

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

**Modelagem de Séries Temporais
Discretas Utilizando Modelo Nebuloso
Takagi-Sugeno**

**Autor: Amilcar José Pucciarelli
Orientador: Prof. Dr. Gilmar Barreto**

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: **Automação**.

Banca Examinadora

Gilmar Barreto, Dr. DMCSI/FEEC/UNICAMP
Luís Antônio Aguirre, Dr. DEELT/UFMG
Celso Pascoli Bottura, Dr. DMCSI/FEEC/UNICAMP
Fernando Antônio Campos Gomide, Dr. DCA/FEEC/UNICAMP

Campinas, SP

17 de Maio de 2005

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

P961m Pucciarelli, Amilcar José
Modelagem de Séries Temporais Discretas
Utilizando Modelo Nebuloso Takagi-Sugeno
Amilcar José Pucciarelli. – Campinas, SP:
[s.n.], 2005.

Orientador: Gilmar Barreto.
Dissertação (mestrado) - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Sistemas difusos. 2. Modelagem de dados.
3. Mínimos quadrados. 4. Séries temporais.
I. Barreto, Gilmar. II. Universidade Estadual de Campinas.
Faculdade de Engenharia Elétrica e de Computação. III.
Título

Título em Inglês: Discrete Time Series Modelling using Takagi-Sugeno Fuzzy Model.

Keywords: Data Modelling, Fuzzy Systems, Least Square, Time Series.

Área de concentração: Automação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Luís Antônio Aguirre, Celso Pascoli Bottura e Fernando Antônio Campos Gomide.

Data da defesa: 17/05/2005

*“ Não tema a crítica. Henry Ford esqueceu-se de colocar
uma engrenagem de ré em seu primeiro automóvel. ”*

Eleanor L. Doan

Aos meus pais, José e Sandra
Aos meus irmãos, Raphael e Giovani
À minha avó Maria e meus familiares

Agradecimentos

Ao meu orientador, Prof. Gilmar Barreto, sou grato pela orientação.

Ao Prof. Celso Pascoli Bottura pela ajuda e conselhos na minha caminhada durante o mestrado.

Aos colegas do laboratório LCSi, Ginalber, Annabel, Maurício, Rogério, Fernando, Felipe, Marcelo, André e Lorena.

Aos demais colegas de pós-graduação, pelas críticas e sugestões.

A minha família pelo apoio durante esta jornada.

À CAPES e ao CNPq, pelo apoio financeiro.

Resumo

Este estudo primeiramente investiga fundamentos teóricos para análise, desenvolvimento e implementação de algoritmos para modelagem de dados de sistemas dinâmicos e de séries temporais com a finalidade de predição.

As séries temporais utilizadas são baseadas em dados reais retirados da literatura. A grande vantagem de se modelar uma série temporal e de se prever um dado futuro é poder tomar ações antecipadas sobre ela o que vem a ser muito útil, por exemplo em controle.

O modelo nebuloso Takagi-Sugeno será utilizado na modelagem das séries temporais onde os conjuntos nebulosos do antecedente e os parâmetros do conseqüente são estimados via métodos de agrupamentos e identificação paramétrica, respectivamente.

Palavras-chave: Séries Temporais, Predição, Sistemas Nebulosos, Takagi-Sugeno, Modelagem, Identificação.

Abstract

This work firstly explores theoretical foundations for analysis, development and implementation of algorithms for data modelling dynamic systems and time series with a prediction goal.

The used time series are based on real data from the literature. The main advantage of time series modelling and forecasting is make anticipated decisions about it, and this becomes very useful, for example, in control.

The Takagi-Sugeno fuzzy model is used for time series modelling where the antecedent fuzzy partitions and the consequent parameters are estimated by clustering methods and parametric identification, respectively.

Keywords: Time Series, Forecasting, Fuzzy Systems, Takagi-Sugeno, Modelling, Identification.

Sumário

Trabalhos Publicados Pelo Autor	xiii
1 INTRODUÇÃO	1
1.1 Motivação	1
1.2 Objetivo	1
1.3 Organização da dissertação	2
2 SÉRIES TEMPORAIS	3
2.1 Introdução	3
2.2 Conceitos de Processos Estocásticos	7
2.2.1 Média e Variância de um Processo Estacionário	7
2.2.2 Autocovariância e Autocorrelação de um Processo Estacionário	8
2.3 Modelos Paramétricos	9
2.3.1 Modelo Autoregressivo (AR)	9
2.3.2 Modelo Média Móvel (MA)	10
2.3.3 Modelo Autoregressivo - Média Móvel (ARMA)	11
2.3.4 Modelo Autoregressivo - Integrado - Média Móvel (ARIMA)	12
3 INTRODUÇÃO À TEORIA DE SISTEMAS NEBULOSOS	14
3.1 Teoria de Sistemas Nebulosos	14
3.2 Conjuntos Nebulosos	16
3.3 Tipos de Funções de Pertinência	17
3.4 Operações com Conjuntos Nebulosos	20
3.4.1 Complemento Nebuloso	20
3.4.2 União Nebulosa: Normas-S	22
3.4.3 Intersecção Nebulosa: Normas-T	24
3.5 Propriedades Algébricas	25
3.6 Relações Nebulosas e o Princípio da Extensão	27
3.6.1 Relações Clássicas e Relações Nebulosas	27
3.6.2 Operações em Relações Nebulosas	29
3.6.3 Composições de Relações Nebulosas	29
3.6.4 Projeções e Extensões Cilíndricas de Relações	30
3.7 Variáveis Lingüísticas	31
3.8 Regras Nebulosas	31

3.8.1	Interpretações de Regras Nebulosas SE-ENTÃO	32
3.9	Base de Regras Nebulosas	34
3.10	Máquina de Inferência Nebulosa	37
3.11	Fuzzificador e Defuzzificador	37
3.11.1	Fuzzificador	37
3.11.2	Defuzzificador	38
3.12	Exemplo - Controlador Adaptativo PI Nebuloso:	39
3.12.1	Modelo de Controle PI Adaptativo Nebuloso	40
3.12.2	Lei Adaptativa Nebulosa	45
3.12.3	Simulações	47
3.13	Modelos Nebulosos Takagi-Sugeno	49
4	AGRUPAMENTO NEBULOSO	52
4.1	Introdução	52
4.2	Agrupamentos Nebulosos	52
4.2.1	FCM - <i>Fuzzy C-Means</i>	52
4.2.2	Agrupamento Gustafson-Kessel	58
4.2.3	Agrupamento Gath-Geva	62
4.2.4	FCRM - <i>Fuzzy c-Regression Model</i>	67
4.3	Agrupamento Não-Nebuloso	69
4.3.1	ECM - <i>Evolving Clustering Method</i>	69
5	IDENTIFICAÇÃO PARAMÉTRICA	72
5.1	Mínimos Quadrados	72
5.1.1	Mínimos Quadrados Recursivo Ponderado	75
5.2	Resíduos e Erros de Predição	80
5.3	Método do Gradiente	81
5.4	Método dos Mínimos Quadrados Recursivo Estendido	88
6	ESTUDO DE CASOS	91
6.1	Introdução	91
6.2	Braço Robótico Flexível	92
6.2.1	FCM - <i>Fuzzy C-Means</i>	93
6.2.2	Agrupamento Gustafson-Kessel	95
6.2.3	Agrupamento Gath-Geva	97
6.2.4	FCRM - <i>Fuzzy c-Regression Model</i>	99
6.3	Processo Neutralizador de pH	100
6.3.1	FCM - <i>Fuzzy C-Means</i>	100
6.3.2	FCRM - <i>Fuzzy c-Regression Model</i>	103
6.3.3	ECM - <i>Evolving Clustering Method</i>	103
6.4	Função Senoidal	105
6.4.1	FCM - <i>Fuzzy C-Means</i>	106
6.4.2	Agrupamento Gustafson-Kessel	108

7	CONCLUSÕES E COMENTÁRIOS	110
	BIBLIOGRAFIA	112

Lista de Figuras

2.1	Importações mensais australianas oriundas do Japão.	4
2.2	Número de grandes terremotos por ano entre 1900 - 1998.	4
2.3	População dos EUA de 1790 a 1990 em intervalos de 10 anos.	5
2.4	Saída de CO ₂ em porcentagem no forno a gás industrial.	5
2.5	Variável aleatória com valores 0 e 1	6
3.1	Conjuntos da teoria clássica: A e B	14
3.2	Conjunto nebuloso A com seus respectivos graus de pertinência.	15
3.3	Esquema em bloco de um sistema nebuloso com fuzzificador e defuzzificador	16
3.4	Função Triangular com $a = -2.5$, $m = 0$ e $b = 2.5$	18
3.5	Função Γ com $k = 2$ e $a = 1$	18
3.6	Função S com $a = -1$ e $b = 3$	19
3.7	Função trapezoidal com $a = -2.5$, $m = 0$, $n = 2.5$ e $b = 5.0$	19
3.8	Função gaussiana com $k = 2$ e $m = 2$	20
3.9	Função em forma de Sino com $a = 2$, $b = 2$ e $m = 2$	20
3.10	Função sigmoide com $k = 1$ e $m = 2$	21
3.11	Função “ <i>exponential-like</i> ” com $k = 2$ e $m = 1$	21
3.12	Conjunto A “-” e seu complemento \bar{A} “- -”.	22
3.13	Conjuntos nebulosos A “-” e B “- -”.	24
3.14	Normas-s mais utilizadas na literatura: máximo “-”, soma algébrica “- -” e soma drástica “.”.	24
3.15	Conjuntos nebulosos A “-” e B “- -”.	26
3.16	Normas-t mais utilizadas na literatura: mínimo “-”, produto algébrico “- -” e produto drástico “.”.	26
3.17	União “- -” e intersecção “-” de um conjunto nebuloso A e seu complemento \bar{A}	27
3.18	Sistema nebuloso com n entradas e m saídas	35
3.19	Esquema em bloco do controle adaptativo	40
3.20	Diagrama de blocos do controlador	41
3.21	Partições nebulosas do espaço de discurso de entrada	42
3.22	Comportamento da resposta no movimento subamortecido	43
3.23	Comportamento do erro no movimento subamortecido	44
3.24	Diagrama em blocos da lei adaptativa	45
3.25	Resposta a referência constante, degrau unitário.	47
3.26	Resposta à uma variação na referência.	48
3.27	Adaptação do ganho proporcional.	48

3.28	Adaptação do ganho integral.	49
3.29	Conjuntos nebulosos para x_1 e x_2	50
3.30	Superfície nebulosa entrada-saída deste modelo.	50
4.1	Saída de CO ₂ em porcentagem no forno a gás industrial.	54
4.2	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	54
4.3	Funções de pertinência.	55
4.4	Curvas de nível das funções de pertinência.	55
4.5	Quantidades de terremotos maior ou igual a 7.0 na escala Richter.	56
4.6	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	56
4.7	Funções de pertinência.	57
4.8	Curvas de nível das funções de pertinência.	57
4.9	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	59
4.10	Funções de pertinência.	60
4.11	Curvas de nível das funções de pertinência.	60
4.12	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	61
4.13	Funções de pertinência.	61
4.14	Curvas de nível das funções de pertinência.	62
4.15	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	64
4.16	Funções de pertinência.	64
4.17	Curvas de nível das funções de pertinência.	65
4.18	Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.	65
4.19	Funções de pertinência.	66
4.20	Curvas de nível das funções de pertinência.	66
4.21	x_1 : cria um novo grupo C_1^0	69
4.22	x_2 : atualiza grupo $C_1^0 \rightarrow C_1^1$; x_3 : cria um novo grupo C_2^0 ; x_4 : não faz nada.	70
4.23	x_5 : atualiza grupo $C_1^1 \rightarrow C_1^2$; x_6 : não faz nada; x_7 : atualiza grupo $C_2^0 \rightarrow C_2^1$; x_8 : cria um novo grupo C_3^0	71
4.24	x_9 : atualiza grupo $C_2^1 \rightarrow C_2^2$; x_{10} : atualiza grupo $C_3^0 \rightarrow C_3^1$	71
5.1	Estimação dos parâmetros \hat{a} “-” e \hat{b} “-” via mínimos quadrados para o exemplo 4.1.	78
5.2	Seqüência de erro do exemplo 4.1.	79
5.3	Estimação de parâmetros \hat{a} “-” e \hat{b} “-” via mínimos quadrados para o exemplo 4.2.	79
5.4	Estimação de parâmetros \hat{a} “-” e \hat{b} “-” via mínimos quadrados para o exemplo 4.3.	80
5.5	Estimação de parâmetros via método do gradiente para o exemplo 4.4	83

5.6	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.5	83
5.7	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.6	84
5.8	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.8	85
5.9	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.8	85
5.10	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.9	86
5.11	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.10	87
5.12	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.11	87
5.13	Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.12	88
5.14	Estimação de parâmetros \hat{a} “-” e \hat{c} “- -” via método dos mínimos quadrados estendido para o exemplo 4.13	89
5.15	Erro do exemplo 4.13	90
6.1	Torque da estrutura do braço robótico flexível.	92
6.2	Aceleração do braço robótico flexível.	93
6.3	Modelagem da série da saída do braço robótico para o FCM com $p = 6$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.	94
6.4	Modelagem da série da saída do braço robótico para o Gustafson-Kessel para $p = 6$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.	96
6.5	Modelagem da série da saída do braço robótico para o Gath-Geva para $p = 6$ e $c = 3$ onde “- -” é a série desejada e “-” é o modelo estimado.	98
6.6	Modelagem da série da saída do braço robótico para o FCRM para $p = 6$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.	99
6.7	Vazao da solução ácida	101
6.8	Vazao da solução básica	101
6.9	pH da solução no tanque	102
6.10	Modelagem da série da saída do neutralizador de pH para o FCM com $p = 2$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.	102
6.11	Modelagem da série da saída do neutralizador de pH para o FCRM com $p = 2$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.	104
6.12	Modelagem da série da saída do neutralizador de pH para o ECM com a distância máxima igual a 5.5 onde “- -” é a série desejada e “-” é o modelo estimado.	105
6.13	Função seno de x	106
6.14	Modelagem da série da função seno para o FCM com $p = 4$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.	107
6.15	Modelagem da série da saída do neutralizador de pH para o GK com $p = 2$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.	109

Lista de Tabelas

3.1	Tabela verdade para $p \rightarrow q$.	33
3.2	Base de Regras de u	44
3.3	Base de Regras de K'_P	46
3.4	Base de Regras de K'_I	46
3.5	Características de K'_P K'_I	46
6.1	Critérios de validação do FCM para o braço robótico com c fixo igual a 4 e p variante.	93
6.2	Critérios de validação do FCM para o braço robótico com c variante e p igual a 6.	94
6.3	Critérios de validação do GK para o braço robótico para c igual a 4 e p variante.	96
6.4	Critérios de validação do GK para o braço robótico para c variante e p igual a 6.	96
6.5	Critérios de validação do GG para o braço robótico com p igual a 6 e c variante.	97
6.6	Critérios de validação do FCRM para o braço robótico com p igual a 6 e c variante.	99
6.7	Critérios de validação do FCM para o neutralizador de pH com p variante e c igual a 4.	102
6.8	Critérios de validação do FCM para o neutralizador de pH com p igual a 2 e c variante.	103
6.9	Critérios de validação do FCRM para o neutralizador de pH	103
6.10	Critérios de validação do ECM para o neutralizador de pH	104
6.11	Critérios de validação do FCM para a função senoidal com p variante e c igual a 4.	106
6.12	Critérios de validação do FCM para a função senoidal com p igual a 4 e c variante.	107
6.13	Critérios de validação do GK para a função senoidal com $p = 2$ e c variante.	108

Trabalhos Publicados Pelo Autor

1. Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Intelligent adaptive data modeling applied to discrete time-series prediction in industrial processes*. INDUSCON 2004 - IEEE VI Conferência Internacional de Aplicações Industriais, Joinville, SC, 2004.
2. Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Adaptive Identification of Takagi-Sugeno Fuzzy Neural Model to Discrete Time Series Prediction*. SBRN 2004 - VIII Simpósio Brasileiro de Redes Neurais, UFMA - São Luís, MA, 2004.
3. Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Identificação Adaptativa de Modelos Fuzzy Takagi-Sugeno para Predição de Séries Temporais Discretas no Tempo*. 3º DINCON - Congresso Temático de Dinâmica e Controle da SBMAC, UNESP - Campus de Ilha Solteira, 31 de Maio a 3 de Junho, 2004.
4. Pucciarelli, Amilcar J. e Serra, Ginalber L.O. *Projeto de Controle PI Fuzzy Adaptativo Baseado no Conhecimento de Especialistas*. 2º DINCON - Congresso Temático de Dinâmica e Controle da SBMAC, ITA - São José dos Campos, 18 a 22 de Agosto, 2003.

Capítulo 1

INTRODUÇÃO

1.1 Motivação

Uma série temporal consiste de um conjunto de observações no tempo e geralmente as observações em instantes de tempo próximos são correlacionadas. São objetivos dos estudos de séries temporais: investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros da série, descrever apenas o comportamento da série ou procurar periodicidade relevante nos dados.

No problema de previsão de séries temporais é necessário a modelagem (identificação) do sistema real. Esta previsão é muito importante principalmente nas áreas de economia, como por exemplo na predição de ações na bolsa de valores, na engenharia, por exemplo previsão dos índices pluviométricos, e outras áreas.

Os sistemas nebulosos têm sido utilizados em diferentes campos do conhecimento e em diversas aplicações, tais como aproximação de funções, classificação de padrões, previsão de séries temporais, filtragem de sinais e identificação e controle de processos dinâmicos, particularmente quando os métodos convencionais não proporcionam resultados satisfatórios.

A solução abordada nesta dissertação enfoca os sistemas nebulosos do tipo Takagi-Sugeno pois apresentam a capacidade de processar informações incertas ou qualitativas com capacidade de aproximação universal.

1.2 Objetivo

O objetivo deste trabalho é o de desenvolver sistemas nebulosos funcionais do tipo Takagi-Sugeno (TS) e aplicá-los na resolução de problemas de modelagem de séries temporais discretas. É realizado também um estudo comparativo para os métodos de agrupamento, geradores dos conjuntos nebulosos, mostrados nesta dissertação.

Para o treinamento *offline* do conseqüente do modelo TS é utilizado o método de mínimos quadrados ponderado onde o peso de ponderação é igual ao grau de ativação da regra onde está ocorrendo a estimação dos parâmetros. Portanto, a abordagem para a estimação dos parâmetros do conseqüente é local, ou seja, para cada regra os parâmetros são estimados separadamente. A partição do espaço de discurso das variáveis de entrada também serão realizados de modo

offline.

O modelo utilizado no conseqüente é do tipo autoregressivo (AR) determinístico. Portanto, não consideramos que haja algum ruído incluso no modelo. A ordem do modelo é determinada por tentativa e erro utilizando os critérios de erro máximo absoluto ($e_{\text{máx}}$), erro quadrático médio (EQM) ou *mean square error* (*MSE*) e variância levada em conta (VLC) ou *variance accounted for* (*VAF*) para seleção.

1.3 Organização da dissertação

Este trabalho está dividido em sete capítulos. Este capítulo apresenta uma visão geral do escopo da tese, assim como a motivação para o desenvolvimento do trabalho e seus objetivos.

O Capítulo 2 apresenta uma revisão dos principais conceitos de séries temporais. Isto inclui as definições de séries temporais, média, variância, autocovariância e autocorrelação quando estas séries apresentam comportamento estacionário. Também são mostrados alguns modelos paramétricos clássicos, muito conhecidos na literatura para a modelagem de sistemas dinâmicos.

O Capítulo 3 apresenta os principais conceitos relacionados a teoria de conjuntos nebulosos: definições, máquina de inferência nebulosa, modelo lingüístico do tipo Mamdani e um enfoque especial no modelo funcional do tipo Takagi-Sugeno. O modelo Takagi-Sugeno é composto em sua premissa por relações nebulosas e no conseqüente apresenta uma função das variáveis numéricas de entrada.

O Capítulo 4 apresenta alguns métodos de agrupamentos nebulosos, como FCM (*Fuzzy C-Means*), Gustafson-Kessel, Gath-Geva e FCRM (*Fuzzy C-Regression Model*), e não nebulosos, como ECM (*Evolving Clustering Method*), para partionar o espaço de discurso das variáveis de entrada, gerando os conjuntos nebulosos.

No Capítulo 5 apresentamos alguns métodos de identificação paramétrica que podem ser utilizados para a estimação dos parâmetros do conseqüente do modelo Takagi-Sugeno. É dado um enfoque especial no método de mínimos quadrados recursivo ponderado, porém comentamos outros métodos como gradiente e mínimos quadrados recursivo estendido.

O Capítulo 6 apresenta alguns resultados obtidos na modelagem de séries temporais utilizando o modelo nebuloso Takagi-Sugeno. O modelo utilizado no conseqüente será o modelo determinístico autoregressivo (AR), apresentado no Capítulo 2. Os conjuntos nebulosos e os parâmetros do modelo AR são determinados pelos métodos de agrupamento FCM, Gustafson-Kessel, Gath-Geva, FCRM e ECM, e a técnica de mínimos quadrados recursivo ponderado, respectivamente.

Finalmente, o Capítulo 7, apresenta as conclusões do trabalho e sugestões para trabalhos futuros.

Capítulo 2

SÉRIES TEMPORAIS

2.1 Introdução

Neste capítulo serão apresentados conceitos básicos sobre a análise, modelagem e predição de séries temporais. O objetivo deste capítulo é apenas mostrar certas características das séries temporais e os métodos mais comentados na literatura Anderson [3], Ballini [5], Box & Jenkins [8], Brockwell & Davis [9], Montgomery & Johnson [28] e Morettin & Toloí [29].

Uma série temporal consiste de um conjunto de observações no tempo, $k = t_1, \dots, t_N$, onde $t_i \in \mathfrak{R}$, $k \in \mathfrak{N}$ de uma variável $y(k)$. Em geral, as observações vizinhas são correlacionadas e o estudo de uma série temporal consiste em analisar e modelar esta dependência. Uma possível aplicação é a predição de um valor futuro da série.

Exemplos de séries temporais ocorrem em variados campos como Engenharia, Geofísica, Meteorologia, Sociologia e Economia:

- Séries temporais econômicas: lucros de empresas no ano, preço diário de mercadorias e preço de ações na bolsa. A Fig.2.1 mostra uma série temporal de importações mensais na Austrália vindas do Japão entre os meses de julho de 1965 a outubro 1993.
- Séries temporais físicas: quantidade de chuva, temperatura medida em horas, dias ou anos. A Fig.2.2 mostra o número de terremotos com valor 7 na escala Richter ou superior entre os anos de 1900 e 1998.
- Séries temporais demográficas: população de uma região, número de desempregados e taxa de mortalidade infantil. A Fig.2.3 mostra a população dos EUA de 1790 a 1990 em intervalos de 10 anos.
- Processos: saída de uma caldeira. A Fig.2.4 mostra a saída de CO_2 em porcentagem em um forno a gás industrial.
- Processos binários: comunicação digital. A Fig.2.5 mostra uma série temporal com valores entre 0 e 1.

São objetivos dos estudos de séries temporais: investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros da série, descrever apenas o comportamento da série

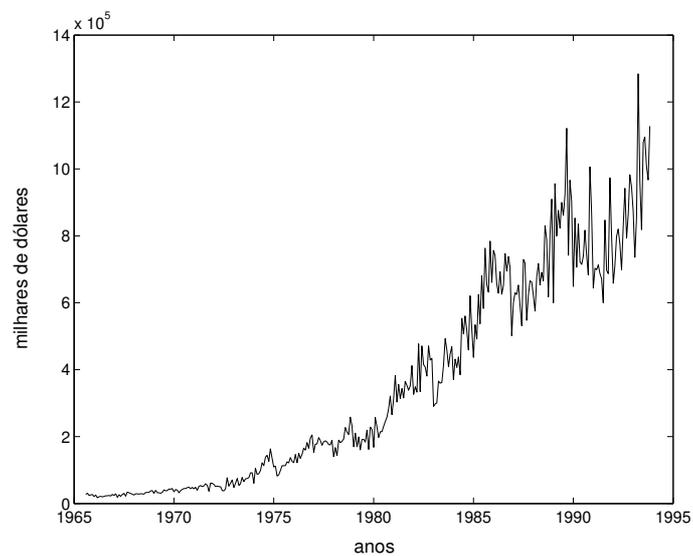


Figura 2.1: Importações mensais australianas oriundas do Japão.

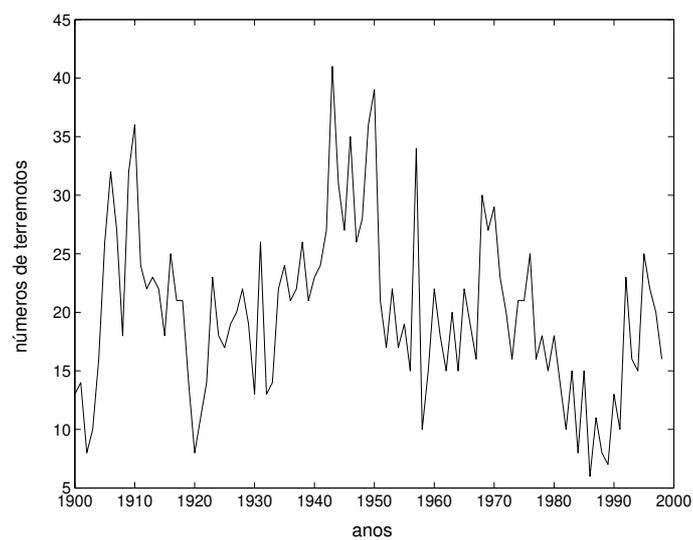


Figura 2.2: Número de grandes terremotos por ano entre 1900 - 1998.

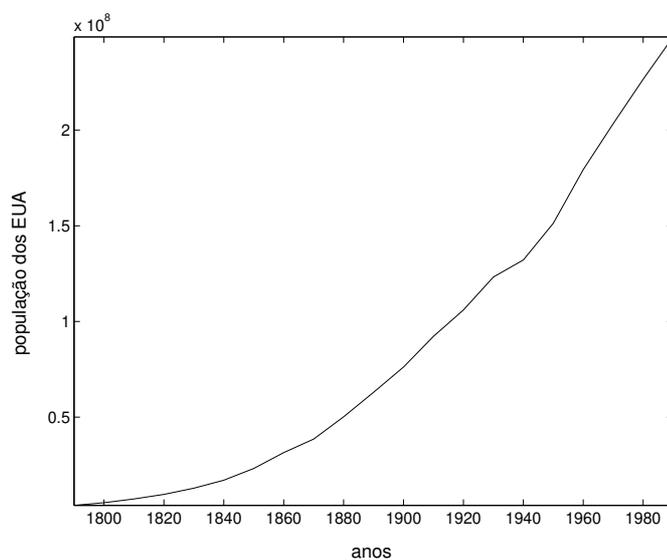


Figura 2.3: População dos EUA de 1790 a 1990 em intervalos de 10 anos.

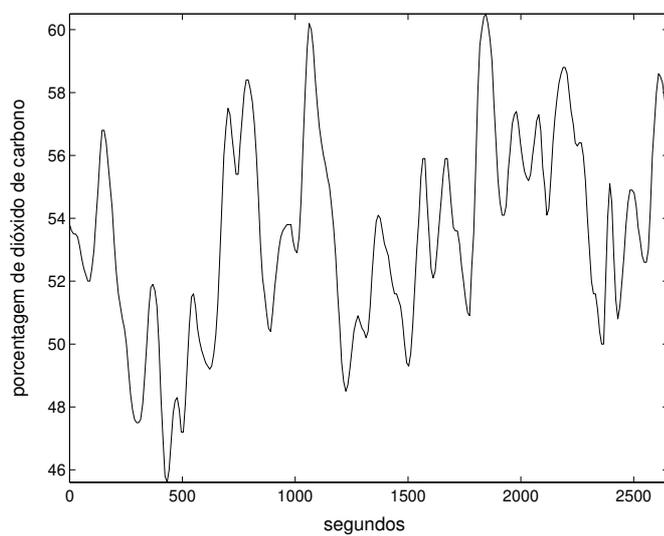


Figura 2.4: Saída de CO₂ em porcentagem no forno a gás industrial.

Observando os casos anteriores, podemos generalizar que a série temporal é representada da seguinte forma:

$$y(k) = b_1 f_1(k) + b_2 f_2(k) + \dots + b_q f_q(k) + \epsilon(k) \quad (2.3)$$

onde b_i são parâmetros, $f_i(k)$ funções de k e q é o número de funções.

Outro modelo utilizado é baseado em valores passados da série ou também chamados auto-regressores. Deste modo, as funções podem ser escritas no caso linear como:

$$f_i(k) = y(k - i) \quad (2.4)$$

e reescrevendo (2.3), teremos:

$$y(k) = b_1 y(k - 1) + b_2 y(k - 2) + \dots + b_q y(k - q) + b_{q+1} + \epsilon(k) \quad (2.5)$$

onde b_{q+1} é uma constante extra somada a equação.

2.2 Conceitos de Processos Estocásticos

Um processo estocástico é uma família $\{y(k), k \in Z_+\}$, onde Z_+ é o conjunto dos inteiros positivos, tal que, para cada $k \in Z_+$, $y(k)$ é uma variável aleatória, Ballini [5].

Em geral, as observações que caracterizam uma série temporal igualmente espaçada podem ser descritas por uma variável aleatória dada por: $\{y(k), 1 \leq k \leq N\}$, com função densidade de probabilidade conjunta $p(y(1), y(2), \dots, y(N))$, Box e Jenkins [8].

Uma série temporal é dita estacionária se não há mudanças sistemáticas na média e na matriz de covariância ao longo do tempo. Ou ainda, se suas propriedades não são alteradas no tempo, de modo que a origem não é importante. Em outras palavras, se o processo estocástico $y(k)$ é estritamente estacionário, a distribuição de probabilidade conjunta com as N observações $y(1), y(2), \dots, y(N)$ para $k = 1, 2, \dots, y(N)$ é a mesma associada com as N observações $y(1 + \tau), y(2 + \tau), \dots, y(N + \tau)$, quando adicionamos uma constante $\tau \in Z_+$ ao tempo, isto é:

$$p(y(1), y(2), \dots, y(N)) = p(y(1 + \tau), y(2 + \tau), \dots, y(N + \tau)). \quad (2.6)$$

O processo estocástico é estacionário no sentido amplo se e somente se:

1. $E[y(k)] = \mu(k) = \mu$, constante para qualquer k ;
2. $Var[y(k)] = \sigma^2(k) = \sigma^2$, constante para qualquer k ;
3. $Cov(y(k), y(k + \tau)) = \gamma(\tau)$, ou seja, é uma função que só depende de τ .

2.2.1 Média e Variância de um Processo Estacionário

Para um processo estocástico unidimensional $y(k)$ tem-se:

- Função valor médio:

$$\mu(k) = E[y(k)] = \int_{-\infty}^{+\infty} y p(y, k) dy \quad (2.7)$$

onde $p(y, k)$ é a densidade de probabilidade da variável aleatória $y(k)$.

- Função variância:

$$\sigma_y^2(k) = E[(y(k) - \mu)(y(k) - \mu)] = \int_{-\infty}^{+\infty} (y - \mu)^2 p(y, k) dy \quad (2.8)$$

Para um processo estacionário a distribuição de probabilidade $p(y, k)$ é a mesma para todo tempo k e pode ser escrita como $p(y)$. Assim, a média é a constante:

$$\mu = E[y(k)] = \int_{-\infty}^{+\infty} y p(y) dy \quad (2.9)$$

e a variância é a constante:

$$\sigma_y^2 = E[y(k) - \mu]^2 = \int_{-\infty}^{+\infty} (y - \mu)^2 p(y) dy \quad (2.10)$$

A média μ do processo estacionário pode ser estimada por:

$$\bar{\mu} = \frac{1}{N} \sum_{k=1}^N y(k) \quad (2.11)$$

sendo denominada média amostral, e a estimativa da variância σ_y^2 , é dada por:

$$\bar{\sigma}_y^2 = \frac{1}{N} \sum_{k=1}^N (y(k) - \bar{\mu})^2 \quad (2.12)$$

denominada variância amostral.

2.2.2 Autocovariância e Autocorrelação de um Processo Estacionário

Em um processo estacionário, a função de distribuição $p(y(k), y(k + \tau))$ é a mesma para todo tempo $k, k + \tau$, a qual é separada por τ . Assim, a covariância entre $y(k)$ e o valor $y(k) + \tau$, separado por τ intervalos é chamada de autocovariância de passo τ sendo definida por:

$$\gamma(\tau) = Cov[y(k), y(k + \tau)] = E[(y(k) - \mu)(y(k + \tau) - \mu)] \quad (2.13)$$

onde μ é a média da série.

Analogamente, a autocorrelação $v(\tau)$ com passo τ é definida como:

$$v(\tau) = \frac{E[(y(k) - \mu)(y(k + \tau) - \mu)]}{\sqrt{E[(y(k) - \mu)^2(y(k + \tau) - \mu)^2]}} = \frac{\gamma(\tau)}{\sigma_y^2} \quad (2.14)$$

onde a variância é igual $\gamma(0)$ tanto no $k + \tau$ -ésimo instante de tempo como no k -ésimo instante de tempo, isto é $v(0) = 1$.

Portanto, a matriz de covariância de um processo estacionário para N observações, é representada por:

$$\Gamma(N) = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(N-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \gamma(N-3) & \cdots & \gamma(0) \end{bmatrix} = \sigma_y^2 \Upsilon(N) \quad (2.15)$$

onde $\Gamma(N)$ é simétrica com elementos constantes nas diagonais e a matriz $\Upsilon(N)$ é chamada de matriz de correlação, dada por:

$$\Upsilon(N) = \begin{bmatrix} 1 & v(1) & v(2) & \cdots & v(N-1) \\ v(1) & 1 & v(1) & \cdots & v(N-2) \\ v(2) & v(1) & 1 & \cdots & v(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v(N-1) & v(N-2) & v(N-3) & \cdots & 1 \end{bmatrix}. \quad (2.16)$$

Ambas as matrizes são definidas positivas para um processo estacionário.

2.3 Modelos Paramétricos

2.3.1 Modelo Autoregressivo (AR)

O modelo autoregressivo é assim chamado porque a representação da observação $y(k)$ é expressa como uma combinação linear de p valores passados $y(k-1), y(k-2), \dots, y(k-p)$ da mesma série temporal:

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \dots + a_p y(k-p) + \varepsilon(k) \quad (2.17)$$

onde a_1, a_2, \dots, a_p são parâmetros do modelo e $\varepsilon(k)$ é o erro. O modelo acima pode ser chamado de modelo autoregressivo de ordem p , ou somente AR(p).

Considerando $B^m y(k) = y(k-m)$, uma nomenclatura muito utilizada também é:

$$y(k) = (a_1 B^1 + a_2 B^2 + \dots + a_p B^p) y(k) + \varepsilon(k) \quad (2.18)$$

ou

$$(1 - a_1 B^1 - a_2 B^2 - \dots - a_p B^p) y(k) = \varepsilon(k) \quad (2.19)$$

Se considerarmos:

$$A_p(B) = 1 - a_1 B^1 - a_2 B^2 - \dots - a_p B^p \quad (2.20)$$

podemos reescrever (2.19) como:

$$A_p(B)y(k) = \varepsilon(k). \quad (2.21)$$

O modelo autoregressivo pode ser usado para representar séries temporais estacionárias e não estacionárias.

Considerando uma série estacionária e um modelo AR(1):

$$y(k) = a_1 y(k-1) + \varepsilon(k), \quad (2.22)$$

onde o ruído $\varepsilon(k)$, $k \geq 0$, é uma seqüência de variáveis aleatórias não correlacionadas com média zero e variância constante, isto é:

$$E[\varepsilon(k)] = 0; \text{Var}[\varepsilon(k)] = \sigma_\varepsilon^2. \quad (2.23)$$

Como $\varepsilon(k)$ é não correlacionada em relação a $y(k-1)$, calculando as variâncias temos:

$$\sigma_y^2 = a_1^2 \sigma_y^2 + \sigma_\varepsilon^2 \quad (2.24)$$

então

$$\sigma_y^2(1 - a_1^2) = \sigma_\varepsilon^2 \quad (2.25)$$

e, para σ_y^2 ser finito e não negativo, requer:

$$-1 < a_1 < 1. \quad (2.26)$$

Esta condição é necessária e suficiente para a série ser estacionária.

Nestas condições, calculando a esperança na equação (2.22):

$$\mu = a_1 \mu + 0 \quad (2.27)$$

e portanto, se $a_1 \neq 1$:

$$\mu = 0. \quad (2.28)$$

No capítulo de Introdução à Teoria de Sistemas Nebulosos, capítulo 3, retornaremos ao modelo AR quando mencionarmos o modelo nebuloso Takagi-Sugeno.

2.3.2 Modelo Média Móvel (MA)

O modelo média móvel de ordem q é dado por:

$$y(k) = \mu + \varepsilon(k) + c_1 \varepsilon(k-1) + c_2 \varepsilon(k-2) + \dots + c_q \varepsilon(k-q) \quad (2.29)$$

onde c_1, c_2, \dots, c_q são parâmetros do modelo, μ é a média da série e $\varepsilon(i)$ é uma variável aleatória com média zero e variância constante:

$$E[\varepsilon(i)] = 0; \text{Var}[\varepsilon(i)] = \sigma_\varepsilon^2. \quad (2.30)$$

O modelo média móvel de ordem q pode ser escrito como MA(q).

Considerando $B^m \varepsilon(k) = \varepsilon(k - m)$, podemos reescrever o modelo média móvel como:

$$y(k) = (1 + c_1 B + c_2 B^2 + \dots + c_q B^q) \varepsilon(k) + \mu \quad (2.31)$$

ou

$$y(k) = C(B) \varepsilon(k) + \mu \quad (2.32)$$

onde $C(B) = 1 + c_1 B + c_2 B^2 + \dots + c_q B^q$ é o operador MA(q).

Considerando o modelo MA(1):

$$y(k) = \mu + \varepsilon(k) + c\varepsilon(k - 1), \quad (2.33)$$

passando μ para o outro lado da equação, elevando ambos os lados ao quadrado e tomando a esperança matemática temos:

$$(y(k) - \mu)^2 = (\varepsilon(k) + c\varepsilon(k - 1))^2, \quad (2.34)$$

ou

$$\sigma_y^2 = (1 + c^2) \sigma_\varepsilon^2. \quad (2.35)$$

No caso de MA(q) teríamos:

$$\sigma_y^2 = (1 + c_1^2 + c_2^2 + \dots + c_q^2) \sigma_\varepsilon^2. \quad (2.36)$$

Portanto, para q finito o processo é sempre estacionário, pois a variância é constante.

2.3.3 Modelo Autoregressivo - Média Móvel (ARMA)

O modelo autoregressivo-média móvel é uma mistura dos modelos autoregressivo (AR) e média móvel (MA). Um modelo ARMA(p, q) é representado como:

$$y(k) = a_1 y(k - 1) + a_2 y(k - 2) + \dots + a_p y(k - p) + \varepsilon(k) + c_1 \varepsilon(k - 1) + c_2 \varepsilon(k - 2) + \dots + c_q \varepsilon(k - q) \quad (2.37)$$

ou

$$A(B)y(k) = C(B)\varepsilon(k) \quad (2.38)$$

onde

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p \quad (2.39)$$

e

$$C(B) = c_1 B + c_2 B^2 + \dots + c_q B^q \quad (2.40)$$

Um caso especial de (2.37) muito utilizado é o modelo ARMA(1, 1):

$$(1 - aB)y(k) = (1 + cB)\varepsilon(k) \quad (2.41)$$

ou

$$y(k) = ay(k-1) + \varepsilon(k) + c\varepsilon(k-1) \quad (2.42)$$

O processo será estacionário para $|a| < 1$, como no caso do AR(1).

As séries temporais analisadas pelos algoritmos desta dissertação serão do tipo estacionárias.

2.3.4 Modelo Autoregressivo - Integrado - Média Móvel (ARIMA)

O modelo autoregressivo - média móvel (ARMA) é um modelo de séries temporais não estacionárias lineares bastante usual.

Uma extensão do modelo ARMA, o modelo autoregressivo - integrado - média móvel (ARIMA), foi proposto para descrever melhor séries temporais não estacionárias, as quais não apresentam médias constantes.

Algumas séries econômicas, por exemplo, são não estacionárias, mas quando diferenciamos tornam-se estacionárias. Por exemplo, $y(k)$ é não estacionário, mas

$$\Delta y(k) = y(k) - y(k-1) = (1 - B)y(k) \quad (2.43)$$

é estacionária.

Uma série pode apresentar várias formas de não estacionariedade. No caso de um modelo AR(1), a condição de estacionariedade é $|a| < 1$. Se $a = 1$ obtemos um processo não estacionário, $\varepsilon(k) = y(k) - y(k-1)$, e se $|a| > 1$ o processo diverge quanto maior for o k . O modelo ARIMA trata destas séries não estacionárias com comportamento não explosivo. Ou seja, através do modelo ARIMA, uma série não estacionária pode ser transformada em uma série estacionária por meio de diferenças.

Definindo o operador diferença Δ como:

$$\Delta y(k) = y(k) - y(k-1) = (1 - B)y(k) \quad (2.44)$$

ou

$$\Delta = 1 - B \quad (2.45)$$

Ordens maiores de diferenças podem ser representadas como:

$$\begin{aligned} \Delta^2 &= (1 - B)^2 \\ \Delta^3 &= (1 - B)^3 \\ \vdots &\quad \quad \quad \vdots \\ \Delta^d &= (1 - B)^d \end{aligned} \quad (2.46)$$

Considerando as premissas anteriores o modelo autoregressivo - integrado - média móvel de ordem (p, d, q) ou ARIMA (p, d, q) , é:

$$A(B)\Delta^d y(k) = C(B)\varepsilon(k). \quad (2.47)$$

Os dois casos do modelo ARIMA mais utilizados são:

- Modelo ARIMA(1, 1, 1):

$$(1 - aB)\Delta y(k) = (1 + cB)\varepsilon(k) \quad (2.48)$$

simplificando temos:

$$y(k) = (1 + a)y(k - 1) - ay(k - 2) + \varepsilon(k) + c\varepsilon(k - 1) \quad (2.49)$$

- Modelo ARIMA(2, 1, 0):

$$(1 - a_1B - a_2B^2)\Delta y(k) = \varepsilon(k) \quad (2.50)$$

ou

$$y(k) = y(k - 1) + a_1(y(k - 1) - y(k - 2)) + a_2(y(k - 2) - y(k - 3)) + \varepsilon(k) \quad (2.51)$$

Observe que se $d = 0$, o modelo ARIMA transforma-se no modelo autoregressivo - média móvel (ARMA).

Capítulo 3

INTRODUÇÃO À TEORIA DE SISTEMAS NEBULOSOS

3.1 Teoria de Sistemas Nebulosos

Na teoria clássica de conjuntos, um elemento do universo de discurso (domínio) pertence ou não pertence ao referido conjunto (valores de pertinência μ de 0 ou 1). Um exemplo de conjuntos clássicos pode ser visto na Fig.3.1.

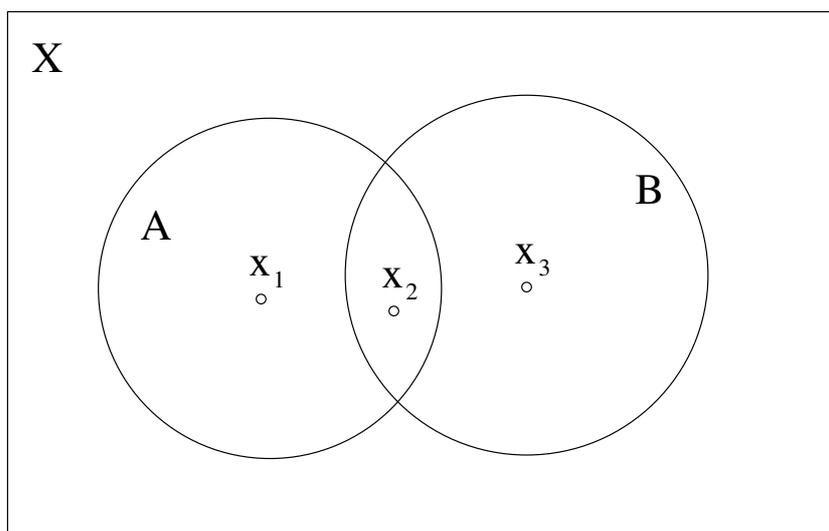


Figura 3.1: Conjuntos da teoria clássica: A e B .

Neste caso, em relação aos elementos x_1 , x_2 e x_3 e aos conjuntos A e B podemos escrever que:

- $x_1 \in A \rightarrow \mu_A(x_1) = 1$ e $x_1 \notin B \rightarrow \mu_B(x_1) = 0$;
- $x_2 \in A \rightarrow \mu_A(x_2) = 1$ e $x_2 \in B \rightarrow \mu_B(x_2) = 1$;
- $x_3 \notin A \rightarrow \mu_A(x_3) = 0$ e $x_3 \in B \rightarrow \mu_B(x_3) = 1$;

Na teoria dos conjuntos nebulosos existe um grau de pertinência, no intervalo de $[0, 1]$, de cada elemento, a um determinado conjunto. Podemos ver na Fig.3.2 os elementos, x_1 , x_2 , x_3 e x_4 , apresentam graus de pertinências em relação ao conjunto nebuloso A de 0.0, 0.3, 0.6 e 1.0, respectivamente.

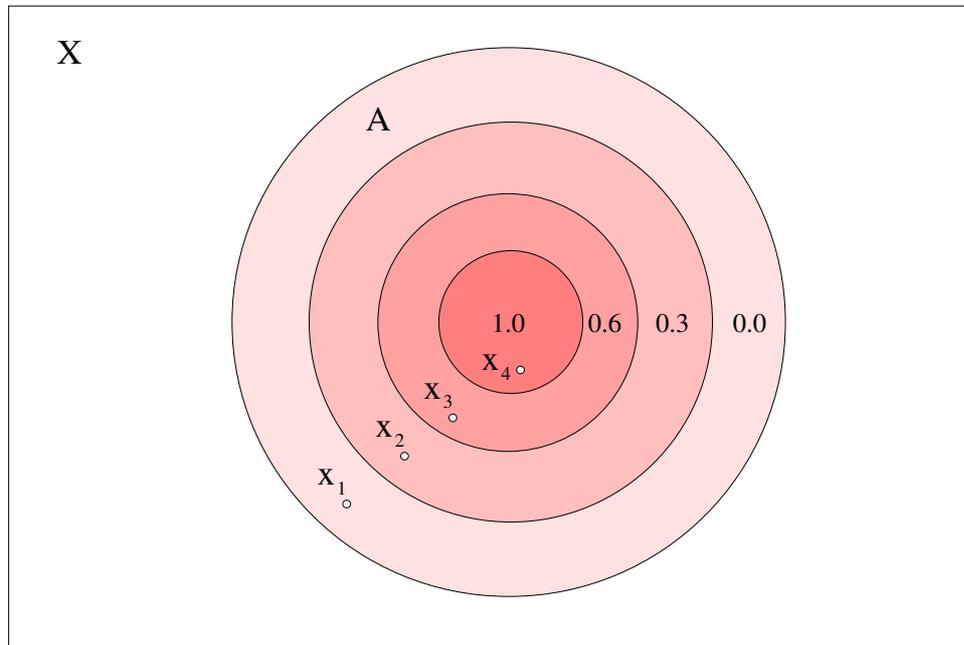


Figura 3.2: Conjunto nebuloso A com seus respectivos graus de pertinência.

A teoria de conjuntos nebulosos pode ser utilizada para traduzir em termos matemáticos a informação imprecisa expressa por um conjunto de regras lingüísticas. O modelo lingüístico composto de regras nebulosas convencionais, isto é, regras da forma SE-ENTÃO relacionando variáveis lingüísticas de entrada (premissa) e saída (conseqüente), pode ser visto como:

$$Ru^l : \text{SE } x_1 \text{ é } A_1^l \dots x_n \text{ é } A_n^l, \text{ ENTÃO } y^l \text{ é } B^l \quad (3.1)$$

As principais vantagens dos sistemas nebulosos são:

- capacidade de aproximação universal;
- é capaz de capturar informações vagas, em geral descritas em uma linguagem natural e convertê-las para um formato numérico, de fácil manipulação pelos computadores;
- a saída de um sistema nebuloso é geralmente contínua e suave com o tempo, muito utilizada para o controle de sistemas;
- na maioria dos casos, obtemos interpretabilidade e simplificação do modelo do processo;
- melhor tratamento das imprecisões inerentes aos sensores utilizados;
- facilidade de incorporação do conhecimento de especialistas humanos;

- não é sempre necessário conhecer o modelo matemático do processo.

Há dois tipos de arquiteturas de sistemas nebulosos que são comumente usadas na literatura:

1. Modelos Nebulosos Lingüísticos do Tipo Mamdani:

As entradas e saídas dos sistemas reais de engenharia são variáveis que possuem valores reais (Ver Fig.3.3). O sistema nebuloso com fuzzificador e defuzzificador permite a transformação dos valores de entradas reais para um conjunto nebuloso, trabalha com estes dados transformados e retorna um valor real na saída. O formato da regra para este modelo pode ser visto em (3.1) onde as variáveis do antecedente e conseqüente apresentam valores lingüísticos. Além de sistemas nebulosos do tipo Mamdani com fuzzificador e defuzzificador podemos ter sistemas nebulosos derivados como: sistemas nebulosos com fuzzificador, sistemas nebulosos com defuzzificador, e sistemas puramente nebulosos.

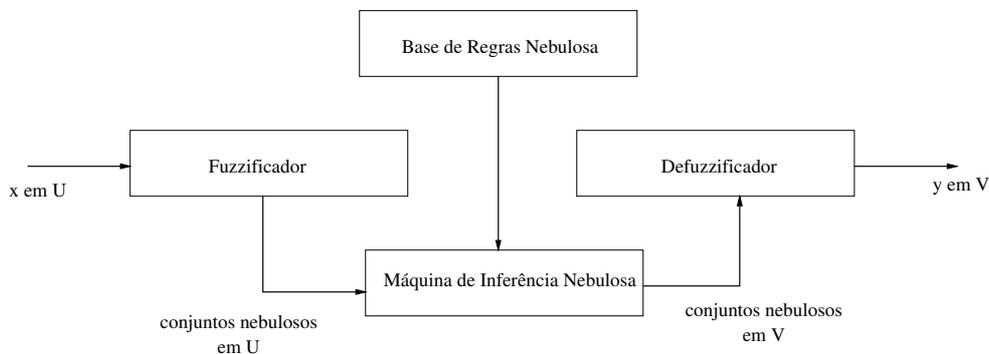


Figura 3.3: Esquema em bloco de um sistema nebuloso com fuzzificador e defuzzificador

2. Modelos Nebulosos Funcionais do Tipo Takagi-Sugeno (TS):

Ao contrário das regras nebulosas SE-ENTÃO apresentadas em (3.1), sistemas TS utilizam regras da seguinte forma:

$$Ru^l : \text{SE } x_1 \text{ é } A_1^l \dots x_n \text{ é } A_n^l, \text{ ENTÃO } y^l = f(\mathbf{x}) \quad (3.2)$$

onde a função de saída de cada regra y^l é função das variáveis numéricas de entrada $\mathbf{x} = (x_1; \dots; x_n)^T$. Por esta razão este modelo não requer o procedimento de defuzzificação.

Cada regra determina um modelo local e os conjuntos nebulosos da sua premissa determinam a sua região de ativação.

3.2 Conjuntos Nebulosos

A teoria de conjuntos nebulosos pode ser vista em diversos livros: Pedrycz e Gomide [31], Wang [42]. Na teoria clássica de conjuntos, um conjunto A consiste de um número finito ou infinito de elementos pertencentes a um conjunto específico chamado universo de discurso, \mathbf{X} . Os elementos do universo de discurso podem ou não pertencer ao conjunto A . Isto pode ser expresso pela função característica (pertinência) μ_A :

$$\mu_A(x) = \begin{cases} 1 & \text{se e somente se } x \in A \\ 0 & \text{se e somente se } x \notin A \end{cases}$$

A função $\mu_A : \mathbf{X} \rightarrow \{0, 1\}$ proporciona limites bem definidos. O conceito base dos conjuntos nebulosos é introduzir valores intermediários no intervalo $[0, 1]$ para a função característica. Assim um conjunto nebuloso A em \mathbf{X} é um conjunto de pares ordenados:

$$A = \{\mu_A(x)/x\} \quad x \in \mathbf{X}$$

onde $\mu_A(x)$ é o valor de pertinência de x em A , no intervalo fechado $[0, 1]$:

$$\mu_A(x) : \mathbf{X} \rightarrow [0, 1]$$

3.3 Tipos de Funções de Pertinência

Uma função de pertinência pode ter diversos formatos e propriedades. A escolha da função é feita dependendo do contexto de sua aplicação. Em alguns casos, a sensibilidade do sistema não é influenciada pelo formato da função. As funções de pertinência mais utilizadas na literatura Pedrycz e Gomide [31], Wang [42] são:

1. Função Triangular:

$$\mu_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x-a}{m-a}, & \text{se } x \in [a, m] \\ \frac{b-x}{b-m}, & \text{se } x \in [m, b] \\ 0, & \text{se } x \geq b, \end{cases} \quad (3.3)$$

onde m é o valor modal, e a e b são os limites superior e inferior, respectivamente, para valores não nulos de $A(x)$, conforme representado na Fig.3.4.

2. Função Γ :

$$\mu_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ 1 - e^{-k(x-a)^2}, & \text{se } x > a \end{cases} \quad (3.4)$$

ou

$$\mu_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{k(x-a)^2}{1+k(x-a)^2}, & \text{se } x > a, \end{cases} \quad (3.5)$$

onde $k > 0$, conforme representado na Fig.3.5.

3. Função S :

$$\mu_A(x) = \begin{cases} 0, & \text{se } x \leq a \\ 2 \left(\frac{x-a}{b-a}\right)^2, & \text{se } x \in [a, m] \\ 1 - 2 \left(\frac{x-b}{b-a}\right)^2, & \text{se } x \in [m, b] \\ 1, & \text{se } x > b, \end{cases} \quad (3.6)$$

Conforme representado na Fig.3.6.

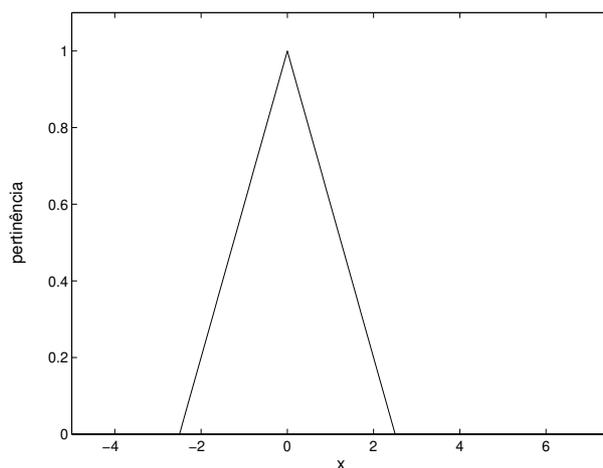


Figura 3.4: Função Triangular com $a = -2.5$, $m = 0$ e $b = 2.5$.

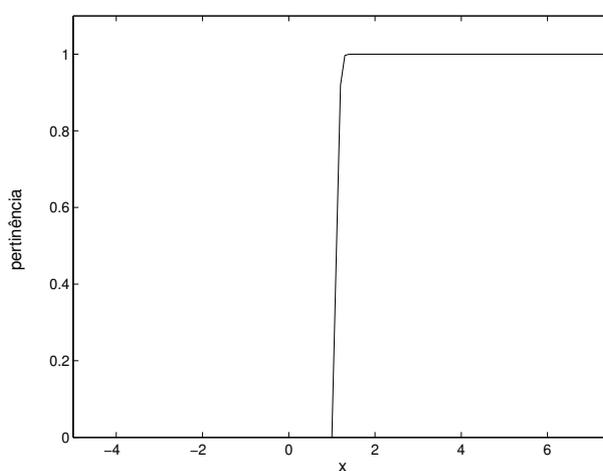


Figura 3.5: Função Γ com $k = 2$ e $a = 1$.

4. Função Trapezoidal:

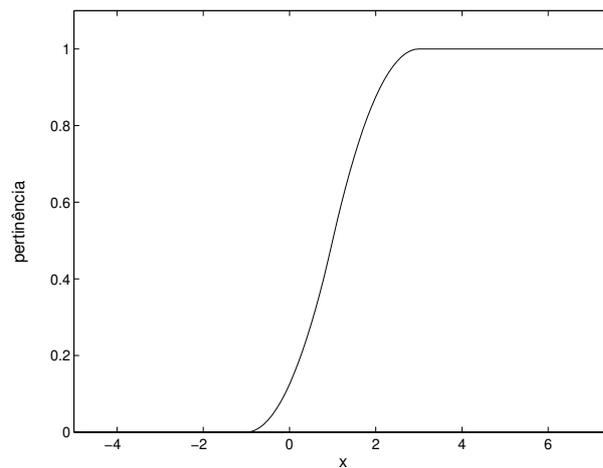
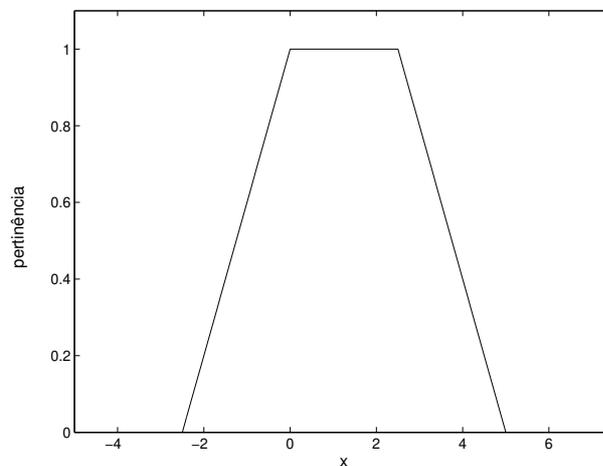
$$\mu_A(x) = \begin{cases} 0, & \text{se } x < a \\ \frac{x-a}{m-a}, & \text{se } x \in [a, m] \\ 1, & \text{se } x \in [m, n] \\ \frac{b-x}{b-n}, & \text{se } x \in [n, b] \\ 0, & \text{se } x > b, \end{cases} \quad (3.7)$$

Conforme representado na Fig.3.7.

5. Função Gaussiana:

$$\mu_A(x) = e^{-k(x-m)^2}, \quad (3.8)$$

onde $k > 0$. Conforme representado na Fig.3.8.

Figura 3.6: Função S com $a = -1$ e $b = 3$.Figura 3.7: Função trapezoidal com $a = -2.5$, $m = 0$, $n = 2.5$ e $b = 5.0$.

6. Função em forma de Sino:

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x-m}{a} \right|^{2b}} \quad (3.9)$$

Conforme representado na Fig.3.9.

7. Função Sigmoide:

$$\mu_A(x) = \frac{1}{1 + e^{-k(x-m)}} \quad (3.10)$$

Conforme representado na Fig.3.10.

8. Função “*Exponential-like*” (similar à exponencial):

$$\mu_A(x) = \frac{1}{1 + k(x-m)^2}, \quad (3.11)$$

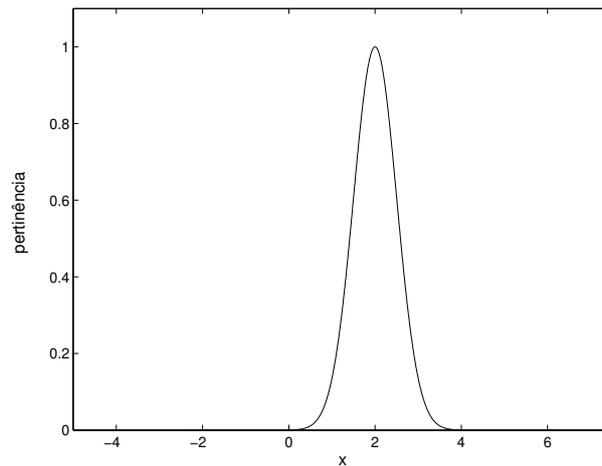


Figura 3.8: Função gaussiana com $k = 2$ e $m = 2$.

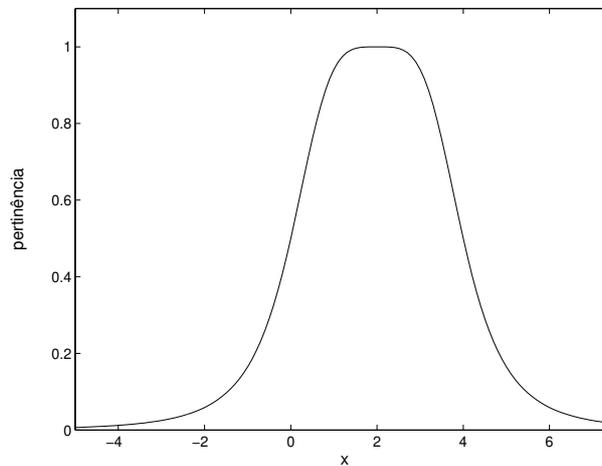


Figura 3.9: Função em forma de Sino com $a = 2$, $b = 2$ e $m = 2$.

$k > 1$, ou

$$\mu_A(x) = \frac{k(x - m)^2}{1 + k(x - m)^2}, \quad (3.12)$$

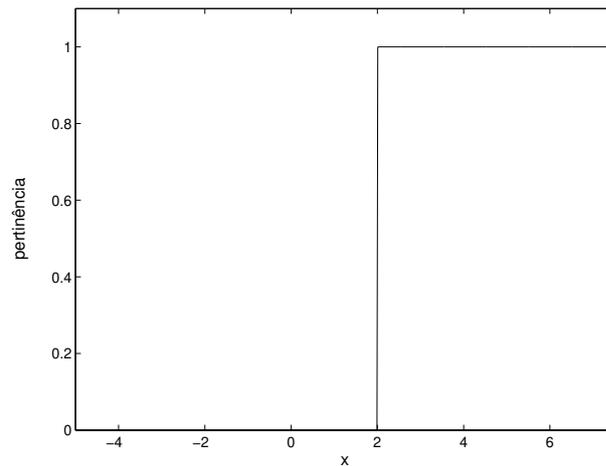
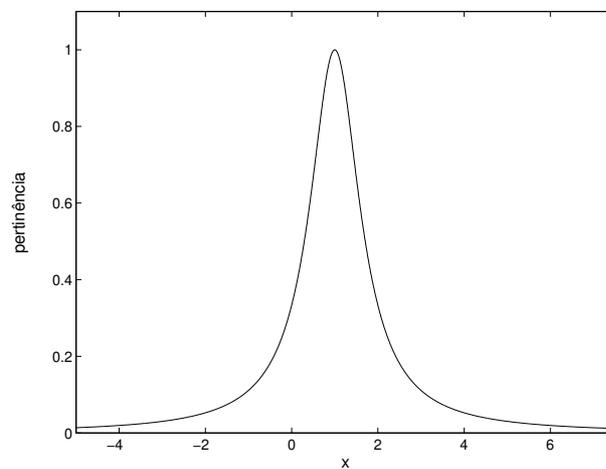
$k > 0$. Conforme representado na Fig.3.11.

3.4 Operações com Conjuntos Nebulosos

3.4.1 Complemento Nebuloso

Seja $c : [0, 1] \rightarrow [0, 1]$ um operador que mapeia os graus de pertinência $\mu_A(x)$ do conjunto nebuloso A para o complemento do conjunto nebuloso A , ou seja:

$$c[\mu_A(x)] = \mu_{\bar{A}}(x) \quad (3.13)$$

Figura 3.10: Função sigmoidal com $k = 1$ e $m = 2$.Figura 3.11: Função “*exponential-like*” com $k = 2$ e $m = 1$.

Neste caso, a função c para ser considerada um complemento deve satisfazer as seguintes propriedades:

- $c(0) = 1$ e $c(1) = 0$;
- Considerando as funções de pertinência $\mu_A, \mu_B \in [0, 1]$, se $\mu_A < \mu_B$, então $c(\mu_A) \geq c(\mu_B)$.

Três exemplos de complementos podem ser vistos abaixo:

1. complemento de 1

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (3.14)$$

2. classe Sugeno

$$\mu_{\bar{A}}(x) = \frac{1 - \mu_A(x)}{1 + \lambda \mu_A(x)} \quad (3.15)$$

onde $\lambda \in (-1, \infty)$

3. classe Yager

$$\mu_{\bar{A}}(x) = (1 - \mu_A(x)^w)^{1/w} \quad (3.16)$$

onde $w \in (0, \infty)$

Um exemplo de complemento mais comum, complemento de 1, pode ser visto na Fig.3.12.

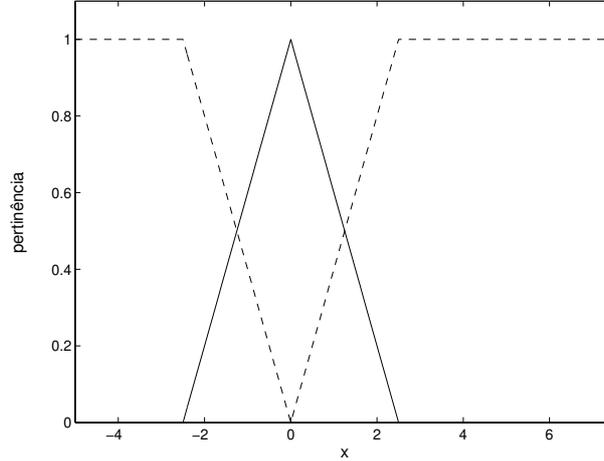


Figura 3.12: Conjunto A “-” e seu complemento \bar{A} “- -”.

3.4.2 União Nebulosa: Normas-S

Seja $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ um operador que mapeia os graus de pertinência dos conjuntos nebulosos A e B em um valor entre 0 e 1 que representa a união de A e B , ou seja,

$$s[\mu_A(x), \mu_B(x)] = \mu_{A \cup B}(x) \quad (3.17)$$

Neste caso, o operador s para ser considerado uma norma-s deve satisfazer as seguintes propriedades:

- $s(1, 1) = 1$, $s(0, \mu_A(x)) = s(\mu_A(x), 0) = \mu_A(x)$;
- $s(\mu_A(x), \mu_B(x)) = s(\mu_B(x), \mu_A(x))$;
- Se $\mu_A(x) \leq \mu_{A'}(x)$ e $\mu_B(x) \leq \mu_{B'}(x)$, então $s(\mu_A(x), \mu_B(x)) \leq s(\mu_{A'}(x), \mu_{B'}(x))$;
- $s(s(\mu_A(x), \mu_B(x)), \mu_C(x)) = s(\mu_A(x), s(\mu_B(x), \mu_C(x)))$.

Alguns exemplos de normas-s podem ser vistos abaixo:

1. máximo:

$$s[\mu_A(x), \mu_B(x)] = \max[\mu_A(x), \mu_B(x)]. \quad (3.18)$$

2. classe Dombi:

$$s[\mu_A(x), \mu_B(x)] = \frac{1}{1 + \left[\left(\frac{1}{\mu_A(x)} - 1 \right)^{-\lambda} + \left(\frac{1}{\mu_B(x)} - 1 \right)^{-\lambda} \right]^{-1/\lambda}} \quad (3.19)$$

onde $\lambda \in (0, \infty)$.

3. classe Dubois-Prade:

$$s[\mu_A(x), \mu_B(x)] = \frac{\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x) - \min(\mu_A(x), \mu_B(x), 1 - \alpha)}{\max(1 - \mu_A(x), 1 - \mu_B(x), \alpha)} \quad (3.20)$$

onde o parâmetro $\alpha \in [0, 1]$.

4. classe Yager:

$$s[\mu_A(x), \mu_B(x)] = \min \left[1, (\mu_A(x)^w + \mu_B(x)^w)^{1/w} \right] \quad (3.21)$$

onde $w \in (0, \infty)$

5. soma algébrica:

$$s[\mu_A(x), \mu_B(x)] = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x) \quad (3.22)$$

6. soma drástica:

$$s[\mu_A(x), \mu_B(x)] = \begin{cases} \mu_A(x), & \text{se } \mu_B(x) = 0 \\ \mu_B(x), & \text{se } \mu_A(x) = 0 \\ 1, & \text{caso contrário} \end{cases} \quad (3.23)$$

7. soma Einstein:

$$s[\mu_A(x), \mu_B(x)] = \frac{\mu_A(x) + \mu_B(x)}{1 + \mu_A(x)\mu_B(x)} \quad (3.24)$$

Teorema 3.1 Para qualquer norma s , isto é, para uma função $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que satisfaz as condições anteriores, tem-se a seguinte desigualdade:

$$\max(\mu_A(x), \mu_B(x)) \leq s(\mu_A(x), \mu_B(x)) \leq s_{sd}(\mu_A(x), \mu_B(x)) \quad (3.25)$$

para $\mu_A(x), \mu_B(x) \in [0, 1]$ e s_{sd} o operador de soma drástica, expressão (3.23).

Prova: Ver Wang [42].

Alguns exemplos mais comuns de normas-s podem ser vistos nas Fig.3.13 e Fig.3.14.

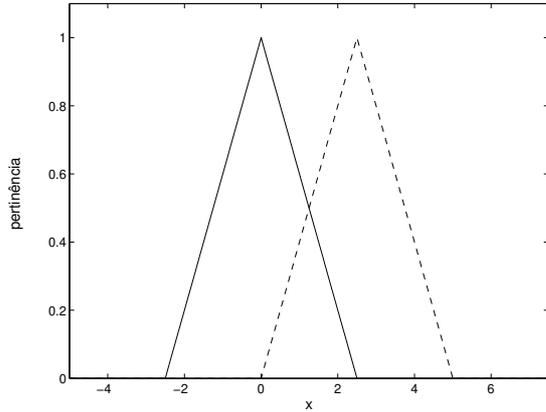


Figura 3.13: Conjuntos nebulosos A “-” e B “-”.

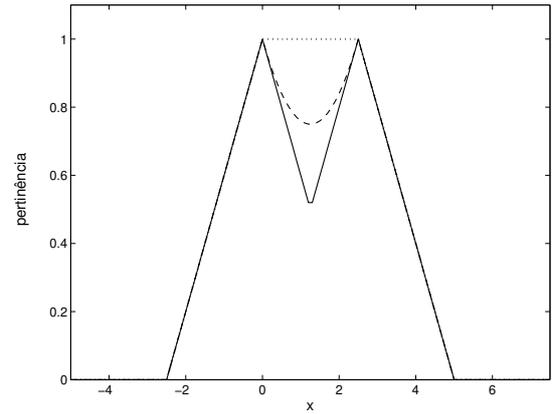


Figura 3.14: Normas-s mais utilizadas na literatura: máximo “-”, soma algébrica “- -” e soma drástica “.”.

3.4.3 Intersecção Nebulosa: Normas-T

Seja $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ um operador que mapeia os graus de pertinência dos conjuntos nebulosos A e B em um valor entre 0 e 1 que representa a intersecção de A e B , ou seja:

$$t[\mu_A(x), \mu_B(x)] = \mu_{A \cap B}(x) \quad (3.26)$$

Neste caso, o operador t para ser considerado uma norma-t deve satisfazer as seguintes propriedades:

- $t(0, 0) = 0$, $t(1, \mu_A(x)) = s(\mu_A(x), 1) = \mu_A(x)$;
- $t(\mu_A(x), \mu_B(x)) = t(\mu_B(x), \mu_A(x))$;
- Se $\mu_A(x) \leq \mu_{A'}(x)$ e $\mu_B(x) \leq \mu_{B'}(x)$, então $t(\mu_A(x), \mu_B(x)) \leq t(\mu_{A'}(x), \mu_{B'}(x))$;
- $t(t(\mu_A(x), \mu_B(x)), \mu_C(x)) = t(\mu_A(x), t(\mu_B(x), \mu_C(x)))$.

Alguns exemplos de normas-t podem ser vistos abaixo:

1. mínimo:

$$t[\mu_A(x), \mu_B(x)] = \min[\mu_A(x), \mu_B(x)]. \quad (3.27)$$

2. classe Dombi:

$$t[\mu_A(x), \mu_B(x)] = \frac{1}{1 + \left[\left(\frac{1}{\mu_A(x)} - 1 \right)^\lambda + \left(\frac{1}{\mu_B(x)} - 1 \right)^\lambda \right]^{1/\lambda}} \quad (3.28)$$

onde $\lambda \in (0, \infty)$.

3. classe Dubois-Prade:

$$t[\mu_A(x), \mu_B(x)] = \frac{\mu_A(x)\mu_B(x)}{\max(\mu_A(x), \mu_B(x), \alpha)} \quad (3.29)$$

onde o parâmetro $\alpha \in [0, 1]$.

4. classe Yager:

$$t[\mu_A(x), \mu_B(x)] = 1 - \min \left[1, ((1 - \mu_A(x))^w + (1 - \mu_B(x))^w)^{1/w} \right] \quad (3.30)$$

onde $w \in (0, \infty)$

5. produto algébrico:

$$t[\mu_A(x), \mu_B(x)] = \mu_A(x)\mu_B(x) \quad (3.31)$$

6. produto drástico:

$$t[\mu_A(x), \mu_B(x)] = \begin{cases} \mu_A(x), & \text{se } \mu_B(x) = 1 \\ \mu_B(x), & \text{se } \mu_A(x) = 1 \\ 0, & \text{caso contrário} \end{cases} \quad (3.32)$$

7. produto Einstein:

$$t[\mu_A(x), \mu_B(x)] = \frac{\mu_A(x)\mu_B(x)}{2 - (\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x))} \quad (3.33)$$

Teorema 3.2 Para qualquer norma-t t , isto é, para uma função $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que satisfaz as condições anteriores, tem-se a seguinte desigualdade:

$$t_{pd}(\mu_A(x), \mu_B(x)) \leq t(\mu_A(x), \mu_B(x)) \leq \min(\mu_A(x), \mu_B(x)) \quad (3.34)$$

para $\mu_A(x), \mu_B(x) \in [0, 1]$ e t_{pd} o operador de produto drástico, expressão (3.32).

Prova: Ver Wang [42].

Alguns exemplos mais comuns de normas-t podem ser vistos nas Fig.3.15 e Fig.3.16.

3.5 Propriedades Algébricas

Utilizando o conceito de união, intersecção e complemento para os conjuntos nebulosos A , B e C definidos no universo de discurso \mathbf{X} , é possível

1. Comutatividade:

$$\begin{aligned} A \cap B &= B \cap A \\ A \cup B &= B \cup A \end{aligned}$$

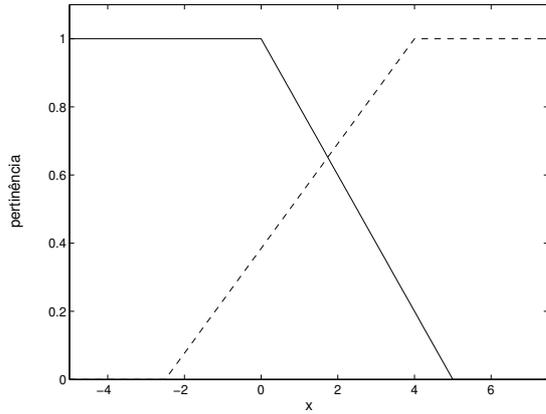


Figura 3.15: Conjuntos nebulosos A “-” e B “-”.

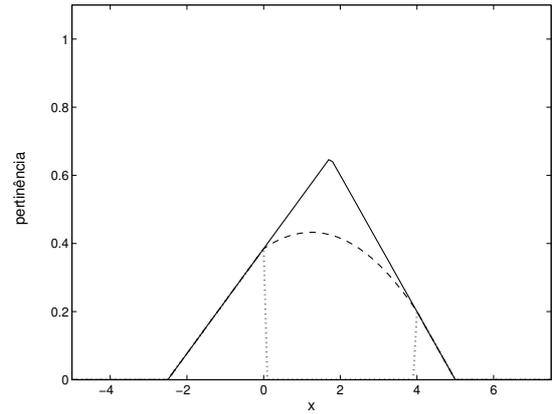


Figura 3.16: Normas-t mais utilizadas na literatura: mínimo “-”, produto algébrico “-” e produto drástico “.”.

2. Associatividade:

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap C = A \cup B \cap C \\ A \cap (B \cup C) &= (A \cap B) \cup C = A \cap B \cup C \end{aligned}$$

3. Idempotência:

$$\begin{aligned} A \cup A &= A \\ A \cap A &= A \end{aligned}$$

4. Distributividade:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

5. Condições de contorno:

$$\begin{aligned} A \cup \emptyset &= A, & A \cup \mathbf{X} &= \mathbf{X} \\ A \cap \emptyset &= \emptyset, & A \cap \mathbf{X} &= \mathbf{A} \end{aligned}$$

6. Involução:

$$\overline{\overline{A}} = A$$

, em que as barras indicam o complemento.

7. Transitividade:

$$\text{se } A \subset B \text{ e } B \subset C \implies A \subset C$$

8. Lei de De Morgan:

$$\begin{aligned} \overline{(A \cap B)} &= \overline{A} \cup \overline{B} \\ \overline{(A \cup B)} &= \overline{A} \cap \overline{B} \end{aligned}$$

, em que as barras indicam o complemento.

As propriedades de conjuntos clássicos (lógica booleana) $A \cap \bar{A} = \emptyset$ e $A \cup \bar{A} = \mathbf{X}$ não se verificam para alguns conjuntos nebulosos:

$$\begin{aligned} \mu_{A \cap \bar{A}}(x) = \mu_A(x) \wedge (1 - \mu_A(x)) \neq 0 &\implies A \cap \bar{A} \neq \emptyset \\ \mu_{A \cup \bar{A}}(x) = \mu_A(x) \vee (1 - \mu_A(x)) \neq 1 &\implies A \cup \bar{A} \neq \mathbf{X} \end{aligned}$$

onde os símbolos \wedge e \vee são operadores min e max, respectivamente. Utilizando o exemplo da Fig.3.12 podemos verificar na Fig.3.17 a união e a intersecção de um conjunto nebuloso A e seu complemento \bar{A} .

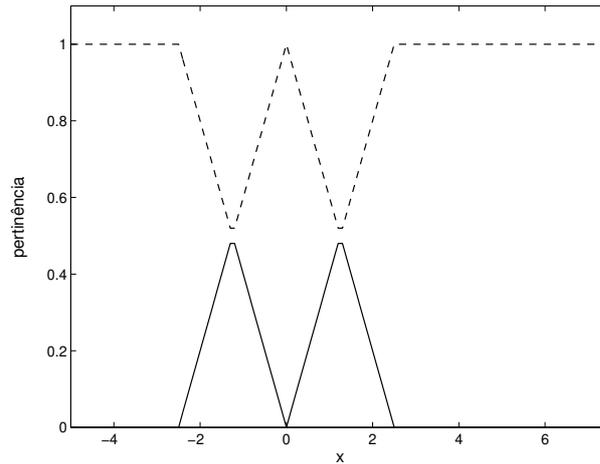


Figura 3.17: União “- -” e intersecção “-” de um conjunto nebuloso A e seu complemento \bar{A} .

3.6 Relações Nebulosas e o Princípio da Extensão

3.6.1 Relações Clássicas e Relações Nebulosas

Sejam \mathbf{U} e \mathbf{V} dois conjuntos clássicos arbitrários. Uma relação R é um subconjunto do produto Cartesiano destes dois conjuntos, definidos como:

$$R : \mathbf{U} \times \mathbf{V} \rightarrow \{0, 1\}.$$

O produto Cartesiano é o conjunto não nebuloso de todos os pares ordenados (u, v) tal que $u \in \mathbf{U}$ e $v \in \mathbf{V}$. Se o valor da relação para u e v é igual a 1:

$$R(u, v) = 1,$$

então estes dois elementos são relacionados. Caso contrário, se $R(u, v) = 0$, dizemos que os dois elementos não são relacionados. A relação pode ser representada como uma função de pertinência da seguinte maneira:

$$\mu_R(u, v) = \begin{cases} 1 & \text{se } (u, v) \in R(\mathbf{U}, \mathbf{V}) \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo 3.1 *Sejam:*

$$\mathbf{U} = \{1, 2, 3\}$$

e

$$\mathbf{V} = \{2, 3, 4\}.$$

então o produto cartesiano de \mathbf{U} e \mathbf{V} é o conjunto:

$$\mathbf{U} \times \mathbf{V} = \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4)\}.$$

Seja $R(\mathbf{U}, \mathbf{V})$ a relação onde os primeiros elementos são menores que os segundos elementos, então:

$$R(\mathbf{U}, \mathbf{V}) = \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 4)\}$$

e também pode ser representada por:

		V		
		2	3	4
U	1	1	1	1
	2	1	1	1
	3	0	0	1

Porém, para certas relações é difícil optar por uma avaliação zero-um. Nestes casos, a relação nebulosa é necessária. Uma relação nebulosa é um conjunto nebuloso definido no produto cartesiano de conjuntos clássicos. A relação nebulosa R em $\mathbf{U} \times \mathbf{V}$ é definida como o conjunto nebuloso:

$$R = \{((u, v), \mu_R(u, v)) \mid (u, v) \in \mathbf{U} \times \mathbf{V}\}$$

onde $\mu_R : \mathbf{U} \times \mathbf{V} \rightarrow [0, 1]$.

Exemplo 3.2 *Sejam:*

$$\mathbf{U} = \{Campinas, Porto Alegre, Manaus\}$$

e

$$\mathbf{V} = \{Campinas, Brasília, Salvador\}.$$

Podemos definir a relação de “muito distante” entre os dois conjuntos de cidades como:

		V		
		Campinas	Brasília	Salvador
U	Campinas	0	0.3	0.6
	Porto Alegre	0.3	0.7	0.95
	Manaus	1	0.8	0.85

3.6.2 Operações em Relações Nebulosas

Considerando R e W como relações em $\mathbf{U} \times \mathbf{V}$ podemos definir as seguintes operações:

- União: $(R \cup W)(u, v) = R(u, v) \text{ s } W(u, v)$.
- Intersecção: $(R \cap W)(u, v) = R(u, v) \text{ t } W(u, v)$.
- Complemento: $\overline{R}(u, v) = 1 - R(u, v)$.

Algumas propriedades que são aplicáveis a conjuntos nebulosos podem ser aplicadas em relações nebulosas:

- Inclusão: $R \subset W \quad R(u, v) \leq W(u, v)$.
- Igualdade: $R = W \quad R(u, v) = W(u, v)$.

para todos u em \mathbf{U} e v em \mathbf{V} .

Se R é representado em forma de matriz (isto é, \mathbf{U} e \mathbf{V} são finitos e discretos), as seguintes operações são evidenciadas:

$$\begin{aligned} (R^T)^T &= R, \\ (\overline{R})^T &= \overline{R^T}. \end{aligned} \tag{3.35}$$

onde R^T é a operação de transposição. Ao aplicar a transposição da matriz podemos escrever a relação R^T como uma nova relação:

$$R^T(x, y) = R(y, x). \tag{3.36}$$

3.6.3 Composições de Relações Nebulosas

Suponha que R , G e W são relações nebulosas definidas conforme o produto cartesiano $\mathbf{U} \times \mathbf{V}$, $\mathbf{U} \times \mathbf{X}$ e $\mathbf{X} \times \mathbf{V}$, respectivamente.

- composição sup-t:

$$R = G \circ W. \tag{3.37}$$

$$R(u, v) = \text{sup}_{x \in X} [G(u, x) \text{ t } W(x, v)].$$

A composição sup-min é um caso particular desta família:

$$R(u, v) = \text{sup}_{x \in X} [G(u, x) \wedge W(x, v)]. \tag{3.38}$$

Neste caso, o operador mais utilizado é a função **max-produto algébrico**.

- composição inf-t:

$$R = G \bullet W. \quad (3.39)$$

$$R(u, v) = \inf_{x \in X} [G(u, x) \wedge W(x, v)].$$

A composição inf-max é um caso particular desta família:

$$R(u, v) = \inf_{x \in X} [G(u, x) \vee W(x, v)]. \quad (3.40)$$

Os operadores de composição, sup-t(max-t) e inf-s(min-s), possuem uma série de propriedades.

- Associativa:

$$R \circ (G \circ W) = (R \circ G) \circ W. \quad (3.41)$$

$$R \bullet (G \bullet W) = (R \bullet G) \bullet W.$$

- Distributiva sobre a união e intersecção:

$$R \circ (G \cup W) = (R \circ G) \cup (R \circ W). \quad (3.42)$$

$$R \bullet (G \cap W) = (R \bullet G) \cap (R \bullet W).$$

- Fracamente distributiva sobre a intersecção e união:

$$R \circ (G \cap W) = (R \circ G) \cap (R \circ W). \quad (3.43)$$

$$R \bullet (G \cup W) = (R \bullet G) \cup (R \bullet W).$$

- Monotonicidade:

$$\text{Se } G \subset W \text{ então } R \circ G \subset R \circ W \text{ E } R \bullet G \supset R \bullet W \quad (3.44)$$

3.6.4 Projeções e Extensões Cilíndricas de Relações

As operações de projeção e extensão são usadas para modificarmos o tamanho da relação nebulosa. Seja R uma relação definida em $\mathbf{U} \times \mathbf{V}$. Então projeção de R em \mathbf{U} é definida como:

$$R_x(x) = Proj_x R(x) = \sup_{y \in \mathbf{V}} R(x, y), \quad (3.45)$$

onde $x \in \mathbf{U}$. Observe que esta operação reduz a dimensão da relação. Geralmente, o operador mais utilizado é a função máximo. Contrariamente às projeções, a extensão cilíndrica eleva a dimensão de uma relação. A extensão cilíndrica em $\mathbf{U} \times \mathbf{V}$ de uma relação R em \mathbf{U} pode ser escrita da seguinte maneira:

$$cil R(x, y) = R(x) \quad \text{onde } \forall y \in \mathbf{V} \quad (3.46)$$

onde $x \in \mathbf{U}$.

3.7 Variáveis Linguísticas

Em nosso dia-a-dia, as palavras são freqüentemente usadas para descrever variáveis. Se uma variável poder ser quantizada por palavras em nossa linguagem natural, ela é chamada de variável lingüística, onde as palavras são representadas por conjuntos nebulosos no universo de discurso na qual a variável é definida. Por exemplo, a variável “temperatura ambiente” pode ser classificada como “fria”, “morna” ou “quente”. A variável “temperatura ambiente” pode ser classificada como uma variável lingüística e os termos {“fria”, “morna”, “quente”} são os conjuntos nebulosos que a caracterizam.

Uma variável lingüística é caracterizada por (X, T, U, M) , onde:

- X é o nome da variável lingüística;
- T é o conjunto de valores lingüísticos que X pode assumir;
- U é o universo dos valores da variável X ;
- M é uma regra semântica que relaciona os valores lingüísticos em T com os conjuntos nebulosos em U .

As variáveis lingüísticas podem ser classificadas por termos lingüísticos compostos, como “não frio”, “mais ou menos quente”, “levemente rápido”, “muito devagar”, etc. Estes termos separadamente podem ser classificados em três grupos:

- Termos primários: “devagar”, “médio”, “rápido”, “frio”, “morno”, “quente”, etc.
- Complementos e conexões: “não”, “e” e “ou”.
- Limitantes: “muito”, “levemente”, “mais ou menos”, etc.

3.8 Regras Nebulosas

Os sistemas nebulosos podem ser representados por regras nebulosas SE-ENTÃO, em analogia com o raciocínio humano onde elas são expressas da seguinte maneira:

$$\text{SE} < \text{proposição nebulosa} >, \text{ENTÃO} < \text{proposição nebulosa} > .$$

Para entendermos as regras nebulosas precisamos saber o que são as proposições nebulosas. Há dois tipos de proposições nebulosas: proposições nebulosas simples e proposições nebulosas compostas. Uma proposição nebulosa simples é uma simples frase:

$$x \text{ é } A$$

onde x é uma variável lingüística, e A é um valor lingüístico de x (função de pertinência). Uma proposição nebulosa composta é uma composição de proposições nebulosas simples utilizando os conectivos “E”, “OU”, e “NÃO” nos quais representam intersecção nebulosa, união nebulosa, e complemento nebuloso, respectivamente.

- Conectivo “E”: O conectivo “E” é utilizado em intersecções nebulosas. Por exemplo, sejam x e y variáveis lingüísticas nos domínios U e V , e A e B os conjuntos nebulosos em U e V , respectivamente, então a proposição nebulosa composta:

$$x \text{ é } A \text{ E } y \text{ é } B$$

é interpretada como a relação nebulosa $A \cap B$ em $U \times V$ com função de pertinência

$$\mu_{A \cap B}(x, y) = t[\mu_A(x), \mu_B(y)]$$

onde $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ é uma norma-t.

- Conectivo “OU”: O conectivo “OU” é utilizado em uniões nebulosas. Por exemplo, sejam x e y variáveis lingüísticas nos domínios U e V , e A e B os conjuntos nebulosos em U e V , respectivamente, então a proposição nebulosa composta:

$$x \text{ é } A \text{ OU } y \text{ é } B$$

é interpretada como a relação nebulosa $A \cup B$ em $U \times V$ com função de pertinência

$$\mu_{A \cup B}(x, y) = s[\mu_A(x), \mu_B(y)]$$

onde $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ é uma norma-s.

- Conectivo “NÃO”: O conectivo “NÃO” é utilizado em complementos nebulosos. Por exemplo, seja x uma variável lingüística no domínios U , e A o conjunto nebuloso em U , então a proposição nebulosa composta:

$$x \text{ é NÃO } A$$

é interpretada como a relação nebulosa \bar{A} em U com função de pertinência

$$\mu_{\bar{A}}(x) = c[\mu_A(x)]$$

onde $c : [0, 1] \rightarrow [0, 1]$ é um complemento.

3.8.1 Interpretações de Regras Nebulosas SE-ENTÃO

As proposições nebulosas podem ser nomeadas de acordo com suas posições dentro da regra nebulosa SE-ENTÃO:

SE < proposição do antecedente >, ENTÃO < proposição do conseqüente >.

Como na lógica clássica ($p \rightarrow q$), a proposição do antecedente será simbolizada por p e a proposição do conseqüente será q . A tabela verdade para $p \rightarrow q$ pode ser vista na Tab.3.1 e é equivalente a:

$$\bar{p} \vee q \tag{3.47}$$

Tabela 3.1: Tabela verdade para $p \rightarrow q$.

p	q	$p \rightarrow q$
V	F	F
F	V	V
F	F	V

e

$$(p \wedge q) \vee \bar{p} \quad (3.48)$$

onde \neg , \vee e \wedge representam os operadores lógicos clássicos “NÃO”, “OU” e “E”, respectivamente.

Analogamente, as regras SE-ENTÃO tem o papel de implicação (“ \rightarrow ”) da lógica clássica. Portanto, supomos que p é uma relação nebulosa definida em $U = U_1 \times \dots \times U_n$, q é uma relação nebulosa definida em $V = V_1 \times \dots \times V_m$, e x e y são variáveis lingüísticas em U e V , respectivamente. Os tipos de implicações mais encontrados na literatura podem ser vistos abaixo:

- **Implicação de Dienes-Rescher:** Se substituirmos os operadores lógicos \neg e \vee na expressão (3.47) pelo complemento nebuloso básico e pela função máximo, respectivamente, então obteremos a chamada implicação de Dienes-Rescher. A regra nebulosa SE-ENTÃO é interpretada como uma relação nebulosa R_D em $U \times V$ com função de pertinência:

$$\mu_{R_D}(x, y) = \max[1 - \mu_p(x), \mu_q(y)]. \quad (3.49)$$

- **Implicação de Lukasiewicz:** Se usarmos a norma-s de Yager com $\omega = 1$ para \vee e complemento nebuloso básico para \neg na expressão (3.47), nós obteremos a implicação de Lukasiewicz. A regra nebulosa SE-ENTÃO é interpretada como uma relação nebulosa R_L em $U \times V$ com função de pertinência:

$$\mu_{R_L}(x, y) = \min[1, 1 - \mu_p(x) + \mu_q(y)]. \quad (3.50)$$

- **Implicação de Zadeh:** Se substituirmos os operadores lógicos \neg , \vee e \wedge na equação (3.48) pelo complemento nebuloso básico, pela função de máximo e mínimo, respectivamente, então obteremos a chamada implicação de Zadeh. A regra nebulosa SE-ENTÃO é interpretada como uma relação nebulosa R_Z em $U \times V$ com função de pertinência:

$$\mu_{R_Z}(x, y) = \max[\min(\mu_p(x), \mu_q(y)), 1 - \mu_p(x)]. \quad (3.51)$$

- **Implicação de Gödel:** A regra nebulosa SE-ENTÃO é interpretada como uma relação nebulosa R_G em $U \times V$ com função de pertinência:

$$\mu_{R_G}(x, y) = \begin{cases} 1 & \text{se } \mu_p(x) \leq \mu_q(y) \\ \mu_q(y) & \text{caso contrário} \end{cases} \quad (3.52)$$

Exemplo 3.3 *Sejam*

$$U = 2, 3, 4$$

e

$$V = 1, 2, 3$$

Suponha que $x \in U$ e $y \in V$. A regra SE-ENTÃO utilizada será:

SE x é **grande**, ENTÃO y é **pequeno**

onde os conjuntos **grande** e **pequeno** são definidos como:

$$\text{grande} = \frac{0.1}{2} + \frac{0.5}{3} + \frac{1}{4}$$

$$\text{pequeno} = \frac{1}{1} + \frac{0.5}{2} + \frac{0.1}{3}$$

Se utilizarmos a implicação de Dienes-Rescher, então a regra SE-ENTÃO é interpretada como a seguinte relação nebulosa R_D em $U \times V$:

$$R_D = \frac{1}{(2,1)} + \frac{0.9}{(2,3)} + \frac{1}{(3,1)} + \frac{0.5}{(3,2)} + \frac{0.5}{(3,3)} + \frac{1}{(4,1)} + \frac{0.5}{(4,2)} + \frac{0.1}{(4,3)}$$

Se utilizarmos a implicação de Lukasiewicz, teremos a seguinte relação R_L :

$$R_L = \frac{1}{(2,1)} + \frac{1}{(2,2)} + \frac{1}{(2,3)} + \frac{1}{(3,1)} + \frac{1}{(3,2)} + \frac{0.6}{(3,3)} + \frac{1}{(4,1)} + \frac{0.5}{(4,2)} + \frac{0.1}{(4,3)}$$

Na implicação de Zadeh, temos:

$$R_Z = \frac{0.9}{(2,1)} + \frac{0.9}{(2,3)} + \frac{0.5}{(3,1)} + \frac{0.5}{(3,2)} + \frac{0.5}{(3,3)} + \frac{1}{(4,1)} + \frac{0.5}{(4,2)} + \frac{0.1}{(4,3)}$$

e na de Gödel, temos:

$$R_G = \frac{1}{(2,1)} + \frac{1}{(2,2)} + \frac{1}{(2,3)} + \frac{1}{(3,1)} + \frac{1}{(3,2)} + \frac{0.1}{(3,3)} + \frac{1}{(4,1)} + \frac{0.5}{(4,2)} + \frac{0.1}{(4,3)}$$

Este exemplo está baseado em Wang, "A Course in Fuzzy Systems and Control", [42].

3.9 Base de Regras Nebulosas

A base de regras nebulosas consiste de um conjunto de regras nebulosas SE-ENTÃO tais como:

$$Ru^l : \text{SE } x_1 \text{ é } A_1^l \text{ E } \dots \text{ E } x_n \text{ é } A_n^l, \text{ ENTÃO } y \text{ é } B^l \quad (3.53)$$

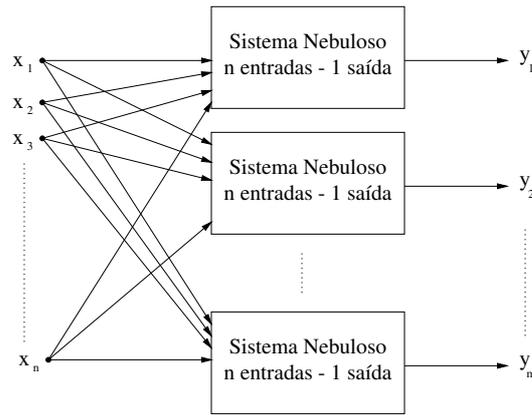


Figura 3.18: Sistema nebuloso com n entradas e m saídas

onde A_i^l e B^l são conjuntos nebulosos em $U_i \subset R$ e $V \subset R$, respectivamente, e $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in U$ e $y \in V$ são variáveis de entrada e saída do sistema nebuloso. Seja M o número de expressões SE-ENTÃO na base de regras nebulosas, tal que $l = 1, 2, \dots, M$. Perceba que para o caso de haver mais de uma saída em nosso sistema nebuloso, estas podem ser desacopladas como aparece na Fig.3.18.

Existem alguns casos especiais de construções das expressões SE-ENTÃO:

1. Regras parciais:

$$\text{SE } x_1 \text{ é } A_1^l \text{ E } \dots \text{ E } x_m \text{ é } A_m^l, \text{ ENTÃO } y \text{ é } B^l \quad (3.54)$$

onde $m < n$ e n é o tamanho do vetor \mathbf{x} . É equivalente a escrever:

$$\text{SE } x_1 \text{ é } A_1^l \text{ E } \dots \text{ E } x_m \text{ é } A_m^l \text{ E } x_{m+1} \text{ é } I \text{ E } \dots \text{ E } x_n \text{ é } I, \text{ ENTÃO } y \text{ é } B^l \quad (3.55)$$

onde I é um conjunto nebuloso em R com $\mu_I(x) \equiv 1$ para todo $x \in R$.

2. Regras ou:

Uma regra ou pode ser representada da seguinte maneira:

$$\text{SE } x_1 \text{ é } A_1^l \text{ E } \dots \text{ E } x_m \text{ é } A_m^l \text{ OU } x_{m+1} \text{ é } I \text{ E } \dots \text{ E } x_n \text{ é } I, \text{ ENTÃO } y \text{ é } B^l. \quad (3.56)$$

Esta regra equivale a escrever as duas regras seguintes:

$$\begin{aligned} \text{SE } x_1 \text{ é } A_1^l \text{ E } \dots \text{ E } x_m \text{ é } A_m^l, \text{ ENTÃO } y \text{ é } B^l \\ \text{SE } x_{m+1} \text{ é } A_{m+1}^l \text{ E } \dots \text{ E } x_n \text{ é } A_n^l, \text{ ENTÃO } y \text{ é } B^l \end{aligned} \quad (3.57)$$

3. Expressão nebulosa sozinha:

$$y \text{ é } B^l. \quad (3.58)$$

É equivalente a:

$$\text{SE } x_1 \text{ é } I \text{ E } \dots \text{ E } x_n \text{ é } I, \text{ ENTÃO } y \text{ é } B^l. \quad (3.59)$$

4. **Regras gradativas:** As regras gradativas são, por exemplo:

$$\text{Quanto menor é } x, \text{ maior é } y. \quad (3.60)$$

Seja o conjunto nebuloso S representando “menor”, por exemplo, $\mu_S(x) = (1 + e^{5(x+2)})$ e B é um conjunto nebuloso “maior”, por exemplo, $\mu_B(x) = 1/(1 + e^{(-5(y-2))})$, então a regra gradativa é equivalente a:

$$\text{SE } x \text{ é } S, \text{ ENTÃO } y \text{ é } B. \quad (3.61)$$

5. **Regras não-nebulosas:** As regras não nebulosas representam a lógica booleana clássica onde as funções de pertinências A_i^l e B^l podem obter somente valores 0 ou 1 (verdadeiro ou falso).

Algumas propriedades no projeto da base de regras devem ser consideradas: quantas regras são necessárias para cobrir todas as possibilidades de ocorrências; e se ocorre algum conflito entre regras.

Um conjunto de regras nebulosas SE-ENTÃO é dita completa se para qualquer $x \in U$, existe ao menos uma regra na base de regra nebulosa, Rw^l , tal que:

$$\mu_{A_i^l} \neq 0$$

para $i = 1, 2, \dots, n$.

Intuitivamente, uma base de regras está completa se para um ponto no espaço de entrada existe ao menos uma regra ativa, tal que, o grau de pertinência da parte SE da regra seja não nulo neste ponto.

Exemplo 3.4 Considere um sistema nebuloso 2 entradas - 1 saída com $U = U_1 \times U_2 = [0, 1] \times [0, 1]$ e $V = [0, 1]$. Defina-se três conjuntos nebulosos P_1, M_1 e G_1 em U_1 , e dois conjuntos nebulosos P_2 e G_2 em U_2 . Deste modo, para uma base de regras nebulosas ser completa, deve conter as seguintes 6 regras nas quais as partes SE constituem todas as possíveis combinações de P_1, M_1 e G_1 com P_2 e G_2 :

$$\begin{aligned} &SE \ x_1 \text{ é } P_1 \ E \ x_2 \text{ é } P_2, \text{ ENTÃO } y \text{ é } B^1 \\ &SE \ x_1 \text{ é } P_1 \ E \ x_2 \text{ é } G_2, \text{ ENTÃO } y \text{ é } B^2 \\ &SE \ x_1 \text{ é } M_1 \ E \ x_2 \text{ é } P_2, \text{ ENTÃO } y \text{ é } B^3 \\ &SE \ x_1 \text{ é } M_1 \ E \ x_2 \text{ é } G_2, \text{ ENTÃO } y \text{ é } B^4 \\ &SE \ x_1 \text{ é } G_1 \ E \ x_2 \text{ é } P_2, \text{ ENTÃO } y \text{ é } B^5 \\ &SE \ x_1 \text{ é } G_1 \ E \ x_2 \text{ é } G_2, \text{ ENTÃO } y \text{ é } B^6 \end{aligned} \quad (3.62)$$

onde B^l ($l = 1, 2, \dots, 6$) são conjuntos nebulosos em V .

Um conjunto nebuloso de regras SE-ENTÃO é consistente se não há regras com as mesmas partes SE mas diferentes partes ENTÃO. O conectivo E é dito contínuo se não existem regras vizinhas nas quais os conjuntos nebulosos das partes ENTÃO têm intersecção vazia.

3.10 Máquina de Inferência Nebulosa

Na máquina de inferência nebulosa, os princípios lógicos nebulosos são usados para combinar as regras nebulosas SE-ENTÃO em um mapeamento de um conjunto A' em U para um conjunto nebuloso B' em V .

Para uma dada regra, as relações nebulosas A' , nas quais representam a premissa \mathbf{x} é A' , e a relação $A \rightarrow B$ em $U \times V$, que representa a regra SE \mathbf{x} é A ENTÃO y é B , um conjunto B' em V , que representa o conseqüente y é B' , pode ser definido como:

$$\mu_{B'}(y) = \sup_{\mathbf{x} \in U^t} [\mu_{A'}(\mathbf{x}), \mu_{A \rightarrow B}(\mathbf{x}, y)] \quad (3.63)$$

Para o problema de M regras nebulosas SE-ENTÃO, para uma mesma base de regras, a solução é a composição destas. Há dois argumentos opostos para interpretar um conjunto de regras:

1º : As regras são expressões condicionais independentes (união).

2º : As regras são expressões condicionais fortemente acopladas (intersecção).

Seja Ru^l uma relação nebulosa em $U \times V$, a qual representa a regra nebulosa SE-ENTÃO, tal que $Ru^l = A_1^l \times \dots \times A_n^l \rightarrow B^l$. Sabemos que $A_1^l \times \dots \times A_n^l$ é uma relação nebulosa em $U = U_1 \times \dots \times U_n$ definida por:

$$\mu_{A_1^l}^l \times \dots \times \mu_{A_n^l}^l(x_1, \dots, x_n) = \mu_{A_1^l}^l(x_1) \star \dots \star \mu_{A_n^l}^l(x_n) \quad (3.64)$$

onde \star é um operador norma-t.

A implicação \rightarrow em Ru^l é definida conforme as várias implicações existentes: Dienes-Rescher, Lukasiewicz, Zadeh e Gödel (Pedrycz e Gomide [31], Wang [42]). Se aceitarmos a primeira condição, então M regras são interpretadas como uma relação única R_M em $U \times V$ definida por:

$$R_M = \bigcup_{l=1}^M Ru^l \quad (3.65)$$

Esta combinação é chamada de combinação Mamdani. Se nós usarmos o símbolo $+$ para representar as normas-s, então a combinação pode ser escrita como:

$$\mu_{R_M}(x, y) = \mu_{Ru^1}(x, y) + \dots + \mu_{Ru^M}(x, y) \quad (3.66)$$

3.11 Fuzzificador e Defuzzificador

3.11.1 Fuzzificador

Os fuzzificadores são definidos como um mapeamento de um valor real $x^* \in U \subset \mathfrak{R}^n$ para um conjunto nebuloso A' em U . Os três tipos de fuzzificadores mais utilizados são [42]:

- **Fuzzificador singleton:**

mapeia um valor real $x^* \in U$ em um conjunto singleton A' em U , no qual tem valor de pertinência 1 no ponto x^* e 0 em todos os outros pontos em U , tal que:

$$\mu_{A'}(x) = \begin{cases} 1 & \text{se } x = x^* \\ 0 & \text{caso contrário} \end{cases} \quad (3.67)$$

- **Fuzzificador gaussiano:**

O fuzzificador mapeia $x^* \in U$ em um conjunto nebuloso A' em U , no qual tem a seguinte função de pertinência Gaussiana:

$$\mu_{A'}(x) = e^{-\left(\frac{x_1 - x_1^*}{a_1}\right)^2} \star \dots \star e^{-\left(\frac{x_n - x_n^*}{a_n}\right)^2} \quad (3.68)$$

onde a_i são parâmetros positivos e a norma-t \star é geralmente escolhida como produto algébrico ou mínimo.

- **Fuzzificador triangular:**

mapeia $x^* \in U$ em um conjunto nebuloso A' em U , no qual tem a seguinte função de pertinência triangular:

$$\mu_{A'}(x) = \begin{cases} \left(1 - \frac{|x_1 - x_1^*|}{b_1}\right) \star \dots \star \left(1 - \frac{|x_n - x_n^*|}{b_n}\right) & \text{se } |x_i - x_i^*| \leq b_i, \quad i = 1, 2, \dots, n \\ 0 & \text{caso contrário} \end{cases} \quad (3.69)$$

3.11.2 Defuzzificador

O defuzzificador é definido como um mapeamento de um conjunto nebuloso B' em $V \subset \mathfrak{R}^1$, o qual é a saída da máquina de inferência nebulosa para um ponto preciso $y^* \in V$. Conceitualmente, a tarefa do defuzzificador é especificar um ponto em V que melhor represente o conjunto nebuloso B' .

- **Defuzzificador de centro de gravidade:**

Obtém y^* como o centro de área da função de pertinência de B' , que é:

$$y^* = \frac{\int_v y \mu_{B'}(y) dy}{\int_v \mu_{B'}(y) dy} \quad (3.70)$$

- **Defuzzificador de Média dos Centros:**

Uma boa aproximação do defuzzificador de centro de gravidade é a média ponderada dos centros dos M conjuntos nebulosos, com os pesos iguais as alturas dos conjuntos nebulosos correspondentes. Especialmente, sejam \bar{y}^l o centro do l -ésimo conjunto nebuloso e w_l sua altura, o defuzzificador média dos centros determina y^* como:

$$y^* = \frac{\sum_{l=1}^M \bar{y}^l w_l}{\sum_{l=1}^M w_l} \quad (3.71)$$

Este tipo de defuzzificador é o mais utilizado pois é implementado facilmente.

- **Defuzzificador Máximo:**

Conceitualmente, o defuzzificador máximo escolhe o y^* como o ponto em V na qual $\mu_{B'}(y)$ atinge seu valor máximo. O conjunto é definido como:

$$hgt(B') = \{y \in V | \mu_{B'}(y) = \sup_{y \in V} \mu_{B'}(y)\} \quad (3.72)$$

onde $hgt(B')$ é o conjunto de todos pontos em V tal que $\mu_{B'}(y)$ atinge seu máximo valor.

3.12 Exemplo - Controlador Adaptativo PI Nebuloso:

Este exemplo apresenta uma aplicação de sistemas nebulosos lingüísticos do tipo Mamdani onde o conhecimento do especialista é utilizado no projeto de um controlador adaptativo PI nebuloso.

O controlador PID convencional é a estratégia de controle mais largamente utilizada em processos industriais por causa da extraordinária efetividade, simplicidade de implementação e vasta aplicabilidade. Embora controladores PID sejam comuns e bem conhecidos, eles são freqüentemente ajustados precariamente. Evidências disto podem ser encontradas em quase todos os processos industriais pois o ajuste é realizado pela experiência do operador [41].

Outro problema dos controladores convencionais é a falta de adaptabilidade a mudanças na planta do sistema. Uma solução para este problema é utilizar um sistema de controle adaptativo onde as características dinâmicas do processo são comparadas com as características desejadas e, de acordo com um índice de desempenho, modifica os parâmetros do controlador para gerar um sinal atuante de tal forma que o desempenho ótimo possa ser mantido independentemente das mudanças na planta. As principais vantagens de um controle adaptativo sobre um não adaptativo são: melhor desempenho pois o controle adaptativo pode se auto-ajustar às mudanças da planta; e menor quantidade de informação sobre a planta é requerida porque a lei de adaptação pode ajustar, ensinar a dinâmica da planta durante a operação em tempo real.

Uma alternativa para estes problemas são os controladores adaptativos nebulosos. Esta seção apresenta um esquema de controle nebuloso adaptativo, conforme a Fig.3.19.

Para as implementações do controlador e da lei adaptativa foram utilizados o erro (e) e a variação do erro ($\delta e = e_k - e_{k-1}$) como variáveis de entrada. A ação do controle (δu) e os ganhos do controlador (K_P e K_I) são as variáveis de saída do controlador e da lei adaptativa, respectivamente. As bases de regras do controlador e da lei adaptativa foram obtidas baseando-se na experiência e conhecimento de especialistas. O método para especialização da base de regras a ser proposto na próxima seção não requer o modelo do processo a ser controlado e é baseado no comportamento dinâmico da resposta ao degrau de um processo linear subamortecido onde o erro é calculado pela diferença entre a referência e a resposta do processo.

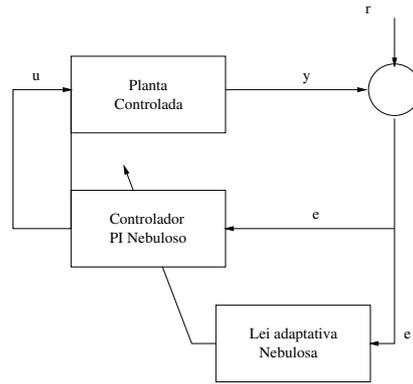


Figura 3.19: Esquema em bloco do controle adaptativo

3.12.1 Modelo de Controle PI Adaptativo Nebuloso

A saída de um controlador PI, no domínio da frequência s , é dado por:

$$u_{PI}(s) = \left(K_P + \frac{K_I}{s} \right) E(s) \quad (3.73)$$

onde K_P e K_I são os ganhos proporcional e integral, respectivamente, e $E(s)$ é a transformada de Laplace do sinal de erro.

Aplicando a transformação bilinear:

$$s = \frac{2(z-1)}{T(z+1)} \quad (3.74)$$

onde $T > 0$ é o tempo de amostragem, resulta a seguinte equação para a saída do controlador PI:

$$u_{PI}(z) = \left(K_P - \frac{K_I T}{2} + \frac{K_I T}{1-z^{-1}} \right) E(z) \quad (3.75)$$

onde $E(z)$ é a transformada- z do sinal de erro.

Considerando $K_p^d = K_P - \frac{K_I T}{2}$ e $K_I^d = K_I T$, e usando a transformada- z inversa, teremos:

$$u_{PI}(k) - u_{PI}(k-1) = K_p^d [e(k) - e(k-1)] + K_I^d e(k) \quad (3.76)$$

Dividindo (3.76) por T , tem-se:

$$\frac{u_{PI}(k) - u_{PI}(k-1)}{T} = K_p^d \left[\frac{e(k) - e(k-1)}{T} \right] + K_I^d e(k) \quad (3.77)$$

e

$$u_{PI}(k) = u_{PI}(k-1) + T \left\{ K_p^d \left[\frac{e(k) - e(k-1)}{T} \right] + K_I^d e(k) \right\} \quad (3.78)$$

Assim, renomeando o termo T e multiplicando a equação (3.78), a lei de controle será:

$$u_{PI}(k) = u_{PI}(k - 1) + K_U^d \left\{ K_P^d \left[\frac{e(k) - e(k - 1)}{T} \right] + K_I^d e(k) \right\} \quad (3.79)$$

onde K_P^d , K_I^d e K_U^d são os ganhos que serão determinados pela lei adaptativa. Por convenção tem-se:

$$e_V^d(k) = \frac{e(k) - e(k - 1)}{T} \quad (3.80)$$

$$e_D^d(k) = e(k) \quad (3.81)$$

$$\Delta_U^d(k) = K_P^d e_V^d(k) + K_I^d e_D^d(k) \quad (3.82)$$

e a lei do controlador PI é dada por:

$$u_{PI}(k) = u_{PI}(k - 1) + K_U^d \Delta_U^d(k). \quad (3.83)$$

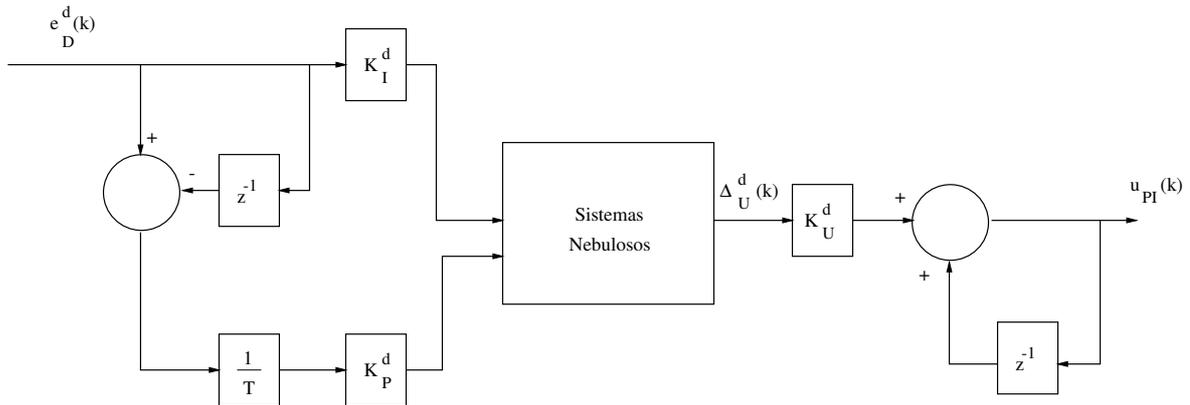


Figura 3.20: Diagrama de blocos do controlador

Fuzzificação

As funções de pertinência da entrada e da saída do controlador PI são mostradas na Fig.3.21.

O controlador PI nebuloso tem duas entradas, o sinal de erro $e_D = ref - y$ e a taxa de variação de erro $e_V = [e(k) - e(k - 1)]/T$, e tem uma saída du . Essas variáveis lingüísticas podem ser classificadas como negativo grande, negativo médio, negativo pequeno, zero, positivo pequeno, positivo médio e positivo grande (NG,NM,NP,Z,PP,PM e PG). Os ganhos K_P^d e K_I^d , utilizados na obtenção da saída do controlador, serão determinados na lei adaptativa. O ganho K_U^d será mantido em um valor unitário.

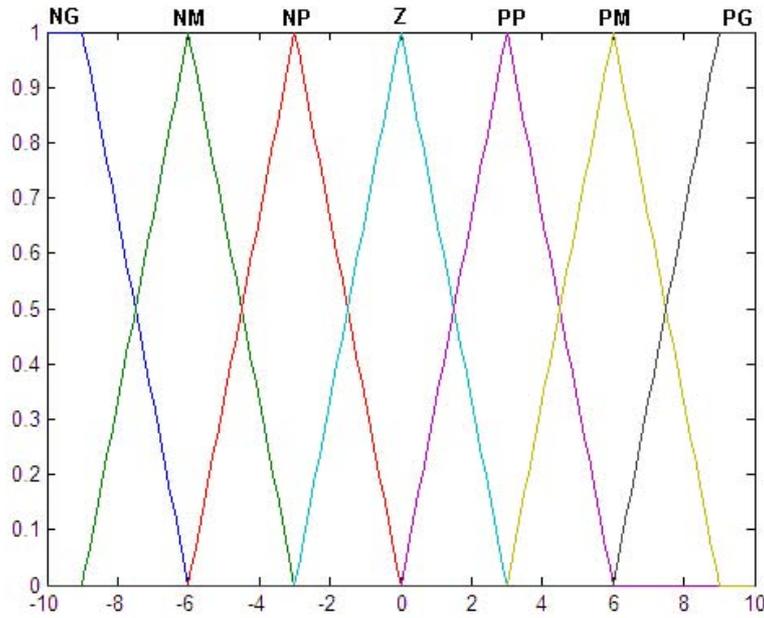


Figura 3.21: Partições nebulosas do espaço de discurso de entrada

Regras de Controle Nebuloso

O objetivo é melhorar critérios como tempo de subida, sobressinal, tempo de acomodação e outros do sinal de saída do processo portanto o método para especificação da base de regras é baseado no comportamento dinâmico da resposta ao degrau de um processo linear subamortecido, como nas Fig.3.22 e Fig.3.23.

A resposta deste processo foi subdividida em áreas, onde a cada uma delas, são atribuídas as regras de controle definidas por um especialista.

As informações fornecidas pelas áreas marcadas justamente com os pontos destacados podem ser caracterizadas como:

$$\left\{ \begin{array}{l} A : e > 0 \quad e \quad \delta e < 0 \\ B : e < 0 \quad e \quad \delta e < 0 \\ C : e < 0 \quad e \quad \delta e > 0 \\ D : e > 0 \quad e \quad \delta e > 0 \\ E : \text{idem } A \end{array} \right. \quad (3.84)$$

$$\left\{ \begin{array}{l} 0 : e > 0 \quad e \quad \delta e < 0 \\ 1 : e = 0 \quad e \quad \delta e < 0 \\ 2 : e < 0 \quad e \quad \delta e = 0 \\ 3 : e = 0 \quad e \quad \delta e > 0 \\ 4 : e > 0 \quad e \quad \delta e = 0 \\ 5 : \text{idem } 1 \end{array} \right. \quad (3.85)$$

em que e é o erro calculado pela diferença entre a referência e a resposta do processo, e de é a

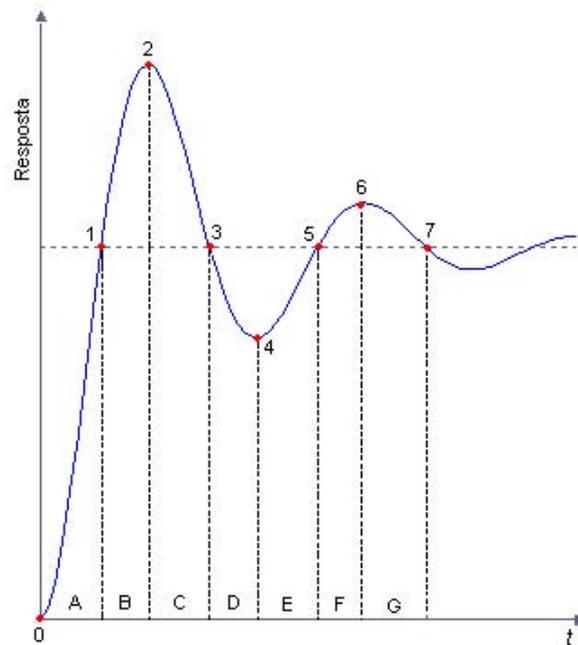


Figura 3.22: Comportamento da resposta no movimento subamortecido

variação do erro calculado por $(e(k) - e(k - 1))/T$ e os termos k e $k - 1$ denotam os instantes atual e anterior respectivamente. A experiência básica de um especialista relata as seguintes informações:

- O incremento na ação de controle du dependerá do sinal e da magnitude de e e de .
 1. Pontos 1 e 3: o sinal de du deverá ser o mesmo de de .
 2. Pontos 2 e 4: o sinal de du deverá ser o mesmo de e .
 3. Área A: As regras correspondentes a esta área tem a função de diminuir o tempo de subida. O sinal de du deverá ser *zero, positivo pequeno, positivo médio ou positivo grande* dependendo da magnitude de e e de .
 4. Área B: As regras correspondentes a esta área têm a função de reduzir o sobressinal. O sinal de du deverá ser *negativo pequeno, negativo médio ou negativo grande* dependendo da magnitude de e e de .
 5. Área C: As regras correspondentes a esta área têm a função de reduzir o *sinal do erro*. O sinal de du deverá ser *zero, positivo pequeno, positivo médio ou positivo grande* dependendo da magnitude de e e de .
 6. Área D: As regras correspondentes a esta área têm a função de reduzir o subsinal. O sinal de du deverá ser *positivo pequeno, positivo médio ou positivo grande* dependendo da magnitude e e de .

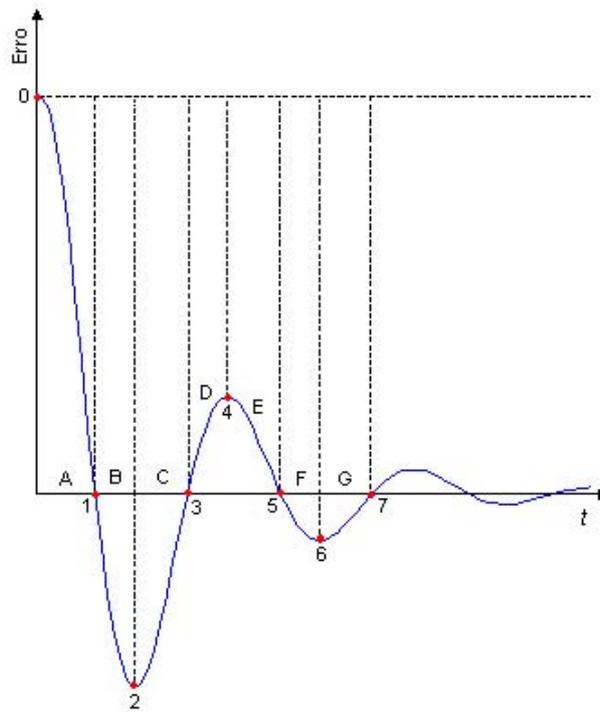


Figura 3.23: Comportamento do erro no movimento subamortecido

- Se e estiver sendo corrigido a uma taxa satisfatória, então mantenha a ação de controle.
- Se e e de são **zero** então mantenha a ação de controle em **zero**.

A base com 49 regras para o cálculo de du pode ser vista na Tabela 3.2.

Tabela 3.2: Base de Regras de u

		e						
		NG	NM	NP	Z	PP	PM	PG
$\frac{de}{dt}$	PG	Z	PP	PP	PM	PM	PG	PG
	PM	NP	Z	PP	PP	PM	PM	PG
	PP	NP	NP	Z	PP	PP	PM	PM
	Z	NM	NP	NP	Z	PP	PP	PM
	NP	NM	NM	NP	NP	Z	PP	PP
	NM	NG	NM	NM	NP	NP	Z	PP
	NG	NG	NG	NM	NM	NP	NP	Z

Defuzzificação

O último passo para o projeto da base de regras é a defuzzificação. O método escolhido é o centro de área:

$$\Delta_U^d(k) = \frac{\sum_{i=1}^n A_i(x) a_i}{\sum_{i=1}^n A_i(x)} \quad (3.86)$$

3.12.2 Lei Adaptativa Nebulosa

Os ganhos K_P^d e K_I^d (chamados K_P e K_I , respectivamente) do controlador nebuloso, como mostrado na equação (3.77), serão obtidos apoiados em análises experimentais e no conhecimento de especialistas. Essas considerações são úteis pois em plantas de segunda ordem podemos considerar, separadamente, certas influências dos ganhos K_P e K_I sobre o *tempo de subida*, *sobressinal* e *tempo de acomodação*. Os ganhos proporcional e integral obtidos na lei adaptativa estão normalizados, ou seja, estão em um intervalo entre 0 e 1. Adotando-se uma faixa de domínio para cada um dos ganhos, podemos obter os ganhos K_P e K_I . Os ganhos normalizados para K_P e K_I serão chamados de K'_P e K'_I . Assim, tem-se as seguintes fórmulas para K_P e K_I :

$$\begin{aligned} K_P^d &= K_P = K'_P(K_{Pmax} - K_{Pmin}) + K_{Pmin} \\ K_I^d &= K_I = K'_I(K_{Imax} - K_{Imin}) + K_{Imin} \end{aligned} \quad (3.87)$$

O diagrama em blocos da lei adaptativa pode ser visto na Fig.3.24.

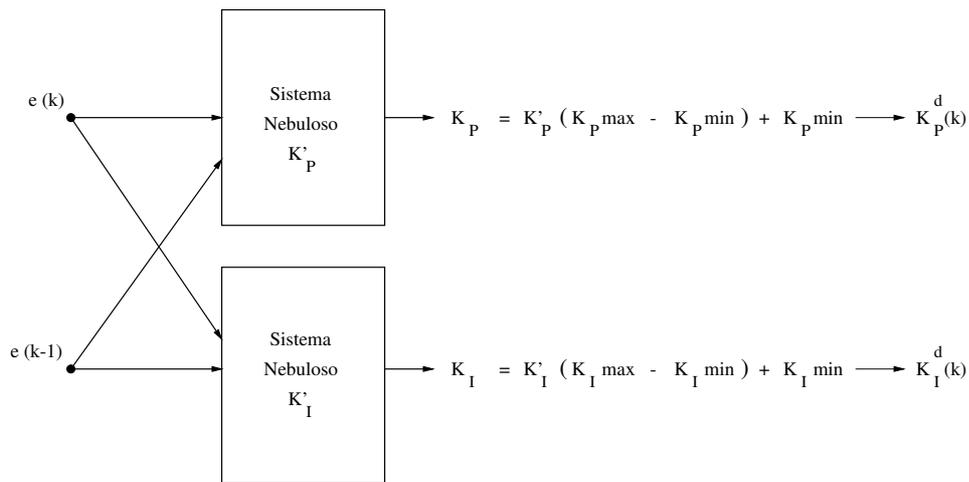


Figura 3.24: Diagrama em blocos da lei adaptativa

Fuzzificação

As funções de pertinência de entrada (erro e taxa de erro) e saída (ganho proporcional e integral) da lei adaptativa são mostrados nas Figs.3 e 7, respectivamente.

Os valores lingüísticos para o erro e a taxa de erro são as mesmas apresentadas no projeto do controlador. Os ganhos proporcional e integral podem receber valores *pequenos* (**P**) ou *grandes* (**G**).

Base de Regras

As bases de regras da lei adaptativa foram determinadas experimentalmente, verificando-se o efeito na resposta do sistema com base na variacao dos ganhos separadamente. Este método é facilitado pois as variáveis lingüísticas K'_P e K'_I podem ter apenas dois valores, *pequeno* ou *grande*. As bases de regras para K'_P e K'_I utilizadas podem ser vistas nas Tabs. 3.3 e 3.4.

Tabela 3.3: Base de Regras de K'_P

		e						
		NG	NM	NP	Z	PP	PM	PG
$\frac{de}{dt}$	PG	G	P	P	P	P	P	G
	PM	G	G	P	P	P	G	G
	PP	G	G	G	P	G	G	G
	Z	G	G	G	G	G	G	G
	NP	G	G	G	P	G	G	G
	NM	G	G	P	P	P	G	G
	NG	G	P	P	P	P	P	G

Tabela 3.4: Base de Regras de K'_I

		e						
		NG	NM	NP	Z	PP	PM	PG
$\frac{de}{dt}$	PG	G	P	P	P	P	P	G
	PM	G	P	P	P	P	P	G
	PP	G	G	P	P	P	G	G
	Z	G	G	G	P	G	G	G
	NP	G	G	P	P	P	G	G
	NM	G	P	P	P	P	P	G
	NG	G	P	P	P	P	P	G

Os valores iniciais para as bases de regras não foram escolhidos aleatoriamente. Podemos generalizar os efeitos do ganho proporcional e integral sobre a resposta do sistema de segunda ordem como mostrado na Tab.3.5. A tabela foi construída para quando se aumenta o ganho proporcional ou integral separadamente.

Tabela 3.5: Características de K'_P K'_I

Ganho	Tempo de Subida	Sobressinal	Tempo de Acomodação
K_P	decrece	crece	não muda
K_I	decrece	crece	crece

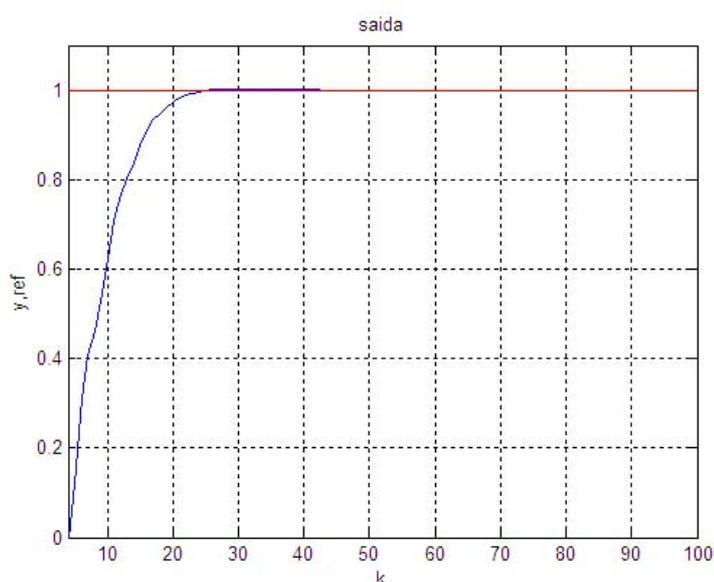


Figura 3.25: Resposta a referência constante, degrau unitário.

Defuzzificação

O método para a defuzzificação na lei adaptativa é o mesmo do controlador, centro de área. A fórmula de defuzzificação pode ser vista na equação (3.86).

3.12.3 Simulações

Para mostrar o desempenho do sistema de controle adaptativo, aqui desenvolvido, será considerado o problema de controle de velocidade de um motor DC de terceira ordem com uma carga inercial, cujo modelo apresenta a seguinte função de transferência no tempo discreto:

$$H = \frac{0.848z^2 - 0.629z + 0.313}{z^3 - 0.986z^2 - 0.888z - 0.368}. \quad (3.88)$$

Na Fig.3.25, observa-se a resposta da planta para uma referência constante, degrau unitário.

Na próxima figura, Fig.3.26, mostra-se a resposta do sistema à uma variação da referência e à uma incerteza paramétrica.

Na Fig.3.27 e Fig.3.28 observa-se o comportamento do ajuste de K_P e K_I .

Neste exemplo, foi proposto e viabilizado um controle adaptativo nebuloso onde a lei adaptativa foi desenvolvida de forma puramente nebulosa. Foi mostrado que é possível desenvolver um sistema nebuloso através de conhecimentos básicos sobre o controlador, sobre a saída do processo, com o apoio de análises experimentais e/ou simulações.

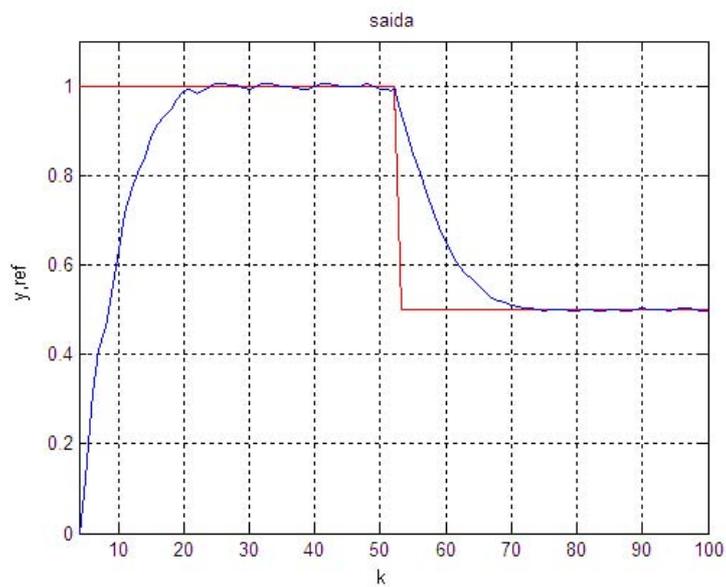


Figura 3.26: Resposta à uma variação na referência.

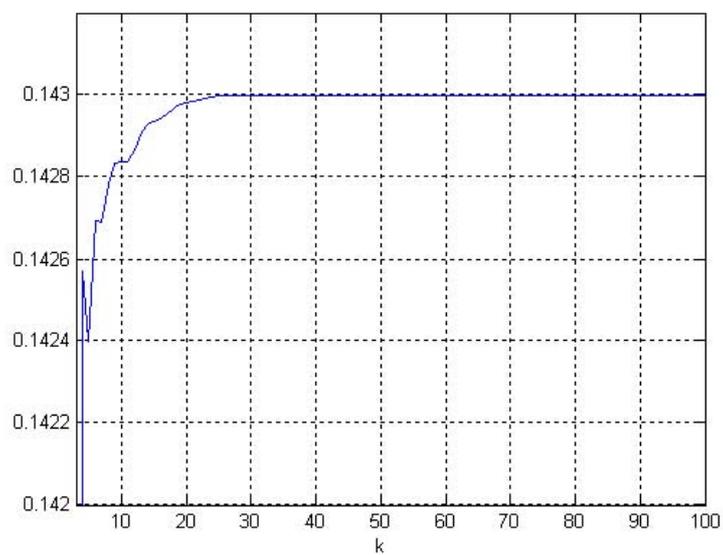


Figura 3.27: Adaptação do ganho proporcional.

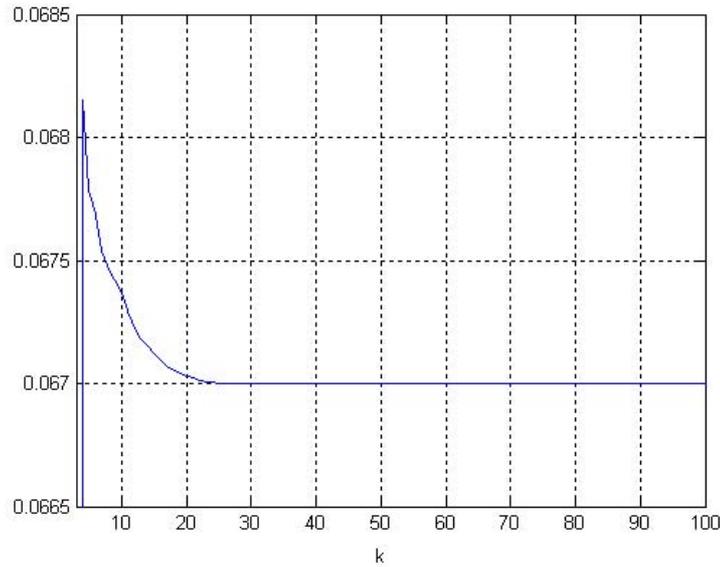


Figura 3.28: Adaptação do ganho integral.

3.13 Modelos Nebulosos Takagi-Sugeno

A idéia de modelos nebulosos locais permite a construção de um modelo como uma coleção de submodelos nos quais as relevâncias são restritas a regiões no espaço de entrada das variáveis:

$$\begin{array}{ll}
 \text{SE } \mathbf{x} \text{ é } F^1, & \text{ENTÃO } y^1 = f^1(\mathbf{x}). \\
 \text{SE } \mathbf{x} \text{ é } F^2, & \text{ENTÃO } y^2 = f^2(\mathbf{x}). \\
 \vdots & \vdots \\
 \text{SE } \mathbf{x} \text{ é } F^N, & \text{ENTÃO } y^N = f^N(\mathbf{x}).
 \end{array} \tag{3.89}$$

onde F^1, F^2 e F^N são relações nebulosas (supondo que \mathbf{x} é um vetor) definidas no espaço de entrada e N é o número de regras e $f^1(\mathbf{x}), f^2(\mathbf{x})$ e $f^N(\mathbf{x})$ são funções lineares ou não-lineares das variáveis numéricas de entrada. Se considerarmos $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_q]^T$, podemos escrever as regras nebulosas da seguinte maneira:

$$\begin{array}{l}
 R^i : \quad \text{SE } x_1 \text{ é } F_1^i, \dots, x_q \text{ é } F_q^i, \quad \text{ENTÃO } y^i = f^i(x_1, \dots, x_q). \\
 \text{onde} \quad \quad \quad i = 1, \dots, N
 \end{array} \tag{3.90}$$

Os q conjuntos nebulosos (F_1^i, \dots, F_q^i) do antecedente da regra R^i formam uma região $F_1^i \times \dots \times F_q^i$ no espaço de entrada de x_1, x_2, \dots, x_q , respectivamente. Esta região é chamada de superfície nebulosa entrada-saída e sua forma está diretamente relacionada com as formas e distribuições dos F^i s.

De uma maneira mais simples, podemos dizer que este modelo decompõe o espaço de entrada em regiões nebulosas e aproxima a saída do sistema em cada região por modelos locais.

Exemplo 3.5 *Supondo $q = 2$ teremos $x = [x_1 \ x_2]^T$ e se dividirmos o espaço de entrada de x_1 e x_2 em 2 conjuntos nebulosos cada.*

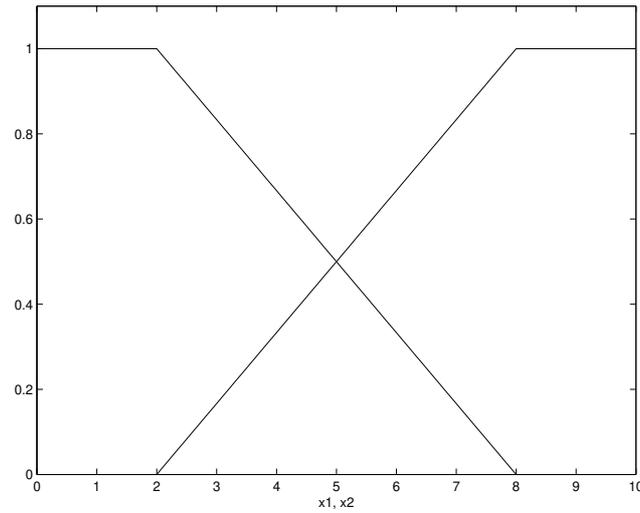


Figura 3.29: Conjuntos nebulosos para x_1 e x_2 .

A superfície nebulosa entrada-saída deste modelo pode ser visto na Fig.3.30.

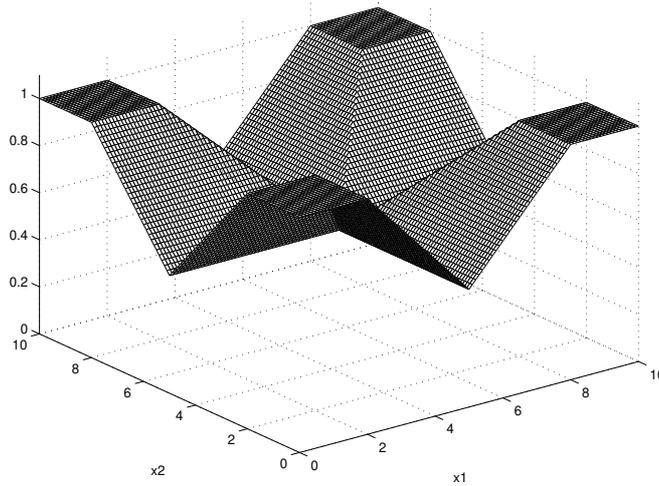


Figura 3.30: Superfície nebulosa entrada-saída deste modelo.

A variável x_j pertence ao conjunto nebuloso k com pertinência $\mu_j^i : R \rightarrow [0, 1]$ para $k = 1, 2, \dots, n_p$ onde n_p é o número de partições nebulosas para a j -ésima variável. Portanto, o grau de ativação (relevância) w^i para a regra i é calculado utilizando uma norma-t representada por $\otimes : [0, 1] \times [0, 1] \rightarrow [0, 1]$, onde:

$$w^i(\mathbf{x}) = \mu_1^i(x_1) \otimes \mu_2^i(x_2) \otimes \dots \otimes \mu_q^i(x_q) \quad (3.91)$$

O grau de ativação normalizado para a regra i é dado por:

$$\gamma^i(\mathbf{x}) = \frac{w^i(\mathbf{x})}{\sum_{r=1}^N w^r(\mathbf{x})} \quad (3.92)$$

Esta normalização implica que $\sum_{r=1}^N \gamma^r(\mathbf{x}) = 1$.

Portanto, a resposta do modelo é uma soma ponderada das funções do conseqüente y^i :

$$y = \sum_{r=1}^N \gamma^r(\mathbf{x}) f^r(\mathbf{x}) = \sum_{r=1}^N \gamma^r(\mathbf{x}) y^r(\mathbf{x}) \quad (3.93)$$

Há uma grande variedade de modelos locais:

1. modelos com funções lineares: as funções locais são lineares do tipo:

$$f^i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x}, \quad (3.94)$$

e \mathbf{a}_i é um vetor de parâmetros;

2. modelos com funções não-lineares: Os modelos polinomiais são os mais utilizados e são expressos por:

$$f^i(\mathbf{x}) = \sum_{j=1}^N a_{ij} x_j + \sum_{j=1}^N b_{ij} x_j^2 + \sum_{j=1}^N c_{ij} x_j^3 \dots \quad (3.95)$$

Também podemos utilizar os modelos harmônicos:

$$f^i(\mathbf{x}) = a_0 \sin(\omega_i^T \mathbf{x}) + b_0 \cos(\omega_i^T \mathbf{x}) \quad (3.96)$$

ou os modelos dinâmicos locais. Neste caso, as regras apresentam equações diferenciais ou a diferenças no conseqüente:

$$\text{SE } \mathbf{x} \text{ é } F^i, \quad \text{ENTÃO } \dot{\mathbf{x}} = f^i(\mathbf{x}, \mathbf{a}) \quad (3.97)$$

ou

$$\text{SE } \mathbf{x} \text{ é } F^i, \quad \text{ENTÃO } \mathbf{x}(k+1) = \mathbf{B}\mathbf{x}(k) + \mathbf{C}\mathbf{u}(k) \quad (3.98)$$

onde k representa o instante atual.

Capítulo 4

AGRUPAMENTO NEBULOSO

4.1 Introdução

A idéia básica do processo de agrupamento de dados (*clustering*) é uma tarefa em que se busca identificar um número finito de categorias ou grupos (*clusters*) para descrever um determinado conjunto de dados, tanto maximizando a homogeneidade de cada grupo como a heterogeneidade entre grupos distintos.

Neste capítulo apresentamos as principais técnicas de agrupamento baseadas em agrupamento nebuloso e novas técnicas baseadas em algoritmos clássicos que são muito utilizados para particionar o espaço de discurso das variáveis lingüísticas dos antecedentes de modelos nebulosos.

De acordo com Babuska no Capítulo 2 em [18], algoritmos de agrupamento podem ser classificados nos seguintes três tipos:

1. Algoritmos utilizando uma distância adaptativa, tal como o algoritmo de Gustafson-Kessel [16] ou o algoritmo de estimação por máxima verossemelhança nebulosa, proposto por I. Gath e A.B. Geva em [13].
2. Algoritmos baseados em formas hiperplanares, chamados também de algoritmo de variações lineares nebulosas ou c-elipsóides nebulosas.
3. Algoritmos c-regressão nebulosos, utilizando formas definidas por funções de regressão, Hathaway [17].

4.2 Agrupamentos Nebulosos

4.2.1 FCM - *Fuzzy C-Means*

O primeiro algoritmo de agrupamento nebuloso, desenvolvido em 1969 por Ruspini [36], é uma extensão do algoritmo C-means Rígido (HCM - *Hard C-Means*), chamado de ISODATA proposto por Ball e Hall [4] (1965). O HCM é um dos mais populares métodos de agrupamento. Ruspini (1969) introduziu a partição nebulosa para descrever estruturas de grupos de um

conjunto de dados e sugeriu um algoritmo computacional que determina uma partição nebulosa. Dunn (1973) generalizou o procedimento de agrupamento de variância mínima para a técnica de agrupamento nebuloso HCM. Bezdek (1981) generalizou a aproximação de Dunn criando o algoritmo *Fuzzy C-Means* (FCM), Silva [37].

Os passos para o algoritmo FCM são os seguintes:

1º **Passo:** Escolher o número de grupos c , $2 \leq c \leq n$, o parâmetro $m > 1$, o critério de parada $\epsilon > 0$, o número máximo $M_{\text{máx}}$ de iterações. Inicializar o contador $M = 0$ e a matriz μ^0 aleatoriamente como:

$$\mu^0 = \begin{bmatrix} \mu_{1,1}^0 & \mu_{1,2}^0 & \cdots & \mu_{1,k}^0 \\ \mu_{2,1}^0 & \mu_{2,2}^0 & \cdots & \mu_{2,k}^0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1}^0 & \mu_{N,2}^0 & \cdots & \mu_{N,k}^0 \end{bmatrix} \quad (4.1)$$

$$0 \leq \mu_{i,j}^0 \leq 1, \text{ onde } 1 \leq i \leq c, 1 \leq j \leq N \quad (4.2)$$

com

$$\sum_{i=1}^c \mu_{i,j}^0 = 1, \quad \forall j = 1, \dots, N. \quad (4.3)$$

Fazer $M = M + 1$.

2º **Passo:** Calcular os c centros dos grupos $\mathbf{C}\mathbf{c}_i^M$, $i = 1, \dots, c$, com a fórmula:

$$\mathbf{C}\mathbf{c}_i^M = \frac{\sum_{k=1}^N (\mu_{k,i}^M)^m \mathbf{x}(k)}{\sum_{k=1}^N (\mu_{k,i}^M)^m} \quad (4.4)$$

3º **Passo:** Calcular $d_{k,i}$ para $1 \leq i \leq c$ e $1 \leq k \leq N$:

$$d_{k,i} = \|\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M\|. \quad (4.5)$$

Atualizar a matriz de pertinência:

$$\mu_{k,i}^M = \frac{1}{\sum_{j=1}^c \left(\frac{d_{k,i}}{d_{k,j}} \right)^{\frac{2}{m-1}}} \quad (4.6)$$

4º **Passo:** Calcular $\Delta = \|\mu^M - \mu^{M-1}\| = \max_{j,i} |\mu_{j,i}^M - \mu_{j,i}^{M-1}|$, $j = 1, \dots, N$, $i = 1, \dots, c$.

5º **Passo:** Se $\Delta > \epsilon$ e $M < M_{\text{max}}$, $M = M + 1$ e retornar para o 2º passo; senão parar.

A seguir apresentamos e comentamos 2 exemplos que utilizamos para ilustrar o algoritmo proposto. Os dados das séries temporais utilizadas podem se encontrados na biblioteca de dados no endereço:

www-personal.buseco.monash.edu.au/~hyndman/TSDL/.

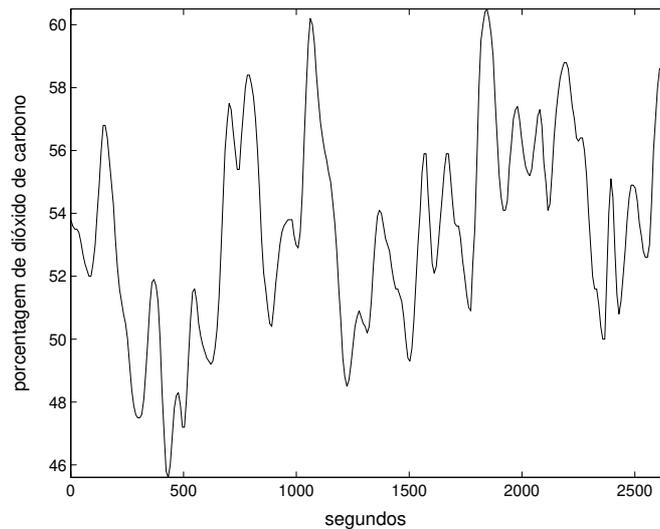


Figura 4.1: Saída de CO_2 em porcentagem no forno a gás industrial.

Exemplo 4.1 Considere a série temporal de um forno a gás industrial onde os dados consistem de 296 amostras de concentração em porcentagem de dióxido de carbono (CO_2) produzida no forno de 9 em 9 segundos. A série pode ser vista na Fig.4.1.

Para esta série temporal utilizamos o algoritmo de agrupamento FCM e determinamos o número de grupos como 2. A Fig.4.2 mostra como o centro de cada grupo se adaptou a partir dos dados da série. Os eixos $y(k-1)$ e $y(k-2)$ são valores de concentração de dióxido de carbono conhecidos e são as variáveis de entrada para o algoritmo de agrupamento FCM.

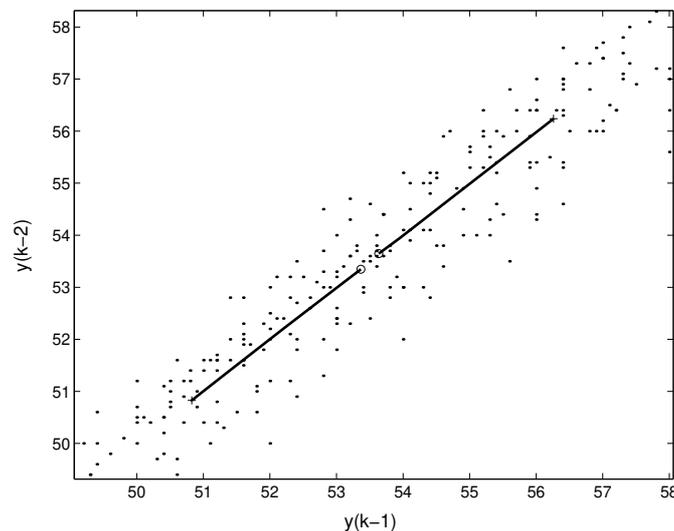


Figura 4.2: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

A superfície na Fig.4.3 representa a pertinência de cada grupo.

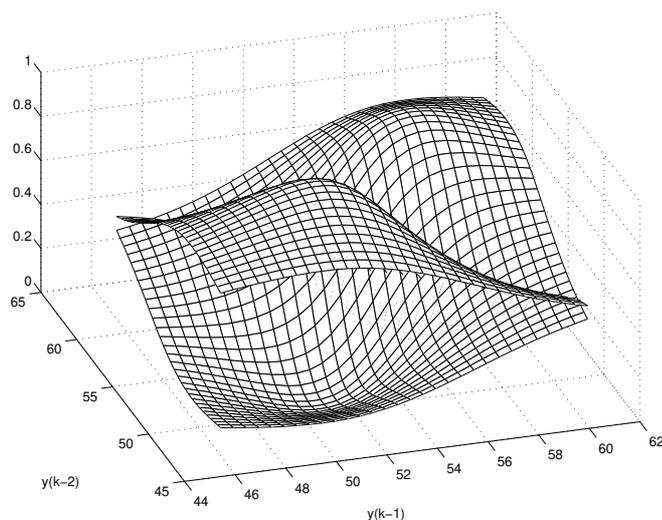


Figura 4.3: Funções de pertinência.

As pertinências dos grupos criados podem ser vistas em forma de cortes como mostrado na Fig.4.4.

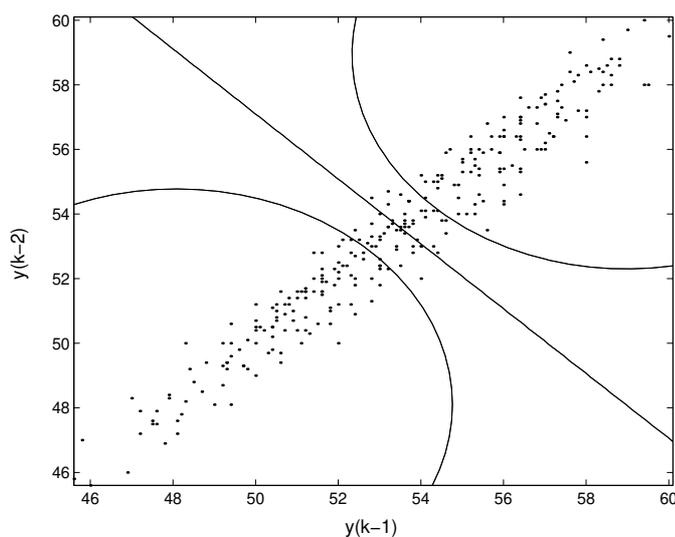


Figura 4.4: Curvas de nível das funções de pertinência.

Exemplo 4.2 A série temporal de terremotos com intensidade maior ou igual a 7.0 na escala Richter entre os anos de 1900 à 1998 (99 amostras) é dada na Fig.4.5.

Para esta série temporal utilizamos o algoritmo de agrupamento FCM e determinamos o número de grupos como 3.

A Fig.4.6 mostra como o centro de cada grupo se adaptou a partir dos dados da série.

A superfície na Fig.4.7 mostra a pertinência de cada grupo.

As pertinências dos grupos criados podem ser vistas em forma de cortes na Fig.4.8.

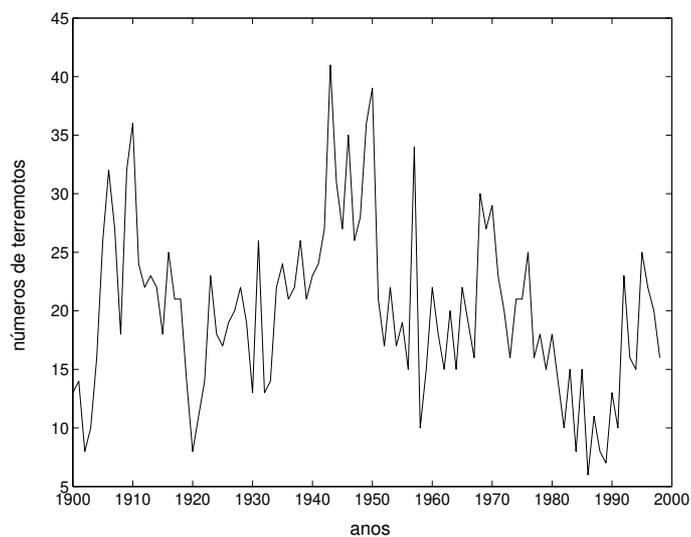


Figura 4.5: Quantidades de terremotos maior ou igual a 7.0 na escala Richter.

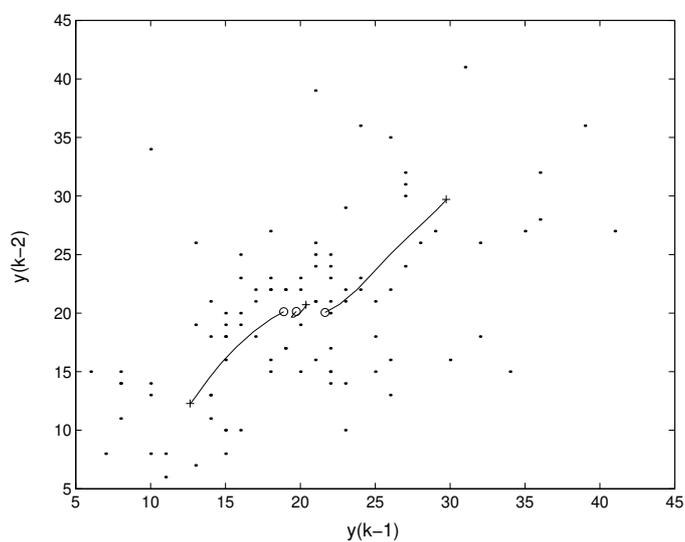


Figura 4.6: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

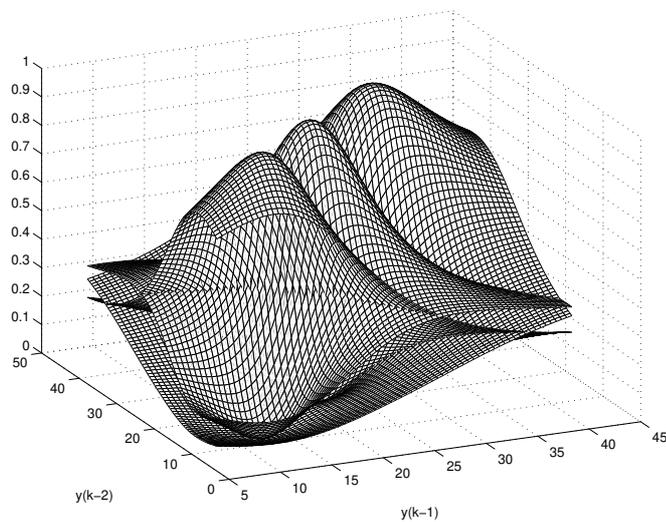


Figura 4.7: Funções de pertinência.

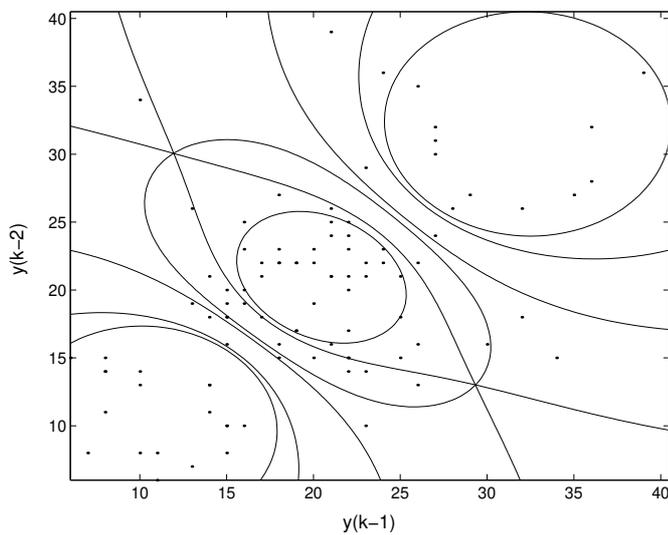


Figura 4.8: Curvas de nível das funções de pertinência.

4.2.2 Agrupamento Gustafson-Kessel

No algoritmo FCM, o agrupamento dos pontos minimiza o critério:

$$J = \sum_{k=1}^N \sum_{i=1}^c \mu_{k,i} d_{k,i}^2 \quad (4.7)$$

onde,

$$d_{k,i} = (\mathbf{x}(k) - Cc_i)^T \mathbf{M}_i (\mathbf{x}(k) - Cc_i) \quad (4.8)$$

e $\mathbf{M}_i = \mathbf{I}$ (matriz identidade).

O algoritmo de Gustafson-Kessel detecta grupos elipsoidais de tamanhos e orientações diferentes utilizando uma função custo onde a matriz \mathbf{M}_i é determinada como:

$$\mathbf{M}_i = \det(F_i)^{\frac{1}{N}} F_i^{-1} \quad (4.9)$$

sendo F_i função da distância $d_{k,i}$ (chamada de distância de Mahalanobis) e da matriz de pertinência μ .

O algoritmo é descrito abaixo:

1º Passo: Escolher o número de grupos c , $2 \leq c \leq n$, o parâmetro $m > 1$, o critério de parada $\epsilon > 0$, o número máximo $M_{\text{máx}}$ de iterações. Inicializar o contador $M = 0$ e a matriz μ^0 aleatoriamente como:

$$\mu^0 = \begin{bmatrix} \mu_{1,1}^0 & \mu_{1,2}^0 & \cdots & \mu_{1,k}^0 \\ \mu_{2,1}^0 & \mu_{2,2}^0 & \cdots & \mu_{2,k}^0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1}^0 & \mu_{N,2}^0 & \cdots & \mu_{N,k}^0 \end{bmatrix} \quad (4.10)$$

$$0 \leq \mu_{i,j}^0 \leq 1, \text{ onde } 1 \leq i \leq c, 1 \leq j \leq N \quad (4.11)$$

com

$$\sum_{i=1}^c \mu_{i,j}^0 = 1, \quad \forall j = 1, \dots, N. \quad (4.12)$$

2º Passo: Calcular os c centros dos grupos Cc_i^M , $i = 1, \dots, c$, com a fórmula:

$$Cc_i^M = \frac{\sum_{k=1}^N (\mu_{k,i}^M)^m \mathbf{x}(k)}{\sum_{k=1}^N (\mu_{k,i}^M)^m} \quad (4.13)$$

3º Passo: Calcular a matriz de covariância para $i = 1, \dots, c$:

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (\mu_{k,i}^M)^m (\mathbf{x}(k) - Cc_i^M) (\mathbf{x}(k) - Cc_i^M)^T}{\sum_{k=1}^N (\mu_{k,i}^M)^m} \quad (4.14)$$

4º **Passo:** Calcular a distância de Mahalanobis para $i = 1, \dots, c$ e $k = 1, \dots, N$:

$$d_{k,i} = (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M)^T \left[\left(\det(\mathbf{F}_i) \right)^{\frac{1}{N}} \mathbf{F}_i^{-1} \right] (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M). \quad (4.15)$$

5º **Passo:** Atualizar a matriz de pertinência:

$$\mu_{k,i}^M = \frac{1}{\sum_{j=1}^c \left(\frac{d_{k,i}}{d_{k,j}} \right)^{\frac{2}{m-1}}} \quad (4.16)$$

6º **Passo:** Calcular $\Delta = \|\mu^M - \mu^{M-1}\| = \max_{ji} |\mu_{j,i}^M - \mu_{j,i}^{M-1}|$, $j = 1, \dots, N$, $i = 1, \dots, c$.

7º **Passo:** Se $\Delta > \epsilon$ e $M < M_{max}$, $M = M + 1$ e retornar para o 2º passo; senão parar.

Para mostrar o comportamento do algoritmo e a forma de suas funções de pertinência são ilustrados dois exemplos.

Exemplo 4.3 Considere a série temporal de um forno a gás industrial descrito no exemplo 4.1. A série pode ser vista na Fig.4.1.

Para esta série temporal utilizamos o algoritmo de agrupamento Gustafson-Kessel e determinamos o número de grupos como 2.

A Fig.4.9 mostra como o centro de cada grupo se adaptou a partir dos dados da série.

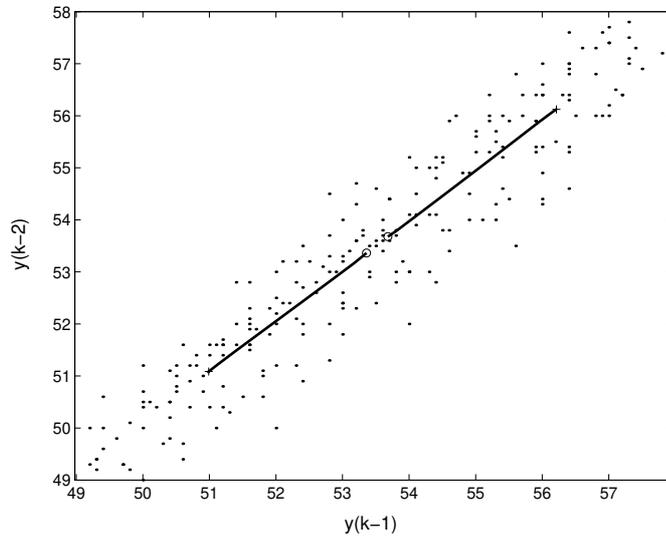


Figura 4.9: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

A superfície na Fig.4.10 mostra a pertinência para cada grupo.

As pertinências dos grupos criados podem ser vistas em forma de cortes como mostrado na Fig.4.11.

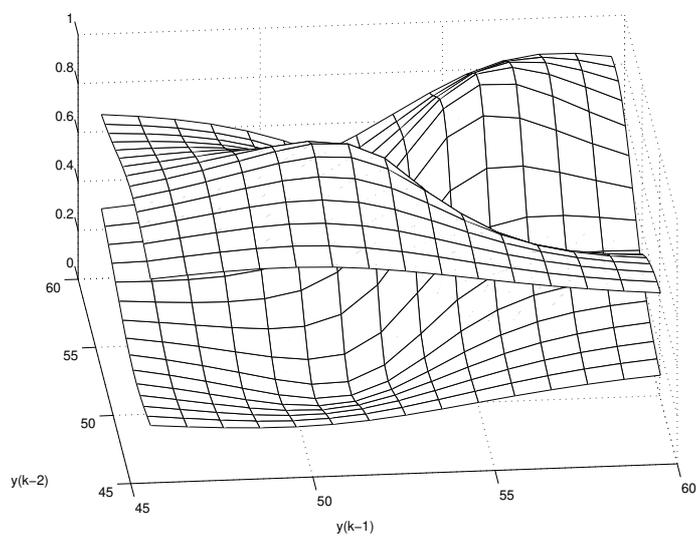


Figura 4.10: Funções de pertinência.

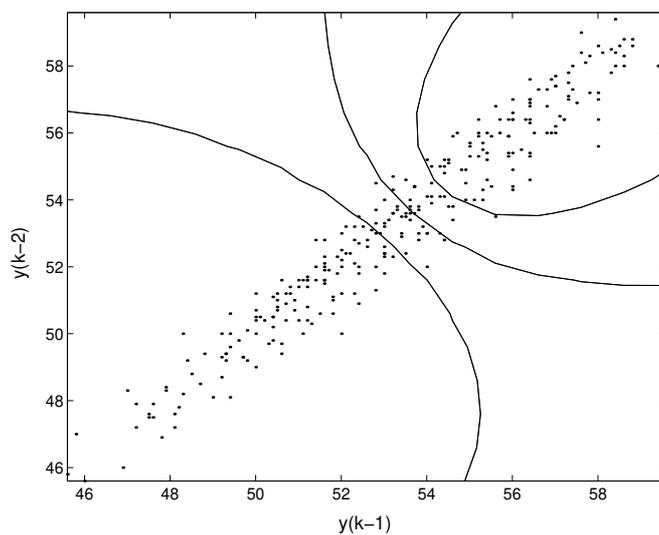


Figura 4.11: Curvas de nível das funções de pertinência.

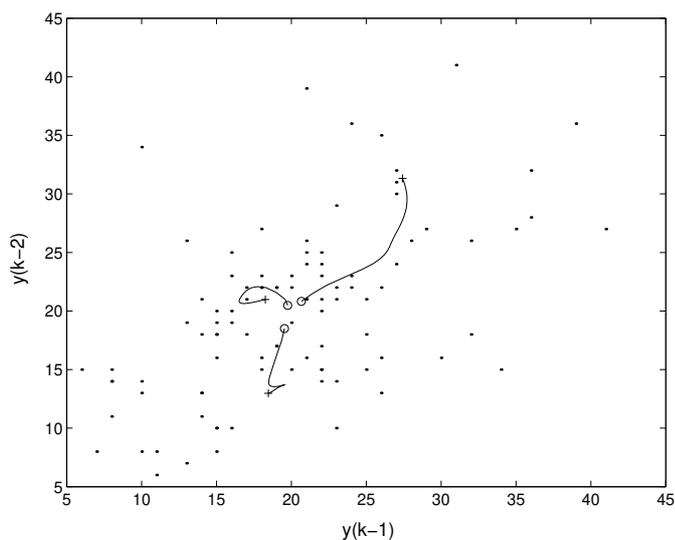


Figura 4.12: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

Exemplo 4.4 Dada a série temporal do exemplo 4.2. A série pode ser vista na Fig.4.5.

Para esta série temporal utilizamos o algoritmo de agrupamento Gustafson-Kessel e determinamos o número de grupos como 3.

A Fig.4.12 mostra como o centro de cada grupo se adaptou a partir dos dados da série.

A superfície na Fig.4.13 mostra a pertinência para cada grupo.

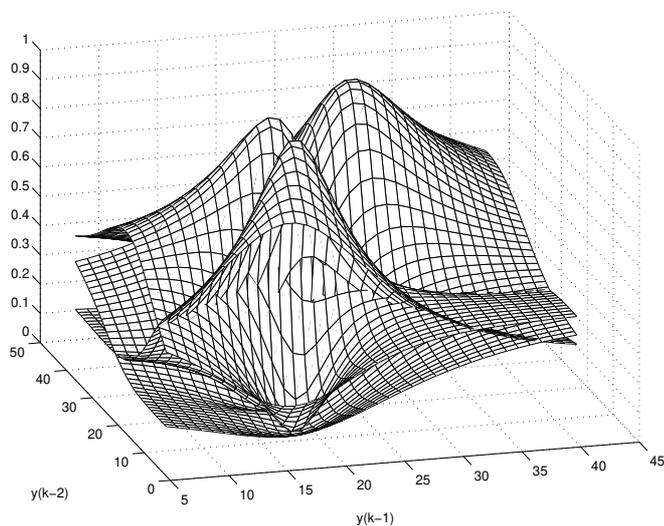


Figura 4.13: Funções de pertinência.

As pertinências dos grupos criados podem ser vistas em forma de cortes como mostrado na Fig.4.14.

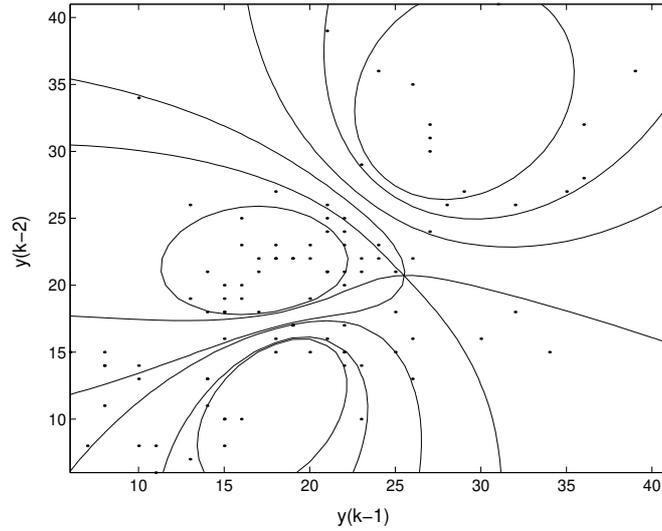


Figura 4.14: Curvas de nível das funções de pertinência.

4.2.3 Agrupamento Gath-Geva

Este algoritmo é derivado da combinação dos algoritmos do c-means nebuloso (FCM) e estimação de máxima verossemelhança nebulosa (FMLE). Este método também pode ser chamado como UFP-ONC (*Unsupervised Fuzzy Partition-Optimal Number of Classes*) pois se desempenha bem em situações de vários formatos de grupos, densidades, e números de pontos em cada grupo, Gath e Geva [13].

O algoritmo é descrito abaixo:

1º Passo: Escolher o número de grupos c , $2 \leq c \leq n$, o parâmetro $m > 1$, o critério de parada $\epsilon > 0$, o número máximo $M_{\text{máx}}$ de iterações. Inicializar o contador $M = 0$ e a matriz μ^0 aleatoriamente como:

$$\mu^0 = \begin{bmatrix} \mu_{1,1}^0 & \mu_{1,2}^0 & \cdots & \mu_{1,k}^0 \\ \mu_{2,1}^0 & \mu_{2,2}^0 & \cdots & \mu_{2,k}^0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1}^0 & \mu_{N,2}^0 & \cdots & \mu_{N,k}^0 \end{bmatrix} \quad (4.17)$$

$$0 \leq \mu_{i,j}^0 \leq 1, \text{ onde } 1 \leq i \leq c, 1 \leq j \leq N \quad (4.18)$$

com

$$\sum_{r=1}^c \mu_{i,r}^0 = 1, \quad \forall j = 1, \dots, N. \quad (4.19)$$

Fazer $M = M + 1$.

2º Passo: Calcular os c centros dos grupos $\mathbf{C}c_i^M$, $i = 1, \dots, c$, com a fórmula:

$$\mathbf{C}c_i^M = \frac{\sum_{k=1}^N (\mu_{k,i}^M)^m \mathbf{x}(k)}{\sum_{k=1}^N (\mu_{k,i}^M)^m} \quad (4.20)$$

3º **Passo:** Calcular a matriz de covariância para $i = 1, \dots, c$:

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (\mu_{k,i}^M)^m (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M) (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M)^T}{\sum_{k=1}^N (\mu_{k,i}^M)^m} \quad (4.21)$$

4º **Passo:** Calcular a média dos dados pertencentes ao grupo i :

$$P_i = \frac{1}{N} \sum_{k=1}^N \mu_{k,i}^M \quad (4.22)$$

5º **Passo:** Calcular a distância para $i = 1, \dots, c$ e $k = 1, \dots, N$:

$$d_{k,i}^2 = \frac{\sqrt{\det(\mathbf{F}_i)}}{P_i} \exp\left(\frac{1}{2} (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M)^T \mathbf{F}_i^{-1} (\mathbf{x}(k) - \mathbf{C}\mathbf{c}_i^M)\right). \quad (4.23)$$

6º **Passo:** Atualizar a matriz de pertinência:

$$\mu_{k,i}^M = \frac{1}{\sum_{j=1}^c \left(\frac{d_{k,i}}{d_{k,j}}\right)^{\frac{2}{m-1}}} \quad (4.24)$$

7º **Passo:** Calcular $\Delta = \|\mu^M - \mu^{M-1}\| = \max_{j,i} |\mu_{j,i}^M - \mu_{j,i}^{M-1}|$, $j = 1, \dots, N$, $i = 1, \dots, c$.

8º **Passo:** Se $\Delta > \epsilon$ e $M < M_{max}$, $M = M + 1$ e retornar para o 2º passo; senão parar.

Como maneira de mostrar o comportamento do algoritmo e a forma de suas funções de pertinência são ilustrados dois exemplos.

Exemplo 4.5 Considere a série temporal de um forno a gás industrial descrito no exemplo 4.1. A série pode ser vista na Fig.4.1.

Para esta série temporal utilizamos o algoritmo de agrupamento Gath-Geva e determinamos o número de grupos como 2.

A Fig.4.15 mostra como o centro de cada grupo se adaptou a partir dos dados da série.

A superfície na Fig.4.16 mostra a pertinência para cada grupo.

As pertinências dos grupos criados podem ser vistas em forma de cortes como mostrado na Fig.4.17.

Exemplo 4.6 Dada a série temporal do exemplo 4.2. A série pode ser vista na Fig.4.5.

Para esta série temporal utilizamos o algoritmo de agrupamento Gath-Geva e determinamos o número de grupos como 3.

A Fig.4.18 mostra como o centro de cada grupo se adaptou a partir dos dados da série.

A superfície na Fig.4.19 mostra a pertinência para cada grupo.

As pertinências dos grupos criados podem ser vistas em forma de cortes como mostrado na Fig.4.20.

O algoritmo FCM normalmente induz grupos circulares devido ao uso da distância Euclidiana, porém em muitas aplicações, em particular no projeto de modelos nebulosos do tipo Takagi-Sugeno, formatos elipsoidais são mais representativos como nos algoritmos Gustafson-Kessel e Gath-Geva. A grande vantagem do FCM sobre os outros é o custo computacional e facilidade implementação.

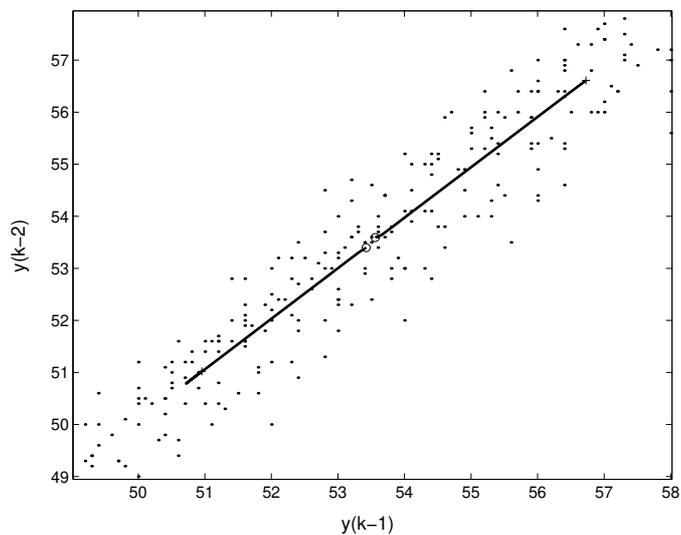


Figura 4.15: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

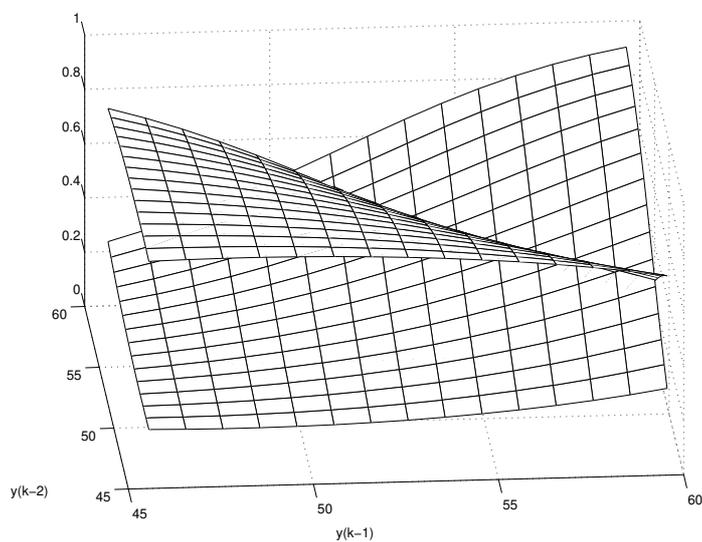


Figura 4.16: Funções de pertinência.

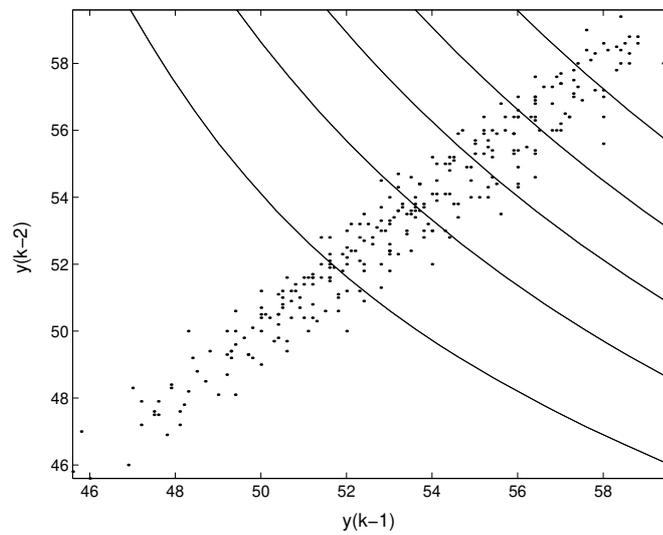


Figura 4.17: Curvas de nível das funções de pertinência.

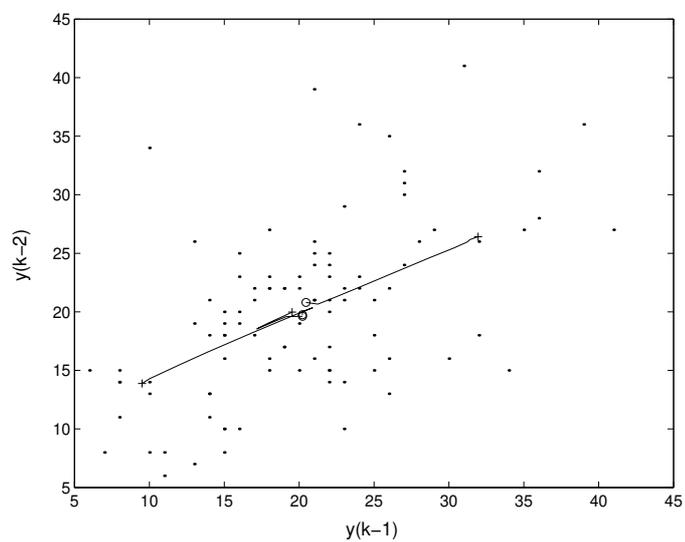


Figura 4.18: Adaptação dos centros de cada grupo, onde “o” são os valores iniciais, “-” a trajetória e “x” os valores finais.

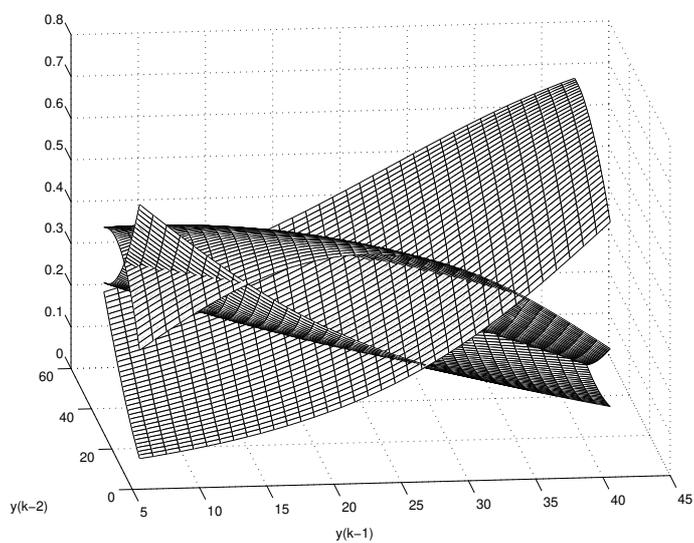


Figura 4.19: Funções de pertinência.

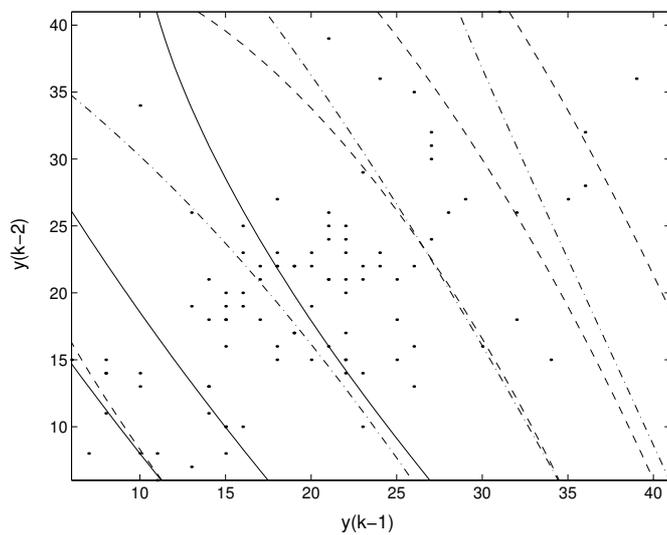


Figura 4.20: Curvas de nível das funções de pertinência.

4.2.4 FCRM - *Fuzzy c-Regression Model*

Este algoritmo de agrupamento é proposto em Kim [22]. Ele considera o modelo do grupo l como:

$$y^l(k+1) = \mathbf{Z}^T(k)\mathbf{A}^l \quad (4.25)$$

onde $\mathbf{Z}^T(k)$ é o vetor de regressores e \mathbf{A}^l é o vetor de parâmetros do grupo l .

O algoritmo FCRM tem os seguintes passos:

1º **Passo:** $M = 0$ e definir a matriz μ de tamanho $N \times k$ como:

$$\mu = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,k} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1} & \mu_{N,2} & \cdots & \mu_{N,k} \end{bmatrix} \quad (4.26)$$

$$0 \leq \mu_{i,j} \leq 1, \text{ onde } 1 \leq i \leq N, 1 \leq j \leq k \quad (4.27)$$

e

$$\sum_{i=1}^N \mu_{i,j} = 1, \quad \forall j = 1, \dots, k. \quad (4.28)$$

Inicializar a matriz μ com valores aleatórios satisfazendo as restrições em (4.27) e (4.28) e a matriz de parâmetros $\mathbf{A}^i = 0$.

2º **Passo:** Defino a função objetiva para o FCRM na M -ésima iteração como:

$$J = \sum_{r=1}^N J^r = \sum_{i=1}^N \sum_{j=1}^k \mu_{i,j} d_{i,j}^2 \quad (4.29)$$

onde $d_{i,j}$ é a norma Euclidiana entre o j -ésimo dado e a saída da i -ésima regra do modelo nebuloso TS:

$$d_{i,j} = \|y(i) - \mathbf{Z}^T(i-1)\mathbf{A}^j\|. \quad (4.30)$$

Para minimizar J satisfazendo a restrição (4.28) reescrevo a seguinte função objetiva:

$$\begin{aligned} \Phi &= J + \sum_{j=1}^k \lambda_j (\sum_{i=1}^N \mu_{i,j} - 1) \\ &= \sum_{i=1}^N \sum_{j=1}^k \mu_{i,j}^m d_{i,j}^2 + \sum_{j=1}^k \lambda_j (\sum_{i=1}^N \mu_{i,j} - 1) \end{aligned} \quad (4.31)$$

onde λ_j são os multiplicadores de Lagrange para as k restrições. Se derivarmos Φ em $\mu_{i,j}$ a condição necessária para minimizarmos (4.29) é:

$$\mu_{i,j} = \frac{1}{\sum_{k=1}^N \left(\frac{d_{i,j}}{d_{k,j}}\right)^{\frac{2}{m-1}}} \quad (4.32)$$

Na M -ésima iteração, calcular a função objetiva J e a matriz μ de acordo com as funções (4.29) e (4.32), respectivamente.

3º **Passo:** Se a função objetiva for menor que um certo valor de tolerância ou se o número de iterações não excedeu o seu limiar então pare. Caso contrário, vá para o 4º passo.

4º **Passo:** Calcular novos grupos na iteração $M+1$, $y^i(k+1) = \mathbf{Z}^T(k)\mathbf{A}^i$ utilizando $\mu_{i,j}$ obtidos no 2º passo e então utilizar o algoritmo dos mínimos quadrados recursivo ponderado:

$$\mathbf{A}^i(k+1) = \mathbf{A}^i(k) + \mathbf{H}(k)[y(k+1) - \mathbf{Z}^T(k)\mathbf{A}^i(k)] \quad (4.33)$$

$$\begin{aligned} \mathbf{H}(k) &= \mathbf{S}(k+1)\mathbf{Z}(k) \\ &= \frac{\mathbf{S}(k)\mathbf{Z}(k)}{\frac{1}{\mu_{i,j}} + \mathbf{Z}^T(k)\mathbf{S}(k)\mathbf{Z}(k)} \end{aligned} \quad (4.34)$$

$$\mathbf{S}(k+1) = [\mathbf{I} - \mathbf{H}(k)\mathbf{Z}^T(k)] \mathbf{S}(k). \quad (4.35)$$

O valor inicial do algoritmo é $\mathbf{S}(0) = \alpha\mathbf{I}$, onde α é um escalar com valor entre 10^2 e 10^6 . O valor escolhido para α reflete o conhecimento *a priori* dos parâmetros, ou seja quanto maior a incerteza sobre o vetor de parâmetros iniciais do sistema, mais alto o valor de α .

5º **Passo:** Vá ao 2º passo e faça $M = M + 1$.

Depois de determinado os parâmetros do conseqüente pelo FCRM, os parâmetros do antecedente podem ser facilmente obtidos usando os valores da matriz μ . Considerando que a função de pertinência escolhida seja do tipo sino temos:

$$\mathbf{F}_j^i(q_{j,1}^i, q_{j,2}^i) = \exp \left\{ - \left(\frac{y(k-j) - q_{j,1}^i}{q_{j,2}^i} \right)^2 \right\}. \quad (4.36)$$

onde $q_{j,1}^i$ e $q_{j,2}^i$ são os parâmetros desejados e são calculados da seguinte maneira:

$$q_{j,1}^i = \frac{\sum_{l=1}^k \mu_{i,l} y(l-j)}{\sum_{l=1}^k \mu_{i,l}} \quad (4.37)$$

e

$$q_{j,2}^i = \sqrt{2 \cdot \frac{\sum_{l=1}^k \mu_{i,l} (y(l-j) - q_{j,1}^i)^2}{\sum_{l=1}^k \mu_{i,l}}}. \quad (4.38)$$

Neste caso, não apresentaremos exemplos como nos algoritmos de agrupamento anteriormente mostrados pois o FRCM é um método completo: estimação dos conjuntos nebulosos das premissas e dos parâmetros dos conseqüentes.

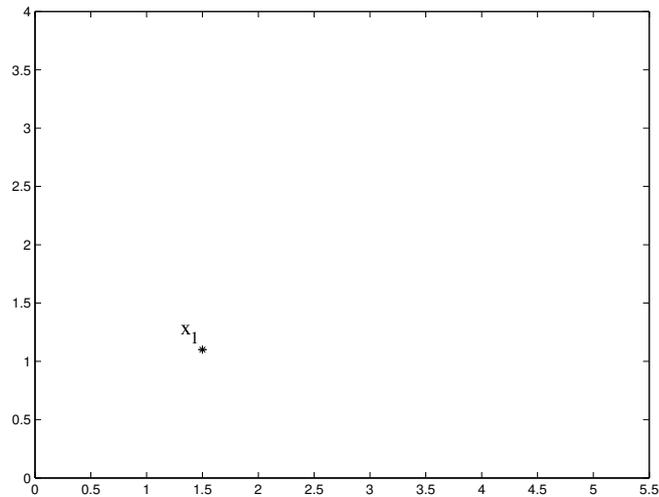


Figura 4.21: x_1 : cria um novo grupo C_1^0 .

4.3 Agrupamento Não-Nebuloso

4.3.1 ECM - *Evolving Clustering Method*

O método ECM é um algoritmo baseado na distância máxima, limitada por uma constante D_{limiar} , entre um centro do grupo e o seu ponto mais afastado, com a vantagem de ser utilizado nos modos *on-line* ou *off-line* com alto desempenho computacional [21].

A formulação on-line é implementada de forma que os centros e raios dos grupos são atualizados a cada nova entrada de dados, de acordo com três possíveis casos:

- um novo grupo é criado. Seu centro, Cc , é alocado com o valor da entrada corrente e seu raio, Ru , é inicialmente zero;
- os grupos são atualizados. Um novo centro é calculado e o novo raio é expandido;
- o grupo não será alterado se a entrada estiver dentro de seu raio.

O algoritmo pode ser resumido como:

1º Passo: Cria o primeiro grupo C_1^0 definindo o primeiro exemplo da entrada como o primeiro centro do grupo Cc_1^0 , e considerando zero o valor do raio do grupo Ru_1 . Como acontece com x_1 e ilustrado na Fig.4.21.

2º Passo: Se todos os exemplos do conjunto de dados tiverem sido processados, então o algoritmo é encerrado. Senão, o exemplo atual, x_i , é utilizado e é calculada a distância entre este dado de entrada e todos os n centros de grupos criados Cc_j , $D_{ij} = \|x_i - Cc_j\|$, onde $j = 1, 2, \dots, n$. Neste trabalho, a distância entre dois vetores \mathbf{x} e \mathbf{y} é representada como a distância Euclidiana:

$$\|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^q |x_i - y_i|^2 \right)^{\frac{1}{2}} / q^{\frac{1}{2}} \quad (4.39)$$

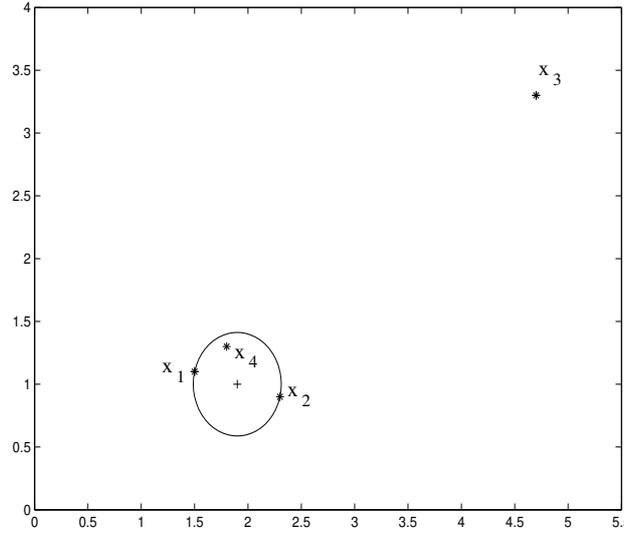


Figura 4.22: x_2 : atualiza grupo $C_1^0 \rightarrow C_1^1$; x_3 : cria um novo grupo C_2^0 ; x_4 : não faz nada.

3º Passo: Se existe algum valor de distância, $D_{ij} = \|x_i - Cc_j\|$, igual ou menor que ao menos um dos raios, Ru_j , $j = 1, 2, \dots, n$, significa que o exemplo x_i , pertence ao grupo C_m com a distância mínima:

$$D_{im} = \|x_i - Cc_m\| = \min(\|x_i - Cc_j\|), \quad (4.40)$$

sujeito a $D_{ij} \leq Ru_j$, $j = 1, 2, \dots, n$. Neste caso, não é criado um novo grupo, os grupos existentes não são atualizados e o algoritmo retorna ao 2º Passo. Como é ilustrado com x_4 na Fig.4.22 e x_6 na Fig.4.23. Caso contrário, vá ao 4º Passo.

4º Passo: Encontra o grupo C_a (com centro Cc_a e raio Ru_a) dos n centros dos grupos existentes através do seguinte cálculo $S_{ij} = D_{ij} + Ru_j$, $j = 1, 2, \dots, n$, e escolhe o centro do grupo Cc_a , com o menor valor, S_{ia} :

$$S_{ia} = D_{ia} + Ru_a = \min(S_{ij}), \quad j = 1, 2, \dots, n. \quad (4.41)$$

5º Passo: Se $S_{ia} > 2 \times D_{limiar}$, o dado x_i não pertence à nenhum grupo existente. Um novo grupo deverá ser criado da mesma maneira como descrito no 1º Passo e o algoritmo retorna ao 2º Passo. Como acontece com x_3 na Fig.4.22 e x_8 na Fig.4.23. Caso contrário, vá ao 6º Passo.

6º Passo: Se $S_{ia} \leq 2 \times D_{limiar}$, o grupo C_a é atualizado deslocando o seu centro, Cc_a , e aumentando o seu raio, Ru_a . O novo raio Ru_a^{new} é igual a $S_{ia}/2$ e seu novo centro Cc_a^{new} é um ponto localizado entre uma linha imaginária que conecta x_i e Cc_a . A distância do novo centro Cc_a^{new} até o ponto x_i é igual à Ru_a^{new} . Como acontece com x_2 na Fig.4.22, x_5 e x_7 na Fig.4.23 e x_9 e x_{10} na Fig.4.24. O algoritmo então retorna ao 2º Passo.

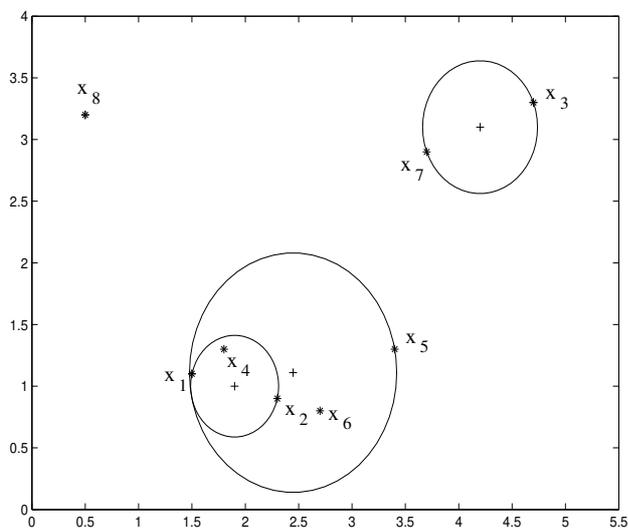


Figura 4.23: x_5 : atualiza grupo $C_1^1 \rightarrow C_1^2$; x_6 : não faz nada; x_7 : atualiza grupo $C_2^0 \rightarrow C_2^1$; x_8 : cria um novo grupo C_3^0 .

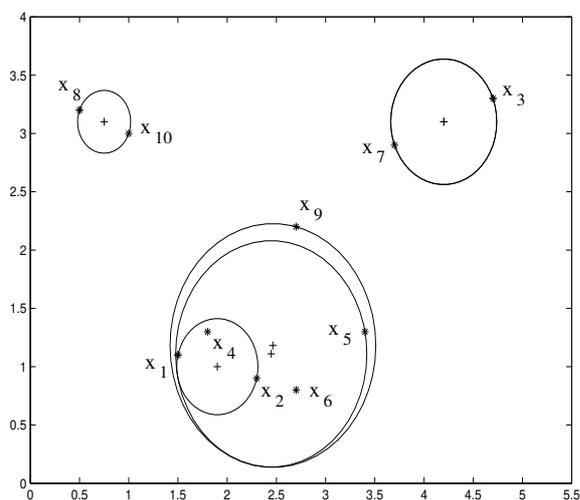


Figura 4.24: x_9 : atualiza grupo $C_2^1 \rightarrow C_2^2$; x_{10} : atualiza grupo $C_3^0 \rightarrow C_3^1$.

Capítulo 5

IDENTIFICAÇÃO PARAMÉTRICA

5.1 Mínimos Quadrados

O método dos mínimos quadrados é a mais antiga e a mais conhecida técnica para a estimação de parâmetros de modelos matemáticos, Caines [10]. O conceito básico desta técnica foi aparentemente elaborado por Carl Friedrich Gauss, matemático e astrônomo alemão (1777-1855) e primeiramente publicado em 1806 por Legendre.

Como exemplo, considere um sistema discreto com uma entrada e uma saída modelado da seguinte maneira:

$$y(k) = b_1u(k-1) + b_2u(k-2) + \dots + b_pu(k-p) + a_1y(k-1) + a_2y(k-2) + \dots + b_qy(k-q) + e(k). \quad (5.1)$$

onde u representa a entrada do sistema e y é a saída. A variável $e(k)$ é o ruído no k -ésimo instante.

Para N medidas temos:

$$\begin{aligned} y(1) &= b_1u(1-1) + b_2u(1-2) + \dots + b_pu(1-p) \\ &\quad + a_1y(1-1) + a_2y(1-2) + \dots + b_qy(1-q) + e(1) \\ y(2) &= b_1u(2-1) + b_2u(2-2) + \dots + b_pu(2-p) \\ &\quad + a_1y(2-1) + a_2y(2-2) + \dots + b_qy(2-q) + e(2) \\ &\quad \vdots \\ y(N) &= b_1u(N-1) + b_2u(N-2) + \dots + b_pu(N-p) \\ &\quad + a_1y(N-1) + a_2y(N-2) + \dots + b_qy(N-q) + e(N) \end{aligned} \quad (5.2)$$

Temos o vetor \mathbf{Y} de saída:

$$\mathbf{Y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}_{N \times 1} . \quad (5.3)$$

A matriz de dados será:

$$\mathbf{X} = \begin{bmatrix} \mathbf{Z}^T(1) \\ \mathbf{Z}^T(2) \\ \vdots \\ \mathbf{Z}^T(N) \end{bmatrix}_{N \times (p+q)} \quad (5.4)$$

onde $\mathbf{Z}(k)$ é o vetor dos regressores no instante k , ou seja:

$$\mathbf{Z}(k) = \begin{bmatrix} u(k-1) \\ \vdots \\ u(k-p) \\ y(k-1) \\ y(k-2) \\ \vdots \\ y(k-q) \end{bmatrix}_{(p+q) \times 1} . \quad (5.5)$$

O objetivo do método de mínimos quadrados é estimar a matriz dos parâmetros que será descrita por:

$$\mathbf{A} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \\ a_1 \\ a_2 \\ \vdots \\ a_q \end{bmatrix}_{(p+q) \times 1} . \quad (5.6)$$

E o vetor de erros será representado como:

$$\epsilon = \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}_{N \times 1} . \quad (5.7)$$

Pode-se reescrever a equação (5.2) como:

$$\mathbf{Y} = \mathbf{XA} + \epsilon \quad (5.8)$$

Considerando a existência do erro, $\epsilon = \mathbf{Y} - \mathbf{XA}$, o problema de mínimos quadrados é determinar \mathbf{A} que minimize o critério \mathbf{J} :

$$\mathbf{J} = \frac{1}{2} \|\epsilon\|^2 = \frac{1}{2} (\mathbf{Y} - \mathbf{XA})^T (\mathbf{Y} - \mathbf{XA}). \quad (5.9)$$

Efetuada a multiplicação teremos:

$$\mathbf{J} = \frac{1}{2} (\mathbf{Y}^T \mathbf{Y} - \mathbf{A}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{XA} + \mathbf{A}^T \mathbf{X}^T \mathbf{XA}). \quad (5.10)$$

Para a minimização da forma quadrática realizamos $\frac{\partial \mathbf{J}}{\partial \mathbf{A}} = 0$:

$$\frac{\partial \mathbf{J}}{\partial \mathbf{A}} = -\mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} + 2\mathbf{A}^T \mathbf{X}^T \mathbf{X} = 0. \quad (5.11)$$

Deste modo, podemos escrever a equação normal:

$$\mathbf{X}^T \mathbf{XA} = \mathbf{X}^T \mathbf{Y}. \quad (5.12)$$

Portanto:

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (5.13)$$

Se $\mathbf{P} = (\mathbf{X}^T \mathbf{X})^{-1}$ temos:

$$\mathbf{A} = \mathbf{PX}^T \mathbf{Y}. \quad (5.14)$$

Para ponderar de maneira diferenciada as observações disponíveis usa-se um tratamento conhecido como mínimos quadrados ponderados:

$$\mathbf{A} = \mathbf{PX}^T \mathbf{WY} \quad (5.15)$$

e

$$\mathbf{P} = (\mathbf{X}^T \mathbf{WX})^{-1}, \quad (5.16)$$

onde:

$$\mathbf{W} = \begin{bmatrix} w(1) & 0 & \cdots & 0 \\ 0 & w(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w(N) \end{bmatrix} \quad (5.17)$$

sendo w o peso de ponderação de cada medida para o cálculo dos parâmetros.

5.1.1 Mínimos Quadrados Recursivo Ponderado

A idéia básica do algoritmo dos mínimos quadrados recursivo é calcular um novo vetor de parâmetros estimados $\mathbf{A}(k+1)$ no instante $k+1$ pela adição de um fator de correção ao vetor de parâmetros estimados anteriormente $\mathbf{A}(k)$ no instante k . Este vetor de correção depende das novas medidas de entrada dos regressores:

$$\mathbf{Z}(k) = [r_1(k), r_2(k), \dots, r_p(k)]^T \quad (5.18)$$

e a saída do processo $y(k)$. No caso do modelo Auto-Regressivo (AR) com uma saída apenas, o vetor $\mathbf{Z}(k)$ seria:

$$\mathbf{Z}(k) = [y(k-1), \dots, y(k-p)]^T \quad (5.19)$$

onde p é o número de regressores.

A estimação para os mínimos quadrados recursivos ponderados é definida por:

$$\mathbf{A}(k) = \mathbf{P}(k)\mathbf{X}^T(k)\mathbf{W}(k)y(k) \quad (5.20)$$

$\mathbf{W}(k)$ é chamada matriz de peso e uma estrutura diagonal:

$$\mathbf{W}(k) = \begin{bmatrix} w(1) & 0 & \dots & 0 \\ 0 & w(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w(k) \end{bmatrix} \quad (5.21)$$

onde:

$$\mathbf{P}(k) = (\mathbf{X}^T(k)\mathbf{W}(k)\mathbf{X}(k))^{-1} \quad (5.22)$$

e $\mathbf{X}(k)$ a matriz de regressão no instante k , logo:

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{Z}^T(1) \\ \mathbf{Z}^T(2) \\ \vdots \\ \mathbf{Z}^T(k) \end{bmatrix} \quad (5.23)$$

Analogamente, a equação (5.20) do estimador para o instante $k+1$ é:

$$\mathbf{A}(k+1) = \mathbf{P}(k+1)\mathbf{Z}^T(k+1)\mathbf{W}(k+1)y(k+1) \quad (5.24)$$

Obviamente, a equação (5.24) pode ser escrita como:

$$\begin{aligned} \mathbf{A}(k+1) &= \mathbf{P}(k+1) \begin{pmatrix} \mathbf{X}(k) \\ \mathbf{Z}^T(k+1) \end{pmatrix}^T \begin{bmatrix} \mathbf{W}(k) & 0 \\ 0 & w(k+1) \end{bmatrix} \begin{pmatrix} y(k) \\ y(k+1) \end{pmatrix} \\ &= \mathbf{P}(k+1) [\mathbf{X}^T(k)\mathbf{W}(k)y(k) + \mathbf{Z}(k+1)w(k+1)y(k+1)] \end{aligned} \quad (5.25)$$

Substituindo $\mathbf{X}^T(k)\mathbf{W}(k)y(k) = \mathbf{P}^{-1}(k)\mathbf{A}(k)$ (da equação (5.20)) na equação (5.24), e adicionando e subtraindo $\mathbf{A}(k)$ no segundo membro da equação, obtemos:

$$\begin{aligned}
\mathbf{A}(k+1) &= \mathbf{A}(k) + \mathbf{P}(k+1)[\mathbf{P}^{-1}(k)\mathbf{A}(k) + \mathbf{Z}(k+1)w(k+1)y(k+1)] - \mathbf{A}(k) \\
&= \mathbf{A}(k) + \mathbf{P}(k+1)\mathbf{P}^{-1}(k)\mathbf{A}(k) - \mathbf{A}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)y(k+1) \\
&= \mathbf{A}(k) + (\mathbf{P}(k+1)\mathbf{P}^{-1}(k) - \mathbf{I})\mathbf{A}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)y(k+1)
\end{aligned} \tag{5.26}$$

onde de acordo com (5.22):

$$\begin{aligned}
\mathbf{P}(k+1) &= \left(\left(\begin{array}{c} \mathbf{X}(k) \\ \mathbf{Z}^T(k+1) \end{array} \right)^T \left[\begin{array}{cc} \mathbf{W}(k) & 0 \\ 0 & w(k+1) \end{array} \right] \left(\begin{array}{c} \mathbf{X}(k) \\ \mathbf{Z}^T(k+1) \end{array} \right) \right)^{-1} \\
&= (\mathbf{X}^T(k)\mathbf{W}(k)\mathbf{X}(k) + \mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T)^{-1} \\
&= (\mathbf{P}(k)^{-1} + \mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T)^{-1}
\end{aligned} \tag{5.27}$$

Calculando a inversa para os dois lados em (5.27), obtemos:

$$\begin{aligned}
\mathbf{P}(k+1)^{-1} &= \mathbf{P}(k)^{-1} + \mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T \\
\mathbf{P}(k)^{-1} &= \mathbf{P}(k+1)^{-1} - \mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T
\end{aligned} \tag{5.28}$$

Substituindo (5.28) em (5.26), a equação recursiva do estimador é obtida:

$$\begin{aligned}
\mathbf{A}(k+1) &= \mathbf{A}(k) + [\mathbf{P}(k+1)(\mathbf{P}(k)^{-1} - \mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T) \\
&\quad - \mathbf{I}] \mathbf{A}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)y(k+1) \\
&= \mathbf{A}(k) + [\mathbf{I} - \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)\mathbf{Z}(k+1)^T - \mathbf{I}] \mathbf{A}(k) \\
&\quad + \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)y(k+1) \\
&= \mathbf{A}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)w(k+1)(y(k+1) - \mathbf{Z}(k+1)^T\mathbf{A}(k))
\end{aligned} \tag{5.29}$$

O algoritmo dos mínimos quadrados recursivo necessita da inversão da matrix \mathbf{P} . Utilizando o seguinte lema da inversão matricial, a inversão da matriz pode ser conseguida com um trabalho computacional menor.

Lema 5.1 *Seja \mathbf{A} , \mathbf{C} e $(\mathbf{A} + \mathbf{BCD})$ matrizes não-singulares, podemos assegurar através do lema de inversão matricial que:*

$$\mathbf{E} = (\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1} \quad (5.30)$$

Prova: Ver Henderson e Searle [19]. Com este lema, a equação (5.29) torna-se:

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \mathbf{P}(k)\mathbf{Z}(k+1) \left(\frac{1}{w(k+1)} + \mathbf{Z}(k+1)^T\mathbf{P}(k)\mathbf{Z}(k+1) \right)^{-1} \mathbf{Z}(k+1)^T\mathbf{P}(k) \quad (5.31)$$

e realizando algumas simplificações obtemos a forma recursiva para $\mathbf{P}(k+1)$:

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \frac{w(k+1)\mathbf{P}(k)\mathbf{Z}(k+1)\mathbf{Z}(k+1)^T\mathbf{P}(k)}{1 + w(k+1)\mathbf{Z}(k+1)^T\mathbf{P}(k)\mathbf{Z}(k+1)} \quad (5.32)$$

Na prática o algoritmo dos mínimos quadrados tem uma taxa de convergência muito alta, mas essa taxa se reduz drasticamente quando a matriz \mathbf{P} tem seus elementos da diagonal pequenos depois de algumas iterações (normalmente depois de 10 ou 20 iterações). Esta propriedade motivou o algoritmo dos Mínimos Quadrados com Restauração da Covariância, onde a matriz \mathbf{P} é restaurada várias vezes. Isto permitirá que o algoritmo mantenha uma taxa de convergência alta. Este algoritmo é muito útil em casos onde o modelo possui parâmetros variantes no tempo. Neste caso, o tempo para que a matriz \mathbf{P} seja restaurada depende da confiabilidade das observações atuais em relação as anteriores.

Seja $\{T_s\} = \{t_1, t_2, t_3, \dots\}$ o conjunto dos instantes nos quais a matriz \mathbf{P} será restaurada. Então para $t \notin \{T_s\}$ a matriz \mathbf{P} é atualizada normalmente como mostrado abaixo:

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \frac{\mathbf{P}(k)\mathbf{Z}(k+1)\mathbf{Z}(k+1)^T\mathbf{P}(k)}{1 + \mathbf{Z}(k+1)^T\mathbf{P}(k)\mathbf{Z}(k+1)} \quad (5.33)$$

Caso contrário, para $t = t_i \in \{T_s\}$, é restaurado $\mathbf{P}(t_i - 1)$:

$$\mathbf{P}(t_i - 1) = k_i\mathbf{I} \quad \text{onde} \quad 0 < k_{min} \leq k_i \leq k_{max} < \infty \quad (5.34)$$

Um modelo muito utilizado em séries temporais é o tipo AR (Auto-Regressivo) onde:

$$y(k) = b_0 + a_1y(k-1) + a_2y(k-2) + \dots + b_qy(k-q) + e(k). \quad (5.35)$$

Neste caso, o vetor dos regressores $\mathbf{Z}(k)$, no instante k , será:

$$\mathbf{Z}(k) = \begin{bmatrix} 1 \\ y(k-1) \\ y(k-2) \\ \vdots \\ y(k-q) \end{bmatrix}. \quad (5.36)$$

A matriz dos parâmetros deste modelo ficará:

$$\mathbf{A} = \begin{bmatrix} b_0 \\ a_1 \\ a_2 \\ \vdots \\ a_q \end{bmatrix}. \quad (5.37)$$

Exemplo 5.1 Considere um sistema de primeira ordem descrito por:

$$y(k) = ay(k-1) + bu(k-1) + e(k)$$

onde os verdadeiros valores dos parâmetros são $a = 0.5$, $b = 1$ e a variância do ruído branco ($e(k)$) é 0.1. A variável u é uma constante unitária, ou seja $u(k) = 1$ para qualquer k .

A Fig.5.1 mostra a variação de dois parâmetros na estimação do algoritmo dos mínimos quadrados recursivo.

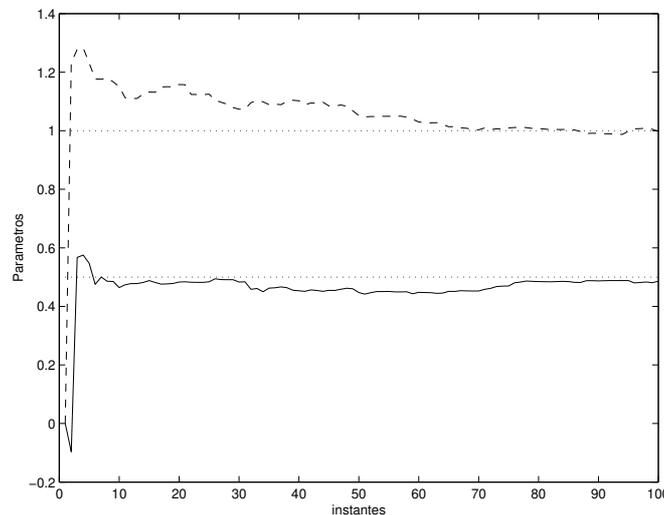


Figura 5.1: Estimação dos parâmetros \hat{a} “-” e \hat{b} “- -” via mínimos quadrados para o exemplo 4.1.

Os parâmetros \hat{a} e \hat{b} são inicializados com valor nulo e a matriz de covariância é $\mathbf{P}(0) = 10\mathbf{I}_2$ onde \mathbf{I}_2 é a matriz identidade com tamanho 2×2 .

O erro é dado por:

$$\epsilon(k) = y(k) - \mathbf{Z}(k)^T \mathbf{A}(k)$$

e pode ser visto na Fig.5.2.

Podemos observar que a estimação de \hat{a} e \hat{b} é rápida pois em menos de 100 iterações os parâmetros estimados estão próximos do real.

Exemplo 5.2 Considere o sistema de primeira ordem descrito no exemplo 5.1. A única diferença, é u ser uma variável aleatória com média zero e variância unitária.

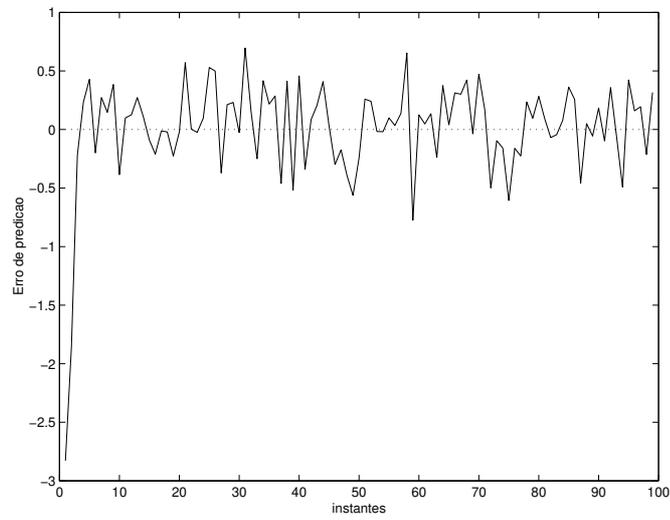


Figura 5.2: Seqüência de erro do exemplo 4.1.

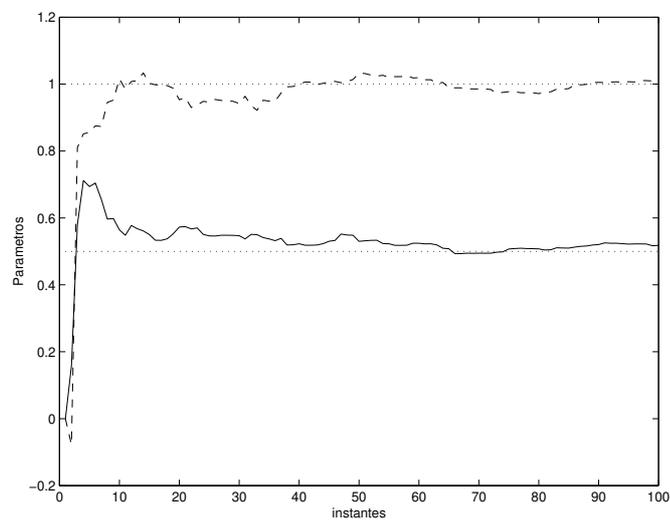


Figura 5.3: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via mínimos quadrados para o exemplo 4.2.

A Fig.5.3 mostra a variação de dois parâmetros na estimação do algoritmo dos mínimos quadrados recursivos.

Os parâmetros \hat{a} e \hat{b} e a matriz de covariância é \mathbf{P} são inicializados conforme o exemplo 5.1.

Como no exemplo anterior também para este sistema a estimação de \hat{a} e \hat{b} é rápida (menos de 100 iterações).

Exemplo 5.3 Considere um sistema de segunda ordem descrito por:

$$y(k) = ay(k-1) + by(k-2) + e(k)$$

onde os verdadeiros valores dos parâmetros são $a = 0.5$, $b = 1$ e a variância do ruído branco ($e(k)$) é 0.1.

A Fig.5.4 mostra a variação de dois parâmetros na estimação do algoritmo dos mínimos quadrados recursivos.

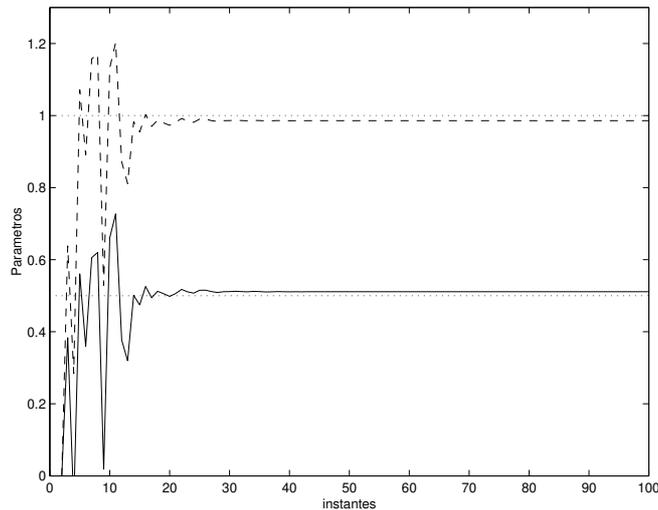


Figura 5.4: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via mínimos quadrados para o exemplo 4.3.

Os parâmetros \hat{a} e \hat{b} são inicializados com valor nulo e a matriz de covariância é $\mathbf{P}(0) = 10\mathbf{I}_2$ onde \mathbf{I}_2 é a matriz identidade com tamanho 2×2 .

Considerando este sistema AR (autoregressivo) de segunda ordem a estimação de \hat{a} e \hat{b} é bastante rápida (foram necessários apenas 20 pontos conhecidos para os parâmetros aproximarem do valor real).

5.2 Resíduos e Erros de Predição

No algoritmo de mínimos quadrados recursivos há duas variáveis, $\epsilon(k)$ e $\eta(k)$ que são sutilmente diferentes. A primeira, $\epsilon(k)$, chamada de erro de predição, é definida como:

$$\epsilon(k) = y(k) - \mathbf{Z}(k)^T \hat{\mathbf{A}}(k-1) \quad (5.38)$$

Em outras palavras, este é o erro entre a saída do sistema, $y(k)$, e a saída predita utilizando os parâmetros estimados $\hat{\mathbf{A}}(k)$ no instante anterior $k-1$. Por esta razão ele é normalmente chamado de erro de predição da saída (*a priori*) [43].

O verdadeiro erro de modelagem (ou resíduo ou erro de predição *a posteriori*) no instante k é:

$$\eta(k) = y(k) - \mathbf{Z}(k)^T \hat{\mathbf{A}}(k) \quad (5.39)$$

onde a única diferença entre $\epsilon(k)$ e $\eta(k)$ é que $\eta(k)$ é baseado no parâmetro estimado $\hat{\mathbf{A}}(k)$ no instante k .

Quando passados vários instantes de tempo, ou seja k é grande, a diferença entre $\epsilon(k)$ e $\eta(k)$ é insignificante, porém nas primeiras recursões a diferença é bastante significativa e os algoritmos de estimação exploram isso.

Podemos escrever $\eta(k)$ em função de $\epsilon(k)$ como:

$$\eta(k) = \epsilon(k) - \mathbf{Z}(k)^T \left(\hat{\mathbf{A}}(k) - \hat{\mathbf{A}}(k-1) \right). \quad (5.40)$$

5.3 Método do Gradiente

A fórmula de atualização do vetor de parâmetros $\hat{\mathbf{A}}$ no algoritmo recursivo do mínimos quadrados é dada por:

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)^T \left(y(k+1) - \mathbf{Z}(k+1)^T \hat{\mathbf{A}}(k) \right) \quad (5.41)$$

ou pode ser escrita como:

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) + \mathbf{P}(k+1)\mathbf{Z}(k+1)^T \epsilon(k+1). \quad (5.42)$$

Considerando o critério:

$$\epsilon^2(k+1) = \frac{1}{2} \left[y(k+1) - \mathbf{Z}(k+1)^T \hat{\mathbf{A}}(k) \right]^2, \quad (5.43)$$

o gradiente em relação a $\hat{\mathbf{A}}(k)$ será:

$$\frac{\partial [\epsilon^2(k+1)]}{\partial \hat{\mathbf{A}}(k)} = - \{ \mathbf{Z}(k+1) \epsilon(k+1) \}, \quad (5.44)$$

Substituindo (5.44) em (5.41), temos:

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) + \mathbf{P}(k+1) \frac{\partial [\epsilon^2(k+1)]}{\partial \hat{\mathbf{A}}(k)} \quad (5.45)$$

onde $\mathbf{P}(k+1)$ pode ser definida como uma matriz definida positiva na qual determina a direção do gradiente.

É factível, portanto, utilizar um ganho $\gamma(k+1)$ no lugar de $\mathbf{P}(k+1)$. Este ganho $\gamma(k+1)$ pode ser um escalar ou uma matriz e pode ser constante ou variante no tempo desde que a convergência de $\hat{\mathbf{A}}(k)$ seja garantida.

Considerando um sistema com um ruído $e(k)$:

$$y(k) = \mathbf{Z}(k)^T \mathbf{A} + e(k), \quad (5.46)$$

significa que $\gamma(k)$ deve decrescer com o tempo pois o ganho determina o tamanho do passo para a convergência de $\hat{\mathbf{A}}$. Se não houver o ruído, $\gamma(k)$ pode ser considerado constante.

Considerando um ganho $\gamma(k)$ que satisfaz a convergência o algoritmo do método do gradiente pode ser simplificado por:

$$\hat{\mathbf{A}}(k+1) = \hat{\mathbf{A}}(k) + \gamma(k+1) \mathbf{Z}(k+1) \epsilon(k+1) \quad (5.47)$$

chamado de algoritmo de aproximação estocástica ou algoritmo do gradiente estocástico quando $\gamma(k)$ é escrito como:

$$\gamma(k+1) = \frac{1}{k} \quad (5.48)$$

As grandes vantagens deste algoritmo do gradiente são a simplicidade e a rapidez computacional. As desvantagens são a baixa taxa de convergência, sua pobre adaptação em sistemas com parâmetros variantes no tempo e a estimação pode ser polarizada.

Alguns ganhos $\gamma(k+1)$ mais utilizados são:

- Algoritmo de aprendizado seqüencial:

$$\gamma(k) = \frac{\alpha}{\beta + \mathbf{Z}(k+1)^T \mathbf{Z}(k+1)} \quad (\beta > 0; 0 < \alpha < 2) \quad (5.49)$$

- Escalar constante:

$$\gamma(k) = \text{escalar constante} \quad (5.50)$$

- Matriz constante:

$$\gamma(k) = \text{matriz constante simétrica definida positiva} \quad (5.51)$$

Exemplo 5.4 Considere o sistema de primeira ordem descrito no exemplo 5.1. A Fig.5.5 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

Os parâmetros a e b são inicializados com valor nulo e o ganho é:

$$\gamma(k) = 1/k$$

conforme o algoritmo do gradiente estocástico. A estimação de a e b é muito lenta em relação aos mínimos quadrados, e mesmo após 2000 instantes de tempo, os valores estimados ainda estão longe do real.

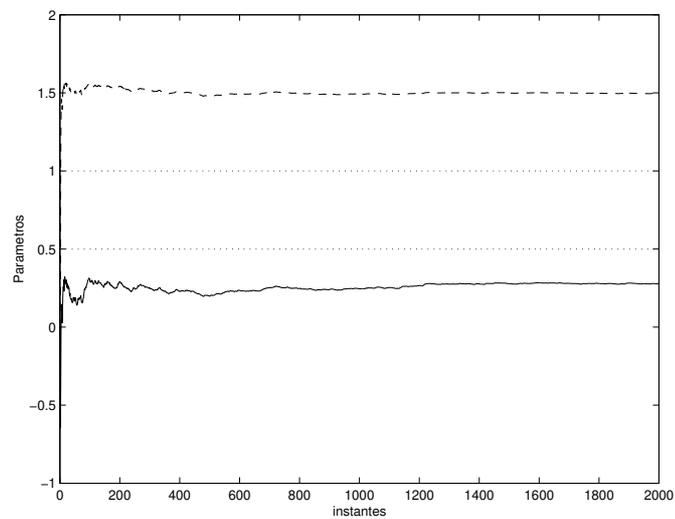


Figura 5.5: Estimação de parâmetros via método do gradiente para o exemplo 4.4

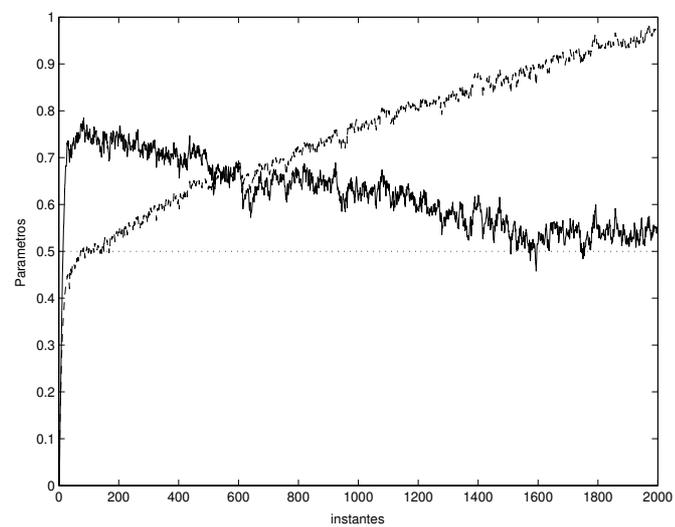


Figura 5.6: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.5

Exemplo 5.5 Considere o sistema descrito no exemplo 5.1. A Fig.5.6 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

Os parâmetros \hat{a} e \hat{b} são inicializados com valor nulo e o ganho é:

$$\gamma(k) = \frac{\alpha}{\beta + \mathbf{Z}(k)^T \mathbf{Z}(k)}$$

onde $\alpha = 0.1$ e $\beta = 1$, conforme o algoritmo do gradiente estocástico.

Diferentemente do exemplo anterior, \hat{a} e \hat{b} conseguem ser estimados em cerca de 2000 instantes de tempo. Há uma variação relativamente grande nos parâmetros em instantes próximos.

Exemplo 5.6 Considerando o sistema do exemplo 5.1. A Fig.5.7 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

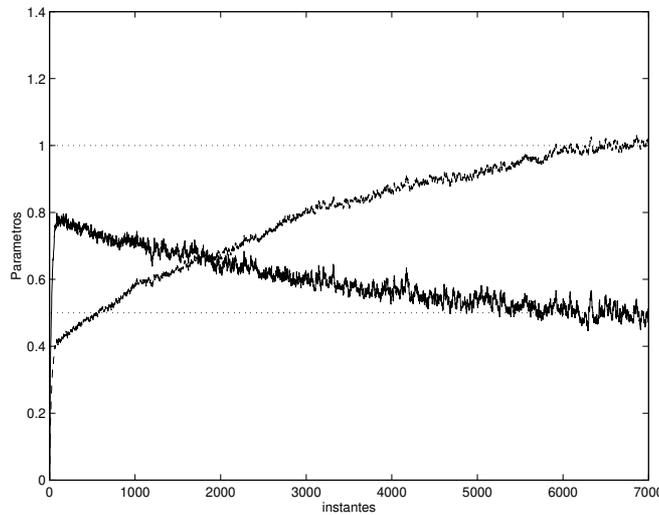


Figura 5.7: Estimação de parâmetros \hat{a} “-” e \hat{b} “-” via método do gradiente para o exemplo 4.6

Os parâmetros \hat{a} e \hat{b} são inicializados com valor nulo e o ganho é constante:

$$\gamma(k) = 0.01.$$

Este exemplo apresentou um resultado um pouco pior que o anterior pois a convergência ocorreu depois de 600 pontos amostrados.

Exemplo 5.7 Considerando o sistema do exemplo 5.2. A Fig.5.8 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.4 (gradiente estocástico).

Para este sistema, os parâmetros se aproximam do valor real em menos de 200 pontos amostrados.

Exemplo 5.8 Considerando o sistema do exemplo 5.2. A Fig.5.9 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

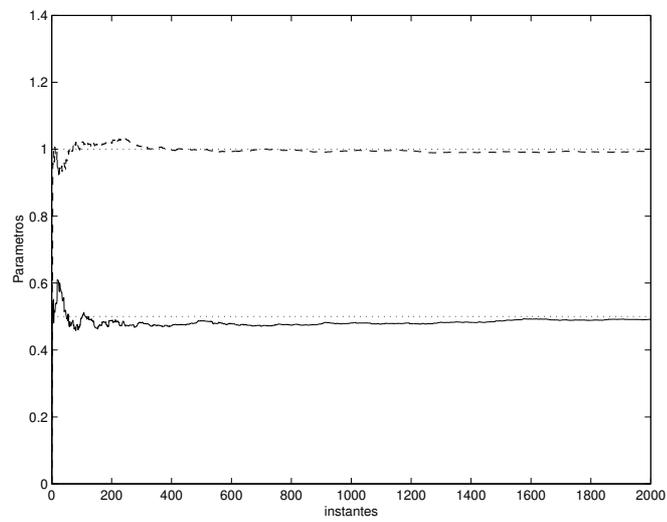


Figura 5.8: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.8

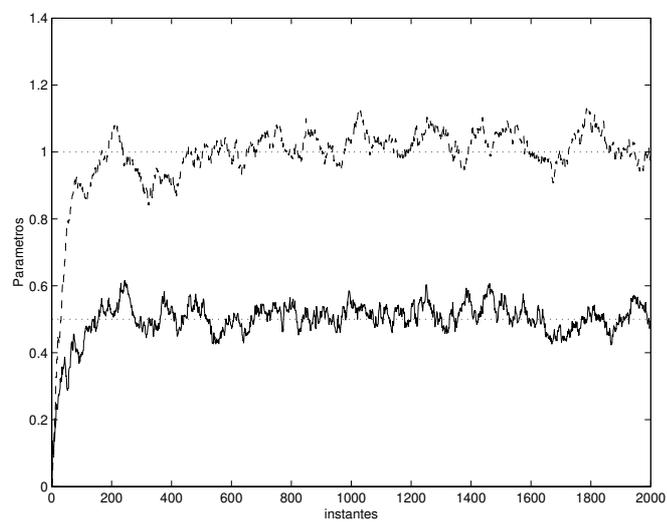


Figura 5.9: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.8

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.5 (aprendizado seqüencial).

Os parâmetros apresentaram valores bem próximos dos reais após 200 iterações porém os valores paramétricos oscilam sem convergirem para os valores.

Exemplo 5.9 Considerando o sistema do exemplo 5.2. A Fig.5.10 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

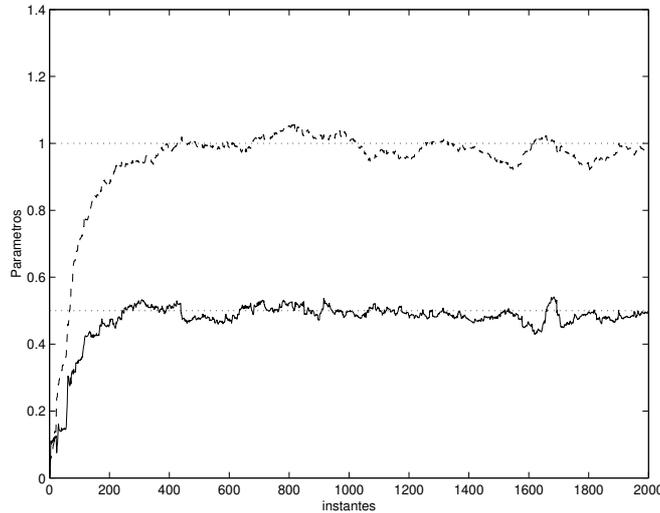


Figura 5.10: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.9

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.6.

Os parâmetros \hat{a} e \hat{b} convergem após 400 instantes de tempo porém não se estabilizam em valores fixos próximos dos reais (há oscilação)

Exemplo 5.10 Considere o sistema de segunda ordem descrito no exemplo 5.3 com os parâmetros a e b iguais a 1.5 e -0.56 . A Fig.5.11 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.4.

A convergência é muito lenta, e depois de 7000 instantes de tempo, os parâmetros ainda não convergiram.

Exemplo 5.11 Considere o sistema de segunda ordem descrito no exemplo anterior. A Fig.5.12 mostra a variação de dois parâmetros na estimação dos parâmetros do método do gradiente.

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.5.

A Fig.5.12 mostra convergência lenta e depois de 2000 pontos os parâmetros oscilam em torno dos valores desejados. Observe que a oscilação é pequena.

Exemplo 5.12 Considere o sistema de segunda ordem descrito no exemplo 5.10. A Fig.5.13 mostra a variação de dois parâmetros na estimação dos parâmetros no método do gradiente.

Os parâmetros \hat{a} e \hat{b} e $\gamma(k)$ são inicializados conforme exemplo 5.6.

Em relação ao exemplo anterior, a estimação é mais lenta porém a oscilação é menor em torno dos valores reais.

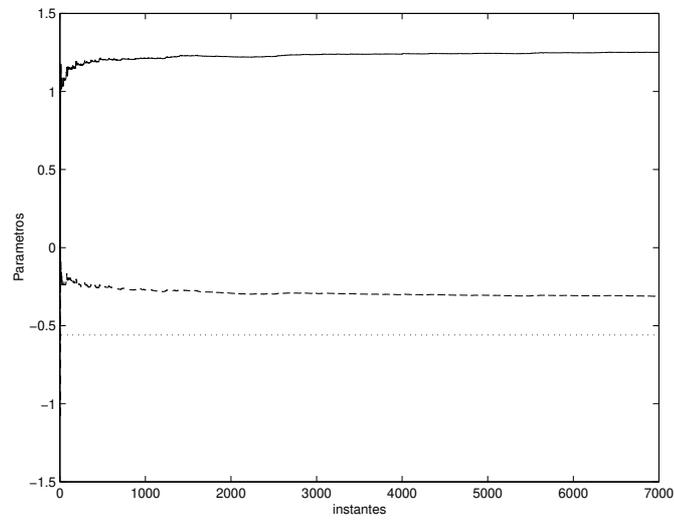


Figura 5.11: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.10

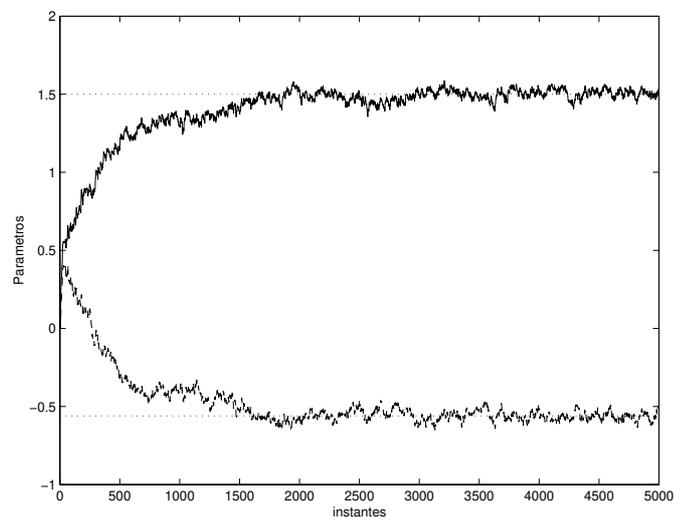


Figura 5.12: Estimação de parâmetros \hat{a} “-” e \hat{b} “- -” via método do gradiente para o exemplo 4.11

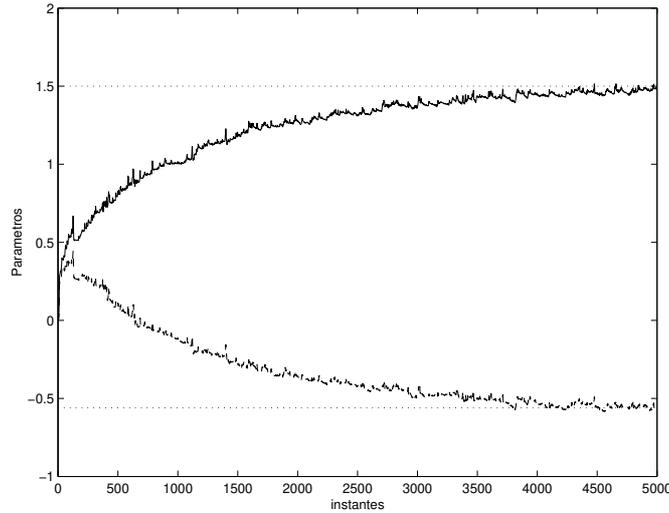


Figura 5.13: Estimação de parâmetros \hat{a} “-” e \hat{b} “-” via método do gradiente para o exemplo 4.12

O método do gradiente apresentou um pior desempenho na estimação dos parâmetros em relação ao algoritmo de mínimos quadrados recursivo. O critério para esta afirmação foi baseado nos resultados experimentais anteriores, pois a convergência para o método do gradiente apresentou-se em média 10 vezes mais lenta.

Métodos utilizando mínimos quadrados combinados com o método do gradiente já são utilizados para a estimação dos parâmetros do consequente de sistemas nebulosos Takagi-Sugeno.

5.4 Método dos Mínimos Quadrados Recursivo Estendido

O algoritmo recursivo dos mínimos quadrados estendido é baseado em sistemas onde a saída é corrompida por um ruído colorido. Uma maneira para solucionar este problema é utilizar o erro de predição calculado no vetor de regressão.

Desta maneira, o vetor de parâmetros é dado por:

$$\mathbf{A} = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}, c_1, \dots, c_{n_c}] \quad (5.52)$$

onde n_a , n_b e n_c são constantes que determinam a ordem do modelo e o vetor de regressão é dado por:

$$\mathbf{Z}_{RELS}(k)^T = [y(k-1), \dots, y(k-n_a), u(k), \dots, u(k-n_b), \epsilon(k-1), \dots, \epsilon(k-n_c)]. \quad (5.53)$$

onde $\epsilon(k)$ é o erro de predição dado por:

$$\epsilon(k) = y(k) - \mathbf{Z}_{RELS}(k)^T \hat{\mathbf{A}}. \quad (5.54)$$

Exemplo 5.13 Considere o sistema descrito por:

$$y(k) = ay(k-1) + e(k) + ce(k-1)$$

onde os parâmetros verdadeiros são $a = 0.7$ e $c = -0.5$. A parcela do sistema $e(k) + ce(k-1)$ representa um ruído colorido agregado.

A Fig.5.14 mostra a variação de dois parâmetros na estimação dos parâmetros do algoritmo dos mínimos quadrados recursivo estendido.

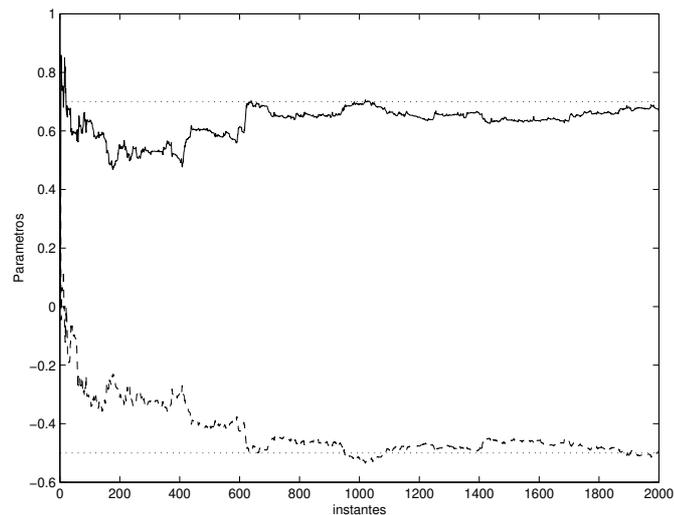


Figura 5.14: Estimação de parâmetros \hat{a} “-” e \hat{c} “- -” via método dos mínimos quadrados estendido para o exemplo 4.13

Os parâmetros a e c são inicializados com valor nulo e a matriz de covariância é $\mathbf{P}(0) = 10\mathbf{I}_2$ onde \mathbf{I}_2 é a matriz identidade com tamanho 2×2 .

O comportamento do erro pode ser visto na Fig.5.15.

Os parâmetros convergem para os valores reais em torno do instante de tempo 2000.

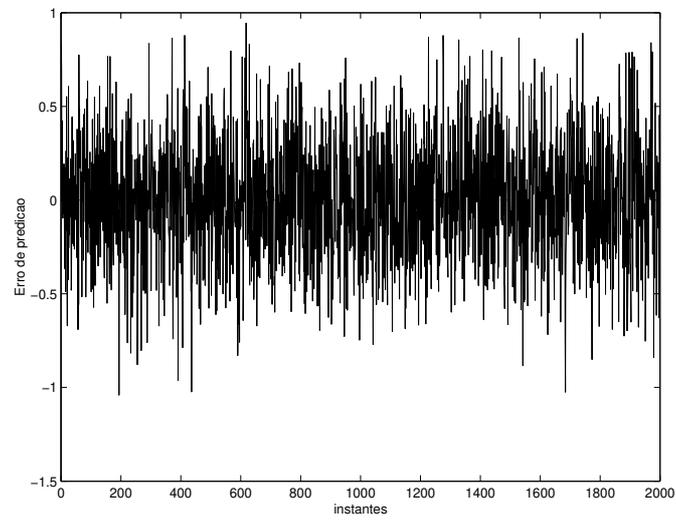


Figura 5.15: Erro do exemplo 4.13

Capítulo 6

ESTUDO DE CASOS

6.1 Introdução

Este capítulo apresenta e discute alguns exemplos com dados reais de sistemas dinâmicos discretos no problema de modelagem de séries temporais utilizando modelo nebuloso Takagi-Sugeno. São utilizadas as técnicas de agrupamento apresentadas no Capítulo 4 e de mínimos quadrados recursivo ponderado no Capítulo 5.

O modelo TS que será utilizado para a modelagem de séries temporais pode ser escrito na forma de regras SE-ENTÃO como:

$$\begin{aligned} R^i : \text{ SE } y(k-1) \text{ é } F_1^i \text{ E } y(k-2) \text{ é } F_2^i \text{ E } \dots \text{ E } y(k-p) \text{ é } F_p^i \\ \text{ ENTÃO } y^i(k) = c^i + a_1^i y(k-1) + a_2^i y(k-2) + \dots + a_p^i y(k-p) \end{aligned} \quad (6.1)$$
$$i = 1, \dots, N$$

onde N é o número de regras, $F_1^i, F_2^i, \dots, F_p^i$ são os conjuntos nebulosos da i -ésima regra das variáveis da premissa determinados via técnica de agrupamento, y^i é a saída da i -ésima regra e c^i, a_1^i, \dots, a_p^i são parâmetros da função linear do conseqüente. O modelo local do conseqüente é representado como o modelo determinístico auto-regressivo de ordem p , $AR(p)$. Este modelo local utiliza apenas dados passados da série para representá-lo.

A abordagem utilizada na estimação dos parâmetros do conseqüente é local, ou seja para cada regra é calculado os parâmetros c^i, a_1^i, \dots, a_p^i separadamente. Outra característica na estimação recursiva utilizando o método dos mínimos quadrado ponderado é que o valor de ponderação w será o grau de ativação da regra, e portanto para cada instante e cada regra teremos valores de ponderação diferentes.

Três critérios, muito comuns em análise experimental de modelagem nebulosa, foram usados com o objetivo de validação do modelo:

- Erro máximo absoluto ($e_{\text{máx}}$):

$$e_{\text{máx}} = \max(|\mathbf{y} - \hat{\mathbf{y}}|) \quad (6.2)$$

onde \mathbf{y} é a saída nominal da série e $\hat{\mathbf{y}}$ é a saída do modelo TS.

- Variância Levada em Conta (VLC) ou *Variance Accounted For (VAF)*:

$$VAF(\%) = \left[1 - \frac{\text{var}(\mathbf{y} - \hat{\mathbf{y}})}{\text{var}(\mathbf{y})} \right] \times 100 \quad (6.3)$$

onde var é a variância;

- Erro Quadrático Médio (EQM) ou *Mean Square Error (MSE)*:

$$MSE = \frac{1}{N} \sum_{k=1}^N (\mathbf{y}(k) - \hat{\mathbf{y}}(k))^2 \quad (6.4)$$

onde N é o número de pontos.

Os valores ideais para $e_{\text{máx}}$ e MSE seriam zero e para o VAF seria cem.

6.2 Braço Robótico Flexível

Esta série temporal de um braço robótico é resultado experimento realizado por Hendrik Van Brussel e Jan Swevers no Laboratório de Produção Industrial e Automação de Katholieke Universiteit Leuven, Swevers [39] e Jankowski [20].

O braço está instalado a motor elétrico e a modelagem da função de transferência é calculada pela reação medida do torque na base da estrutura para a aceleração do braço flexível, aplicando uma entrada com sinal senoidal com frequência variável. Os dados de entrada (torque da estrutura) e saída (aceleração do braço) podem ser vistas nas Fig.6.1 e Fig.6.2, respectivamente.

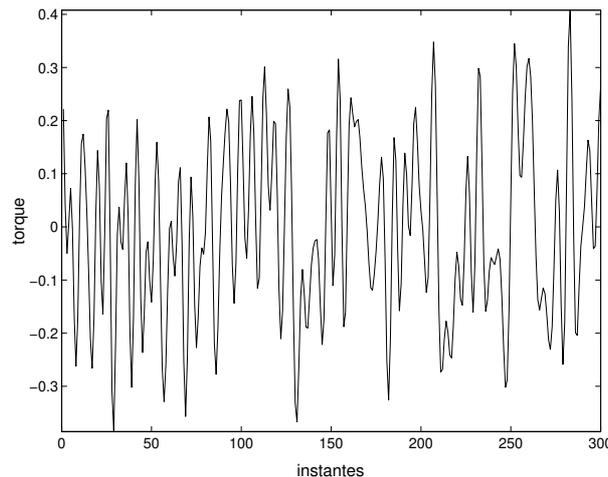


Figura 6.1: Torque da estrutura do braço robótico flexível.

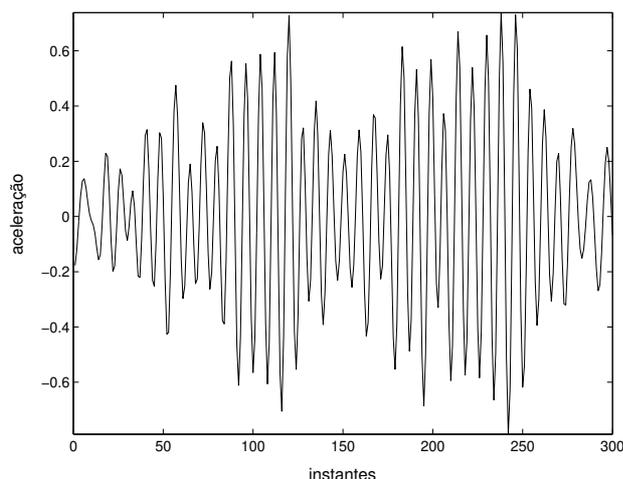


Figura 6.2: Aceleração do braço robótico flexível.

6.2.1 FCM - *Fuzzy C-Means*

Considerando a aceleração do braço robótico como uma série temporal utilizamos o algoritmo de agrupamento *Fuzzy C-Means* para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados recursivo ponderado para ajuste dos parâmetros do conseqüente dos submodelos, onde a ponderação é feita pela matriz de pertinência μ descrita no Capítulo 4. Foram realizadas simulações para determinar a ordem do modelo local do conseqüente. Para isto foi fixado o número de grupos (c , igual ao número de regras) em 4 e variou a ordem do modelo auto-regressivo (p) entre 2 e 8, como mostra a Tab.6.1. Os critérios utilizados para avaliar a melhor ordem são o erro máximo absoluto ($e_{\text{máx}}$), a variância levada em conta (VLC) ou *variance accounted for* (VAF) e o erro quadrático médio (EQM) ou *mean square error* (MSE), já descritos acima.

Tabela 6.1: Critérios de validação do FCM para o braço robótico com c fixo igual a 4 e p variante.

Número de auto-regressores	$e_{\text{máx}}$	MSE	VAF
2	0.5795	0.0714	39.9645
3	0.5708	0.0689	42.0348
4	0.5618	0.0684	42.4599
5	0.5578	0.0684	42.4453
6	0.5563	0.0679	42.8852
7	0.5570	0.0680	42.8196
8	0.5761	0.0679	42.8671

A melhor ordem para o modelo local auto-regressivo do conseqüente, segundo os critérios adotados, foi o de ordem igual a 6. Para todos os critérios esta ordem foi a vencedora. Não será realizada nenhuma análise de estabilidade para este modelo.

Determinada a melhor ordem, pelos critérios $e_{\text{máx}}$, MSE e VAF , é determinado o número ótimo de grupos segundo os mesmos critérios, ver Tab.6.2.

Tabela 6.2: Critérios de validação do FCM para o braço robótico com c variante e p igual a 6.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.5561	0.0679	42.8639
3	0.5560	0.0679	42.8801
4	0.5560	0.0679	42.8844
5	0.5560	0.0679	42.8588

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de agrupamentos c igual a 4 e ordem p igual a 6. Os resultados mostrados nas tabelas acima são os melhores encontrados depois de várias simulações com os parâmetros c e p fixos. O resultado gráfico pode ser visto na Fig.6.3 onde os 150 primeiros instantes de amostragem foram utilizados para a estimação do antecedente (conjuntos nebulosos) e conseqüente (parâmetros do modelo auto-regressivo).

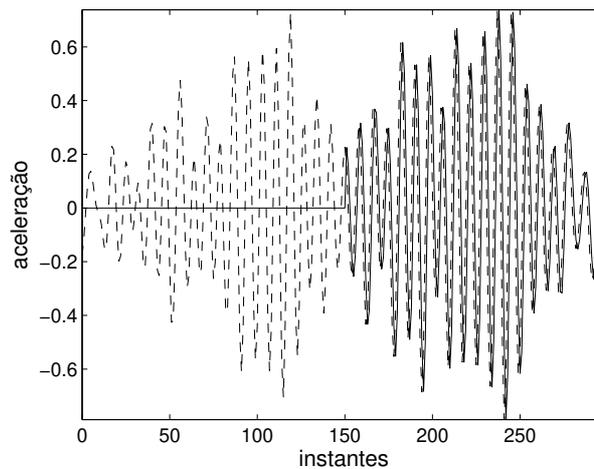


Figura 6.3: Modelagem da série da saída do braço robótico para o FCM com $p = 6$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.

Pela Fig.6.3 podemos observar que o modelo apresentou um comportamento semelhante ao desejado, porém os valores para os critérios $e_{\text{máx}}$, MSE e VAF ficaram abaixo do esperado.

O modelo pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
R^1 : \text{ SE } y(k-1) \text{ é } F_1^1(-0.0968) \text{ E } y(k-2) \text{ é } F_2^1(-0.2600) \text{ E } y(k-3) \text{ é } F_3^1(-0.2677) \\
\text{ E } y(k-4) \text{ é } F_4^1(-0.1148) \text{ E } y(k-5) \text{ é } F_5^1(0.1068) \text{ E } y(k-6) \text{ é } F_6^1(0.2646) \\
\text{ ENTÃO } y^1(k) = 0.0003 + 4.2368y(k-1) - 8.8107y(k-2) + 10.9043y(k-3) \\
- 8.5102y(k-4) + 3.9535y(k-5) - 0.9008y(k-6)
\end{aligned}$$

$$\begin{aligned}
R^2 : \text{ SE } y(k-1) \text{ é } F_1^2(-0.2738) \text{ E } y(k-2) \text{ é } F_2^2(-0.1408) \text{ E } y(k-3) \text{ é } F_3^2(0.0806) \\
\text{ E } y(k-4) \text{ é } F_4^2(0.2583) \text{ E } y(k-5) \text{ é } F_5^2(0.2861) \text{ E } y(k-6) \text{ é } F_6^2(0.1485) \\
\text{ ENTÃO } y^2(k) = 0.0003 + 4.3716y(k-1) - 9.1965y(k-2) + 11.4444y(k-3) \\
- 8.8973y(k-4) + 4.0948y(k-5) - 0.9090y(k-6)
\end{aligned}$$

$$\begin{aligned}
R^3 : \text{ SE } y(k-1) \text{ é } F_1^3(0.1386) \text{ E } y(k-2) \text{ é } F_2^3(0.2996) \text{ E } y(k-3) \text{ é } F_3^3(0.2888) \\
\text{ E } y(k-4) \text{ é } F_4^3(0.1122) \text{ E } y(k-5) \text{ é } F_5^3(-0.1253) \text{ E } y(k-6) \text{ é } F_6^3(-0.2829) \\
\text{ ENTÃO } y^3(k) = 0.0002 + 4.3554y(k-1) - 9.0043y(k-2) + 10.9531y(k-3) \\
- 8.2605y(k-4) + 3.6691y(k-5) - 0.7746y(k-6)
\end{aligned}$$

$$\begin{aligned}
R^4 : \text{ SE } y(k-1) \text{ é } F_1^4(0.2700) \text{ E } y(k-2) \text{ é } F_2^4(0.1281) \text{ E } y(k-3) \text{ é } F_3^4(-0.0886) \\
\text{ E } y(k-4) \text{ é } F_4^4(-0.2496) \text{ E } y(k-5) \text{ é } F_5^4(-0.2595) \text{ E } y(k-6) \text{ é } F_6^4(-0.1155) \\
\text{ ENTÃO } y^4(k) = 0.0002 + 4.2777y(k-1) - 8.8707y(k-2) + 10.8999y(k-3) \\
- 8.3872y(k-4) + 3.8253y(k-5) - 0.8436y(k-6)
\end{aligned}$$

(6.5)

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parenteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

A vantagem deste modelo está na facilidade implementação, porém a solução, conjuntos nebulosos do modelo, pode variar mesmo fixando c e p pois a matriz μ é inicializada aleatoriamente, ver Capítulo 4.

6.2.2 Agrupamento Gustafson-Kessel

Considerando o mesmos dados da aceleração do braço robótico como uma série temporal, descritos anteriormente, utilizamos agora o algoritmo de agrupamento Gustafson-Kessel (GK) para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados recursivo ponderado para ajuste dos parâmetros do conseqüente dos submodelos. Como no caso do FCM foram realizadas simulações para determinar a ordem do modelo local do conseqüente. Para isto foi fixado o número de grupos c em 4 e variou a ordem do modelo auto-regressivo (p) entre 2 e 7, como mostra a Tab.6.3. Os critérios utilizados para avaliar a melhor ordem são os mesmo do FCM, $e_{\text{máx}}$, VAF e MSE .

Como para o FCM a melhor ordem para o modelo local auto-regressivo do conseqüente, segundo os critérios adotados, foi o de ordem igual a 6. Para todos os critérios esta ordem foi a vencedora. Não será realizada nenhuma análise de estabilidade para este modelo ou modelos seguintes.

Determinada a melhor ordem, pelos critérios $e_{\text{máx}}$, MSE e VAF , é determinado o número ótimo de grupos segundo os mesmos critérios, ver Tab.6.4.

Tabela 6.3: Critérios de validação do GK para o braço robótico para c igual a 4 e p variante.

Número de auto-regressores	$e_{\text{máx}}$	MSE	VAF
2	0.5760	0.0716	39.7961
3	0.5711	0.0686	42.0504
4	0.5619	0.0685	42.3749
5	0.5579	0.0684	42.4610
6	0.5563	0.0679	42.8570
7	0.5571	0.0680	42.7987

Tabela 6.4: Critérios de validação do GK para o braço robótico para c variante e p igual a 6.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.5563	0.0679	42.8575
3	0.5563	0.0679	42.8570
4	0.5563	0.0679	42.8570
5	0.5563	0.0679	42.8570

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de agrupamentos c igual a 2 e ordem p igual a 6, porém para c igual a 3, 4 e 5 os resultados foram bem próximos. Voltamos enfatizar que os resultados mostrados nestes exemplos acima são os melhores encontrados depois de várias simulações com os parâmetros c e p fixos. O resultado gráfico pode ser visto na Fig.6.4 onde os 150 primeiros instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

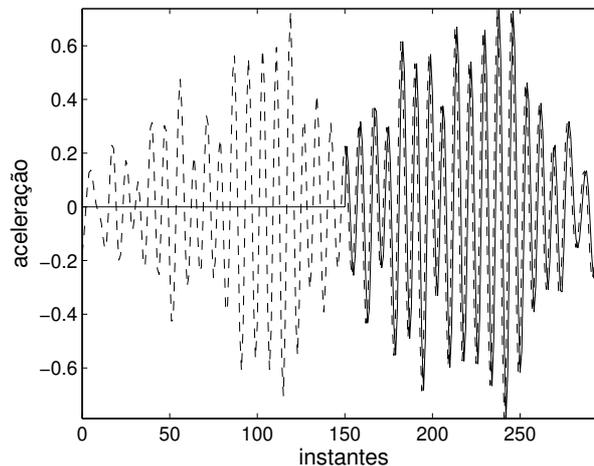


Figura 6.4: Modelagem da série da saída do braço robótico para o Gustafson-Kessel para $p = 6$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.

O modelo nebuloso TS via GK e mínimos quadrado recursivo ponderado pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
R^1 : & \text{ SE } y(k-1) \text{ é } F_1^1(0.0031) \text{ E } y(k-2) \text{ é } F_2^1(0.0026) \text{ E } y(k-3) \text{ é } F_3^1(0.0028) \\
& \text{ E } y(k-4) \text{ é } F_4^1(0.0047) \text{ E } y(k-5) \text{ é } F_5^1(0.0074) \text{ E } y(k-6) \text{ é } F_6^1(0.0093) \\
& \text{ ENTÃO } y^1(k) = 0.0002 + 4.3144y(k-1) - 8.9822y(k-2) + 11.0682y(k-3) \\
& \quad - 8.5293y(k-4) + 3.8941y(k-5) - 0.8592y(k-6) \\
R^2 : & \text{ SE } y(k-1) \text{ é } F_1^2(-0.0005) \text{ E } y(k-2) \text{ é } F_2^2(0.0016) \text{ E } y(k-3) \text{ é } F_3^2(0.0046) \\
& \text{ E } y(k-4) \text{ é } F_4^2(0.0060) \text{ E } y(k-5) \text{ é } F_5^2(0.0042) \text{ E } y(k-6) \text{ é } F_6^2(-0.0002) \\
& \text{ ENTÃO } y^2(k) = 0.0002 + 4.3115y(k-1) - 8.9702y(k-2) + 11.0480y(k-3) \\
& \quad - 8.5115y(k-4) + 3.8861y(k-5) - 0.8579y(k-6)
\end{aligned} \tag{6.6}$$

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parênteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

A vantagem deste modelo está na forma das funções de pertinência criadas pois apresentam um formato elipsoidal, ou seja, depende da necessidade de dispersão do grupo nebuloso em cada dimensão. Esta diferenciação em relação ao FCM ocorre no cálculo da distância onde se utiliza a matriz de covariância. Porém a solução pode variar mesmo fixando c e p dependendo como a matriz μ é inicializada, semelhante ao FCM. Outro problema pode ocorrer quando calculamos a distância pois a matriz de covariância não pode ser singular, ou seja, não pode ter determinante nulo.

Como para o FCM os valores para os critérios $e_{\text{máx}}$, MSE e VAF ficaram abaixo do esperado.

6.2.3 Agrupamento Gath-Geva

Ainda considerando a saída do sistema do braço robótico utilizamos o algoritmo de agrupamento Gath-Geva para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados para ajuste dos parâmetros do conseqüente dos submodelos AR(6). Portanto, consideramos que para este caso a melhor ordem é 6, como no caso do algoritmos FCM e GK.

O número ótimo de grupos segundo os mesmos critérios $e_{\text{máx}}$, MSE e VAF podem ser vistos Tab.6.5.

Tabela 6.5: Critérios de validação do GG para o braço robótico com p igual a 6 e c variante.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.5579	0.0681	18.2360
3	0.5560	0.0678	18.4816
4	0.5580	0.0680	18.3125
5	0.5575	0.0678	18.4778

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de agrupamentos c igual a 3 e ordem p igual a 6, porém para c igual a 5 o resultado foi bem próximo. O resultado gráfico pode ser visto na Fig.6.5 onde os 150 primeiros instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

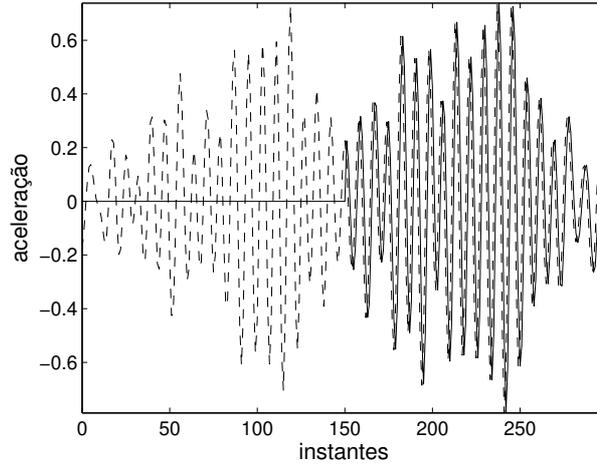


Figura 6.5: Modelagem da série da saída do braço robótico para o Gath-Geva para $p = 6$ e $c = 3$ onde “- -” é a série desejada e “-” é o modelo estimado.

O modelo nebuloso TS via GG e mínimos quadrado recursivo ponderado pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
 R^1 : \text{SE } y(k-1) \text{ é } F_1^1(-0.0268) \text{ E } y(k-2) \text{ é } F_2^1(0.0383) \text{ E } y(k-3) \text{ é } F_3^1(0.0893) \\
 \text{E } y(k-4) \text{ é } F_4^1(0.1060) \text{ E } y(k-5) \text{ é } F_5^1(0.0852) \text{ E } y(k-6) \text{ é } F_6^1(0.0414) \\
 \text{ENTÃO } y^1(k) = 0.0022 + 4.6220y(k-1) - 10.0906y(k-2) + 12.8789y(k-3) \\
 - 10.1642y(k-4) + 4.6983y(k-5) - 1.0369y(k-6)
 \end{aligned}$$

$$\begin{aligned}
 R^2 : \text{SE } y(k-1) \text{ é } F_1^2(-0.0168) \text{ E } y(k-2) \text{ é } F_2^2(-0.0387) \text{ E } y(k-3) \text{ é } F_3^2(-0.0392) \\
 \text{E } y(k-4) \text{ é } F_4^2(-0.0285) \text{ E } y(k-5) \text{ é } F_5^2(-0.0206) \text{ E } y(k-6) \text{ é } F_6^2(-0.0232) \\
 \text{ENTÃO } y^2(k) = -0.0002 + 4.2489y(k-1) - 8.8018y(k-2) + 10.8461y(k-3) \\
 - 8.4270y(k-4) + 3.9042y(k-5) - 0.8891y(k-6)
 \end{aligned}$$

$$\begin{aligned}
 R^3 : \text{SE } y(k-1) \text{ é } F_1^3(0.0185) \text{ E } y(k-2) \text{ é } F_2^3(0.0094) \text{ E } y(k-3) \text{ é } F_3^3(-0.0015) \\
 \text{E } y(k-4) \text{ é } F_4^3(0.0002) \text{ E } y(k-5) \text{ é } F_5^3(0.0184) \text{ E } y(k-6) \text{ é } F_6^3(0.0440) \\
 \text{ENTÃO } y^3(k) = -0.0000 + 4.1806y(k-1) - 8.4520y(k-2) + 10.1025y(k-3) \\
 - 7.5567y(k-4) + 3.3542y(k-5) - 0.7252y(k-6)
 \end{aligned}$$

(6.7)

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parênteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

A vantagens e desvantagens do GG são semelhantes do algoritmo GK. Porém os resultados mostrados para esta série foram piores.

Como para o FCM e GK os valores para os critérios $e_{\text{máx}}$, MSE e VAF ficaram abaixo do esperado.

6.2.4 FCRM - *Fuzzy c-Regression Model*

Como nos casos anteriores, FCM, GK e GG, considerando a saída do sistema do braço robótico utilizamos o algoritmo de agrupamento FCRM para ajustar os conjuntos nebulosos do antecedente e conseqüente. Uma grande vantagem deste método é que o algoritmo inclui o método dos mínimos quadrados para a estimação do conseqüente. O modelo do conseqüente será do tipo AR(6) e para determinar o melhor números de regras foram feitas várias simulações para vários números de regras. Os resultados dos critérios de validação obtidos para a modelagem da série podem ser vistos na Tab.6.6.

Tabela 6.6: Critérios de validação do FCRM para o braço robótico com p igual a 6 e c variante.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.0055	4.3415×10^{-6}	99.9964
3	0.0055	4.3415×10^{-6}	99.9964
4	0.0055	4.3415×10^{-6}	99.9964

O algoritmo apresentou resultados semelhantes para todos as simulações pois os conjuntos nebulosos e parâmetros do conseqüente são iguais. Portanto o sistema nebuloso pode ser substituído pelo modelo local do conseqüente. O resultado gráfico para dois grupos pode ser visto na Fig.6.6 onde os 150 primeiros instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

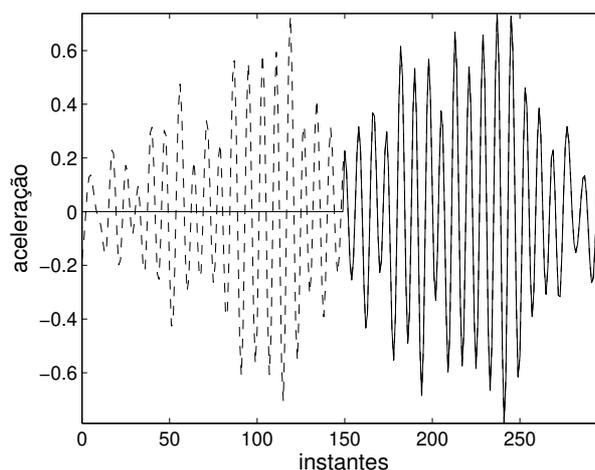


Figura 6.6: Modelagem da série da saída do braço robótico para o FCRM para $p = 6$ e $c = 2$ onde “-” é a série desejada e “-” é o modelo estimado.

O modelo nebuloso TS via FCRM para $c = 2$ pode ser descrito na forma de regras SE-

ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
 R^1 : & \text{ SE } y(k-1) \text{ é } F_1^1(0.0020) \text{ E } y(k-2) \text{ é } F_2^1(0.0022) \text{ E } y(k-3) \text{ é } F_3^1(0.0018) \\
 & \text{ E } y(k-4) \text{ é } F_4^1(0.0009) \text{ E } y(k-5) \text{ é } F_5^1(-0.0002) \text{ E } y(k-6) \text{ é } F_6^1(-0.0009) \\
 \text{ ENTÃO } & y^1(k) = 0.0002 + 4.3130y(k-1) - 8.9765y(k-2) + 11.0587y(k-3) \\
 & \quad - 8.5211y(k-4) + 3.8905y(k-5) - 0.8586y(k-6)
 \end{aligned} \tag{6.8}$$

$$\begin{aligned}
 R^2 : & \text{ SE } y(k-1) \text{ é } F_1^2(0.0020) \text{ E } y(k-2) \text{ é } F_2^2(0.0022) \text{ E } y(k-3) \text{ é } F_3^2(0.0018) \\
 & \text{ E } y(k-4) \text{ é } F_4^2(0.0009) \text{ E } y(k-5) \text{ é } F_5^2(-0.0002) \text{ E } y(k-6) \text{ é } F_6^2(-0.0009) \\
 \text{ ENTÃO } & y^2(k) = 0.0002 + 4.3130y(k-1) - 8.9765y(k-2) + 11.0587y(k-3) \\
 & \quad - 8.5211y(k-4) + 3.8905y(k-5) - 0.8586y(k-6)
 \end{aligned}$$

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parênteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

A vantagem deste modelo está no cálculo da distância que minimiza a função custo do erro quadrático para cada regra, onde o erro de cada regra é a saída desejada subtraída da saída do modelo local. Deste modo, se a série apresentar um comportamento linear, os modelos locais para todas as regras serão semelhantes. Outra vantagem é que este algoritmo apresenta os mínimos quadrados incluso determinando os parâmetros do conseqüente também.

Como para o FCRM os valores para os critérios $e_{\text{máx}}$, MSE e VAF ficaram acima do esperado e em relação aos algoritmos anteriores, o FCRM apresentou um rendimento bem melhor para todos critérios.

6.3 Processo Neutralizador de pH

Os dados da simulação deste processo neutralizador de pH em um tanque com mexedor e volume constante foi realizado por Jairo Espinosa da Katholieke [27].

Como descrição do processo sabemos que o tanque tem 1100 litros de volume, a solução ácida (HAC) tem concentração 0.0032 Mol/l e a solução básica (NaOH) tem concentração 0.05 Mol/l. O tempo de amostragem é de 10 segundos. As variáveis de entrada podem ser vistas na Fig.6.7 e Fig.6.8. A saída é mostrada na Fig.6.9

Os dados de entrada do processo são a vazão da solução ácida em litros e a vazão da solução básica em litros. A saída é o pH da solução no tanque. Este processo é altamente não-linear.

6.3.1 FCM - *Fuzzy C-Means*

Considerando a saída do processo neutralizador de pH, pH da solução, como uma série temporal utilizamos o algoritmo de agrupamento *Fuzzy C-Means* (FCM) para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados recursivo ponderado para ajuste dos parâmetros do conseqüente dos submodelos. Como no caso anterior do braço robótico, foram realizadas simulações para determinar a ordem do modelo local do conseqüente. Para isto foi fixado o número de grupos c em 4 e variou a ordem do modelo auto-regressivo p entre 2 e 5, como mostra a Tab.6.7. Os critérios utilizados para avaliar a melhor ordem são os mesmos já descritos acima.

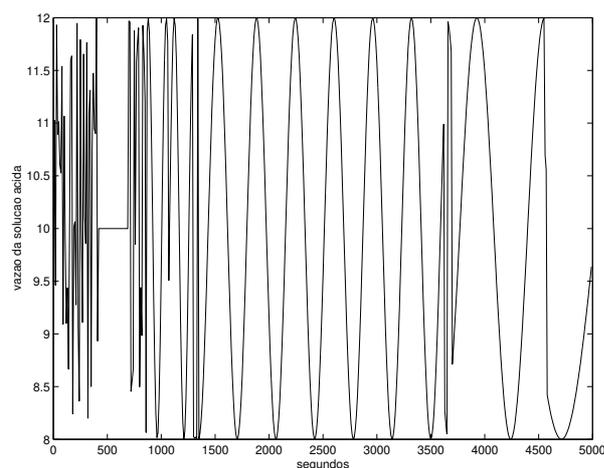


Figura 6.7: Vazao da solução ácida

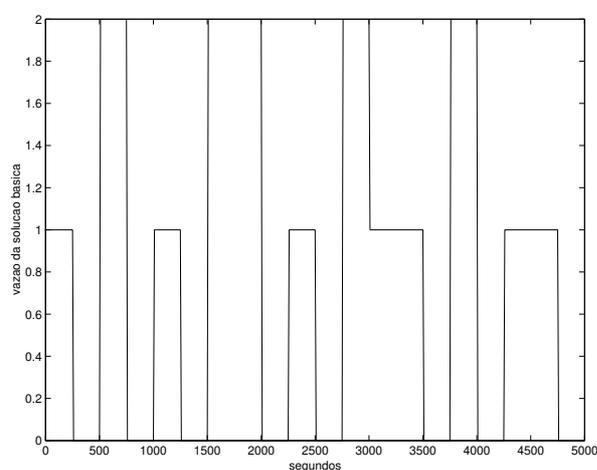


Figura 6.8: Vazao da solução básica

A melhor ordem para o modelo local auto-regressivo do conseqüente, segundo os critérios adotados, foi o de ordem igual a 2. Para todos os critérios esta ordem foi a vencedora. Como no caso do braço robótico não será realizada nenhuma análise de estabilidade para os modelos locais.

Determinada a melhor ordem, pelos critérios $e_{\text{máx}}$, MSE e VAF , é determinado o número ótimo de grupos segundo os mesmos critérios, ver Tab.6.8.

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de agrupamentos igual a 4. O resultado gráfico pode ser visto na Fig.6.10 onde 250 instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

Como no caso do braço robótico para o FCM, GK e GG, a Fig.6.10 apresentou um comportamento semelhante ao desejado, porém os valores para os critérios $e_{\text{máx}}$, MSE e VAF ficaram abaixo do esperado.

O modelo TS via FCM para $p = 2$ e $c = 4$ (número de regras) pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

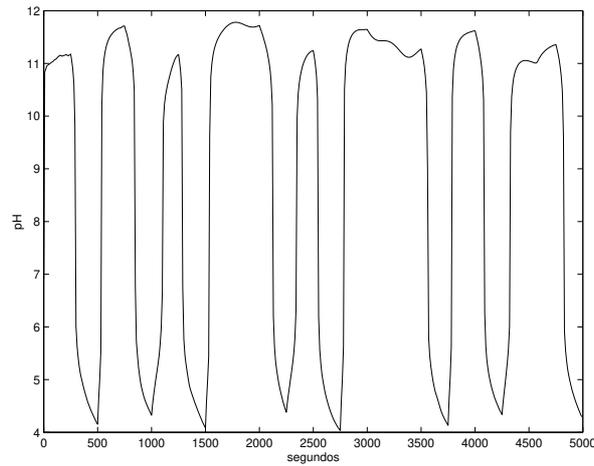
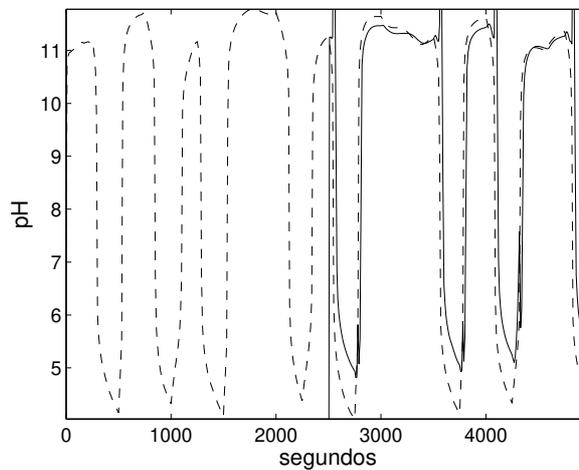


Figura 6.9: pH da solução no tanque

Tabela 6.7: Critérios de validação do FCM para o neutralizador de pH com p variante e c igual a 4.

Número de auto-regressores	$e_{\text{máx}}$	MSE	VAF
2	7.3978	3.2866	69.3163
3	8.4101	4.5971	56.0753
4	8.5874	6.2970	41.4781
5	10.2531	10.0428	10.2917

Figura 6.10: Modelagem da série da saída do neutralizador de pH para o FCM com $p = 2$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.

$$R^1 : \text{SE } y(k-1) \text{ é } F_1^1(4.7497) \text{ E } y(k-2) \text{ é } F_2^1(4.7568) \\ \text{ENTÃO } y^1(k) = 0.5157 + 0.5903y(k-1) + 0.5606y(k-2)$$

$$R^2 : \text{SE } y(k-1) \text{ é } F_1^2(4.9431) \text{ E } y(k-2) \text{ é } F_2^2(4.9531) \\ \text{ENTÃO } y^2(k) = 0.5393 + 0.6200y(k-1) + 0.5878y(k-2)$$

$$R^3 : \text{SE } y(k-1) \text{ é } F_1^3(7.7703) \text{ E } y(k-2) \text{ é } F_2^3(7.2593) \\ \text{ENTÃO } y^3(k) = 0.2974 + 0.6233y(k-1) + 0.4932y(k-2)$$

$$R^4 : \text{SE } y(k-1) \text{ é } F_1^4(11.2866) \text{ E } y(k-2) \text{ é } F_2^4(11.2784) \\ \text{ENTÃO } y^4(k) = 0.3913 + 0.5062y(k-1) + 0.4603y(k-2)$$

(6.9)

Tabela 6.8: Critérios de validação do FCM para o neutralizador de pH com p igual a 2 e c variante.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	9.0071	5.1062	54.7317
3	8.6833	3.5425	65.2325
4	7.3978	3.2866	69.3163
5	7.7789	3.2010	68.8894

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parênteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

Para este caso do neutralizador de pH o modelo apresentou as mesma vantagem de facilidade implementação, e o mesmo problema de inicialização da matriz μ .

6.3.2 FCRM - *Fuzzy c-Regression Model*

No mesmo problema de modelagem da série temporal da saída do neutralizador de pH utilizamos o algoritmo de agrupamento *Fuzzy c-Regression Model* para ajustar os conjuntos nebulosos do antecedente e os parâmetros do conseqüente dos submodelos AR(2). Portanto, como para o FCM, estamos supondo que a melhor ordem é igual a 2 (p). Foram feitas simulações para vários números de regras e os resultados dos critérios de validação obtidos na modelagem podem ser vistos na Tab.6.9.

Tabela 6.9: Critérios de validação do FCRM para o neutralizador de pH

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	5.0006	0.7368	92.3474
3	4.9628	0.7587	92.2024
4	4.9271	0.7805	92.1258
5	4.8938	0.8043	92.0762

O melhor resultado nas simulações para MSE e VAF foi encontrada para o número de grupos igual a 2. O resultado gráfico pode ser visto na Fig.6.11 onde 250 instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

Como no caso do braço robótico, este caso do neutralizador de pH para o FCRM também apresentou regras AR(2) determinístico como mostra abaixo:

$$y(k) = 0.2558 + 0.5479y(k-1) + 0.4313y(k-2) \quad (6.10)$$

Este modelo apresentou resultados superiores ao modelo FCM, porém o tempo computacional é maior.

6.3.3 ECM - *Evolving Clustering Method*

Ainda considerando a saída do processo neutralizador de pH como uma série temporal utilizamos o algoritmo de agrupamento *Evolving Clustering Method* para ajustar os conjuntos

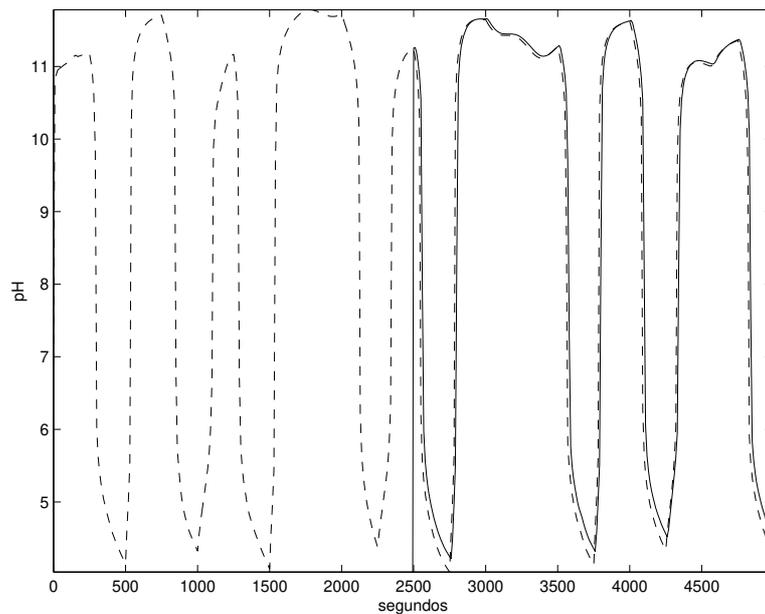


Figura 6.11: Modelagem da série da saída do neutralizador de pH para o FCRM com $p = 2$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.

nebulosos do antecedente e o método dos mínimos quadrados para ajuste dos parâmetros do conseqüente dos submodelos AR(2). Para o ECM não é possível fixar um número de agrupamentos mas é possível determinar uma distância máxima de projeto. foi determinada por tentativa e erro para que o número de conjuntos nebulosos fosse próximo dos exemplos acima. A relação entre o tamanho da distância de projeto e o número de grupos c não é linear. Este é um grande problema deste algoritmo. Um resultado satisfatório, segundo os critérios de validação, obtido na modelagem da série pode ser visto na Tab.6.10.

Tabela 6.10: Critérios de validação do ECM para o neutralizador de pH

Distância máxima	Número de grupos	$e_{máx}$	MSE	VAF
5.5	4	4.9721	1.6713	83.5456

Podemo ver na Fig.6.12 o resultado conseguido por tentativa e erro onde $c = 4$.

O modelo TS via ECM com $c = 4$ (número de regras) pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

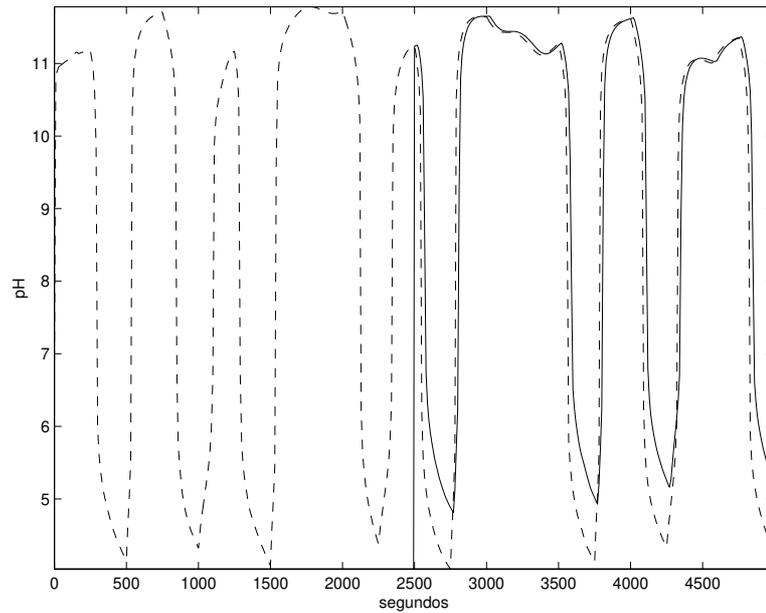


Figura 6.12: Modelagem da série da saída do neutralizador de pH para o ECM com a distância máxima igual a 5.5 onde “- -” é a série desejada e “-” é o modelo estimado.

$$\begin{aligned}
 R^1 : \text{ SE } y(k-1) \text{ é } F_1^1(8.0580) \text{ E } y(k-2) \text{ é } F_2^1(5.0868) \\
 \text{ ENTÃO } \tilde{y}^1(k) &= 0.1693 + 0.5738y(k-1) + 0.4124y(k-2) \\
 R^2 : \text{ SE } y(k-1) \text{ é } F_1^2(8.0580) \text{ E } y(k-2) \text{ é } F_2^2(4.9963) \\
 \text{ ENTÃO } \tilde{y}^2(k) &= 0.4312 + 0.6242y(k-1) + 0.5472y(k-2) \\
 R^3 : \text{ SE } y(k-1) \text{ é } F_1^3(7.9736) \text{ E } y(k-2) \text{ é } F_2^3(5.0868) \\
 \text{ ENTÃO } \tilde{y}^3(k) &= 0.5446 + 0.5448y(k-1) + 0.5448y(k-2) \\
 R^4 : \text{ SE } y(k-1) \text{ é } F_1^4(7.9736) \text{ E } y(k-2) \text{ é } F_2^4(4.9963) \\
 \text{ ENTÃO } \tilde{y}^4(k) &= 0.5425 + 0.5426y(k-1) + 0.5426y(k-2)
 \end{aligned} \tag{6.11}$$

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parenteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

A grande vantagem do algoritmo ECM é que ele pode ser implementado no modo *online* de identificação do modelo. Além disso o custo computacional é menor do que qualquer outro algoritmo implementado até aqui. O resultado obtido ficou um pouco abaixo do esperado.

6.4 Função Senoidal

Foram realizadas 201 amostragens de uma função senoidal dada pela seguinte equação:

$$y = \sin(x) \tag{6.12}$$

onde x varia de 0 à 20. A Fig.6.13 ilustra esta função.

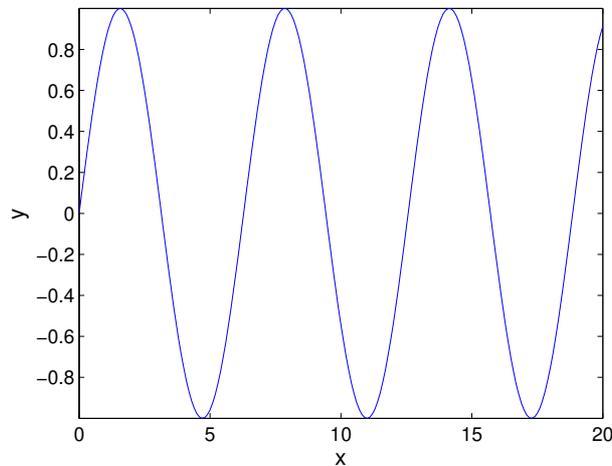


Figura 6.13: Função seno de x .

6.4.1 FCM - *Fuzzy C-Means*

Considerando a função seno, Fig.6.13, como uma série temporal utilizamos o algoritmo de agrupamento *Fuzzy C-Means* para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados para ajuste dos parâmetros dos submodelos do conseqüente. Como no estudos de casos anteriores, foram realizadas simulações para determinar a ordem do modelo local do conseqüente. Para isto foi fixado o número de grupos c em 4 e variou a ordem do modelo auto-regressivo p entre 2 e 5, como mostra a Tab.6.11. Os critérios utilizados para avaliar a melhor ordem são os mesmos já descritos acima.

Tabela 6.11: Critérios de validação do FCM para a função senoidal com p variante e c igual a 4.

Número de auto-regressores	$e_{\text{máx}}$	MSE	VAF
2	0.1819	0.0118	97.6862
3	0.1058	0.0033	99.3200
4	0.1155	0.0017	99.6602
5	0.2011	0.0066	98.9219

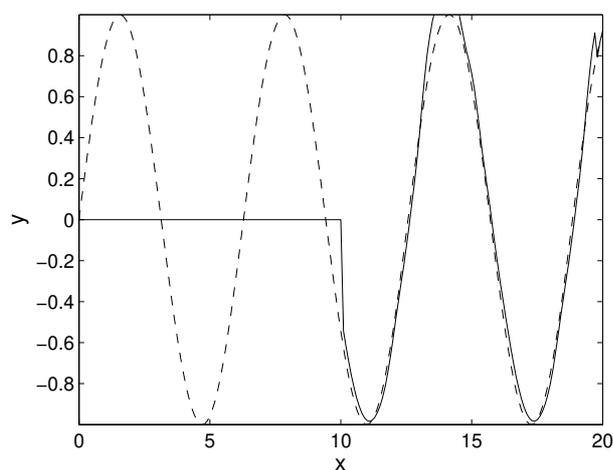
A melhor ordem para o modelo local auto-regressivo do conseqüente, segundo os critérios adotados, foi o de ordem igual a 4. Para dois critérios esta ordem foi a vencedora (MSE e VAF).

Determinada a melhor ordem, pelos critérios $e_{\text{máx}}$, MSE e VAF , é determinado o número ótimo de grupos segundo os mesmos critérios, ver Tab.6.12.

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de grupos igual a 4. O resultado gráfico pode ser visto na Fig.6.14 onde 100 instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

Tabela 6.12: Critérios de validação do FCM para a função senoidal com p igual a 4 e c variante.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.0791	0.0028	99.4291
3	0.0958	0.0033	99.3352
4	0.1155	0.0017	99.6602
5	0.2807	0.0135	97.4392

Figura 6.14: Modelagem da série da função seno para o FCM com $p = 4$ e $c = 4$ onde “- -” é a série desejada e “-” é o modelo estimado.

O modelo pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
R^1 : & \text{ SE } y(k-1) \text{ é } F_1^1(-0.8114) \text{ E } y(k-2) \text{ é } F_2^1(-0.8197) \\
& \text{ E } y(k-3) \text{ é } F_3^1(-0.8198) \text{ E } y(k-4) \text{ é } F_4^1(-0.8117) \\
\text{ENTÃO } & y^1(k) = -0.0767 + 1.0101y(k-1) + 0.4856y(k-2) \\
& -0.0438y(k-3) - 0.5727y(k-4) \\
R^2 : & \text{ SE } y(k-1) \text{ é } F_1^2(-0.0534) \text{ E } y(k-2) \text{ é } F_2^2(-0.0283) \\
& \text{ E } y(k-3) \text{ é } F_3^2(-0.0030) \text{ E } y(k-4) \text{ é } F_4^2(0.0224) \\
\text{ENTÃO } & y^2(k) = -0.0014 + 0.9401y(k-1) + 0.4570y(k-2) \\
& -0.0307y(k-3) - 0.5181y(k-4) \\
R^3 : & \text{ SE } y(k-1) \text{ é } F_1^3(0.9111) \text{ E } y(k-2) \text{ é } F_2^3(0.9202) \\
& \text{ E } y(k-3) \text{ é } F_3^3(0.9201) \text{ E } y(k-4) \text{ é } F_4^3(0.9109) \\
\text{ENTÃO } & y^3(k) = -0.0290 + 1.1449y(k-1) + 0.5536y(k-2) \\
& -0.0433y(k-3) - 0.6397y(k-4) \\
R^4 : & \text{ SE } y(k-1) \text{ é } F_1^4(0.5380) \text{ E } y(k-2) \text{ é } F_2^4(0.5427) \\
& \text{ E } y(k-3) \text{ é } F_3^4(0.5421) \text{ E } y(k-4) \text{ é } F_4^4(0.5360) \\
\text{ENTÃO } & y^4(k) = -0.0815 + 1.5168y(k-1) + 0.7074y(k-2) \\
& -0.1091y(k-3) - 0.9245y(k-4)
\end{aligned} \tag{6.13}$$

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parênteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

6.4.2 Agrupamento Gustafson-Kessel

Considerando a função seno dada pela Fig.6.13 como uma série temporal utilizamos o algoritmo de agrupamento Agrupamento Gustafson-Kessel para ajustar os conjuntos nebulosos do antecedente e o método dos mínimos quadrados para ajuste dos parâmetros do conseqüente dos submodelos de ordem 2. A ordem é diferente do exemplo para o FCM pois o algoritmo apresentou problemas na singularidade da matriz de covariância. Foram realizadas simulações para vários números de regras. Os resultados dos critérios de validação obtidos na modelagem da série podem ser vistos na Tab.6.13.

Tabela 6.13: Critérios de validação do GK para a função senoidal com $p = 2$ e c variante.

Número de grupos	$e_{\text{máx}}$	MSE	VAF
2	0.1939	0.0132	97.7187
3	0.2498	0.0249	94.9198
4	0.4996	0.0657	90.7693
5	0.9445	0.221	65.3652

O melhor resultado nas simulações para esta série temporal foi encontrada para o número de grupos igual a 2. O resultado gráfico pode ser visto na Fig.6.15 onde 100 primeiros instantes de amostragem foram utilizados para a estimação do antecedente e conseqüente.

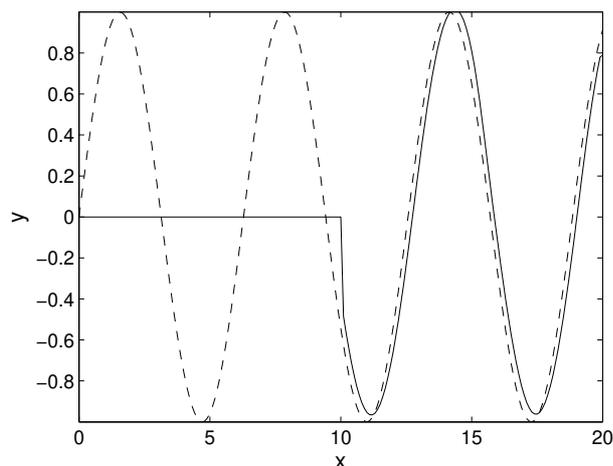


Figura 6.15: Modelagem da série da saída do neutralizador de pH para o GK com $p = 2$ e $c = 2$ onde “- -” é a série desejada e “-” é o modelo estimado.

O modelo TS via FCM para $p = 2$ e $c = 2$ (número de regras) pode ser descrito na forma de regras SE-ENTÃO como pode ser visto abaixo:

$$\begin{aligned}
 R^1 : & \text{ SE } y(k-1) \text{ é } F_1^1(0.3612) \text{ E } y(k-2) \text{ é } F_2^1(0.4463) \\
 & \text{ ENTÃO } \tilde{y}^1(k) = -0.1439 + 0.5840y(k-1) + 0.4291y(k-2) \\
 R^2 : & \text{ SE } y(k-1) \text{ é } F_1^2(0.1321) \text{ E } y(k-2) \text{ é } F_2^2(0.1035) \\
 & \text{ ENTÃO } \tilde{y}^2(k) = 0.0143 + 1.7199y(k-1) - 0.7214y(k-2)
 \end{aligned} \tag{6.14}$$

onde $F_i^j(\cdot)$ é o conjunto nebuloso no espaço de discurso de $y(k-i)$ na j -ésima regra e o número entre os parenteses é o centro do conjunto nebuloso e y^j é a saída do modelo da j -ésima regra.

Os resultados para o FCM e GK foram satisfatórios. O GK apresentou um desempenho inferior, prejudicado pela redução na ordem do modelo por causa da singularidade da matriz de covariância.

Capítulo 7

CONCLUSÕES E COMENTÁRIOS

No capítulo 6, foi apresentado o modelo nebuloso Takagi-Sugeno para a modelagem de séries temporais. Foi realizada uma análise comparativa dos métodos de agrupamento para três estudos de casos nos quais estes métodos influenciam no desempenho do modelo e na identificação dos parâmetros do conseqüente. Os métodos de agrupamentos utilizados nesta dissertação foram: *Fuzzy C-Means* (FCM), Gustafson-Kessel, Gath-Geva, *Fuzzy c-Regression Model* (FCRM) e *Evolving Clustering Method* (ECM).

Os algoritmos foram testados em três casos diferentes: braço robótico flexível, neutralizador de pH e função senoidal. Nestes casos foram utilizadas como série temporal apenas a saída dos sistema, ou seja, foi suposto que não conhecíamos a entrada.

Os resultados obtidos que avaliam a modelagem são: gráfico da saída modelada sobre a saída real, o erro máximo absoluto ($e_{\text{máx}}$), MSE (*Mean Square Error*) e VAF (*Variance Accounted For*).

Sobre os algoritmos de agrupamento foram tiradas as seguintes conclusões:

- O FCM tem a vantagem de facilidade implementação, porém a solução, conjuntos nebulosos do modelo, pode variar mesmo fixando c (número de grupos) e p (ordem do modelo do conseqüente, no caso o $AR(p)$) pois a matriz μ é inicializada aleatoriamente a cada nova iteração, ver Capítulo 4.
- Os algoritmos GK e GG apresentam vantagem na forma das funções de pertinência criadas pois apresentam um formato elipsoidal, ou seja, mudam o tamanho da dispersão do grupo nebuloso em cada dimensão de acordo com a necessidade. Esta diferenciação em relação ao FCM ocorre no cálculo da distância onde se utiliza a matriz de covariância. Porém a solução pode variar mesmo fixando c e p dependendo como a matriz μ é inicializada, semelhante ao FCM. Outro problema pode ocorrer quando calculamos a distância pois a matriz de covariância não pode ser singular, ou seja, não pode ter determinante nulo. Este problema pode ser visto para o algoritmo GK no Capítulo 6 no caso da função senoidal.
- O agrupamento FCRM apresentou os resultados mais satisfatórios pois o algoritmo calcula a distância que minimiza a função custo do erro quadrático para cada regra, onde o erro de cada regra é a saída desejada subtraída da saída do modelo local. Deste modo, se

a série apresentar um comportamento linear, os modelos locais para todas as regras serão semelhantes. Outra vantagem é que este algoritmo apresenta os mínimos quadrados incluso determinando os parâmetros do conseqüente também. O único problema encontrado foi o custo computacional.

- O agrupamento ECM tem como grande vantagem que ele pode ser implementado no modo *online* de identificação do modelo. Além disso o custo computacional é menor do que qualquer outro algoritmo implementado até aqui. Para o ECM não é possível fixar um número de agrupamentos mas é possível determinar uma distância máxima de projeto. O problema é que esta distância é determinada por tentativa e erro para que o número de conjuntos nebulosos não seja muito grande. Esta relação entre o tamanho da distância de projeto e o número de grupos c não é linear. Uma solução seria a normalização dos dados.

As formas das funções de pertinências para FCM, GK e GG podem ser vistas no Capítulo 4.

Não foi utilizada nenhuma técnica de normalização dos dados de entrada (auto-regressores) ou retirada de tendência dos dados da série. Utilizar estes métodos podem trazer melhores resultados na modelagem.

O método dos mínimos quadrados recursivo ponderado foi utilizado para estimar os parâmetros dos modelos locais. Foi suposto que não havia dados corrompidos por ruídos correlacionados ou que não fossem do tipo ruído branco. Além disso, não foi realizada nenhuma análise de estabilidade para estes modelos locais encontrados. A análise de estabilidade dos modelos TS serão estudados detalhadamente e mostrados em trabalhos futuros.

Os resultados apresentados mostram que os modelos nebulosos TS são ferramentas que fornecem bom desempenho quando tratamos o problema de modelagem de séries temporais.

Para trabalhos futuros, o estudo de redes neurais nebulosas não recorrentes e recorrentes para a predição da série seria interessante. Poderia ser explorado também a predição de vários passos a frente. Um outro possível ajuste nos algoritmos anteriores seria utilizar ferramentas como algoritmo genético para determinar um número ótimo de regras ou para, no caso de estudo de redes neurais, determinar a estrutura da rede.

Referências Bibliográficas

- [1] Aguirre, L. A. e Billings, S.A., *Nonlinear Chaotic Systems: Approaches and Implications for Science and Engineering - A Review*. Applied Signal Processing, Inglaterra, v. 1995, n. 2, pp. 224-248, 1996.
- [2] Aguirre, Luís Antônio. *Introdução à Identificação de Sistemas - Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. UFMG, 2004.
- [3] Anderson, O.D.. *Time Series Analysis and Forecasting - The Box-Jenkins approach*. Butterworths, 1976.
- [4] Ball, G.H. e Hall, D.J.. *A Novel Method of Data Analysis and Pattern Classification*. Technical report, Stanford Research Institute, Menlo Park, CA, 1965.
- [5] Ballini, Rosângela. *Análise e Previsão de Vazões Utilizando Modelos de Séries Temporais, Redes Neurais e Redes Neurais Nebulosas*. Tese de Doutorado, FEEC - UNICAMP, Brasil, 2000.
- [6] Barreto, Gilmar. *Modelagem Computacional Distribuída e Paralela de Sistemas e de Séries Temporais Multivariáveis no Espaço de Estado*. Tese de Doutorado, FEEC - UNICAMP, Brasil, 2002.
- [7] Bottura, Celso Pascoli. *Análise Linear de Sistemas*. Guanabara Dois, 1982.
- [8] Box, G.E.P. e Jenkins, G.M.. *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden Day, 1970.
- [9] Brockwell, Peter J. e Davis, Richard A.. *Introduction to Time Series and Forecasting*. Springer, 1996.

- [10] Caines, Peter E.. *Linear Stochastic Systems*. Wiley, 1988.
- [11] Chatfield, C.. *The Analysis of Time Series - An Introduction*. Chapman & Hall, 4ª Edição, 1989.
- [12] Chen, Bor-Sen; Peng, Sen-Chueh e Wang, Ku-Chen. *Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks: A Fuzzy AR Approach*. IEEE Transactions on Fuzzy Systems, Vol.8, No.5, October, 2000.
- [13] Gath, I. e Geva, A.B.. *Unsupervised Optimal Fuzzy Clustering*. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.11, No.7, July, 1989.
- [14] Goodwin, Grahan C. e Payne, Robert L.. *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, 1977.
- [15] Goodwin, Grahan C. e Sin, Kwai Sang. *Adaptive Filtering Prediction and Control*. Prentice Hall, 1984.
- [16] Gustafson, D.E. e Kessel, W.C.. *Fuzzy Clustering With a Fuzzy Covariance Matrix*. Proc. IEEE CDC San Diego, CA, pp.761-766, 1979.
- [17] Hathaway, Richard J. e Bezdek, James C.. *Switching Regression Models and Fuzzy Clustering*. IEEE Transactions on Fuzzy Systems, Vol.1, No.3, August, 1993.
- [18] Hellendoorn, H. e Driankov, D. (Eds.). *Fuzzy Model Identification - Selected Approaches*. Springer, 1997.
- [19] Henderson, H.V. e Searle, S.R.. *On Deriving the Inverse of a Sum of Matrices*. SIAM Review - Volume 23. No.1, 53-60, January, 1981.
- [20] Jankowski, K.P.; Van Brussel, H.. *An approach to discrete inverse dynamics control of flexible-joint robots*. IEEE Transactions on Robotics and Automation, Volume: 8 , Pages:651 - 658, Issue: 5 , Oct. 1992.
- [21] Kasabov, Nikola K. e Song, Qun. *DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction*. IEEE Transactions on Fuzzy

- Systems, Vol.10, No.2, April, 2002.
- [22] Kim, Euntai; Park, Minkee; Ji, Seunghwan e Park, Mignon. *A New Approach to Fuzzy Modeling*. IEEE Transactions on Fuzzy Systems, Vol.5, No.3, August, 1997.
- [23] Lyung, Lennart. *System Identification - Theory for the User*. Prentice Hall, 2ª edição, 1999.
- [24] Lughofer, Edwin e Groissböck, Werner. *Generating Correlation and Regression Models from High Dimensional Measurement Data - Advanced Aspects, Strategies and Validation*. Technical Report FLLL-TR-0212, Johannes Kepler Universität Linz, 2002.
- [25] Lughofer, Edwin e Groissböck, Werner. *Online Adaptation of Correlation and Regression Models*. Technical Report FLLL-TR-0213, Johannes Kepler Universität Linz, 2002.
- [26] Lughofer, Edwin. *Online Adaptation of Takagi-Sugeno Fuzzy Inference Systems*. Technical Report FLLL-TR-0217, Johannes Kepler Universität Linz, 2003.
- [27] Mc Avoy, T.J., Hsu, E. and Lowenthal, S.. *Dynamics of pH in controlled stirred tank reactor*. Ind.Eng.Chem.Process Des.Develop, 11, pp.71-78, 1972.
- [28] Montgomery, Douglas C. e Johnson, Lynwood A. . *Forecasting and Time Series Analysis*. McGraw-Hill Book Company, 1976.
- [29] Morettin, Pedro Alberto e Toloi, Clélia Maria de Castro. *Modelos para Previsão de Séries Temporais*. Instituto de Matemática Pura e Aplicada, 1981.
- [30] Nobre, F.S.M. e Palhares, A.G.B.. *Informações Quantitativas e Qualitativas no Projeto de Controladores Nebulosos*. Controle & Automação, Vol.8, No.2, pp.77-93, Maio a Agosto de 1997.
- [31] Pedrycz, Witold e Gomide, Fernando. *An Introduction to Fuzzy Sets - Analysis and Design*. MIT Press, 1998.
- [32] Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Intelligent adaptive data modeling applied to discrete time-series prediction in industrial processes*. INDUSCON

- 2004 - IEEE VI Conferência Internacional de Aplicações Industriais, Joinville, SC, 2004.
- [33] Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Adaptive Identification of Takagi-Sugeno Fuzzy Neural Model to Discrete Time Series Prediction*. SBRN 2004 - VIII Simpósio Brasileiro de Redes Neurais, UFMA - São Luís, MA, 2004.
- [34] Pucciarelli, Amilcar J.; Barreto, Gilmar e Serra, Ginalber L.O. *Identificação Adaptativa de Modelos Fuzzy Takagi-Sugeno para Predição de Séries Temporais Discretas no Tempo*. 3º DINCON - Congresso Temático de Dinâmica e Controle da SBMAC, UNESP - Campus de Ilha Solteira, 31 de Maio a 3 de Junho, 2004.
- [35] Pucciarelli, Amilcar J. e Serra, Ginalber L.O. *Projeto de Controle PI Fuzzy Adaptativo Baseado no Conhecimento de Especialistas*. 2º DINCON - Congresso Temático de Dinâmica e Controle da SBMAC, ITA - São José dos Campos, 18 a 22 de Agosto, 2003.
- [36] Ruspini, E.H.. *A New Approach to clustering*. Information and Control, Volume:15, Number:1, 22-32, July 1969.
- [37] Silva, Leila Roling Scariot da. *Aprendizagem Participativa em Agrupamento Nebuloso de Dados*. Dissertação de Mestrado, FEEC - UNICAMP, Brasil, 2003.
- [38] Söderström, Torsten e Stoica, Petre. *System Identification*. Prentice Hall, 1989.
- [39] Swevers, J.; Torfs, D.; Adams, M.; De Schutter, J.; Van Brussel, H.. *Comparison of control algorithms for flexible joint robots implemented on a KUKA IR 161/60 industrial robot*. Fifth International Conference on Advanced Robotics, vol.1, pages:120 - 125, 1991.
- [40] Takagi, T. e Sugeno, M.. *Fuzzy identification of systems and its application to modeling and control*. IEEE Trans. Systems, Man and Cybernetics, vol.15, no.1, pp.116-132, 1985.
- [41] Vega, P.; Prada, C.; Alexandre, V.. *Self-tuning Predictive PID Controller*, IEEE Proceedings-D, vol.138, no.3, May 1991.
- [42] Wang, Li-Xin. *A Course in Fuzzy Systems and Control*. Prentice Hall, 1996.

- [43] Wellstead, P.E. e Zarrop M.B.. *Self-Tuning Systems - Control and Signal Processing*. Wiley, 1991.
- [44] Young, Peter. *Recursive Estimation and Time-Series Analysis - An Introduction*. Springer-Verlag, 1984.