
	6.1112 ±1.5163	8.5274 ±7.1925	7.2257 ±3.6272
3	10.0389 ±3.8887	5.2167 ±3.4704	4.8022 ±2.1240
	6.0616 ±3.0833	6.4510 ±4.7069	6.1691 ±2.5043
4	9.6916 ±4.7279	7.4234 ±4.4944	5.0443 ±2.6607
	5.8004 ±1.1480	7.0782 ±3.0639	6.7976 ±4.6523

Next, in a second scenario we analyze a more complex situation, with a linear sub-system with higher number of coefficients, i.e. $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and the same cubic root nonlinear distortion. The Hammerstein polynomial model is kept fixed, with respect to the previous scenario. Table 4.3 presents the results. This case was particularly interesting for the reason that for a long channel the autocorrelation criterion obtained the best results for both distributions with one delay in the FIR part as well as in the IIR part, i.e. $M = 1$ and $Q = 1$. The autocorrelation criterion achieved better results in most cases with the uniform distribution and with the Laplace distribution, nevertheless, the correntropy-based algorithm is fairly similar with respect to the autocorrelation-based algorithm.

TABLE 4.3: PERFORMANCE RESULTS FOR THE SECOND SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Uniform			
1	N/A	14.3675 ±1.3090	12.1697 ±3.9542
		10.0483 ±2.6356	10.9514 ±2.4465
2	10.3487 ±4.8042	4.2215 ±2.6087	5.6993 ±3.9722
	12.2270 ±1.9237	7.1159 ±5.7061	5.8529 ±3.1429
3	7.9271 ±5.4106	4.1403 ±2.1130	4.6832 ±2.6312
	9.4325 ±3.5489	4.1403 ±4.8972	7.5669 ±5.5669
4	5.8392 ±3.6569	3.1586 ±1.7876	4.2030 ±2.0286
	5.9517 ±3.1271	4.9308 ±3.3318	9.5028 ±3.5657
Laplace			
1	N/A	12,8232 ±2.5805	10.6611 ±3.8431
		9,0241 ±1.7864	7.9572 ±1.0392
2	10.8289 ±4.5698	6,9777 ±4.7980	6.9542 ±3.9372
	9.3756 ±1.4639	10,7657 ±4.8344	6.4196 ±2.5645
3	9.0718 ±4.1196	7,1028 ±3.7207	6.1031 ±4.1582
	7.6368 ±2.9561	7,2779 ±3.7073	6.2302 ±4.3222
4	6.7983 ±3.0400	5,7620 ±4.1957	6.5560 ±4.2270
	5.7902 ±1.2705	6,9577 ±3.8080	4.1066 ±3.0345

As observed, the general behavior of the algorithm can be considered similar among all possible configurations, for both distributions, since the SNR values were relatively close with a small difference between each other. Here the inclusion of a feedback loop was relevant to enhance the inversion performance.

Following, in the third scenario we modify the cubic root function to represent a harder nonlinear distortion, $f[v] = \tanh(3v)$. The linear channel remains as $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and we increment the flexibility of the polynomial model of the Hammerstein system by setting $k = 5$. The results presented in Table 4.4, show that the autocorrelation-based algorithm achieved better results for both distributions, however, the results were comparatively close, the autocorrentropy criterion was slightly inferior. For the uniform distribution the IIR structure played a relevant role to highlight the inversion performance and the SNR values obtained with the FIR configuration can be considered fair for the Laplace distributed signal.

TABLE 4.4: PERFORMANCE RESULTS FOR THE THIRD SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Uniform			
1	N/A	6.6408 \pm 0.4852	7.1136 \pm0.9926
		8.4698 \pm1.1950	7.5135 \pm 0.8659
2	5.3037 \pm 1.7731	3.5200 \pm 2.2806	3.4118 \pm 3.9057
	5.6166 \pm 0.7256	4.6532 \pm 2.1918	4.3149 \pm 2.0068
3	5.9823 \pm 1.7657	2.7248 \pm 1.9703	2.7801 \pm 2.0592
	4.1090 \pm 1.3381	3.1194 \pm 1.6949	3.1099 \pm 2.2759
4	4.0638 \pm 1.9985	2.8662 \pm 2.2546	2.2321 \pm 1.5921
	4.3067 \pm 0.7420	3.5884 \pm 1.5798	3.9430 \pm 2.1250
Laplace			
1	N/A	5.5722 \pm 0.9826	5.6373 \pm 1.7084
		6.9036 \pm 0.9193	7.0176 \pm1.2008
2	5.8718 \pm0.8484	3.2309 \pm 2.9480	3.6815 \pm 2.2600
	6.8963 \pm 0.6402	5.9120 \pm 2.1063	5.2697 \pm 1.8783
3	5.4386 \pm 1.1977	4.7709 \pm 2.3995	2.7367 \pm 2.5375
	5.6730 \pm 1.1698	4.4340 \pm 2.6679	4.4865 \pm 2.8044
4	4.5855 \pm 1.1322	1.4664 \pm 1.1923	1.7542 \pm 2.2162
	4.2183 \pm 1.2551	3.9375 \pm 1.8202	4.2981 \pm 2.4042

Finally, in the fourth scenario, the Wiener system is described by $H(z) = 1 + 0.5z^{-1}$, and the same nonlinear distortion, $f[v] = \tanh(3v)$ and the polynomial model is kept to $k = 5$. Table 4.5 presents the results. As noticed, in overall, when the Wiener system contains the nonlinear distortion $f[v] = \tanh(3v)$, the algorithm suffers an evident depreciation. A possible explanation for this behavior may lie on the parametric polynomial that our technique adopts which may not be flexible enough to invert the nonlinear mapping. Notwithstanding, the autocorrelation criterion attained slightly better results for both distributions.

TABLE 4.5: PERFORMANCE RESULTS FOR THE FOURTH SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Uniform			
1	N/A	6.3965 ±0.2584	4.3140 ±1.0687
		6.5296 ±3.5231	6.1625 ±4.8719
2	2.8539 ±0.7036	2.3227 ±1.3197	3.3606 ±1.2090
	6.0263 ±3.4061	5.3250 ±5.2356	5.4100 ±4.8041
3	3.1804 ±2.1984	2.4631 ±1.7538	2.6322 ±1.7173
	5.1739 ±3.3396	5.8894 ±4.1099	5.2675 ±3.6377
4	2.6404 ±1.3311	2.9340 ±1.9044	3.0756 ±1.4558
	6.2219 ±3.6980	5.8681 ±2.3961	4.3480 ±5.1689
Laplace			
1	N/A	4.8299 ±0.5993	3.6216 ±0.7221
		4.9733 ±0.6191	4.6350 ±0.8511
2	2.5883 ±1.3352	3.7914 ±1.3024	1.6851 ±2.1414
	4.1779 ±1.0477	3.9551 ±1.3525	4.4669 ±0.8409
3	2.7929 ±1.8486	1.7669 ±1.8518	1.1340 ±1.8968
	3.8905 ±0.9037	4.0634 ±1.5235	3.4007 ±1.7619
4	2.4208 ±1.3721	1.2515 ±1.5480	1.7168 ±1.8169
	3.9706 ±0.6846	4.4175 ±1.3796	1.9301 ±1.3848

As manifested, in cases where the linear channel $H(z)$ is composed with a large number of coefficients and the nonlinear distortion remains as $f[v] = \tanh(3v)$, the performance of the algorithm is better than when the linear channel is $H(z) = 1 + 0.5z^{-1}$. This may happen due to the fact that the longest channel generates a signal at the input of the nonlinear distortion $f[v] = \tanh(3v)$, with more states than the shorter channel. On the other hand, in most of the cases in the four scenarios with continuously distributed signals, the IIR structure presented better results with respect to the SNR values obtained with purely FIR configurations.

4.2.2 Discrete Case

For the discrete case, we consider the input is an i.i.d. signal with samples drawn from the alphabet $\{-1, +1\}$ submitted to the same linear precoder $P(z)$, whose analytical autocorrentropy functions was already given by the Eq. 4.1, in the preliminary experiments of Section 4.1, and its autocorrelation function is given by

$$R_s(m) = \begin{cases} 2\sigma_s^2, & m = 0 \\ \sigma_s^2, & m = \pm 1, \\ 0, & |m| \geq 2 \end{cases}$$

where σ_s^2 is the variance of $s(n)$. We also consider the Alternate Mark Inversion (AMI) source [60], whose dependent symbol sequence is drawn from the alphabet $\{-1, 0, +1\}$, its autocorrentropy function is analytically given by [46]

$$V_s(m) = \begin{cases} \kappa(0), & m = 0 \\ \frac{1}{4}\kappa(0) + \frac{1}{2}\kappa(1) + \frac{1}{4}\kappa(2), & m = \pm 1, \\ \frac{3}{8}\kappa(0) + \frac{1}{2}\kappa(1) + \frac{1}{8}\kappa(2), & |m| > 1 \end{cases}$$

and autocorrelation is defined by

$$R_s(m) = \begin{cases} \sigma_s^2, & m = 0 \\ -\frac{1}{2}\sigma_s^2, & m = \pm 1. \\ 0, & |m| \geq 2 \end{cases}$$

The number of lags used in the $J_{cor}(\cdot)$ as well as in the $J_R(\cdot)$ cost function for equalization is $P = 10$.

Once in the continuous scenarios we tested the performance of the autocorrentropy-based algorithm with respect to the kernel size, we repeat the same experiments presented in Section 4.2.1, this time with the discrete distributions previously mentioned. Figures 4.10 and 4.11 indicate that the algorithm obtained better results whit the kernel size $\sigma = 0.4$ for both cases.

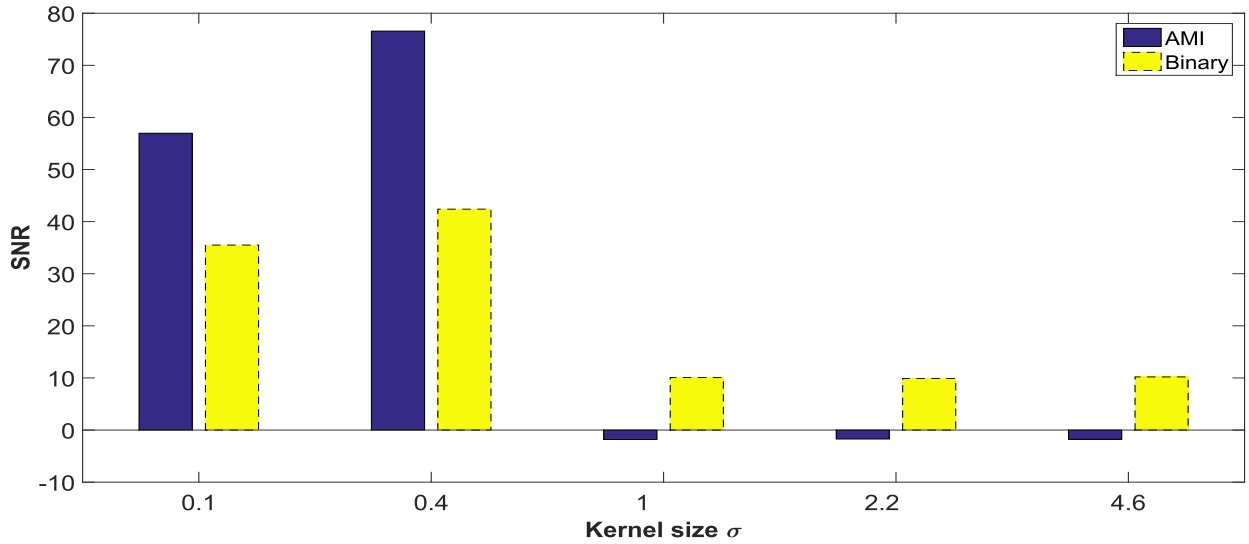


Figure 4.10: Kernel performance for the discrete case in the first scenario with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v) \sqrt[3]{|v|}$.

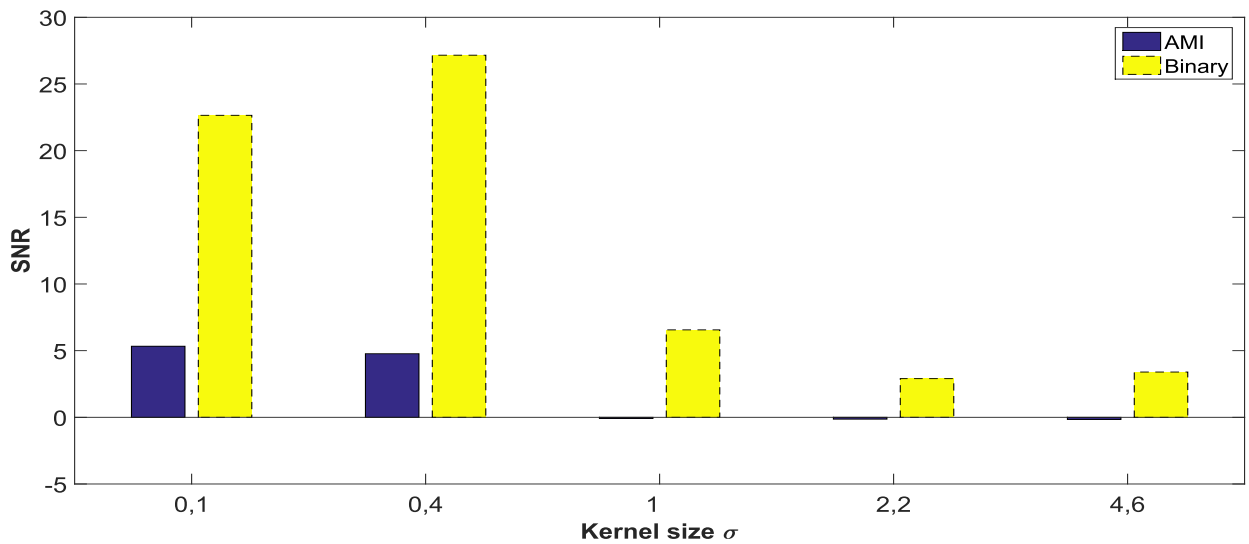


Figure 4.11: Kernel performance for the discrete case in the third scenario with parameters $M = 1$, and $Q = 1$ and minimum phase system $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and nonlinear distortion $f[v] = \tanh(3v)$.

Now, the same scenarios of the continuous case are considered in this case, as well. In the first one, the Wiener system is modeled by a minimum phase system with coefficients $H(z) = 1 + 0.5z^{-1}$ and nonlinear distortion $f[v] = \text{sign}(v) \sqrt[3]{|v|}$, the polynomial model is set to $k = 3$. Table 4.6 shows the mean and standard deviation of SNR, in dB, for the two distributions. In all cases the results obtained by the autocorrentropy criterion indicate that the desired signal is recovered with a significant degree of accuracy, this is remarkably noticeable in the duobinary $\{+1, -1\}$ signal code, where the autocorrelation was inferior. For the AMI line code, the autocorrentropy criteria

TABLE 4.6: PERFORMANCE RESULTS FOR THE FIRST SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Binary			
1	N/A	42.3971 ±22.9748	34.1152 ±11.8529
		8.3664 ±0.6832	6.0335 ±0.3755
2	15.4638 ±5.1540	15.0029 ±14.5091	33.4532 ±9.9356
	9.3447 ±1.5647	7.2518 ±0.6168	8.7425 ±0.7005
3	24.5195 ±0.9949	26.0103 ±10.3711	27.8254 ±7.8177
	10.1324 ±2.0326	6.6126 ±1.6287	7.8646 ±0.7984
4	24.9277 ±2.4238	29.1586 ±5.6757	29.5466 ±5.3453
	9.0018 ±0.9443	7.1526 ±2.3802	6.9511 ±2.3962
AMI			
1	N/A	76.5697 ±17.1204	36.7313 ±26.5264
		14.8651 ±2.4182	16.5883 ±4.4528
2	5.2648 ±0.2427	15.6042 ±25.0736	11.5516 ±30.3120
	6.5509 ±0.1529	14.7809 ±3.2623	10.7013 ±3.6170
3	18.4655 ±0.0399	5.6490 ±16.1036	11.8754 ±18.4088
	7.9706 ±3.9259	8.4627 ±4.0956	10.2360 ±4.4483
4	24.6631 ±0.2584	15.0663 ±0.2584	6.9703 ±14.1611
	10.1143 ±2.9990	8.9963 ±7.6855	6.0951 ±2.5593

present better results than autocorrelation, one can note that the linear part of the Wiener system was total and perfectly inverted with the IIR filter that corresponds to $M = 1$, $Q = 1$. For the autocorrelation optimization criteria in both distributions, the IIR configuration was relevant to raise the inversion. Although the autocorrelation criteria show the best results in most of the cases, the results with the autocorrelation criteria in both distributions can be considered good.

The second scenario analyzes the algorithm in a more complex situation, as the Wiener system is formed by $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$ and the same nonlinear distortion $f[v] = \text{sign}(v) \sqrt[3]{|v|}$ with the polynomial model kept to $k = 3$. Table 4.7 presents the results. For the duobinary $\{+1, -1\}$ signal, it is remarkable that, the IIR structure is more efficient from the standpoint inversion system with the autocorrelation criteria, which illustrates the potential advantages of using a recurrent filter in the Hammerstein system. On the other hand, for the AMI line code, note that the performance of the algorithm with the autocorrelation criterion presented inferior results with respect to the

TABLE 4.7: PERFORMANCE RESULTS FOR THE SECOND SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Binary			
1	N/A	14.9085 \pm 0.2570	18.8647 \pm 0.5799
		11.5885 \pm0.9077	6.4354 \pm 0.7243
2	14.8598 \pm 0.2019	2.5390 \pm 0.3407	16.7106 \pm 0.2174
	7.9064 \pm 2.0373	6.8221 \pm 0.5538	8.2442 \pm 0.3275
3	20.1459 \pm 0.8004	21.2914 \pm 1.0035	22.3770 \pm1.5154
	7.2140 \pm 0.6360	6.1165 \pm 1.4536	7.1126 \pm 1.4768
4	20.5990 \pm 0.6557	11.0622 \pm 11.3577	13.5573 \pm 10.8867
	6.2557 \pm 0.8187	7.3473 \pm 1.5524	8.1772 \pm 1.2897
AMI			
1	N/A	4.5801 \pm 0.8405	4.3686 \pm 1.0847
		7.4608 \pm0.7446	6.8574 \pm 0.9419
2	3.5905 \pm 1.4828	4.8299 \pm0.2775	3.5960 \pm 1.9341
	5.6506 \pm 3.8338	7.8432 \pm 7.4359	5.8440 \pm 4.7018
3	3.3818 \pm 1.7300	4.2166 \pm 1.4110	3.0845 \pm 1.0944
	6.4164 \pm 2.6637	3.2790 \pm 1.4989	7.2741 \pm 4.7657
4	3.3337 \pm 2.5240	3.0204 \pm 2.3695	2.7375 \pm 2.0971
	4.9267 \pm 1.4839	5.9401 \pm 3.2289	3.8521 \pm 1.9609

duobinary $\{+1, -1\}$ signal, also, although the autocorrelation criterion was better, the general behavior was inferior with respect to the duobinary $\{+1, -1\}$ signal.

We can see that the algorithm with the duobinary $\{+1, -1\}$ signal for both criteria and both scenarios and the AMI line code for both criteria in the first scenario present such a better behavior with respect to the continuous case in these same scenarios. On the other hand, the standard deviation is larger for AMI line code in the first scenario, which may be a subject of analysis in the future. Also, it is visible that the AMI line code in the second scenario for both criteria presented an inferior performance with respect to the first one, and it may indicate that these criteria with this particular signal, is more sensitive to domains with larger search spaces.

Next, the third scenario is composed of a Wiener system with a harder nonlinear distortion $f[v] = \tanh(3v)$ and we increment the flexibility of the polynomial model of the Hammerstein system by setting $k = 5$. The linear channel is kept fixed, with respect to the previous scenario. In Table 4.8 we can see the results. As we can observe, for both criteria, the precoded duobinary

signal $\{-1, +1\}$ provided better results than the AMI line code, note that the IIR structure with $M = 2$, $Q = 2$ provided the best performance for the autocorrentropy criterion, which led better results in comparison with the autocorrelation criterion. Then again, the autocorrentropy criterion with the AMI line code is limited for searching an ideal solution inversion.

TABLE 4.8: PERFORMANCE RESULTS FOR THE THIRD SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Binary			
1	N/A	27.1582 \pm 2.3291	26.3274 \pm 3.18161
		11.4609 \pm1.3548	6.8779 \pm 0.5564
2	27.2379 \pm 2.6070	27.0057 \pm 3.8345	28.3819 \pm4.3323
	5.6971 \pm 0.1138	6.1671 \pm 0.3077	7.9212 \pm 0.3867
3	22.3939 \pm 3.4233	25.6450 \pm 10.3711	26.1593 \pm 2.8519
	6.3156 \pm 0.4363	6.4826 \pm 1.9191	8.7940 \pm 0.2716
4	26.4549 \pm 2.3703	24.7444 \pm 2.4619	25.2451 \pm 1.9930
	7.1988 \pm 0.6260	6.4536 \pm 1.6520	7.2348 \pm 1.6392
AMI			
1	N/A	5.6105 \pm3.9239	1.4308 \pm 4.3520
		8.7712 \pm0.2614	8.6935 \pm 0.7230
2	0.7853 \pm 0.0859	0.1777 \pm 0.9319	3.6273 \pm 1.3206
	4.1404 \pm 0.5087	4.2678 \pm 4.3620	5.3170 \pm 3.4674
3	3.4488 \pm 1.4149	4.1253 \pm 0.1393	3.0500 \pm 1.1166
	3.1774 \pm 0.6927	4.0420 \pm 2.2997	5.3144 \pm 3.1668
4	3.7339 \pm 0.7518	3.6607 \pm 10.3711	3.5601 \pm 0.9476
	3.6231 \pm 1.2135	3.6399 \pm 1.5418	3.9534 \pm 2.5417

Finally, in the fourth scenario the Wiener system is described by $H(z) = 1 + 0.5z^{-1}$, and the same nonlinear distortion, $f[v] = \tanh(3v)$ and the polynomial model is kept to $k = 5$. Table 4.9 exhibits the results. A highlighted value corresponds to the best result for each criterion. As we can see, once again, the performance of the algorithm experiences a depreciation in its performance when the linear channel is shorter with respect to the previous scenario.

TABLE 4.9: PERFORMANCE RESULTS FOR THE FOURTH SCENARIO. TOP VALUES OF EACH CELL CORRESPOND TO THE FRAMEWORK WITH $J_{cor}(\cdot)$ AND THE LOWER VALUES WITH $J_R(\cdot)$.

M	Q		
	0	1	2
Binary			
1	N/A	4.9167 \pm 0.2174	3.1764 \pm 2.4140
		5.5307 \pm 0.2880	7.5144 \pm 0.1467
2	4.9772 \pm0.3420	2.2236 \pm 1.3044	3.4316 \pm 4.1865
		9.5335 \pm 0.3517	9.1858 \pm 0.3202
3	4.5591 \pm 0.1248	3.2175 \pm 2.3347	2.4382 \pm 2.1489
		9.7029 \pm0.5925	7.0029 \pm 2.8907
4	4.5109 \pm 0.1253	5.8063 \pm 2.9019	3.9344 \pm 2.8939
		7.1639 \pm 0.6990	7.4731 \pm 1.3221
AMI			
1	N/A	4.6721 \pm3.6613	1.5216 \pm 0.1286
		13.4194 \pm0.9629	5.6747 \pm 5.7263
2	1.6445 \pm 0.9317	1.2171 \pm 0.1080	1.5887 \pm 0.3240
		8.9709 \pm 0.2696	5.4681 \pm 5.6308
3	1.9804 \pm 1.2897	1.5110 \pm 0.2060	1.7387 \pm 0.9394
		7.1056 \pm 2.5710	7.0846 \pm 4.9281
4	3.6749 \pm 4.8210	0.9921 \pm 1.4930	1.2119 \pm 1.3420
		6.6887 \pm 4.2382	7.5635 \pm 3.0442

In view of the repeated behavior of the algorithm, when the linear channel $H(z)$ is composed with a larger number of coefficients presents better results than when the linear channel is shorter, we made a straightforward experiment to analyze closer this kind of conduct. For this, we consider the duobinary $\{+1, -1\}$ signal in the third and fourth scenarios with the autocorrentropy criterion. Figures 4.12 and 4.13 confirm what we suggested previously in the continuous case, i.e. the longest channel generates a signal at the input of the nonlinear distortion $f[v] = \tanh(3v)$, with more states than the shorter channel, these “more diversified” states may cause the output of the nonlinear distortion to be not saturated at $+1/-1$, and possibly facilitating the nonlinear inversion task for the polynomial model.

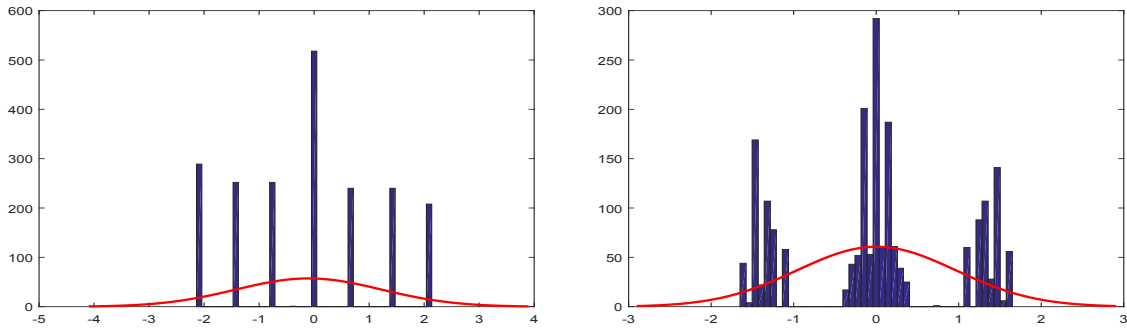


Figure 4.12: Experiment with the duobinary $\{+1, -1\}$ signal with the autocorrentropy criterion. The Figure of the left correspond to the signal after passing through the linear channel $H(z) = 1 + 0.5z^{-1}$ and the other one through $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$.

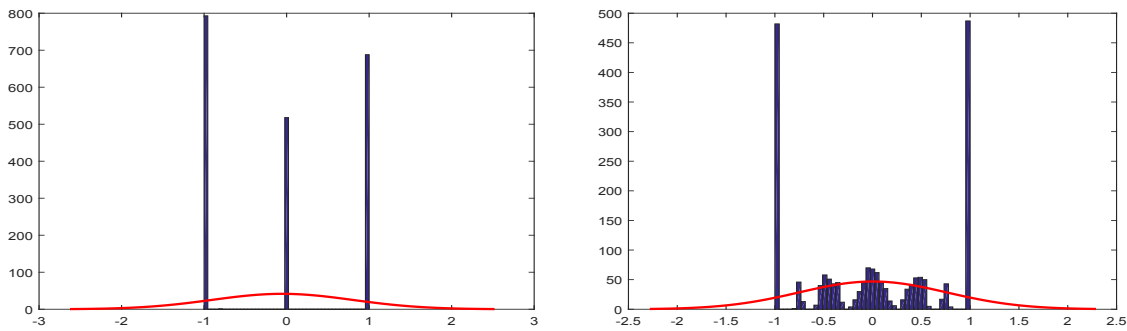


Figure 4.13: Filtered duobinary $\{+1, -1\}$ signal after passing through the nonlinear distortion $f[v] = \tanh(3v)$. The Figure of the left correspond to the signal after passing through the linear channel $H(z) = 1 + 0.5z^{-1}$ and the other one through $H(z) = 1 - 0.0919z^{-1} + 0.2282z^{-2} - 0.1274z^{-3} + 0.1408z^{-4} - 0.0189z^{-5} + 0.0173z^{-6} - 0.0072z^{-7} + 0.0038z^{-8}$.

In a general perspective, the autocorrelation and the correntropy-based criterion provided reasonable results that validate the idea of using dependence as criterion to invert the original system. One can see that the correntropy-based criterion presented the best results: for all cases in the continuous scenario and for the precoded duobinary signal $\{-1, +1\}$ in the first, second and third scenarios. Furthermore, it is possible to see that the feedback loop in the linear filter was pertinent to build up the inversion performance, since most of the top scores, for both criteria, were obtained with $Q \geq 1$. However, the correntropy-based criterion showed an inferior overall performance for the AMI line code in the third scenario, where predominantly autocorrelation-based criterion performed better.

Finally, for the sake of illustration, we show in Figure 4.14 the algorithm results of the precoded binary signal in the third scenario using the correntropy criteria and with $M = 2$, and $Q = 2$. The figures confirm that the algorithm inverted almost perfectly the desired signal, this can be observed in the error signal illustrated in the figure of the top. Also, in the second figure the estimated correntropy of the output signal $V_y(m)$ practically matches the analytical correntropy of $V_s(m)$. Moreover, the last figure shows the convolution between $h(n)$ and $w(n)$.

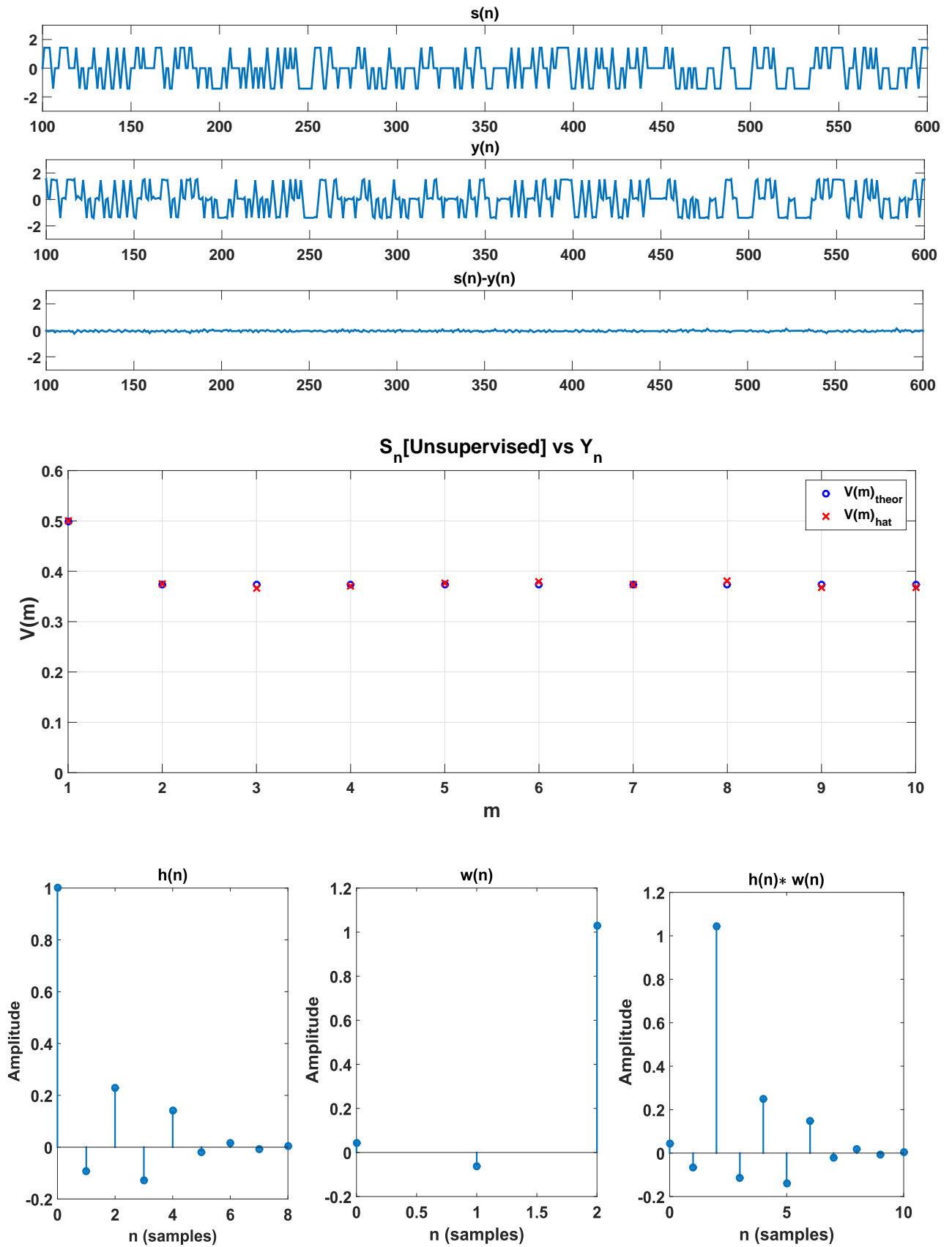


Figure 4.14: Example of algorithm results with duobinary signal source $\{+1, -1\}$, $M = 2$, and $Q = 2$ for third scenario.

4.3 Summary

In this chapter we presented a set of experiments in order to evaluate the framework proposed in Chapter 3. The first experiment was allowing to show us whether the based-autocorrentropy criterion has the potential to be applied in full conditions in the context of blind inversion of Wiener systems. The partial success in the preliminary test yielded the realization of an experimental analysis of the new framework for the Wiener-Hammerstein problem, and the results indicated that the framework is feasible, ensuring a valid performance of the method.

The next chapter concludes this dissertation, with the final considerations of the presented work and the future outlook of research and study.

Chapter 5

Conclusion

Nonlinear models can provide an accurate description and prediction of physical systems that have a nonlinear behavior. Wiener systems and Hammerstein systems are nonlinear models that are used in many domains for their simplicity and physical meaning. Particularly, blind identification of nonlinear systems has become an important issue with many practical applications. However, most of the identification/inversion approaches assume the availability of input of the system which in most of the cases is considered i.i.d..

Given the characteristics of the mentioned problem, it was developed a new framework, incorporating an immune-inspired strategy for searching and the use of dependence measures, particularly the new generalized correlation function as well as the autocorrelation, as new criteria for the inversion task. The proposed approach was shown to be able to deal with non-independent signals and to perform considerably good in performing the inversion task. As examples of the flexibility of the proposed approach, two different dependence-based criteria were presented on completely different scenarios, the first scenario where continuously distributed input signals were tested and the second experiment considered input signals that are discretely distributed.

After the introduction chapter, Chapter 2 presents a description of the historical aspects and definitions of Information Theoretic Learning, and later Chapter 2 links this primary subject with the adaptive signal processing and machine learning theories. Although this description may be tedious for some readers it is important for the full understanding of all the work presented in this dissertation.

Prior works in the context of inversion of Wiener systems assume originally i.i.d. signals as the input of the system, although different blind inversion methods are proposed. Although in a real-world situation non-independent sources are common, the inversion task for these signals still demands a deeper research. This leads to problems unanswered in the literature. The first part of the Chapter 3 presents the main contributions of the related work in the context of blind inversion of Wiener systems. The second part of the Chapter presents this dissertation proposal: a framework based on CLONALG algorithm and dependence-based criteria as objective functions of the parameter optimization problem. In the structure of this framework, linear stages with and without feedback (FIR or IIR) enhancing the flexibility on modeling the Hammerstein system were

considered.

Finally, the Chapter 4 presents the experimental results of the proposed approach. We performed a qualitative study of the performance of the autocorrentropy criterion in the context of a Wiener system which provided evidence of the algorithm feasibility for a non-linear setting. A series of simulations were carried out, in different scenarios, which showed that the new method has satisfactory performance in the scenarios that were evaluated. It is not conclusive which one of the two possibilities of criteria is preferable. Notwithstanding, considering the previously known methods, both criteria indicate that the exploration of the temporal structure of the input signals can lead to the inversion of Wiener systems.

Future Work

Given the results, there are a number of possible future work to be done, and we have listed below some of them.

- It is interesting the study of more pliable nonlinearity models, e.g. monotonic neural networks, in order to search for further improvement of the algorithm.
- It may be interesting to apply the proposed framework to real systems where the input signal is non-independent, for instance, a piece of a song.
- Perform a comparative analysis of existing methodologies for involving a higher number of channels scenarios, source signals and nonlinearities.

Author Publications

- [1*] S. A. Fernandez, R. R. F. Attux, D. G. Fantinato, J. R. Montalvão, D. G. Silva, "An Immune-Inspired, Dependence-Based Approach to Blind Inversion of Wiener Systems," accepted in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning ESANN 2016*, 2016.

REFERENCES

- [1] B. Widrow and E. Walach, *Adaptive Inverse Control: A Signal Processing Approach. Reissue Edition*. John Wiley & Sons, 2008.
- [2] W. Greblicki, “Nonlinearity Estimation in Hammerstein Systems Based on Ordered Observations,” *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1224–1233, 1996.
- [3] S. Prakriya and D. Hatzinakos, “Blind Identification of Linear Subsystems of LTI-ZMNL-LTI Models with Cyclostationary Inputs,” *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2023–2036, 1997.
- [4] R. Bars, I. Bèzi, B. Pilipar, and B. Ojhelyi, “Nonlinear and Long Range Control of a Distillation Pilot Plant,” *Identification and System Parameter Estimation. 9th IFAC/IFORS Symposium*, pp. 848–853, 1990.
- [5] R. C. Brinker, “A Comparison of Results from Parameter Estimations of Impulse Responses of the Transient Visual System,” *Biological Cybernetics*, vol. 61, pp. 139–151, 1989.
- [6] I. W. Hunter, “Frog Muscle Fiber Dynamic Stiffness Determined Using Nonlinear System Identification Techniques,” *Biophysical Journal*, vol. 49, p. 81a, 1985.
- [7] —, “The Identification of Nonlinear Biological Systems: Wiener and Hammerstein Cascade Models,” *Biological Cybernetics*, vol. 55, pp. 135–144, 1986.
- [8] Y. Zhu, “Identification of Hammerstein Models for Control Using ASYM,” *International Journal of Control*, vol. 73, no. 18, pp. 1692–1702, 2000.
- [9] K. Abed-Meraim, W. Qiu, and Y. Hua, “Blind System Identification,” *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.
- [10] R. K. Pearson, *Discrete-Time Dynamic Models*. Oxford University Press, USA, 1999.
- [11] A. Taleb, J. Solé-Casals, and C. Jutten, “Quasi-nonparametric Blind Inversion of Wiener Systems,” *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 917–924, 2001.
- [12] J. Sole, C. Jutten, and A. Taleb, “Parametric Approach to Blind Deconvolution of Nonlinear Channels,” *Neurocomputing*, vol. 48, no. 1, pp. 339–355, 2002.
- [13] M. Babaie-Zadeh, J. Solé-Casals, and C. Jutten, “Blind Inversion of Wiener System Using a Minimization-Projection (MP) Approach,” *ICA [S.l.: s.n.]*, pp. 681–686, 2003.

- [14] L. Chan, “Practical Method for Blind Inversion of Wiener Systems,” in *IEEE International Joint Conference on Neural Networks*, vol. 3, 2004, pp. 2163–2168.
- [15] J. Solé-Casals, C. Jutten, and D. Pham, “Fast Approximation of Nonlinearities for Improving Inversion Algorithms of PNL Mixtures and Wiener Systems,” *Transactions on Signal processing*, vol. 85, no. 9, pp. 1780–1786, 2005.
- [16] F. Rojas, J. Solé-Casals, and C. G. Puntonet, “An Evolutionary Approach for Blind Inversion of Wiener Systems,” in *Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 260–267.
- [17] J. Solé-Casals and C. Caiafa, “A Fast Gradient Approximation for Nonlinear Blind Signal Processing,” *Cognitive Computation*, vol. 5, no. 4, pp. 483–492, 2013.
- [18] D. Silva, J. Montalvão, R. Attux, and L. Coradine, “An immune-Inspired, Information-Theoretic Framework for Blind Inversion of Wiener Systems,” *Signal Processing*, vol. 113, pp. 18–31, 2015.
- [19] J. Príncipe, D. Xu, and J. Fisher, “Information Theoretic Learning,” *Unsupervised adaptive filtering*, vol. 1, pp. 265–319, 2000.
- [20] D. Erdogmus and J. Príncipe, “An Error-Entropy Minimization Algorithm for Supervised Training of Nonlinear Adaptive Systems,” *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1780–1786, 2002.
- [21] K. E. Hild, D. Erdogmus, and J. Príncipe, “Blind Source Separation Using Renyi’s Mutual Information,” *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 174–176, 2001.
- [22] D. Fantinato, R. Attux, A. Neves, R. Suyama, and J. Romano, “Blind Deconvolution of Correlated Sources Based on Second-Order Statistics,” in *SBrT 2013*. Sociedade Brasileira de Telecomunicações, 2013, pp. 1–5.
- [23] B. V. K. Kumar, D. W. Carlson, and A. Mahalanobis, “Optimal Trade-Off Synthetic Discrimination Function Filters for Arbitrary Devices,” *Optics Letters*, vol. 19, no. 19, pp. 1156–1158, 1994.
- [24] T. Nitta, “Orthogonality of Decision Boundaries in Complex-Valued Neural Networks,” *Neural Computation*, vol. 16, no. 1, pp. 73–97, 2004.
- [25] S. Buchholz and G. Sommer, “A Hyperbolic Multilayer Perceptron,” in *Proceedings of the International Joint Conferences on Neural Networks*, vol. 2, no. 12, pp. 129–133, 2000.
- [26] S. M. Moser and P. Chen, *A Student’s Guide to Coding and Information Theory*. Cambridge University Press, 2012.
- [27] C. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

- [28] R. d. R. Van Steveninck and S. Laughlin, “The Rate of Information Transfer at Graded-Potential Synapses,” *Nature*, vol. 379, no. 6566, pp. 642–645, 1996.
- [29] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
- [30] A. R enyi, “On Measures of Entropy and Information,” in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 547–561.
- [31] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [32] A. Navot, “On the Role of Feature Selection in Machine Learning,” Ph.D. dissertation, Hebrew University, 2006.
- [33] R. Battiti, “Using Mutual Information for Selecting Features in Supervised Neural Net Learning,” *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.
- [34] H. Peng, F. Long, and C. Ding, “Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [35] F. Rossi, A. Lendasse, D. Fran ois, V. Wertz, and M. Verleysen, “Mutual Information for the Selection of Relevant Variables in Spectrometric Nonlinear Modelling,” *Chemometrics and Intelligent Laboratory Systems*, vol. 80, no. 2, pp. 215–226, 2006.
- [36] S. Haykin, “Adaptive Filters,” *Signal Processing Magazine*, vol. 6, pp. 1–6, 1999.
- [37] ———, “Adaptive Filter Theory,” *Prentice Hall*, vol. 2, pp. 478–481, 2002.
- [38] A. K. Ziarani and A. Konrad, “A nonlinear adaptive method of elimination of power line interference in ECG signals,” *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 6, pp. 540–547, 2002.
- [39] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*. Artech House Norwood, 2005, vol. 46.
- [40] N. Wiener, *Nonlinear Problems in Random Theory*. MIT, 1958.
- [41] B. Widrow, J. McCool, and B. Medoff, “Adaptive Control by Inverse Modeling,” in *Twelfth Asilomar Conference on Circuits, Systems, and Computers*, 1978, pp. 90–94.
- [42] G. A. Darbellay and I. Vajda, “Estimation of the Information by an Adaptive Partitioning of the Observation Space,” *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1315–1321, 1999.
- [43] T. M. Dias, R. Attux, J. M. Romano, and R. Suyama, *Blind Source Separation of Post-Nonlinear Mixtures Using Evolutionary Computation and Gaussianization*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, vol. 5441.

- [44] R. Jenssen, D. Erdogmus, J. Principe, and T. Eltoft, “Towards a Unification of Information Theoretic Learning and Kernel Methods,” in *Machine Learning for Signal Processing*. IEEE, 2004, pp. 93–102.
- [45] P. A. Viola, N. N. Schraudolph, and T. J. Sejnowski, “Empirical Entropy Manipulation for Real-World Problems,” vol. 8. The MIT Press, Cambridge, MA, 1996, pp. 851–857.
- [46] I. Santamaría, P. Pokharel, and J. Principe, “Generalized Correlation Function: Definition, Properties, and Application to Blind Equalization,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.
- [47] S. Abhishek and J. Principe, “Information Theoretic Learning with Adaptive Kernels,” *Signal Processing*, vol. 91, no. 2, pp. 203 – 213, 2011.
- [48] W. Liu, P. Pokharel, and J. Principe, “Correntropy: Properties and Applications in Non-Gaussian Signal Processing,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 5286–5298, Nov. 2007.
- [49] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2233–2246, 2012.
- [50] X. Tang and Y. Hua, “Optimal Design of Non-Regenerative MIMO Wireless Relays,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1398–1407, 2007.
- [51] G. Zhou, Z. Yang, S. Xie, and J. Yang, “Online Blind Source Separation Using Incremental Nonnegative Matrix Factorization with Volume Constraint,” *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 550–560, 2011.
- [52] D. G. e Silva, “Aprendizado de Máquina Baseado na Teoria da Informação: Contribuições a Separação de Sinais em Corpos Finitos e Inversão de Sistemas de Wiener,” Ph.D. dissertation, Universidade Estadual de Campinas, 2013.
- [53] D. Pham, “Blind Separation of Instantaneous Mixture of Sources Based on Order Statistics,” *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 363–375, 2000.
- [54] B. Peeters and G. De Roeck, “Stochastic System Identification for Operational Modal Analysis: A Review,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 123, no. 4, pp. 659–667, 2001.
- [55] J. Principe, *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*. Springer Science & Business Media, 2010.
- [56] F. M. Burnet, “Clonal Selection and After,” *Theoretical Immunology*, pp. 63–85, 1978.
- [57] L. N. de Castro and F. J. Von Zuben, “Learning and Optimization Using the Clonal Selection Principle,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, Jun 2002.

- [58] C. Wada, D. M. Consolaro, R. Ferrari, R. Suyama, R. Attux, and F. J. Von Zuben, *Non-linear Blind Source Deconvolution Using Recurrent Prediction-Error Filters and an Artificial Immune System*. Springer Berlin Heidelberg, 2009, vol. 5441.
- [59] J. M. T. Romano, R. Attux, C. C. Cavalcante, and R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.
- [60] J. G. Proakis, *Intersymbol Interference in Digital Communication Systems*. Wiley Online Library, 2001.