

Laboratório Nacional de Computação Científica
Coordenação de Formação de Recursos Humanos

PROCESSAMENTO DE PERFIS METABÓLICOS

Marco Antônio Vilela

Petrópolis, Março de 2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

PROCESSAMENTO DE PERFIS METABÓLICOS

Marco Antônio Mattos Vilela

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM MODELAGEM COMPUTACIONAL.

Aprovada por:

Profª. Ana Tereza Ribeiro de Vasconcelos, D.Sc.
(Presidente)

Prof. Jonas Almeida, Ph.D

Prof. Marcelo Dutra Fragoso, Ph.D

Prof. Marcello Luiz Rodrigues de Campos, Ph.D

PETRÓPOLIS, RJ - BRASIL

MARÇO DE 2007

VILELA, MARCO ANTÔNIO

Processamento de Perfis Metabólicos

[Petrópolis] 2007

VI, 70 p., 29,7 cm (MCT/LNCC, M.Sc.,
Modelagem Computacional, 2007)

Tese – Laboratório Nacional de Computação
Científica, LNCC

- 1 Processamento de Perfis Metabólicos
 - 2 Suavização e Extração da Derivada temporal
- I. LNCC/MCT II. Título (série)

À minha maravilhosa Mãe - que não me alimenta mais, não cuida mais de mim, não paga minhas contas mas ainda assim escreve minhas dedicatórias.

Agradecimentos

Aos Profs. Ana Tereza e Jonas, pela confiança, oportunidade e suporte na realização desse trabalho.

Aos Profs. Voit, Cristiano e Grivet pelas discussões, comentários e correções que aprimoraram este trabalho.

A todos os meus amigos, incluindo Thaís, por ter me colocado nos agradecimentos da tese dela.

Agradecimento especial a minha avó Maria.

Resumo da Tese apresentada ao LNCC/MCT como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PROCESSAMENTO DE PERFIS METABÓLICOS

Marco Antônio Vilela

Março/2007

Orientadores: Ana Tereza Ribeiro Vasconcelos
Jonas S. Almeida

Modelagem Computacional

Nos últimos 30 anos, a Teoria dos Sistemas Bioquímicos (Biochemical System Theory - BST) tem fornecido uma fundação concreta para o estudo da dinâmica de sistemas biológicos, por exemplo, Sistemas-S (S-systems) usados em engenharia reversa de vias metabólicas (Savageau, 1969; Savageau, 1970; Voit, 2000). Uma característica marcante desse tipo de modelo é que os parâmetros não só quantificam as interações entre os componentes da rede metabólica, mas também fornecem a sua topologia de regulação. Procedimentos automáticos para a parametrização dos Sistemas-S a partir de séries temporais biológicas vêm sendo desenvolvidos por vários pesquisadores, onde se assume que a série temporal e sua derivada temporal são livres de ruído. Entretanto, perfis metabólicos livres de ruído não são realistas em cenários de experimentos de biologia molecular. Técnicas como Redes Neurais Artificiais (RNA), Máquinas de Vetores de Suporte (MVP) e filtro de Savitzky-Golay foram propostas como solução do problema de suavização dos perfis metabólicos com a vantagem da obtenção da derivada temporal simbólica (Almeida and Voit, 2003; Borges, et al., 2006; Borges, et al., 2004; Voit and Almeida, 2004). Entretanto, essas soluções apresentaram alguns artefatos problemáticos na derivada até mesmo quando nenhum problema é visualmente detectado no dado suavizado, deixando aberto um “espaço vazio” na questão de um método automático para a parametrização dos Sistemas-S a partir de dados experimentais. O algoritmo apresentado neste trabalho propõe preencher esse espaço com uma ferramenta robusta para a extração de sinal e de sua derivada temporal a partir de séries temporais ruidosas.

Abstract of Thesis presented to LNCC/MCT as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

METABOLIC PROFILES PROCESSING

Marco Antônio Vilela

March/2007

Advisors: Ana Tereza Ribeiro Vasconcelos
Jonas S. Almeida

Computational Modeling

During the past 30-years, Biochemical System Theory (BST) has been provided a concrete foundation for the study of the dynamic biological systems, e.g. S-systems models for reverse engineering of metabolic networks (Savageau, 1969; Savageau, 1970; Voit, 2000). One of the remarkable characteristics of these models is its parameters not only quantify the interactions between the components of the network, but also elucidate the network's topology. Automatic procedures for S-system parameterization from biological time series have been developed by many researches, where they assume a noise-free time series and a true estimated first derivative in their methodologies (Chou, et al., 2006; Kikuchi, et al., 2003). Nevertheless, this noise-free data is not a realistic scenario of the real biological experimental world. Methods as artificial neural network (ANN), Support Vectors Machines (SVM) and Savitzsky-Golay filter were proposed to overcome the denoising time series problem with the advantage of a closed form output which allowed determining the first derivative symbolically (Almeida and Voit, 2003; Borges, et al., 2006; Borges, et al., 2004; Voit and Almeida, 2004). However, these solutions showed some problematic artifacts in its first derivative even when they are not visually apparent in the smoothed data, leaving a gap on the issue of a fully automatic method for S-system parameterization from experimental data. The algorithm presented in this work is a proposal to fill this gap up providing an unbiased robust tool for signal extraction and first derivative estimation from noisy time series.

Sumário

Introdução	3
1.1 Modelagem Matemática de Vias Metabólicas.....	4
1.1.1 Sistemas-S – (S-systems).....	5
1.1.2 Métodos de Parametrização	7
1.2 Objetivos.....	9
1.3 Apresentação do Trabalho	10
Análise de Séries Temporais.....	11
2.1 Introdução.....	11
2.2 Séries Temporais Estacionárias	12
2.3 Métodos Descritivos	13
2.3.1 Regressão pelos mínimos quadrados.....	13
2.3.2 Regressão Linear pelos mínimos quadrados	14
2.3.3 Suavizador de Média Móveis	15
2.3.4 Regressão por Spline	15
2.3.5 Filtro Savitzky-Golay	16
2.3.6 Redes Neurais Artificiais.....	17
2.3.7 Máquinas de Vetores de Suporte	18
2.3.8 Filtro Whittaker	19
2.4 Erro de Validação Cruzada.....	23
Entropia.....	25
3.1 Entropia de Shannon.....	25
3.2 Entropia Conjunta.....	28
3.3 Entropia Condicional.....	29
3.4 Propriedades da Entropia Condicional e Conjunta.....	30
3.5 Informação Mútua	30
3.6 Entropia de Rényi	31
3.7 “Distância” de Kullback-Leibler	32

3.8	Divergente de Rényi	33
3.9	Aprendizado Teórico de Informação	33
3.9.1	Aprendizado de Máquina Baseado em Teoria da Informação	34
3.9.2	Estimador não-paramétrico de função de densidade de probabilidade	36
3.9.3	Entropia Quadrática de Rényi.....	38
	Suavizador Automático Aplicado a Séries Temporais Biológicas.....	41
4.1	Introdução.....	41
4.2	Entropia do Erro de Validação Cruzada.....	44
4.4	Função Custo Alternativa para o Gradiente	50
4.5	Algoritmo de Segmentação Baseado na Entropia do Erro de Validação Cruzada...	54
4.6	Derivada da curva suavizada	57
4.7	Interface Gráfica	60
	Conclusão	62
5.1	Perspectivas Futuras	64
	Apêndice A – Operações de Hadamard.....	65
A.1	Propriedades do produto de Hadamard	65
	Apêndice B.....	67
	Referências Bibliográficas	68

Capítulo 1

Introdução

Vários componentes dos processos de manutenção e sobrevivência celular vêm sendo elucidados com o avanço da genética e da biologia molecular. Grande quantidade de dados a partir de experimentos de biologia molecular é gerada (expressão de genes em microarranjos; seqüências de genomas completos entre outros), desafiando pesquisadores no desenvolvimento de ferramentas computacionais que facilitem a análise desses dados. O estudo dos processos bioquímicos celulares, como por exemplo as vias metabólicas, vem sendo profundamente beneficiados com diversos tipos de dados moleculares, onde o mecanismo de interatividade entre biomoléculas tem um papel fundamental na análise desses dados. Vias metabólicas são seqüências de reações bioquímicas celulares que tem o intuito de obter energia ou de sintetizar biomoléculas e são diferenciadas em vias catabólicas e anabólicas respectivamente. Estas vias podem ser representadas por grafo de fluxo onde os nós são os componentes da via (chamados de metabólitos) e as arestas representam as interações entre esse componente (Voit 2000). Essa representação é chamada de **mapa metabólico ou mapa de regulação metabólica** (Figura 1.1) e é usualmente obtida por uma série de experimentos biológicos. Uma nova medida experimental chamada de perfis metabólicos, gerada a partir de observações temporais simultâneas das concentrações dos metabólitos que compõem uma via metabólica (Voit and

Almeida 2004), vem se tornando um grande atrativo para o estudo desses processos bioquímicos sendo também esta o foco deste trabalho.

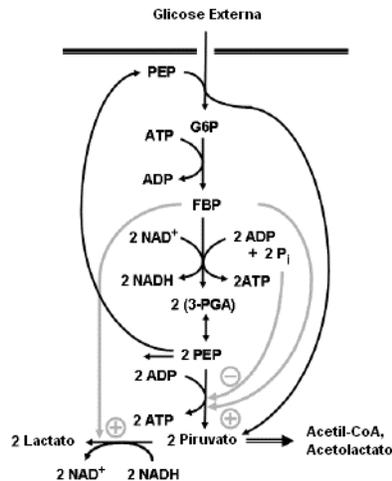


Figura 1.1 – Exemplo de mapa de regulação metabólica da glicose na bactéria *Lactococcus lactis* (Ramos, Neves et al. 2002).

1.1 Modelagem Matemática de Vias Metabólicas

Existem na literatura vários modelos matemáticos que traduzem a informação de um mapa de uma via metabólica em equações com diferentes níveis de simplificação e diferentes pressuposições de uma biologia complexa. O modelo escolhido deve descrever e generalizar o fenômeno em questão estudado e ignorar outros aspectos do sistema de menor relevância na descrição deste fenômeno. Outra característica fundamental dos modelos descritivos é a possibilidade de análise dos parâmetros do modelo, que devem apresentar um significado biológico do processo ou fenômeno estudado. Este trabalho é direcionado ao estudo do comportamento dinâmico da concentração dos componentes de uma via metabólica, sendo assim este o escopo do ferramental matemático estudado.

1.1.1 Sistemas-S – (S-systems)

Os Sistemas-S são sistemas de equações diferenciais não-lineares de primeira ordem que modelam a dinâmica das concentrações dos componentes de uma rede biológica (metabólitos em uma via metabólica; genes em uma rede de regulação gênica) baseados no formalismo de leis da potência (Savageau 1969a; Savageau 1969b; Savageau 1970; Voit 2000). O nome “S-systems” ou Sistemas-S deve-se a duas características fundamentais dos sistemas biológicos em geral: sinergia e saturação. Essas duas características são de grande importância para os sistemas biológicos e não podem ser inferidas a partir de um sistema de equações lineares. Assim, para uma via metabólica representada pela concentrações X_i de n metabólitos dependentes do tempo e m metabólitos independentes do tempo, tem-se que o Sistema-S associado a essa via é dado por (Voit 2000):

$$\dot{X}_i(t) = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}, i = 1, 2, \dots, n \quad (1.1)$$

O termo α_i é um fator constante da produção de X_i e o termo β_i é também um fator constante mas com relação a degradação de X_i . Esses fatores podem assumir valores reais não-negativos. O termo α_i ou β_i é nulo quando a produção ou a degradação da concentração de X_i é interrompida respectivamente. Os parâmetros g_{ij} e h_{ij} são chamados de ordens cinéticas e quantificam a relação de um determinado metabólito X_j na produção e degradação da concentração do metabólito X_i respectivamente. Estudos mostram que reações enzimáticas são governadas por essa mesma lei da potência onde a ordem cinética pode ser qualquer número real, não necessariamente inteiro (Voit 2000). Os Sistemas-S têm sempre a mesma estrutura, diferenciando somente nos valores dos parâmetros, o que

possibilita uma relação direta com a via metabólica modelada. Esse é um aspecto importante dos Sistemas-S, pois as ordens cinéticas não só quantificam as interações como também representam a topologia do mapa metabólico (Figura 1.2). O conhecimento desse mapa permite a identificação de enzimas e reações bioquímicas que são cruciais para a sobrevivência de organismos, exercendo o papel de direcionador para potenciais alvos de drogas. Outra motivação para o estudo dos mapas metabólicos é que várias doenças neurológicas e relacionadas com câncer são na verdade falhas na interação de algum metabólito, resultando no mau funcionamento ou inativação de uma via metabólica. Essas falhas podem ser reconhecidas no modelo como a ausência de uma determinada ordem cinética, ou conseqüentemente, como a independência da concentração de um determinado metabólito no comportamento dinâmico de outro metabólito. O mapa metabólico junto com o modelo dinâmico pode fornecer informações como tempo de consumo, quantidades que devem ser fornecidas assim como efeitos colaterais de drogas usadas no tratamento de doenças.

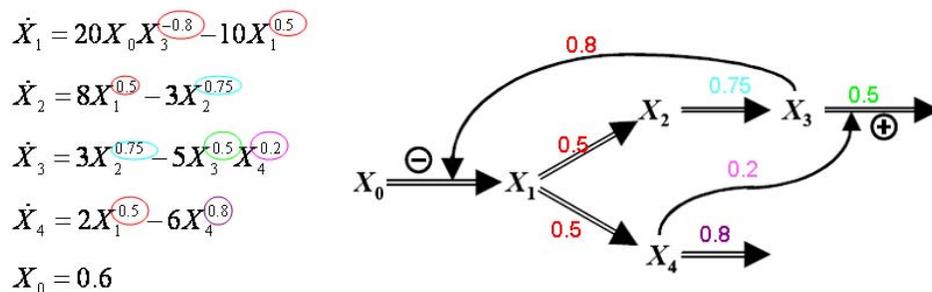


Figura 1.2 – Mapa da via metabólica extraído a partir do S-system (Voit and Almeida 2004). As ordens cinéticas determinam a topologia do mapa.

Experimentos de Ressonância Nuclear Magnética (RNM) possibilitam a medição *in vivo* das concentrações de metabólitos de determinadas vias metabólicas ao longo do tempo (Ramos, Neves et al. 2002). Essas séries temporais são chamadas de perfis metabólicos ou metabolômicos (Figura 1.3), e são obtidas sem um conhecimento prévio do mapa de regulação metabólico. A tradução desses dados em um mapa metabólico tornou-se uma questão fundamental para o avanço da análise das vias metabólicas e vários métodos de parametrização dos Sistemas-S vêm sendo desenvolvidos na tentativa de responder essa questão.

1.1.2 Métodos de Parametrização

A parametrização dos Sistemas-S é baseada em métodos que usam dados do estado estacionário do sistema e métodos que usam dados dinâmicos do sistema (Voit 2000). A estimação a partir do estado estacionário é baseada na reposta do sistema neste estado quando pequenas perturbações nas concentrações dos metabólitos são impostas. Já a estimação com dados dinâmicos é feita utilizando concentrações ao longo do tempo de todos os metabólitos da via (variáveis dependentes X_i). Vários métodos (como algoritmos genéticos, regressão não-linear, recozimento simulado (Simulated Annealing)) vêm sendo aplicados na parametrização a partir de dados dinâmicos (Kikuchi, Tominaga *et al.* 2003; Voit and Almeida 2004; Gonzalez, Kuper *et al.* 2006; Marino and Voit 2006). Infelizmente, esses métodos apresentam restrições de aplicação, como por exemplo o número de variáveis (metabólitos) usados na estimativa. Outro problema apresentado é o long tempo de convergência quando heurísticas de procura são usadas (Voit, Almeida *et al.* 2006). Recentemente, um método de regressão chamado Regressão Alternativa (Alternative

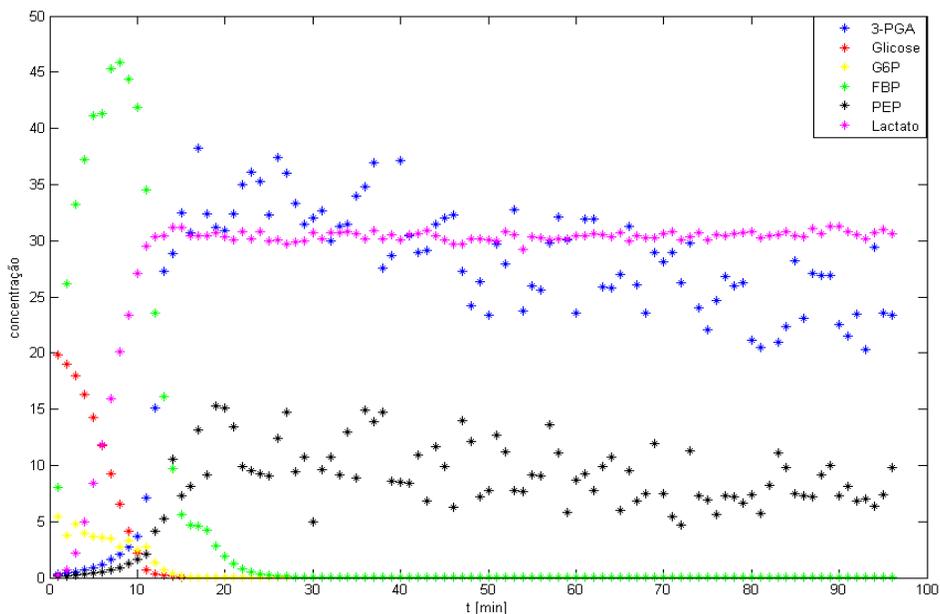


Figura 1.3 – Perfis metabólicos. Concentrações da Glicose, Glicose 6-fosfato (G6P), Frutose 1,6-bisfosfato (FBP), 3-Fosfoglicerato (3-PGA), Fosfoenolpiruvato (PEP) e Lactato do metabolismo da glicose na bactéria *Lactococcus lactis* (Ramos, Neves et al. 2002).

1.2 Objetivos

O avanço crescente nas técnicas de biologia molecular traz consigo uma grande quantidade de dados, o que muitas vezes impossibilita o trabalho humano para a análise desses dados. Acredita-se que em um futuro próximo, séries temporais de concentrações de metabólitos vão seguir essa linha natural de grandes massas de dados moleculares, fazendo necessária a automação da análise.

A proposta deste trabalho é desenvolver uma ferramenta automática e robusta em que uma série temporal suavizada e sua derivada em relação ao tempo possam ser extraídas dos perfis metabólicos obtidos a partir de experimentos de RMN. Essa ferramenta permitirá a utilização de técnicas de parametrização de Sistemas-S já desenvolvidas (Kikuchi,

Tominaga *et al.* 2003; Chou, Martens *et al.* 2006; Marino and Voit 2006) e conseqüentemente a possibilidade de obtenção dos mapas de regulação metabólica.

1.3 Apresentação do Trabalho

Várias técnicas de processamento de sinais foram estudadas neste trabalho como tentativa de obter uma solução para a extração de sinal dos perfis metabólicos. O conteúdo desta dissertação é a seguir descrito:

Alguns métodos e conceitos clássicos de análise de séries temporais tais como; condições de estacionariedade de séries temporais e técnicas de filtragem (filtro Savitzky-Golay e filtro Whittaker) são apresentados no capítulo 2. Uma ênfase é dada ao filtro Whittaker pois este é a base do trabalho desenvolvido.

No capítulo 3 é feita uma introdução dos conceitos de entropia de Shannon e Rényi, de medidas de “distâncias” entre funções densidade de probabilidade e de teoria de informação, que forma a base teórica do algoritmo de otimização do filtro Whittaker apresentado neste trabalho.

O capítulo 4 apresenta o desenvolvimento das equações dos gradientes da otimização dos parâmetros do filtro Whittaker e o método de segmentação de sinal. Os resultados obtidos usando dados reais e simulados são também apresentados neste capítulo. A conclusão e perspectivas futuras são apresentadas no capítulo 5. O Apêndice A introduz algumas propriedades matriciais que foram usadas no desenvolvimento das equações do algoritmo de otimização do filtro Whittaker.

Capítulo 2

Análise de Séries Temporais

2.1 Introdução

Como este trabalho é aplicado a solução da extração de sinal de séries temporais ruidosas, alguns conceitos da teoria de séries temporais e de técnicas de processamento de sinais serão apresentados neste capítulo.

Registros de dados temporais geofísicos, biológicos, econômicos e outras áreas carregam implicitamente em seus pontos uma grande quantidade de informação sobre a dinâmica do sistema em questão que é de extrema importância para pesquisadores da área. Com o objetivo de extrair essa informação implícita a partir de dados temporais, vários métodos de análise de séries temporais vem sendo desenvolvidos. Assim, define-se uma série temporal de tamanho N como a quantificação de um fenômeno por meio de uma seqüência de pontos $\{X_t\}$ observados em instantes de tempo $t=1,2,\dots,N$ (*Fan and Yao 2005*). Basicamente, existem três grandes objetivos na análise de séries temporais: entender a dinâmica do sistema, prever para onde o sistema está evoluindo e controlar o futuro do sistema a partir de intervenções(*Fan and Yao 2005*). É com o primeiro objetivo que este trabalho foi desenvolvido.

2.2 Séries Temporais Estacionárias

Uma série temporal $\{X_t, t=\pm 1, \pm 2, \dots\}$ vista como um processo estocástico é definida como estacionária se as duas condições abaixo forem satisfeitas (Fan and Yao 2005):

i) O valor esperado $E(X_t^2) < \infty$ para cada t e $E(X_t)$ é uma constante independente de t ;

ii) A covariância $Cov(X_t, X_{t+k})$ é independente de t para todo k .

A mesma série temporal $\{X_j\}$ é dita estritamente estacionária se (X_1, \dots, X_n) e $(X_{1+k}, \dots, X_{n+k})$ têm a mesma distribuição conjunta para qualquer inteiro $n \geq 1$ e para qualquer inteiro k . A estacionariedade de uma série temporal é interpretada como a invariância de suas propriedades estatísticas (momentos de primeira e segunda ordem) em relação ao tempo. Séries que não satisfazem as condições de estacionariedade são chamadas não-estacionárias. Uma definição mais abrangente de estacionariedade é dada por:

Seja $\{X_t; t = \pm 1, \pm 2, \dots\}$ uma série temporal estacionária de ordem r . Então para todo $n \leq r$ com $r_1 + r_2 + \dots + r_n \leq r$ e s pertencente ao conjunto dos naturais, tem-se:

$$E \left[\prod_{j=1}^n X_j^{r_j} \right] = E \left[\prod_{j=1}^n X_{j+s}^{r_j} \right] \quad (2.1)$$

Uma série estacionária de ordem $r=1$ é dita estacionária na média e uma série estacionária de ordem $r=2$ é dita estacionária na covariância.

2.3 Métodos Descritivos

Dada uma série temporal $\{X_t; t = \pm 1, \pm 2, \dots\}$, pode-se efetuar a seguinte decomposição:

$$X_t = \mu_t + \varepsilon_t \quad (2.2)$$

onde μ_t é dito componente regular ou tendência e ε_t é a componente irregular ou ruído. A seguir são apresentados alguns métodos para a estimação de μ_t .

2.3.1 Regressão pelos mínimos quadrados

Sejam X e Y dois vetores aleatórios de distribuição conjunta $f_{x,y}(x, y)$. Suponha que uma particular observação do vetor Y é feita, por exemplo $Y=y$. É esperado que o conhecimento do valor dessa observação $Y=y$ dê alguma informação do valor não medido do vetor X , ou seja, uma redução da incerteza sob o valor de X . Assim, o seguinte problema pode ser definido:

Considerando os vetores X e Y descritos acima com respectivas dimensões n e m , e que o vetor $Y=y$, encontrar um estimador \hat{X} para o vetor X tal que a média condicional da norma quadrada do erro de estimação $E\{|X - \hat{X}|^2 | Y = y\}$ seja mínima.

A resposta do problema acima recai na simples prova a seguir (Rhodes 1971):

$$\begin{aligned} E\{|X - \hat{X}|^2 | Y = y\} &= E\{X'X - 2\hat{X}'X + \hat{X}'\hat{X} | Y = y\} \\ &= E\{X'X | Y = y\} - 2\hat{X}'E\{X | Y = y\} + \hat{X}'\hat{X} \end{aligned} \quad (2.3)$$

Derivando a Equação 2.3 em relação a \hat{X} e igualando a zero, temos:

$$\hat{X} = E\{X | Y = y\} \quad (2.4)$$

Para o vetor estimado \hat{X} , o menor erro correspondente é dado por:

$$\begin{aligned} E\left\{|X - \hat{X}|^2|Y = y\right\} &= E\left\{|X|^2|Y = y\right\} - |E\{X|Y = y\}|^2 \\ &= E\left\{|X|^2|Y = y\right\} - |\hat{X}|^2 \end{aligned} \quad (2.5)$$

A Equação 2.5 é a variância condicional do erro de estimação (Rhodes 1971).

2.3.2 Regressão Linear pelos mínimos quadrados

Dado o problema de regressão descrito em 2.3.1, pode-se substituir o estimador \hat{Y} de Y da forma $g(X)$, tal que a função $g(\cdot)$ tenha uma característica linear. Este problema pode ser escrito da seguinte maneira:

Dados os vetores $X = [x_1, x_2, \dots, x_n]^t; Y = [y_1, y_2, \dots, y_n]^t$, encontrar a função linear $Y = f(X) = aX + b$ tal que $E\{|X - F(X)|^2\}$ seja mínima. Definindo as seguintes variáveis matriciais:

$$M = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, Z = \begin{bmatrix} a \\ b \end{bmatrix} \quad (2.6)$$

Temos:

$$E\{|X - F(X)|^2\} = E\{|X - MZ|^2\} = E\{(X - MZ)^t(X - MZ)\} \quad (2.7)$$

Dado um subespaço vetorial $W = \{MK_{2 \times 1} : K \in R^2\}$ de R^n , tem-se que pelo Teorema da Projeção, o vetor $v = X - MZ$ de norma mínima é aquele ortogonal à W , ou seja:

$$(MK)^t(X - MZ) = K^t(M^tX - M^tMZ) = 0 \quad (2.8)$$

Como K é um vetor não nulo, a Equação 2.8 implica em:

$$(M^t X - M^t MZ) = 0 \Rightarrow Z = (M^t M)^{-1} M^t Z \quad (2.9)$$

As componentes de Z são os próprios regressores, ou seja, as constantes a e b da equação da reta que minimiza o erro quadrático 2.7.

2.3.3 Suavizador de Média Móveis

A idéia do suavizador de média móvel é substituir o ponto central de uma dada janela da série temporal pela respectiva média da janela. Dada uma série temporal $\{X_t; t = 1, 2, \dots, n\}$, chama-se média móvel de X a série temporal Y dada por:

$$Y_t = \sum_{j=t-\text{int}(p/2)}^{t+\text{int}(p/2)} w_j X_j \quad (2.10)$$

onde p é número ímpar que representa o tamanho da janela, $\text{int}(\cdot)$ é a função número inteiro do argumento e w é um conjunto de pesos. Sucessivas aplicações de média móvel produzem médias móveis de ordem superior .

2.3.4 Regressão por Spline

Seja S a classe de função das funções contínuas de R em R com primeira e segunda derivadas contínuas. Pode-se encontrar uma função chamada spline $s \in C^{1,2}$, para qual s é uma suavização de uma dada série temporal $\{X_t; t = 1, 2, \dots, n\}$. Essa função s é obtida pela minimização do seguinte funcional:

$$\min \left\{ \sum_{i=1}^n (x_i - s(t_i))^2 + \alpha \int_{-\infty}^{\infty} s''(t)^2 dt \right\} \quad (2.11)$$

onde α é o parâmetro de suavização. Quanto maior o valor de α mais suave é a curva spline resultante.

2.3.5 Filtro Savitzky-Golay

Suponha um série temporal $\{X_t; t = 1, 2, \dots, n\}$. O princípio do filtro Savitzky-Golay (SG) consiste em substituir a observação central de uma determinada janela da série X_t por um ponto dado pelo polinômio regressor ordem p que suaviza a série na própria janela. Para um comprimento de janela w , onde w é um inteiro ímpar, o algoritmo segue os seguintes passos:

- 1) Ajustar um polinômio de ordem p nos pontos da janela;
- 2) Substituir o ponto central da janela pelo ponto dado pelo polinômio naquela posição;
- 3) Mover a janela um ponto para “frente” (na direção crescente da amostragem) e aplicar os passos 1 e 2 novamente.

O algoritmo termina com $n-w+1$ iterações, o que indica que a curva suavizada é sempre menor que a série temporal por $w-1$ pontos. Os parâmetros w e p controlam o grau de suavidade da saída do filtro.

2.3.6 Redes Neurais Artificiais

Redes neurais artificiais (RNA) são modelos conexionistas motivados pela bioquímica das interações entre neurônios. Conhecidas como regressores universais de funções, as RNA's são compostas por unidade básicas de processamento chamadas neurônios, originalmente concebidos no afamado modelo de McCulloch e Pitts (Figura 2.1).

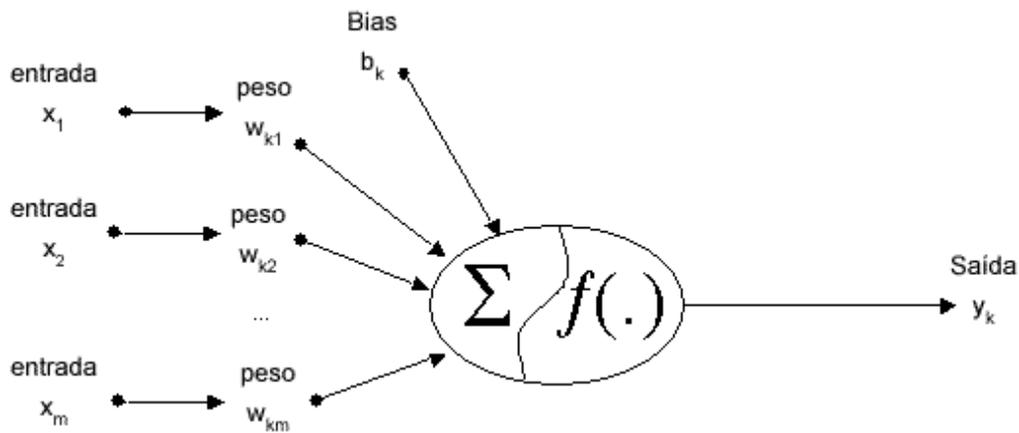


Figura 2.1 – Modelo de neurônio de McCulloch e Pitts.

O objetivo da RNA é achar os valores dos pesos w associados aos neurônios tal que o erro entre o mapeamento y das variáveis de entrada x e a saída desejada seja mínimo. O processo de procura dos pesos da RNA é chamado de aprendizado da rede. O algoritmo de aprendizado comumente utilizado é chamado retro-propagação (backpropagation), que utiliza o gradiente do erro mínimo quadrado entre a saída da rede e o seu respectivo alvo (Hornik, Stinchcombe *et al.* 1989). O RNA vem sendo vastamente utilizado na modelagem de dados biológicos (Almeida 2002), bem como na suavização de perfis metabólicos (Almeida and Voit 2003; Voit and Almeida 2004).

2.3.7 Máquinas de Vetores de Suporte

Introduzida por Vapnik como uma técnica de classificação, máquinas de vetores de suporte (MVS) vêm sendo vastamente aplicadas em problemas de regressão (Smola and Scholkopf 2004). A idéia básica das MVS é determinar uma função de forma $f(t) = \langle w, t \rangle + b$ a partir de uma série temporal X_i tal que o maior desvio entre $f(t)$ e seu respectivo alvo X_i seja dado pelo valor ε e que a função $f(t)$ seja a mais suave possível. Assim, o problema pode ser formulado como (Smola and Scholkopf 2004):

$$\begin{aligned} \min \quad & \frac{1}{2} |w|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{sujeito a} \quad & X_i - \langle w, t \rangle - b \leq \varepsilon + \xi_i \\ & \langle w, t \rangle + b - X_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{2.12}$$

Na Equação 2.12 ξ_i, ξ_i^* são chamadas variáveis de folga, w é um vetor de pesos, b é o ajuste para o viés, ε é a faixa de insensibilidade, C é o parâmetro de regularização da suavidade e $|\cdot|^2$ representa a norma quadrada. A operação representada por $\langle \cdot, \cdot \rangle$ é o produto escalar. Aplicada na suavização de perfis metabólicos (Borges, Vilela *et al.* 2006), as MVS obtiveram um bom resultado na suavização, porém com uma difícil tarefa na estimação dos parâmetros ε e C , tornando o problema ainda maior para uma suavização automática.

2.3.8 Filtro Whittaker

Publicado em 1923 como um algoritmo de suavização(Whittaker 1923), o filtro Whittaker permaneceu desde então na literatura ao ponto da mesma idéia ser republicada em 1997 por Hodrick e Prescott no contexto de séries temporais econômicas(Hodrick and Prescott 1997). A solução proposta por Whittaker para o problema de extração de uma curva suave de uma série temporal foi colocada da seguinte maneira(Whittaker 1923):

Dada uma série temporal y_i de comprimento N com incerteza de medições, pode-se extrair uma nova série temporal suavizada z_i tal que o seguinte problema de otimização seja satisfeito

$$\min \left\{ \sum_{i=1}^N [y_i - z_i]^2 + \lambda^2 \sum_{i=1}^{N-d} [\Delta^d z_i]^2 \right\} \quad (2.13)$$

onde d é a ordem do filtro, λ é o parâmetro de suavização e Δ^d é o operador de diferenças de ordem d , definido como segue:

$$\begin{aligned} d = 1; & \quad \Delta z_i = z_i - z_{i-1} \\ d = 2; & \quad \Delta^2 z_i = (z_i - z_{i-1}) - (z_{i-1} - z_{i-2}) \\ d = 3; & \quad \Delta^3 z_i = (z_i - z_{i-1}) - (z_{i-1} - z_{i-2}) - (z_{i-2} - z_{i-3}) \end{aligned} \quad (2.14)$$

A idéia do filtro é determinar um ponto “ótimo” onde as duas partes conflitantes da função minimizada (Equação 2.13) – fidelidade da curva suavizada z para com a série temporal y representada por $(y_i - z_i)^2$; medida de suavização de z $[\Delta^d z_i]^2$ – sejam balanceados.

O problema de otimização do filtro proposto pela Equação 2.13 consiste na minimização de uma função não-linear. Com o intuito de simplificar o problema, Moscheiov e Raveh (Mosheiov and Raveh 1997) propuseram uma nova função custo a ser minimizada, dada por:

$$\min \left\{ \alpha \sum_{i=1}^N |y_i - z_i| + (1 - \alpha) \sum_{i=1}^{N-2} |z_{i+2} - 2z_{i+1} + z_i| \right\} \quad (2.15)$$

tal que : $0 \leq z_i \leq z_{i+1}$ ou $z_i \geq z_{i+1}$

Fazendo a mudança de variáveis:

$$\begin{aligned} U_i - V_i &= y_i - z_i & U_i &\geq 0, V_i \geq 0 \\ X_i - G_i &= z_{i+2} - 2z_{i+1} - z_i & X_i &\geq 0, G_i \geq 0, i = 1, \dots, N-2 \end{aligned} \quad (2.16)$$

pode-se reescrever 2.15 como uma minimização em programação linear dada por:

$$\min \left\{ \alpha \sum_{i=1}^N (U_i + V_i) + (1 - \alpha) \sum_{i=1}^{N-2} (X_i + G_i) \right\}$$

tal que : $z_i \leq z_{i+1}$ ou $z_i \geq z_{i+1}$

$$\begin{aligned} z_i &\geq 0 & U_i &\geq 0, V_i \geq 0 \\ U_i - V_i &= y_i - z_i & X_i &\geq 0, G_i \geq 0, i = 1, \dots, N-2 \\ X_i - G_i &= z_{i+2} - 2z_{i+1} - z_i \end{aligned} \quad (2.17)$$

onde α é um numero real pertencente ao intervalo $[0,1]$ e selecionado pelo usuário. Esta solução impõe uma característica de monotonicidade no sinal suavizado, o que muitas vezes não se adequa à certos problemas. Recentemente, Paul Eilers publicou uma versão matricial do filtro onde a função minimizante da Equação 2.13 foi escrita como(Eilers 2003):

$$Q = |y - z|^2 + \lambda |Dz|^2 \quad (2.18)$$

onde a norma quadrática $|V|^2 = \sum_i V_i^2$ e D é uma matriz de dimensão $(N-d) \times N$ tal que $|Dz| = \sum \Delta^d z$. Por exemplo, para o valor de ordem $d=1$ e um vetor de pontos de comprimento $N=5$, temos:

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad (2.19)$$

Multiplicando a matriz D pelo vetor z , temos:

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix} = \begin{bmatrix} z_2 - z_1 \\ z_3 - z_2 \\ z_4 - z_3 \\ z_5 - z_4 \end{bmatrix} \quad (2.20)$$

Aplicando a norma quadrática $\|v\|^2 = \sum_i v_i^2$ no lado direito da Equação 2.20 tem-se exatamente a medida de suavização definida na Equação 2.13 por Whittaker. Derivando a Equação 2.18 em relação a z e igualando à zero, temos:

$$\begin{aligned} \partial Q / \partial z &= -2(y - z) + 2\lambda D^T D z = 0 \\ z &= (I + \lambda D^T D)^{-1} y \end{aligned} \quad (2.21)$$

Na Equação 2.21, I representa a matriz identidade de ordem N . Para a solução do sistema linear apresentado nessa equação, dois parâmetros são necessários: o peso da suavização λ e a ordem do filtro d . Estes dois parâmetros têm importância fundamental no resultado do filtro, devendo assim serem escolhidos cuidadosamente. A ordem d é um inteiro que controla o quão flexível é o filtro para seguir os dados de entrada (Figura 2.1). O parâmetro lambda – λ pesa o grau de suavização, fazendo o resultado mais ou menos rugoso (Figura 2.2). Uma propriedade deste filtro é que para altos valores de λ , o filtro comporta-se como um polinômio de grau $d-1$ onde o erro quadrático do filtro é minimizado. Uma característica fundamental da formulação do filtro Whittaker dada por Eilers é a facilidade de cálculo do erro de validação cruzada (cross-validation error - cvE) usando a matriz de

mapeamento do filtro ou comumente chamada de matriz chapéu (“hat” matrix)(Eilers 2003).

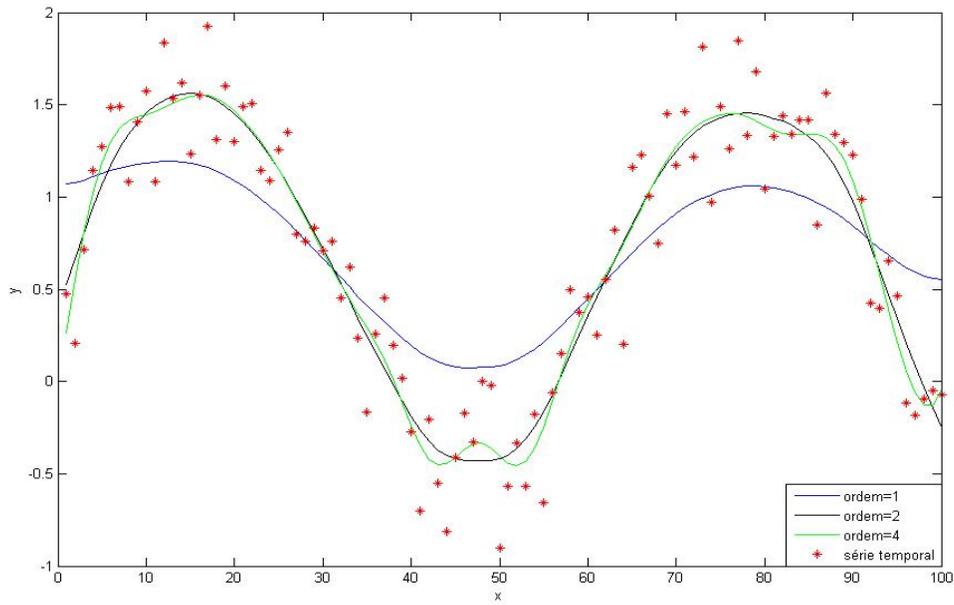


Figura 2.1 – Efeito da ordem no sinal suavizado.

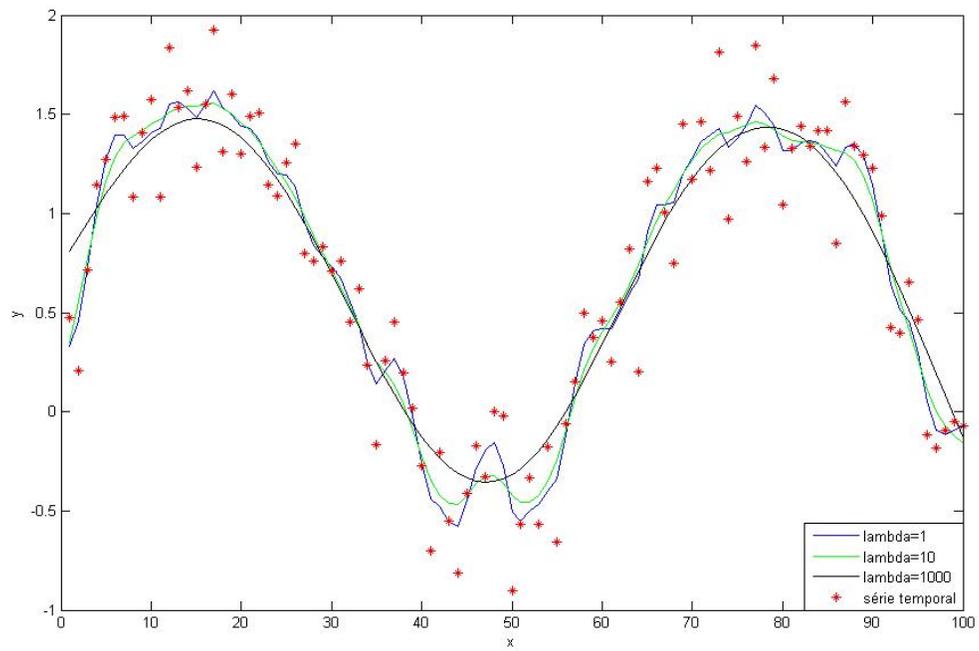


Figura 2.2- Efeito do valor de lambda no sinal suavizado utilizando ordem $d=2$.

2.4 Erro de Validação Cruzada

O erro de validação cruzada (Evc) foi introduzido com intuito de criar modelos robustos na predição de novos dados, fazendo com que o modelo tenha uma medida dessa predição com os mesmos dados usados na aprendizagem. Assim, definindo um algoritmo de aprendizagem como um “estudante”, existe um “professor” que fornece ao “estudante” um grupo de dados de treinamento. A meta do “estudante” é construir um modelo que melhor represente os dados que lhe foram fornecidos. Para cada modelo que o estudante gera, o “professor” dá uma nota de acordo com erro do modelo (o “professor” sabe o que deve ser a saída do modelo), e retorna essa nota ao “estudante”. Este por vez tenta adaptar o modelo novamente tal que sua nota sempre cresça. O esquema de aprendizado apresentado aqui é chamado de supervisionado, pelo fato do sistema de aprendizado (“estudante”) ser dirigido por um supervisor ou professor. Os algoritmos de aprendizagem supervisionados diferem na maneira que o “supervisor” mede o erro do “estudante” e em como o “estudante” atualiza seu modelo. Um dos métodos de medição do erro é o Evc, onde o “supervisor” remove uma das componentes do grupo de treinamento, fornece esse grupo ao “estudante” e mede o erro de predição do modelo para aquela componente que foi retirada do grupo de dados. A componente retirada na etapa anterior é colocada novamente no grupo de treinamento e outra componente é extraída. Esse processo é repetido para todas as componentes do grupo de treinamento e o erro de predição é avaliado para cada uma dessas componentes.

Assim, para um sistema do tipo $y = Wx$, onde y representa o vetor de saída do sistema de dimensão m , x os dados de entrada e W a matriz chapéu do sistema (“hat matrix”)(Hoaglin

and Welsch 1977), pode-se calcular o Evc do ponto estimado \hat{y}_{-i} tal que seu respectivo alvo a_i foi retirado do grupo de treinamento pela Equação 2.21:

$$a_i - \hat{y}_{-i} = \frac{a_i - y_i}{1 - W_{ii}} \quad (2.21)$$

$$Evc = \sqrt{\frac{\sum_i (a_i - \hat{y}_{-i})^2}{m}}$$

onde W_{ii} são os elementos da diagonal de W . Reescrevendo a equação do filtro Whittaker (Equação 2.12), temos:

$$z = (I + \lambda D^T D)^{-1} y = Hy \quad (2.22)$$

onde H é a matriz chapéu do filtro. Analogamente, o Evc do filtro Whittaker pode ser facilmente calculado pela Equação 2.21. A relação entre o Evc e os parâmetros do filtro pode ser vista na Figura 1 abaixo, o que sugere um método automático para a obtenção dos parâmetros.

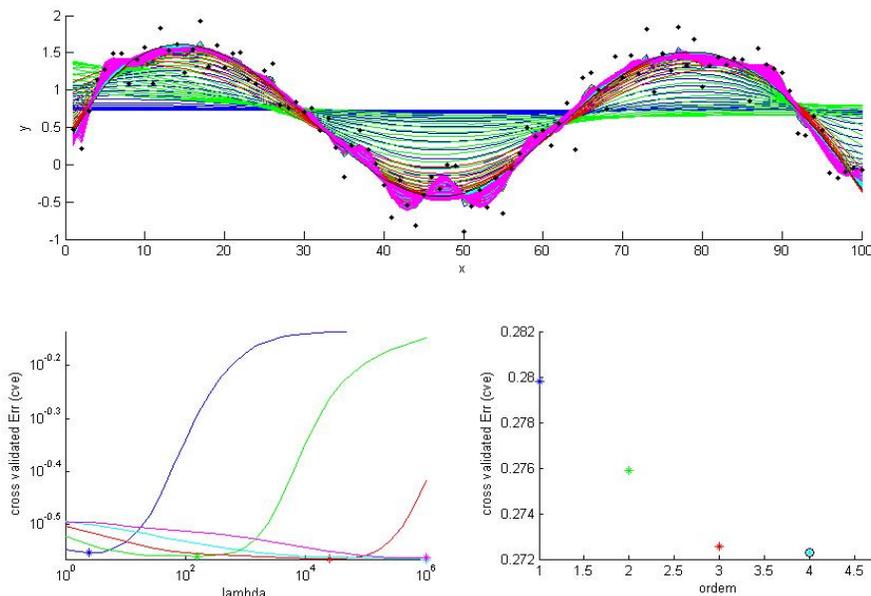


Figura 1 – Relação dos parâmetros do filtro Whittaker com o erro de validação cruzada. O resultado mostra que o menor Evc foi obtido com um valor de ordem igual a 4 e lambda 10^6 .

Capítulo 3

Entropia

Esse capítulo tem como objetivo introduzir conceitos de entropia, medidas de “distâncias” entre funções de densidade de probabilidade e principalmente apresentar as técnicas matemáticas utilizadas no aprendizado de máquina para sistemas adaptativos baseado na teoria da informação.

3.1 Entropia de Shannon

Introduzido como um conceito de desordem no estudo de comportamento termodinâmico dos gases, o termo entropia também foi usado por Claude Shannon em 1948 como uma medida de informação e/ou incerteza em canais de comunicação de dados(Shannon 1948). Do ponto de vista matemático, entropia pode ser definida de várias formas. Assim, para uma variável aleatória discreta que assume valores no conjunto $\{x_1, x_2, \dots, x_n\}$ com probabilidades $\{p_1, p_2, \dots, p_n\}$, a medida de entropia $H(p_1, p_2, \dots, p_n)$ deve satisfazer as seguintes propriedades(Shannon 1948; Martins 2005):

- i) H deve ser contínua em cada variável p_i para todo i .
- ii) Se os valores da variável aleatória são equiprováveis, ou seja, $p_i = \frac{1}{n}$, H cresce monotonicamente com n .

iii) Para uma variável aleatória X que assume valores $\{x_1, x_2, \dots, x_m\}$ com probabilidades $\{p_1, p_2, \dots, p_m\}$, sua entropia pode ser escrita como a soma ponderada da entropia de subconjuntos dessas probabilidades. Assim, para as probabilidades w_k assim definidas:

$$w_1 = \sum_{i=1}^{n_1} p_i; w_2 = \sum_{i=n_1+1}^{n_2} p_i; \dots; w_k = \sum_{i=n_1+n_2+\dots+1}^{n_k} p_i \quad (3.1)$$

pode-se (onde $n_k=m$) escrever a entropia como:

$$H(X) = H(w) + \sum_{j=1}^k w_j H\left(\sum_{i=1}^{n_j} p_i / w_j\right) \quad (3.2)$$

Satisfazendo todas as propriedades axiomáticas acima, Shannon definiu entropia como:

$$H_{sh}(X) = H(p_1, p_2, \dots, p_N) = -\sum_{i=1}^N p_i \log_2 p_i \quad (3.3)$$

A definição acima permite entender entropia como um valor real que é função da distribuição de probabilidade de uma variável aleatória e quantifica a sua incerteza. No caso da entropia de Shannon em que a base do logaritmo é 2, a entropia é medida em bits, tipicamente usada em sistemas de comunicação de dados. A entropia mede a quantidade média de “informação” transmitida por uma variável aleatória X através de um canal de comunicação, ou alternativamente, a quantidade de informação perdida nesse canal quando somente a distribuição de probabilidade é conhecida (Principe, Xu *et al.* 1999). Essa formulação foi a base para um novo campo chamado Teoria da Informação (TI), o qual ganhou tremenda consideração no desenvolvimento de sistemas de comunicação devido ao fato de permitir que duas perguntas fundamentais fossem respondidas: Qual é o menor “código” (mapeamento finito de uma variável aleatória em um grupo finito de caracteres, como por exemplo, uma determinada representação binária) para permitir a representação

de dados e qual é a capacidade máxima de “informação” que um canal de comunicação pode transferir.

Exemplo 3.1: Dada uma variável aleatória X com a seguinte distribuição de probabilidades:

$$\begin{aligned}\Pr(X = 1) &= 1/2, \\ \Pr(X = 2) &= 1/4, \\ \Pr(X = 3) &= 1/8, \\ \Pr(X = 4) &= 1/8.\end{aligned}\tag{3.1.1}$$

Pode-se enviar a informação de qual é o valor da variável X simplesmente codificando cada uma das possibilidades com 2 bits, tal que:

$$\begin{aligned}C(X = 1) &= 00, \\ C(X = 2) &= 01, \\ C(X = 3) &= 10, \\ C(X = 4) &= 11.\end{aligned}\tag{3.1.2}$$

Pode-se definir o comprimento médio de um código de uma variável aleatória X com distribuição de probabilidade $p(x)$ como:

$$L(C(x)) = \sum_x p(x)l(x)\tag{3.1.3}$$

onde $l(x)$ é o comprimento do carácter que representa x . Para o exemplo acima, $L(C(X))=2$ bits usando o código 3.1.2. Por outro lado, se os valores de X forem codificados pelo seguinte código:

$$\begin{aligned}C(X = 1) &= 0, \\ C(X = 2) &= 10, \\ C(X = 3) &= 110, \\ C(X = 4) &= 111.\end{aligned}\tag{3.1.4}$$

onde os valores de maior probabilidade são codificados com um menor número de bits, o comprimento médio desse código será $L(C(X))=1.75$ bits. Calculando a entropia da variável aleatória X , temos:

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = 1.75 \quad (3.1.5)$$

Neste caso, o valor da entropia é o mesmo valor do comprimento médio do código usado. Pode-se provar que entropia de Shannon associada a uma variável aleatória é o menor número médio de bits requeridos para representar a variável (Cover and Thomas 1991). A partir da entropia de Shannon, várias medidas relacionadas à entropia foram definidas, *e.g.* informação mútua. Algumas dessas medidas são descritas a seguir.

3.2 Entropia Conjunta

Sejam (X, Y) um par de variáveis aleatórias discretas caracterizadas pela sua distribuição de probabilidade conjunta $p(x, y)$. Define-se entropia conjunta desse par como:

$$H(X, Y) = -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j) \quad (3.4)$$

Assim, a entropia conjunta expressa incerteza sobre o resultado de um experimento conjunto relativo as duas variáveis aleatórias X e Y .

3.3 Entropia Condicional

Sejam X e Y as mesmas variáveis aleatórias discretas definidas previamente em 3.2. A probabilidade de $X=x_i$ condicionada pela ocorrência de $Y=y_j$ é denotada por $p(x_i|y_j)$ tal

que $\sum_{i=1}^N p(x_i|y_j) = 1$. A entropia de X condicionada a um valor da variável, por exemplo

$Y=y_j$, é definida por:

$$H(X|Y = y_j) = -\sum_{i=1}^N p(x_i|y_j) \log p(x_i|y_j) \quad (3.5)$$

De modo global, a entropia de X condicionada a Y é definida como a soma das entropias de X dado um valor de $Y=y_j$ ponderada pela probabilidade $p(y_j)$.

$$H(X|Y) = \sum_j p(y_j) H(X|Y = y_j) \quad (3.6)$$

Substituindo 3.5 em 3.6 temos:

$$H(X|Y) = -\sum_i \sum_j p(y_j) p(x_i | y_j) \log p(x_i | y_j) \quad (3.7)$$

A probabilidade x_i condicionada por y_j pode ser escrita em função da probabilidade conjunta do par (x_i, y_j) tal que:

$$p(x_i | y_j) = \frac{p(x_i, y_j)}{p(y_j)} \quad (3.8)$$

Rearranjando a Equação 3.8 e substituindo em 3.7 temos:

$$H(X|Y) = -\sum_{i=1}^N \sum_{j=1}^M p(y_j) p(x_i | y_j) \log p(x_i | y_j) \quad (3.9)$$

A entropia condicional é a incerteza sobre o resultado de X tal que o valor de Y é conhecido.

3.4 Propriedades da Entropia Condicional e Conjunta

Definidas em 3.2 e 3.3, as entropias conjunta e condicional apresentam as seguintes propriedades:

- P1. $H(X, Y) \geq 0$ e $H(X | Y) \geq 0$
- P2. $H(X, Y) = H(X) + H(Y | X) = H(Y) + H(X | Y)$
- P3. $H(X, Y) \leq H(X) + H(Y)$
- P4. $H(X | Y) \leq H(X)$
- P5. $H(X, Y) \geq \max\{H(X), H(Y)\}$
- P6. $H(X, Y) = H(X) + H(Y)$ se, e só se, X e Y são variáveis independentes
- P7. $H(X | Y) = H(X)$ se, e só se, X e Y são variáveis independentes

3.5 Informação Mútua

Dado que a entropia condicional $H(X|Y)$ definida pela Equação 3.9 é a entropia de X quando se conhece Y , define-se a redução da incerteza de X devido Y como informação mútua $I(X;Y)$, dada pela Equação 3.10.

$$I(X;Y) = H(X) - H(X | Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3.10)$$

A informação mútua obedece à propriedade de simetria, é não-negativa e igual a zero se e somente se Y e X são independentes. Vários algoritmos usam a informação mútua para a extração de informação de sinais, como por exemplo no método da análise das

componentes independentes (ACI). ACI é um método estatístico que extrai as componentes de um vetor aleatório que são mais estatisticamente independentes possíveis. Um dos possíveis critérios que o método ACI pode usar é a minimização da informação mútua (Comon 1992).

3.6 Entropia de Rényi

Definida como a generalização matemática da entropia de Shannon, a forma geral da entropia de Rényi é dada pela Equação 3.11 (Rényi 1976):

$$H_{R\alpha} = \frac{1}{1-\alpha} \ln \int_{-\infty}^{\infty} f(x)^\alpha dx \quad (3.11)$$

Na Equação 3.11, $\alpha \geq 0, \alpha \neq 1$ define a ordem da entropia e $f(x)$ é a função densidade de probabilidade (fdp) da variável aleatória X . A entropia de Rényi carrega toda a fundamentação teórica definida pela a entropia de Shannon descrita na seção anterior, fazendo desnecessária uma repetição da descrição. Para uma variável aleatória de distribuição discreta, a entropia de Rényi é definida pela Equação 3.12:

$$H_{R\alpha} = \frac{1}{1-\alpha} \ln \sum_i p_i^\alpha \quad (3.12)$$

onde p_i são os valores da função densidade de probabilidade da variável aleatória. Redefinida como um caso particular da entropia de Rényi, a entropia de Shannon pode ser vista como o limite de $H_{R\alpha}$ quando $\alpha \rightarrow 1$ (Martins 2005):

$$\begin{aligned}
\lim_{\alpha \rightarrow 1} H_\alpha &= \lim_{\alpha \rightarrow 1} \frac{1}{1-\alpha} \ln \sum_i p_i^\alpha \\
&= \lim_{\alpha \rightarrow 1} \frac{\frac{\partial}{\partial \alpha} \sum_i p_i^\alpha}{\frac{\partial}{\partial \alpha} \frac{1}{1-\alpha}} \\
&= \lim_{\alpha \rightarrow 1} \frac{\sum_i p_i^\alpha \ln p_i}{-1} \\
&= -\lim_{\alpha \rightarrow 1} \frac{\sum_i p_i^\alpha \ln p_i}{\sum_i p_i^\alpha} \\
&= -\frac{\sum_i p_i \ln p_i}{\sum_i p_i} = -\sum_i p_i \ln p_i
\end{aligned} \tag{3.13}$$

3.7 “Distância” de Kullback-Leibler

Definida como uma semi-distância (não apresenta a propriedade de simetria) dentro da classe de Ali-Silvey de medidas de teoria da informação, a “distância” de Kullback-Leibler, também conhecida como entropia relativa, é um quantificador da diferença entre duas funções de densidade de probabilidade. Assim, para as seguintes funções de densidade de probabilidade $g(x)$ e $f(x)$, pode-se escrever a distância KL como (Kullback and Leibler 1951):

$$KL = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} d(x) \tag{3.14}$$

Essa medida tem como propriedade fundamental ser sempre positiva e ser zero somente quando $f(x)=g(x)$. Vastamente usada em sistemas de aprendizagem estatística, a distância de Kullback-Leibler pode ser vista como um caso particular da distância de Csiszár (dada pela Equação 3.15) entre duas fdp’s quando a função $\Psi(.) = -\log(.)$ (Erdogmus and Principe 2002) (Csiszár and Körner 1981) :

$$D_C = \int_{-\infty}^{\infty} f(x) \Psi \left(\frac{g(x)}{f(x)} \right) dx \quad (3.15)$$

3.8 Divergente de Rényi

Definido como uma família de medidas de “distância” entre funções de densidade de probabilidade, o divergente de Rényi é uma extensão da distância KL assim como a entropia de Rényi é uma generalização da entropia de Shannon. Assim, para as seguintes funções de densidade de probabilidade $g(x)$ e $f(x)$ o divergente de Rényi de ordem α é dado como (Erdogmus, 2002):

$$D_{R\alpha}(g, f) = \frac{1}{\alpha - 1} \log \int_{-\infty}^{\infty} g(x) \left(\frac{g(x)}{f(x)} \right)^{\alpha-1} dx \quad (3.16)$$

3.9 Aprendizado Teórico de Informação

Introduzido por Principe (Principe, Xu *et al.* 1999), o termo Aprendizado Teórico de Informação (Information-Theoretic Learning - ITL) foi usado como um subcampo da teoria da informação aplicado em sistemas adaptativos, e.g. algoritmos de projeções para problemas de separação cega de fontes (o sinal principal é composto por mais de um subsinal totalmente desconhecido pelo receptor e este tenta separá-lo) e treinamento de redes neurais artificiais (RNA) (Erdogmus and Principe 2002), onde a informação usada pelo sistema é extraída diretamente dos dados com o intuito de promover o aprendizado do sistema. A idéia de integrar a entropia de Rényi de segunda ordem (Rényi 1976) com janelas de Parzen para estimação da função de densidade de probabilidade (Parzen 1962) é

a alma da aplicação da ITL em sistemas adaptativos e conseqüentemente na proposta deste trabalho, sendo necessário um aprofundamento do assunto neste capítulo.

3.9.1 Aprendizado de Máquina Baseado em Teoria da Informação

Considere um mapeamento linear/não-linear paramétrico $g: R^k \rightarrow R^l$ de uma vetor aleatório $X \in R^k$ descrito pela seguinte Equação 3.17:

$$Y = g(X, W) \quad (3.17)$$

tal que Y também é uma vetor aleatório tal que $Y \in R^l$ (onde em geral k é diferente de l) e W é uma matriz real de parâmetros. Máquinas de aprendizado (e.g. redes neurais artificiais, filtros adaptativos) têm como objetivo ajustar a matriz de parâmetros W de acordo com um critério definido, por exemplo, a minimização do erro médio quadrático (Mean Square Error - MSE) entre a saída do sistema y_i e o seu respectivo alvo a_i para onde o sistema é direcionado.

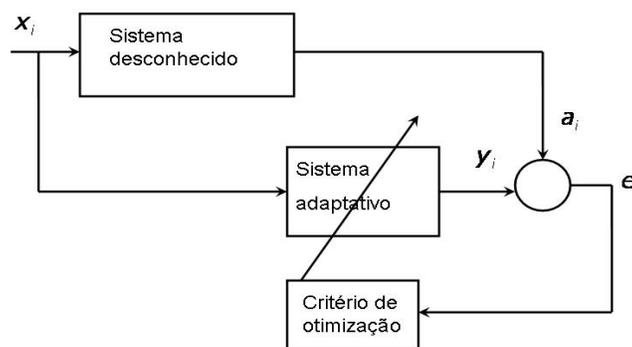


Figura 3.1- Sistema Adaptativo genérico

Assim, para o sistema adaptativo ilustrado na Figura 3.1 com N amostras de dados, pode-se calcular o MSE como:

$$e_i = a_i - y_i$$

$$MSE = \frac{\sum_{i=1}^N e_i^2}{N} \quad (3.18)$$

O critério MSE é equivalente ao de máxima verossimilhança quando o erro do sistema apresenta uma distribuição gaussiana de probabilidade. A máxima verossimilhança é aquela em que a saída do sistema apresenta máxima probabilidade de ocorrência quando condicionada aos respectivos dados de entrada apresentados ao sistema. Denotando por $f(e_i)$ a função densidade de probabilidade do erro, pode-se usar a máxima verossimilhança como critério de otimização pelo qual os parâmetros do mapeamento $g(X, W)$ são estimados. Definida como o produto das probabilidades associadas aos valores dos erros, a verossimilhança de um modelo é dada por:

$$V = \prod_{i=1}^N f(e_i) \Delta \quad (3.19)$$

onde Δ é uma vizinhança infinitesimal contínua da função de densidade de probabilidade ao redor de e_i resultante da discretização (Papoulis 1991; Liano 1996). Por ser constante essa vizinhança pode ser ignorada e a maximização da verossimilhança pode ser vista como a minimização do seu logaritmo negativo, como segue:

$$\text{Minimize} \left(\sum_{i=1}^N -\log f(e_i) \right) \quad (3.20)$$

Assumindo que o erro tem distribuição gaussiana definida pela Equação 3.21:

$$f(e) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-e^2}{2\sigma^2}\right) \quad (3.21)$$

Substituindo a Equação 3.21 em 3.20, temos:

$$\text{Minimize} \left(\sum_{i=1}^N \frac{1}{2} e_i^2 \right) + \frac{N}{2} \log 2\pi \quad (3.22)$$

Que é equivalente a:

$$\text{Minimize} \left(\sum_{i=1}^N e_i^2 \right) \quad (3.23)$$

Como visto na Equação 3.23, a minimização do MSE pode ser visto como um método paramétrico onde se assumiu que o erro tem distribuição gaussiana (Papoulis 1991). A não-gaussianidade é uma característica comum em problemas de dados reais (*e.g.* séries temporais biológicas), onde momentos de alta ordem são fundamentais na caracterização da distribuição dos dados (Papoulis 1991; Erdogmus and Principe 2002; Jenssen, Erdogmus *et al.* 2005). Tendo em vista a limitação do método de minimização do MSE, a ITL propõe uma metodologia generalizada baseada na entropia de Rényi onde não é feita nenhuma prévia suposição sobre função de densidade de probabilidade dos dados e/ou erro. A estimação da função de densidade de probabilidade dos dados é feita utilizando o método não-paramétrico de Parzen, onde uma função kernel $K(\cdot)$ projeta a variável aleatória X no chamado espaço de características.

3.9.2 Estimador não-paramétrico de função de densidade de probabilidade

O método mais simples e conhecido para a estimação uma função de densidade de probabilidade é o histograma. Construído a partir de uma tabela de freqüência, o histograma é uma ferramenta visual usada na análise da tendência central e na dispersão dos dados, mas pouco preciso quando utilizado em um pequeno conjunto de dados. Um

procedimento mais sofisticado para estimação não-paramétrica da fdp é o método das janelas de Parzen, onde uma função kernel que satisfaz a condição

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad (3.24)$$

é usada para estimar a densidade dos dados. Assim, considerando um grupo de amostras geradas por variáveis aleatórias i.i.d. (independentes e identicamente distribuídas) x_1, x_2, \dots, x_N com uma função de densidade de probabilidade $f(\cdot)$, o método das janelas de Parzen consiste em estimar a fdp por meio da seguinte função:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (3.25)$$

onde $K(\cdot)$ é a função kernel e h é a largura da sua janela ou simplesmente largura do kernel, também conhecido como parâmetro de suavização. No caso de um kernel gaussiano, a equação da janela de Parzen pode ser escrita como (Parzen 1962):

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi\sigma N}} \sum_{i=1}^N \exp\left\{-\frac{\|x - x_i\|^2}{2\sigma^2}\right\} \quad (3.26)$$

O parâmetro σ controla a dispersão da função gaussiana centrada na variável x_i , fazendo o resultado da estimação drasticamente diferente para diferentes valores desse parâmetro. Ao fazer σ muito pequeno, o método modela o ruído dos dados resultando em uma estimação “rugosa” da fdp (Figura 2a). Por outro lado, fazer σ grande faz com que a distribuição seja muito suave, perdendo informação dos dados na estimação (Figura 2c). Um método usado aqui para obter o valor da largura do kernel σ é baseado nos próprios dados do modelo, que no caso unidimensional é dado por (Silverman 1986):

$$\sigma = \sigma_x \left[\frac{4}{3N} \right]^{\frac{1}{5}} \quad (3.27)$$

onde σ_x é a covariância dos dados. A largura do kernel é um parâmetro relacionado com a convergência de vários algoritmos adaptativos descritos (Erdogmus and Principe 2002; Santos, Sa *et al.* 2005), e é parte crucial deste processo em que é usado.

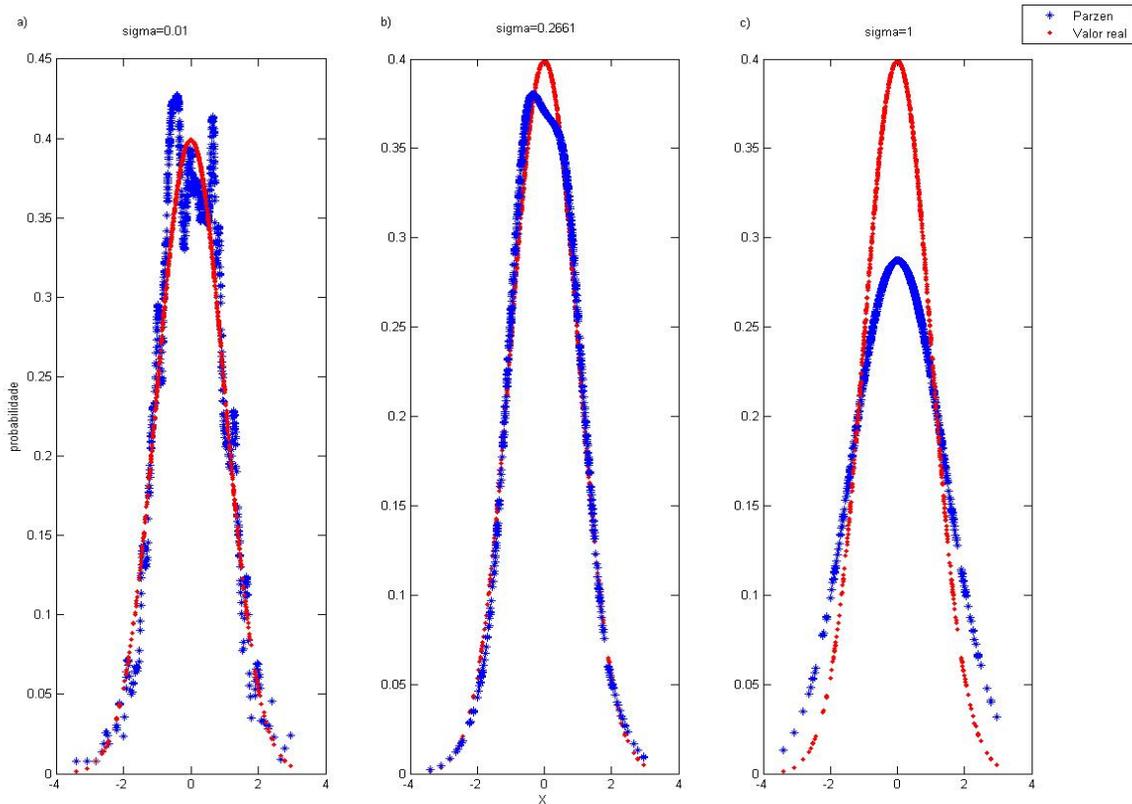


Figura 2. Efeito do comprimento do kernel – Estimação da função densidade de probabilidade de um vetor aleatório de distribuição gaussiana. a) Sigma subestimado, onde artefatos do ruído são refletidos para a fdp. b) Sigma calculado para uma distribuição gaussiana Eq. 15. c) Sigma superestimado, fazendo com que a fdp seja suavizada.

3.9.3 Entropia Quadrática de Rényi

A entropia quadrática de Rényi associada com uma função de densidade de probabilidade $f(x)$ é dado pela Equação 3.28 abaixo (Rényi 1976).

$$H_{R2} = -\ln \int_{-\infty}^{\infty} f(x)^2 dx \quad (3.28)$$

Suponha que $\{x_1, x_2, \dots, x_N\}$ sejam observações independentes da variável aleatória X . Usando o método das janelas de Parzen descrito anteriormente, pode-se reescrever a Equação 3.28 da seguinte forma (Principe, Xu *et al.* 1999):

$$H_{R2} = -\ln \int \frac{1}{N} \sum_{i=1}^N K_{\sigma^2}(x, x_i) \frac{1}{N} \sum_{i=1}^N K_{\sigma^2}(x, x_i) dx \quad (3.29)$$

O argumento do logaritmo natural da Equação 3.29 pode ser visto como a convolução das duas funções gaussianas de mesmo desvio padrão, e que se reduz a uma função gaussiana com desvio padrão dobrado. Assim, para o argumento do logaritmo da Equação 3.29 tem-se:

$$\int \frac{1}{N} \sum_{i=1}^N K_{\sigma^2}(x, x_i) \frac{1}{N} \sum_{i=1}^N K_{\sigma^2}(x, x_i) dx = \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N K_{2\sigma^2}(x_j, x_i) \quad (3.30)$$

Substituindo 3.30 em 3.29, tem-se o valor estimado para a entropia quadrática de Rényi.

$$H_{R2} = -\ln \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N K_{2\sigma^2}(x_j, x_i) = -\ln IP(x) \quad (3.31)$$

O argumento do logaritmo natural da Equação 3.31 $IP(x)$ é chamado de Informação Potencial devido a analogia com energia potencial gerada por um conjunto de partículas em um sistema físico (Erdogmus and Principe 2002). Por meio da Equação 3.31, pode-se calcular a entropia quadrática de Rényi de um conjunto de dados onde nenhuma suposição prévia sobre a distribuição de densidade dos dados é feita. Métodos supervisionados e não-supervisionados de aprendizado de máquina (e.g. treinamento de redes neural artificiais e algoritmos de análise de componentes independentes – Independent Component Analysis -

ICA) usando o gradiente da Equação 3.31 vêm sendo vastamente usados em problemas de processamento de sinais e aprendizado de máquina e podem ser vistos nos seguintes trabalhos (Principe, Xu *et al.* 1999; Erdogmus and Principe 2000; Erdogmus and Principe 2002; Jenssen, Erdogmus *et al.* 2005; Santos, Sa *et al.* 2005). Estes algoritmos partem do princípio que a minimização da entropia do erro do sistema pode ser vista como a minimização do divergente de Rényi entre a fdp da entrada e a fdp da saída (Erdogmus and Principe 2002).

Capítulo 4

Suavizador Automático Aplicado a Séries Temporais Biológicas

4.1 Introdução

As técnicas de parametrização de Sistemas-S a partir dos perfis metabólicos assumem que estes perfis são livres de incerteza de medição (incerteza chamada daqui em diante neste trabalho de ruído), e conseqüentemente as suas derivadas em relação ao tempo (Kikuchi, Tominaga *et al.* 2003; Chou, Martens *et al.* 2006). Como comentado na introdução deste trabalho, perfis metabólicos obtidos em experimentos *in vivo* de ressonância nuclear magnética não satisfazem esta suposição, sendo assim necessário um tratamento do sinal antes da tentativa de parametrização do sistema (Voit and Almeida 2004). O problema de extração de uma curva suave a partir dos pontos de uma série temporal metabólicas vem sendo tratado nos últimos quatro anos onde várias técnicas como redes neurais, filtro Savitzky-Golay e máquinas de vetores de suporte – MVS (Support Vectors Machines) foram aplicadas (Almeida and Voit 2003; Borges, Voit *et al.* 2004; Voit and Almeida 2004; Borges, Vilela *et al.* 2006). Todas essas técnicas esbarraram em algum particular

empecilho, impedindo o uso em uma suavização automática. Em geral, foram encontrados artefatos na derivada gerados pela função de transferência imposta por cada método, e.g. a função sigmoïdal de uma rede neural. Além disso, o problema da suavização é agravado pela descontinuidade da variância do erro de medição presente nos perfis metabólicos (Imoto and Konishi 2003) causada pela maquinaria molecular quando a concentração de um dos metabólicos presente no processo de medição tende para zero (Figura 4.1). Assim, para aplicar um suavizador automático em séries temporais metabólicas, dois problemas básicos devem ser resolvidos: a parametrização automática do suavizador, seja ele qual for, e um tratamento da heterogeneidade do erro de medição, impedindo que o sinal suavizado apresente distorções na parte em que o ruído do perfil metabólico tem baixa variância. Esta distorção é causada por um viés na suavização proveniente de partes em que o perfil metabólico tem uma maior variância do ruído (Figura 4.2).

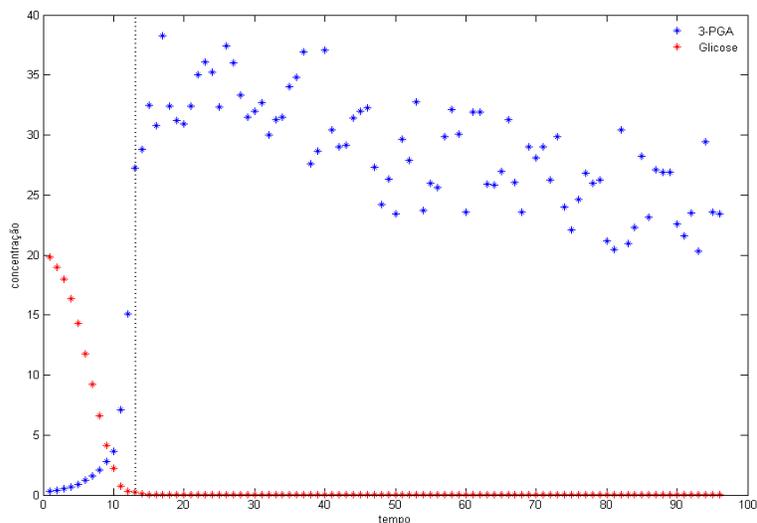


Figura 4.1 – Descontinuidade da variância do ruído. O ruído da série temporal 3-Gliceraldeído fosfatado (3-PGA) aumenta a variância no momento em que a concentração da Glicose tende à zero.

Uma tentativa de resolver o problema da suavização automática foi desenvolvida utilizando um algoritmo que procura janelas na série temporal onde o erro de validação cruzada do

filtro Whittaker é mínimo. Os parâmetros do filtro são determinados a partir de uma busca exaustiva dentro de intervalos determinados empiricamente. Esta solução mostrou falhas em alguns dos perfis metabólicos por duas principais razões: a dependência do erro de validação cruzada em relação a amplitude do sinal e a otimização do parâmetro lambda do filtro Whittaker. Após este trabalho concluiu-se que uma medida de erro independente da escala do sinal e uma otimização dos parâmetros do filtro Whittaker eram necessárias e é nesta direção que parte do trabalho aqui apresentado foi desenvolvido.

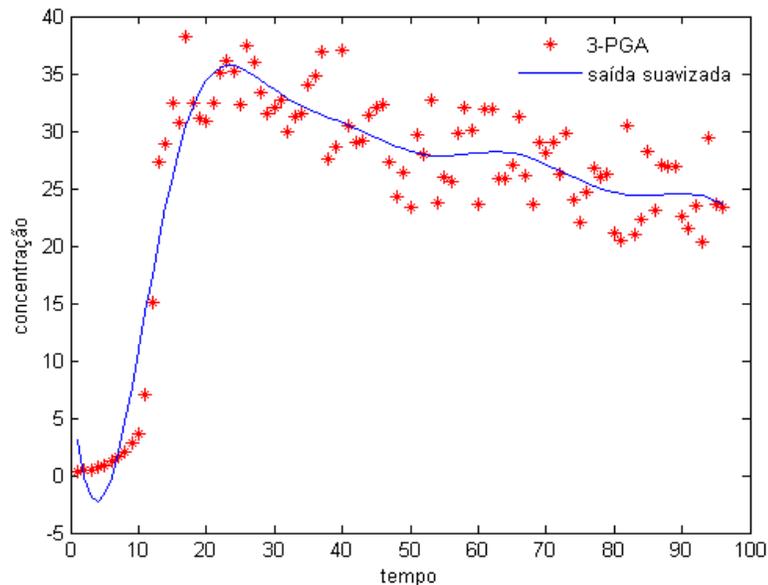


Figura 4.2 – Viés paramétrico na suavização causado pela escolha dos parâmetros do filtro Whittaker em função de uma parte do sinal com maior variância do ruído.

Os tópicos a seguir apresentam dois algoritmos de otimização dos parâmetros do filtro Whittaker baseados no gradiente de duas diferentes funções custo do erro de validação cruzada e um algoritmo de segmentação de sinal baseado em janelas com a entropia do erro mínima.

4.2 Entropia do Erro de Validação Cruzada

Com o intuito de superar a limitação em relação a escala que o erro de validação cruzada proporciona e enriquecer a saída do suavizador com estatísticas de ordem superior, o método proposto utiliza a entropia quadrática de Rényi do Evc como função objetivo no processo de otimização dos parâmetros do filtro Whittaker. Similar ao algoritmo MEE - Minimization of Error Entropy (Minimização da Entropia do Erro) – usado em treinamento de redes neurais artificiais (Principe, Xu *et al.* 1999; Erdogmus and Principe 2002; Santos, Sa *et al.* 2005), o MCVEE – Minimization of Cross-Validation Error Entropy (Minimização da Entropia do Erro de Validação Cruzada) usa o gradiente da informação potencial (Equação 3.30) do erro de validação cruzada no processo de minimização da entropia deste erro. A minimização da entropia do erro de saída de um sistema pode ser vista como a retirada da informação do sinal de erro, maximizando a informação da saída desejada do sistema em relação à sua entrada (Erdogmus and Principe 2000). Como visto no capítulo 2, a entropia quadrática de Rényi de uma variável aleatória $X = \{x_1, x_2, \dots, x_N\}$ é dada pela Equação 4.1 onde a função densidade de probabilidade é estimada pelo método das Janelas de Parzen.

$$H_{R2} = -\ln \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N K_{2\sigma^2}(x_j, x_i) \quad (4.1)$$

onde o kernel gaussiano K é dado por:

$$K_{2\sigma^2}(x_j, x_i) = \frac{1}{\sqrt{2\pi\sigma N}} \sum_{i=1}^N \exp\left\{ \frac{-\|x_j - x_i\|^2}{2\sigma^2} \right\} \quad (4.2)$$

Como o logaritmo é uma função monotônica, o processo de minimização da entropia quadrática de Rényi pode ser atingido pela maximização do argumento do logaritmo

neperiano da Equação 4.1, argumento este já definido como informação potencial (Equação 4.3). Como a ordem d é um parâmetro implícito na equação do filtro Whittaker, o gradiente é somente obtido na direção de lambda λ .

$$IP(Ecv) = \frac{1}{\sqrt{2\pi\sigma N^2}} \sum_{j=1}^N \sum_{i=1}^N \exp\left\{-\frac{\|Ecv_j - Ecv_i\|^2}{2\sigma^2}\right\} \quad (4.3)$$

Derivando a Equação 4.3 em relação à λ tem-se:

$$\frac{\partial IP(Ecv)}{\partial \lambda} = \frac{1}{2\pi N^2 \sigma^2} \sum_{i=1}^N \sum_{j=1}^N (Ecv_j - Ecv_i) K_{2\sigma^2}(Ecv_i, Ecv_j) \left[\frac{\partial Ecv_i}{\partial \lambda} - \frac{\partial Ecv_j}{\partial \lambda} \right] \quad (4.4)$$

Como descrito no capítulo 2, o erro de validação cruzada do filtro Whittaker pode ser facilmente calculado por meio da matriz chapéu H , dada por (Eilers 2003):

$$H = [I + \lambda D^T D]^{-1} \quad (4.5)$$

Assim, tem-se que cada componente do vetor do erro de validação cruzada em que uma amostra de treinamento é excluída por vez (leave-one-out) é dado pela Equação 4.6 abaixo.

$$Evc_i = \frac{y_i - z_i}{1 - H_{ii}} \quad (4.6)$$

O próximo passo é fundamental no equacionamento do algoritmo, que permite expressar a derivada do erro de validação cruzada em relação ao parâmetro λ . Assim, pode-se escrever a Equação 4.6 em sua forma matricial, dada por:

$$\begin{aligned} Evc &= [y - z] \circ [1 - dgH]^{-1} \\ &= [y - [I + \lambda D^T D]^{-1} y] \circ [1 - dg([I + \lambda D^T D]^{-1})]^{-1} \end{aligned} \quad (4.7)$$

Derivando o vetor do Evc (Equação 4.7) em relação a λ , tem-se:

$$\begin{aligned}
\frac{\partial Evc}{\partial \lambda} &= \\
&= \frac{\partial [y - z]}{\partial \lambda} \circ [1 - dgH]^{b-1} + \frac{\partial [1 - dgH]^{b-1}}{\partial \lambda} \circ [y - z] = \\
&= \frac{\partial [y - [I + \lambda D^T D]^{-1} y]}{\partial \lambda} \circ [1 - dg[I + \lambda D^T D]^{-1}]^{b-1} + \\
&\quad \frac{\partial [1 - dg[I + \lambda D^T D]^{-1}]^{b-1}}{\partial \lambda} \circ [y - [I + \lambda D^T D]^{-1} y] = \\
&= -\frac{\partial [[I + \lambda D^T D]^{-1} y]}{\partial \lambda} \circ [1 - dg[I + \lambda D^T D]^{-1}]^{b-1} + \\
&\quad \left(\frac{\partial [1 - dg[I + \lambda D^T D]^{-1}]^{b-1}}{\partial \lambda} \right) \circ [y - [I + \lambda D^T D]^{-1} y]
\end{aligned} \tag{4.8}$$

Resolvendo separadamente as derivadas da Equação 4.8 utilizando a propriedade matricial

PA.10 (Apêndice A) tem-se as seguintes equações 4.9 e 4.10

$$\frac{\partial [[I + \lambda D^T D]^{-1} y]}{\partial \lambda} = -\left([I + \lambda D^T D]^{-1} \left(\frac{\partial [I + \lambda D^T D]}{\partial \lambda} \right) [I + \lambda D^T D]^{-1} \right) y \tag{4.9}$$

$$\begin{aligned}
\frac{\partial [1 - dg[I + \lambda D^T D]^{-1}]^{b-1}}{\partial \lambda} &= \\
&= -[1 - dg[I + \lambda D^T D]^{-1}]^{b-2} \circ \frac{\partial [1 - dg[I + \lambda D^T D]^{-1}]}{\partial \lambda} \\
&= -[1 - dg[I + \lambda D^T D]^{-1}]^{b-2} \circ dg \left(\frac{\partial [I + \lambda D^T D]^{-1}}{\partial \lambda} \right) \\
&= [1 - dg[I + \lambda D^T D]^{-1}]^{b-2} \circ dg \left([I + \lambda D^T D]^{-1} \frac{\partial [I + \lambda D^T D]}{\partial \lambda} [I + \lambda D^T D]^{-1} \right)
\end{aligned} \tag{4.10}$$

Para acelerar o processo de otimização, pode-se substituir o λ por 10^{λ} (Eilers 2003). Assim,

tem-se que

$$\frac{\partial [I + 10^\lambda D^T D]}{\partial \lambda} = 10^\lambda \ln 10 (D^T D) \quad (4.11)$$

Substituindo 4.11 e 4.10 tem-se que

$$\begin{aligned} \frac{\partial [1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-1}}{\partial \lambda} &= \\ &= 10^\lambda \ln 10 \left([1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-2} \circ dg \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) \right) \end{aligned} \quad (4.12)$$

Substituindo 4.11 em 4.9 tem-se que

$$\frac{\partial [I + 10^\lambda D^T D]^{-1} y}{\partial \lambda} = 10^\lambda \ln 10 \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) y \quad (4.13)$$

Com todas as derivadas resolvidas, substitui-se 4.13 e 4.12 em 4.8

$$\begin{aligned} \frac{\partial Ecv}{\partial \lambda} &= \\ &= 10^\lambda \ln 10 \left\{ \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) y \circ [1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-1} + \right. \\ &\quad \left. + \left([1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-2} \circ dg \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) \right) \circ [y - [I + 10^\lambda D^T D]^{-1} y] \right\} \end{aligned} \quad (4.14)$$

Pela propriedade PA.7, tem-se que:

$$[1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-2} = [1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-1} \circ [1 - dg[I + 10^\lambda D^T D]^{-1}]^{p-1} \quad (4.15)$$

Substituindo 4.15 em 4.14 e rearranjando a expressão de acordo com as propriedades PA.1 e PA.3, tem-se que a derivada do erro de validação cruzada em relação à lambda é:

$$\begin{aligned}
\frac{\partial E_{cv}}{\partial \lambda} &= \\
&= 10^\lambda \ln 10 \left\{ \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) y \circ [1 - dg [I + 10^\lambda D^T D]^{-1}]^{-1} + \right. \\
&\quad \left. + \left([1 - dg [I + 10^\lambda D^T D]^{-1}]^{-1} \circ dg \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) \right) \circ \right. \\
&\quad \left. \left([y - [I + 10^\lambda D^T D]^{-1} y] \circ [1 - dg [I + 10^\lambda D^T D]^{-1}]^{-1} \right) \right\}
\end{aligned} \tag{4.16}$$

Finalmente, substituindo 4.7 em 4.16, tem-se:

$$\begin{aligned}
\frac{\partial E_{cv}}{\partial \lambda} &= \\
&= 10^\lambda \ln 10 \left[1 - dg [I + 10^\lambda D^T D]^{-1} \right]^{-1} \circ \left\{ \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) y + \right. \\
&\quad \left. + \left(dg \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) \circ E_{cv} \right) \right\}
\end{aligned} \tag{4.17}$$

Substituindo o vetor de derivada do erro de validação cruzada no gradiente da informação potencial, tem-se que a atualização do lambda pode ser dada por:

$$\lambda_{i+1} = \lambda_i + \eta \frac{\partial E_{vc}}{\partial \lambda} \tag{4.18}$$

onde η é uma constante no intervalo $(0, 1]$ chamada taxa de aprendizagem. Uma procura exaustiva em um intervalo viável é feita para a ordem d do filtro. Experimentos com o filtro mostraram que a ordem ótima flutua no intervalo $[1, 6]$, sendo este o intervalo usado nas avaliações deste trabalho. Para cada valor de ordem d avaliado, lambda é otimizado de acordo com a regra delta apresentada acima (Equação 4.18) até que um dos critérios seja atingido: aumento da entropia, numero máximo de “épocas” ou um valor gradiente mínimo previamente definido. A superfície da informação potencial e de sua respectiva entropia para uma dada série temporal (Figura 4.4) pode ser vista na Figura 4.3.

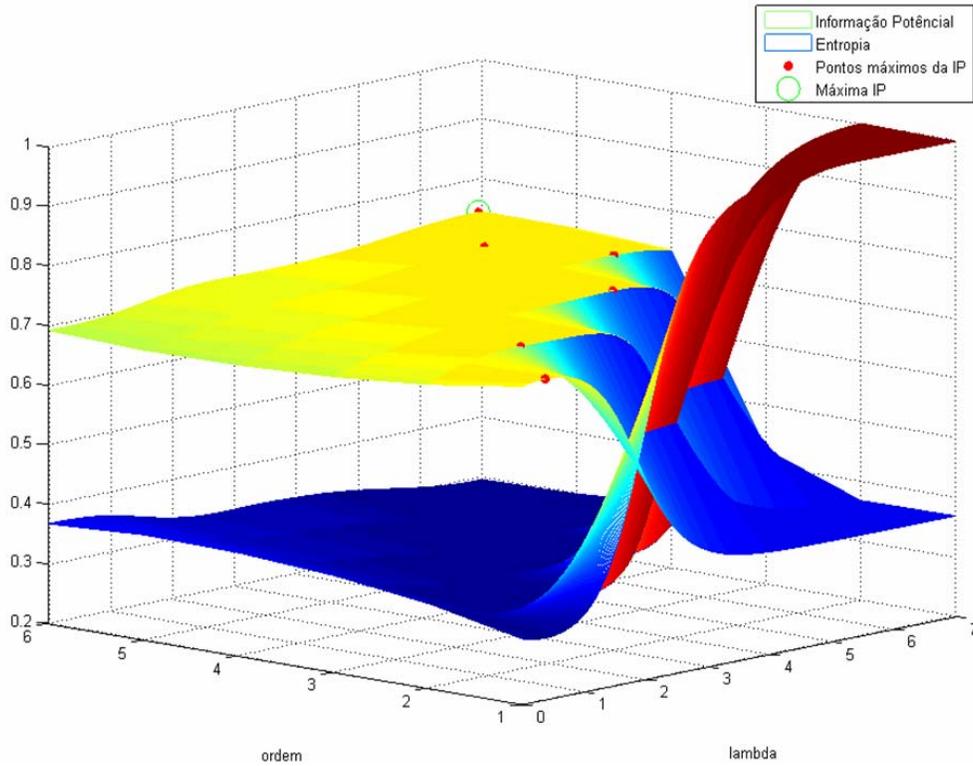


Figura 4.3 – Superfície da informação potencial e da entropia. Os pontos em vermelho representam os valores máximos da informação potencial para cada valor de ordem avaliado.

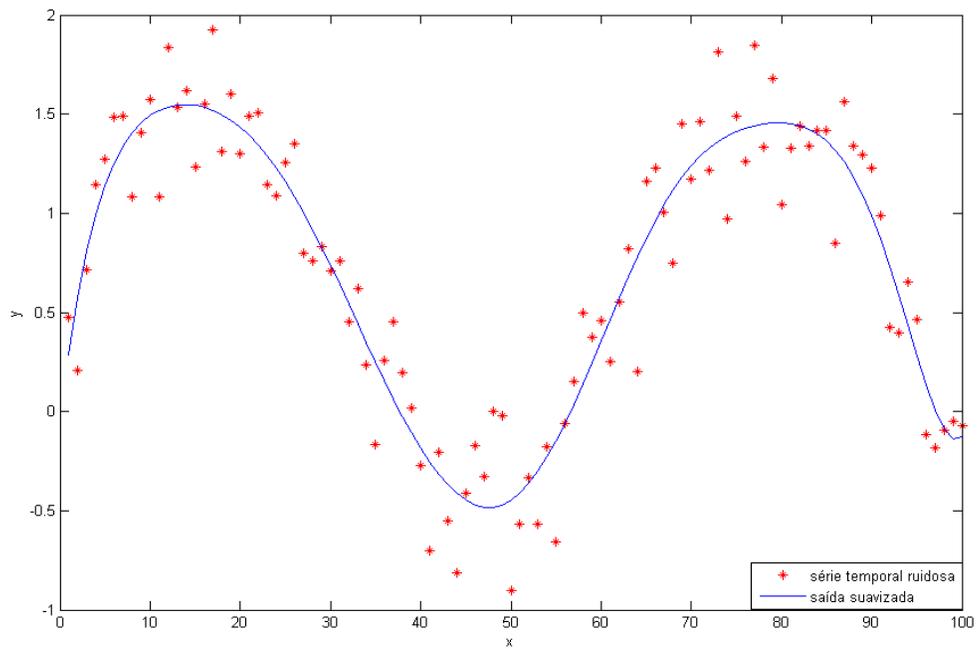


Figura 4.4 – Série temporal e curva suavizada usando o ponto ótimo dos parâmetros do filtro Whittaker (em verde – Figura 4.4)

A parada precoce do algoritmo em um ponto de mínimo local é um problema típico de algoritmos baseados em gradiente. Esse problema é aqui ampliado pelo fato do gradiente da entropia do vetor do Evc ser dependente da largura do kernel σ . Como discutido no capítulo 3, esse parâmetro é de fundamental importância na estimação da densidade do erro e pode comprometer o resultado dessa estimação. O problema da largura do kernel em algoritmos adaptativos vem sendo atacado por vários autores sem uma solução definitiva (Erdogmus and Principe, 2002; Santos, *et al.*, 2005). Outro particular problema comentado por Eilers em (Eilers 2003) é o desempenho do erro de validação cruzada quando uma correlação serial está presente no sinal. Neste caso, a otimização do parâmetro lambda pode ser comprometida porque somente com uma fraca suavização (baixo valor de lambda) os resíduos serão independentes (Eilers 2003). Apesar dos dois problemas, o algoritmo MCVEE foi testado em um grupo de séries temporais com resultado satisfatório. Em alguns casos a suavização não obteve um resultado esperado, o que mostra a necessidade da modificação da função custo no sentido da independência de parâmetros e robustez na suavização.

4.4 Função Custo Alternativa para o Gradiente

Como comentado na seção anterior, o método do gradiente desenvolvido neste trabalho apresenta dois inconvenientes: a dependência do algoritmo em relação a largura do kernel e o problema da fraca suavização na presença de um sinal correlacionado. O problema da largura do kernel vem sendo tratado por vários pesquisadores e diversas questões sobre o valor ótimo desse parâmetro permanecem em aberto. Assim, uma alternativa para a função

custo na otimização dos parâmetros do filtro Whittaker independente de parâmetros pode ser dada pela Equação 4.19:

$$F = |Evc|^2 + |Dz|^2 \quad (4.19)$$

onde a parcela $|Dz|^2$ impede uma fraca suavização na reposta do filtro quando o Evc não responde à resíduos correlacionados. Pode-se reescrever a Equação 4.19 como:

$$F = |Evc|^2 + \left| D[I + 10^\lambda D^T D]^{-1} y \right|^2 \quad (4.20)$$

Derivando F em relação à lambda, tem-se:

$$\frac{\partial F}{\partial \lambda} = 2 \left(\frac{\partial Evc}{\partial \lambda} \right)^T (Evc) + 2 \left(\frac{\partial D[I + 10^\lambda D^T D]^{-1} y}{\partial \lambda} \right)^T (Dz) \quad (4.21)$$

A derivada do segundo termo do lado direito da Equação 4.21 é dada por:

$$\frac{\partial D[I + \lambda D^T D]^{-1} y}{\partial \lambda} = -10^\lambda \ln 10 D \left([I + \lambda D^T D]^{-1} (D^T D) [I + \lambda D^T D]^{-1} \right) y \quad (4.22)$$

Substituindo a Equação 4.22 e 4.17 em 4.21, tem-se que:

$$\begin{aligned} \frac{\partial F}{\partial \lambda} = & 10^\lambda \ln 10^2 \left(\left[1 - dg[I + 10^\lambda D^T D]^{-1} \right]^{-1} \circ \left\{ \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) y \right. \right. \\ & - \left. \left. \left(dg \left([I + 10^\lambda D^T D]^{-1} (D^T D) [I + 10^\lambda D^T D]^{-1} \right) \circ Ecv \right) \right\}^T (Evc) \right. \\ & \left. - \left(D \left([I + \lambda D^T D]^{-1} (D^T D) [I + \lambda D^T D]^{-1} \right) y \right)^T (Dz) \right) \end{aligned} \quad (4.23)$$

A Equação 4.23 é livre de parâmetros independentes e apresentou surpreendentes resultados na suavização das séries temporais testadas. As Figuras 4.5 e 4.6 apresentam uma comparação entre o método do gradiente usando a entropia do Evc (Equação 4.4), a função alternativa desenvolvida nesta seção (Equação 4.23) e somente o Ecv (Equação 4.17).

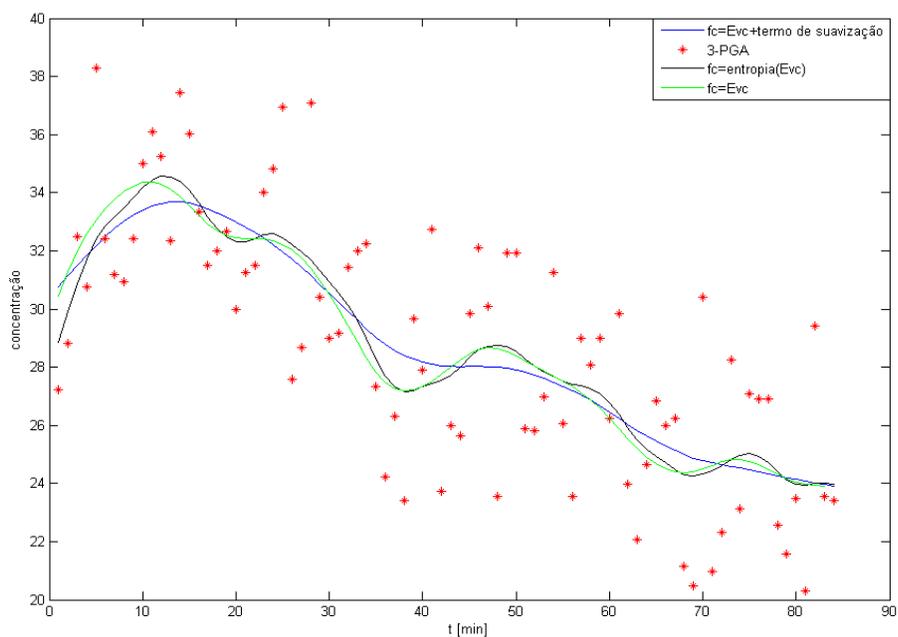


Figura 4.5 – Comparação da suavização do perfil metabólico 3-PGA usando diferentes funções custo.

A superfície da função custo alternativa da otimização dos parâmetros para a curva mostrada na Figura 4.6 é apresentada na Figura 4.7.

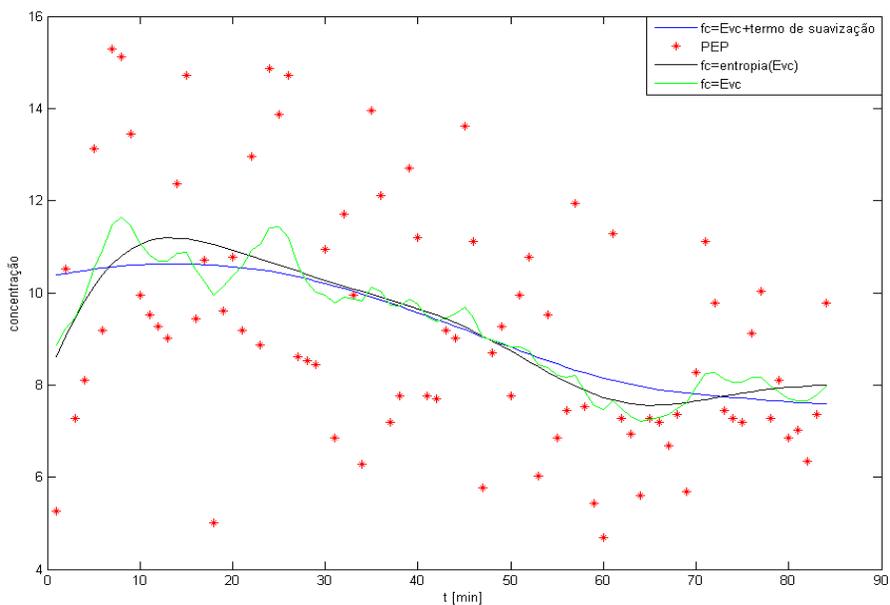


Figura 4.6 - Comparação da suavização do perfil metabólico Fosfoenolpiruvato (PEP) usando diferentes funções custo .

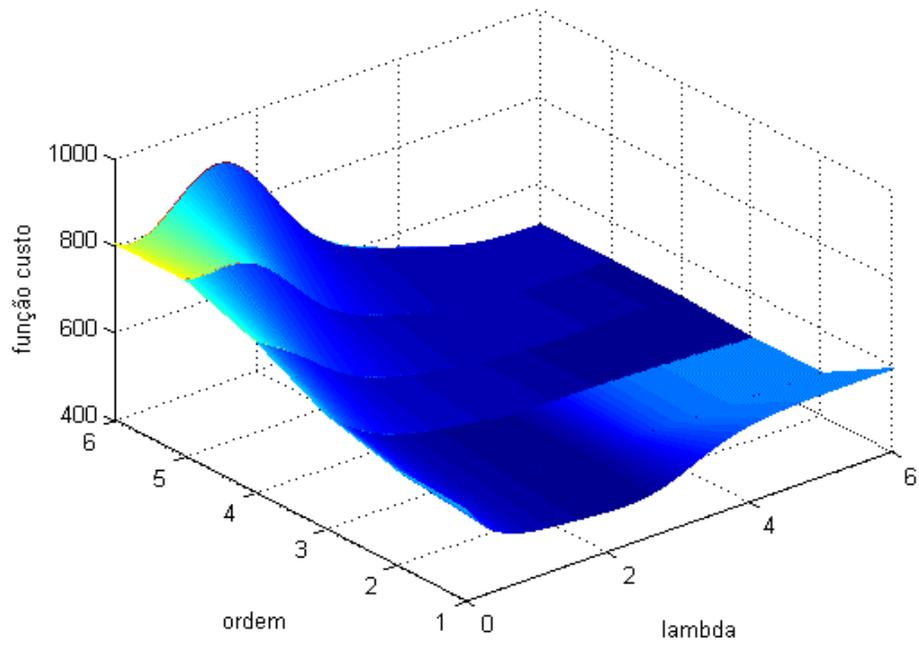


Figura 4.7 – Superfície da função custo alternativa em relação as parâmetros do filtro Whittaker

O comportamento do algoritmo ao longo das iterações pode ser visto na Figura 4.8.

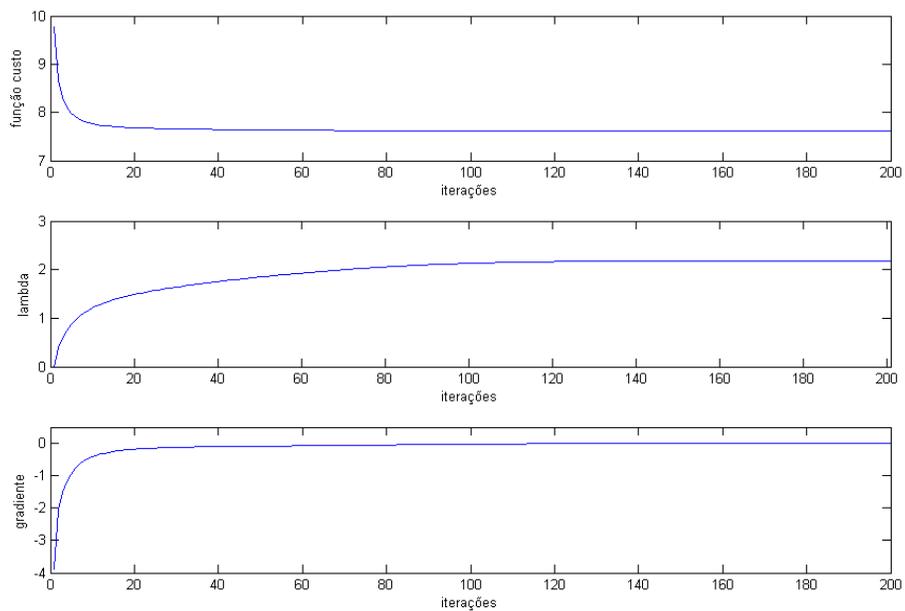


Figura 4.8 – Comportamento assintótico do algoritmo.

4.5 Algoritmo de Segmentação Baseado na Entropia do Erro de Validação Cruzada

Uma alternativa para resolver o problema da suavização de sinais com descontinuidade da variância do ruído é dividir o sinal em janelas onde essa variância seja constante e suavizá-las, ou seja, decompor o sinal em partes com características estacionárias onde o viés de outros segmentos com diferentes estruturas de ruído não prejudique o resultado da suavização. O filtro Whittaker assume que um ruído invariante ao longo do sinal na sua solução (Whittaker 1923), o que estabelece uma condição de estacionariedade do ruído no sinal. Assim, um algoritmo de segmentação baseado na entropia mínima do vetor do erro de validação cruzada foi desenvolvido na tentativa de superar o problema da descontinuidade do ruído nos perfis metabólicos.

O algoritmo começa suavizando toda a série temporal e avaliando a entropia do vetor do erro de validação cruzada, fazendo essa entropia a entropia corrente. A fdp desse erro é calculada pelo método das Janelas de Parzen com uma largura do kernel σ_s estimada pela Equação 3.26 usando a própria série temporal. Essa largura tem um importante papel no algoritmo, pois é usada no processo de parada da segmentação. A série temporal é então dividida em duas janelas: a janela 1 (à esquerda) com os quatro primeiros pontos da série e a janela 2 (à direita), com o conjunto de pontos complementar. Os pontos de cada uma dessas janelas são suavizados com o filtro Whittaker usando o algoritmo adaptativo descrito na seção 4.4 para a obtenção dos parâmetros ótimos. Conseqüentemente a entropia do vetor do Evc do filtro Whittaker dessas duas janelas são calculadas e a seguinte função custo é avaliada para cada janela.

$$fc_i^w = \left| H_{\sigma_w}^w - H_{\sigma_w}^0 \right|, i = 1, 2, \dots, N - 7 \quad (4.24)$$

Na Equação 4.24, $H_{\sigma_w}^w$ é a entropia da janela $w=[1,2]$ calculada com uma largura do kernel σ_w e $H_{\sigma_w}^0$ é a entropia mínima possível calculada com a mesma largura do kernel σ_w . No próximo passo o ponto em seqüência é adicionado na janela 1 e conseqüentemente este ponto é subtraído da janela 2. Este mesmo processo de suavização dos pontos e avaliação da função custo fc (Equação 4.24) é novamente realizado, e assim por diante. Após $N-7$ iterações, todo o sinal é varrido e a função custo fc para cada dupla de janela ($w=1,2$) é obtida (Figura 4.8). O ponto mínimo dos mínimos da função custo das duas janelas é identificado assim como todos os pontos de uma vizinhança contínua tal que os valores de entropia desses pontos pertençam uma certa faixa (faixa determinada por um limiar, chamado aqui de parâmetro de segmentação ps - Equação 4.25). O algoritmo reconhece essa vizinhança selecionada como uma janela de entropia mínima de um erro com variância constante que junto com sua janela complementar, é escolhida para o teste de entropia mínima.

$$faixa \leq \min(fc) + ps \quad (4.25)$$

No caso em que a entropia do vetor do erro de validação cruzada formado por essas duas janelas for menor que a entropia corrente (calculada no primeiro passo do algoritmo), a série temporal é definitivamente segmentada, caso contrário a série permanece íntegra. Essa entropia é calculada pelo mesmo método usado no cálculo da entropia corrente e com a mesma largura do kernel σ_s . Se a série temporal for segmentada, o algoritmo “enxerga” as novas duas janelas como séries independentes e todo o processo de suavização e varredura do sinal é novamente realizado dentro dessas janelas do modo que toda a informação de uma janela não interfere mais na outra. Assim, no final de cada processo de varredura o algoritmo cria duas novas janelas caso a entropia do vetor do Evc decresça ou para o

processo de segmentação em caso contrário. Essa estrutura computacional do algoritmo é conhecida como árvore de procura.

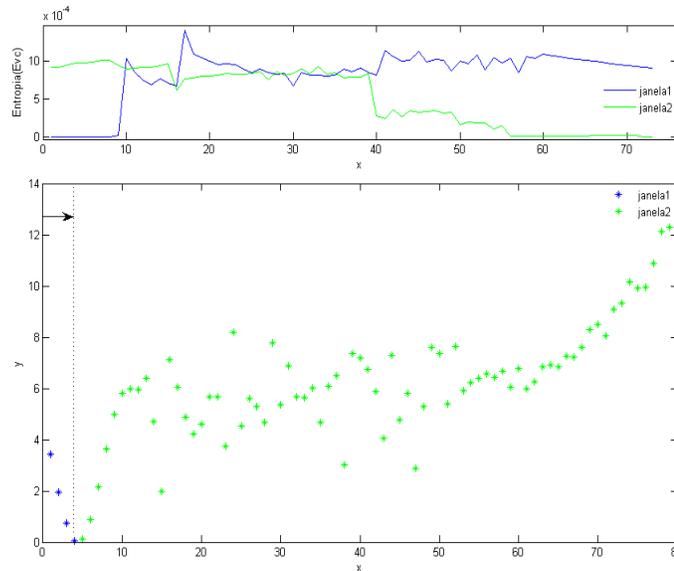


Figura 4.8 – Processo de Varredura do sinal. O processo de varredura começa com a janela 1 (em azul à esquerda) com os 4 primeiros pontos da série e janela 2 (em verde) com os pontos complementares. O fim do processo de varredura ocorre com a janela 2 com os últimos 4 pontos da série e a janela 1 com os pontos complementares.

Com o objetivo de testar o algoritmo de segmentação, uma série temporal com três funções diferentes (senóide+constante+exponencial) foi criada onde diferentes níveis de ruído gaussiano foram adicionados em cada um dos sinais que compõem a série. O resultado da segmentação e suavização é mostrado na Figura 4.9.

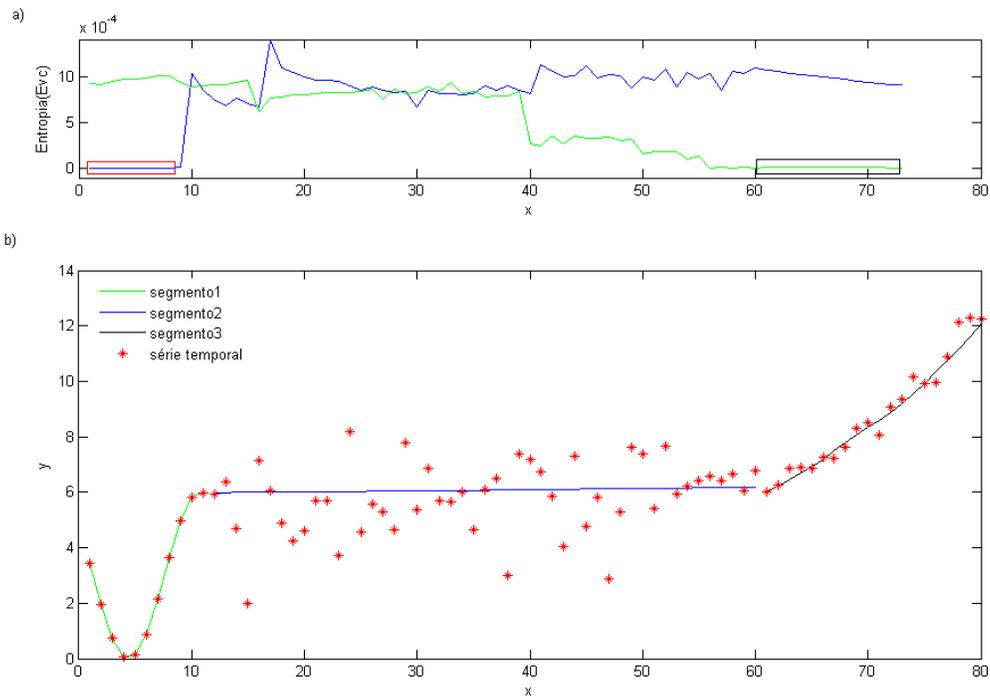


Figura 4.9 – Escolha da janela de segmentação. a) Função da entropia do Evc de todas as janelas avaliadas. Janelas escolhidas para segmentação após a primeira varredura em vermelho e após a segunda varredura em preto. b) Resultado da suavização com três segmentos com diferentes estruturas de ruído.

4.6 Derivada da curva suavizada

A derivada do perfil metabólico tem um papel tão importante quanto a própria série nos métodos de parametrização dos Sistemas-S. Métodos como redes neurais e máquinas de vetores de suporte impõem artefatos na derivada provenientes da função de transferência (e.g. sigmoidal) e da função kernel (e.g. gaussiana) respectivamente. O filtro Savitzky-Golay é um caso em que se obtêm um bom resultado na suavização da série temporal, mas a derivada do polinômio é extremamente distorcida nos pontos da série a medida que o grau do polinômio escolhido aumenta. Após a tentativa de uso de todas as técnicas citadas, concluiu-se que a curva suavizada não deveria ser uma composição de uma função

previamente definida e sim uma função dos pontos da própria série temporal, e é neste contexto que o filtro Whittaker se encaixa. Com isso, dado uma curva suavizada z obtida pelo filtro Whittaker a partir de uma série temporal ruidosa y (Equação 2.17), tem-se que a derivada de z em relação ao tempo é dada por:

$$\frac{dz}{dt} = \frac{dy}{dt} \frac{dz}{dy} = \left(\frac{dy}{dt} \right)^T \left([I + \lambda D^T D]^{-1} \right)^Y \quad (4.26)$$

A derivada do sinal de saída do filtro Whittaker é na verdade uma suavização da derivada do sinal de entrada, dado que o filtro é independente da variável tempo. Isso faz com que a qualidade da derivada do filtro Whittaker seja dependente não só da suavização, mas também da qualidade da derivada numérica aplicada na série temporal. O método de suavização assim como a extração da derivada temporal foram testado em dados reais de experimentos de RNM *in vivo* com a bactéria *Lactococcus lactis* (Ramos, Neves *et al.* 2002), e os resultados são ilustrados nas figuras 4.10,4.11 e 4.12.

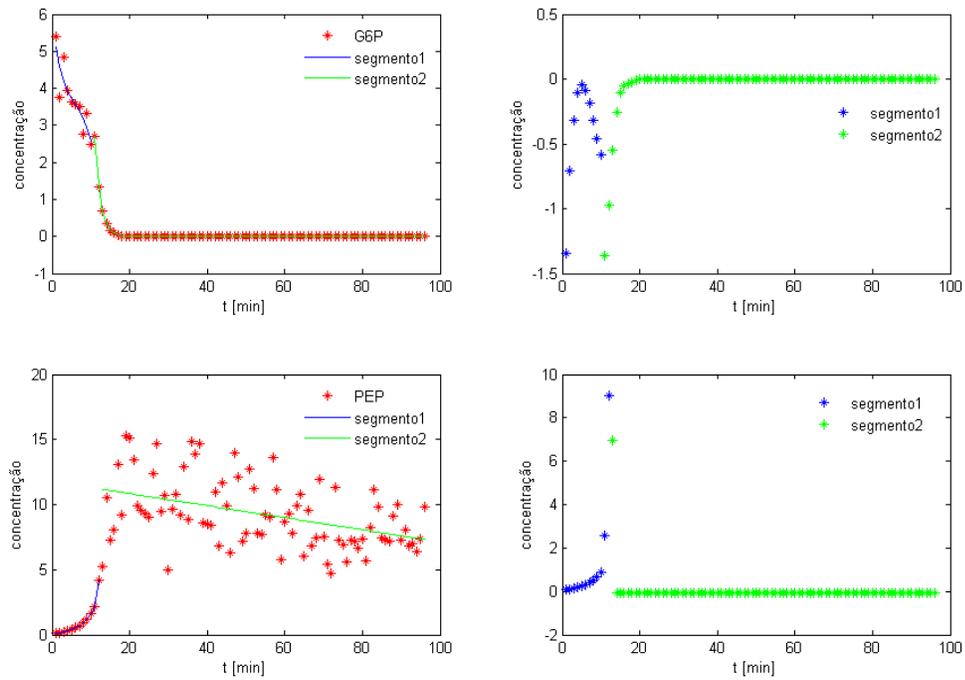


Figura 4.10 – Ilustração do método aplicado aos dados da bactéria *Lactococcus lactis* para a Glicose 6-Fostato (G6P) e Fosfoenolpiruto (PEP). Os segmentos suavizados estão apresentados em diferentes cores bem como na derivada.

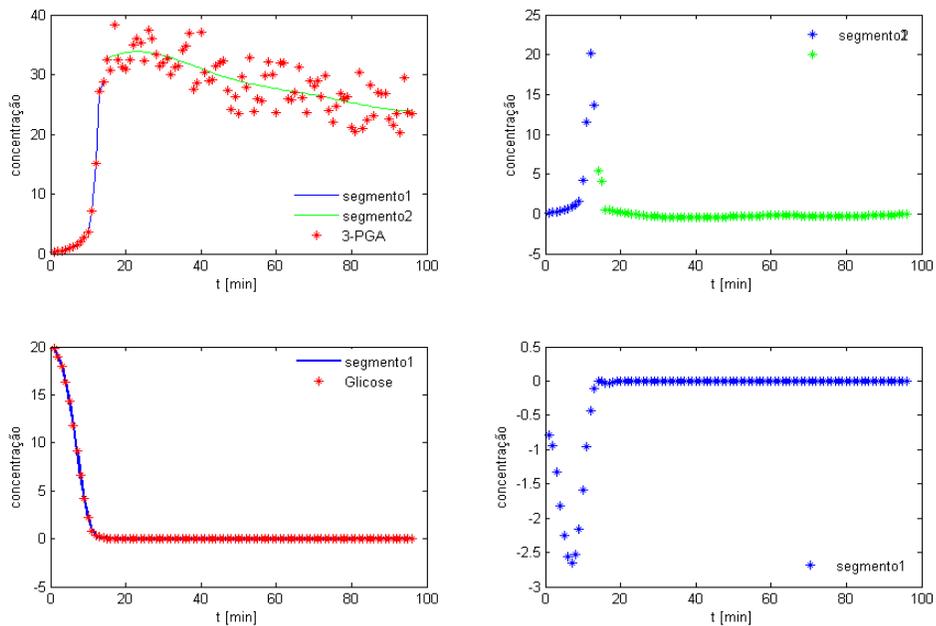


Figura 4.11 -- Ilustração do método aplicado aos dados da bactéria *Lactococcus lactis* para a Glicose e 3 Fosfoglicetato (3-PGA). Os segmentos suavizados estão apresentados em diferentes cores bem como na derivada.

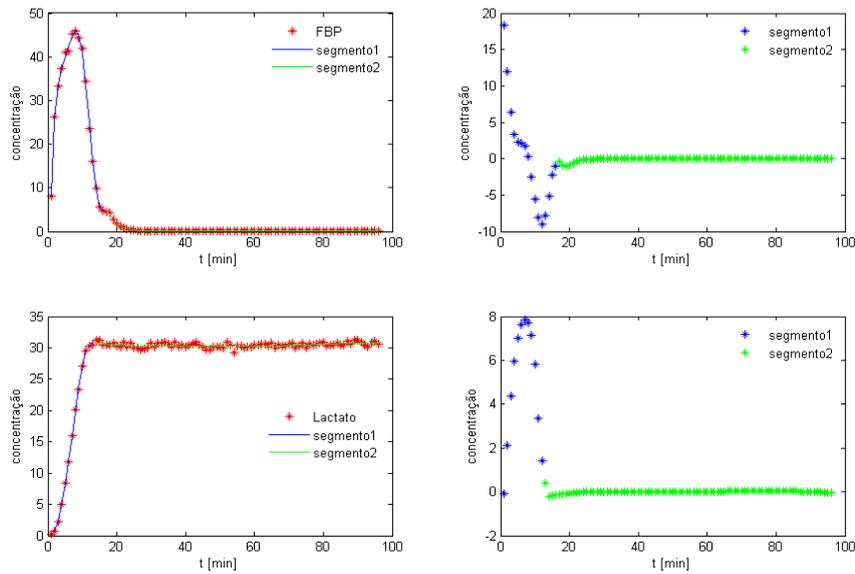


Figura 4.12 - – Ilustração do método aplicado aos dados da bactéria *Lactococcus lactis* para o Lactato e Frutose 1,6 Bifosfato (FBP). Os segmentos suavizados estão apresentados em diferentes cores bem como na derivada.

4.7 Interface Gráfica

A partir dos dois algoritmos descritos anteriormente (parametrização do filtro Whittaker e método de segmentação), foi criada uma interface gráfica executável para o suavizador implementado no ambiente MATLAB®, que pode ser obtida por *download* através do *website* <http://autosmooth.sourceforge.net>. Essa interface é um dos módulos de uma aplicação independente chamada *bioinformaticStation*, podendo também ser obtida pelo mesmo *website*. As aplicações são livres de licença e podem ser executadas em máquinas sem MATLAB®.

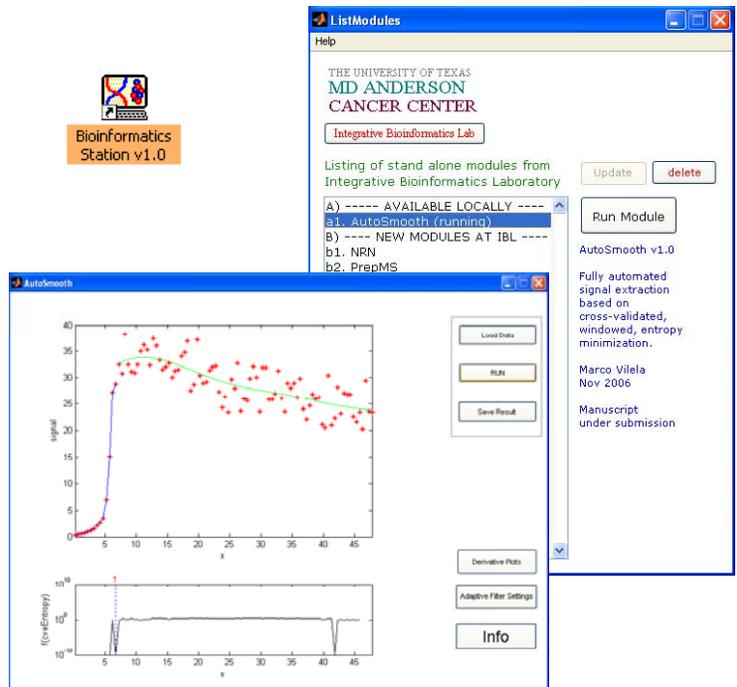


Figura 4.13 – AutoSmoother. Parte da Estação de Bioinformática da Universidade do Texas MDAnderson.

Capítulo 5

Conclusão

A parametrização dos Sistemas-S a partir de perfis metabólicos é vista como um método promissor de extração de conhecimento da complexidade e funcionamento de uma via metabólica. Usual em dados de biologia molecular, os perfis metabólicos apresentam incertezas de medições o que faz necessário a aplicação de técnicas de processamento de sinais em uma etapa anterior a análise com o intuito de eliminar ou amenizar a propagação dessas incertezas. A extração do sinal suavizado e da derivada temporal são fatores imprescindíveis para as técnicas de parametrização dos Sistemas-S, sendo o resultado dessa parametrização extremamente dependente da qualidade desses dois fatores(Savageau 1969a; Savageau 1969b; Savageau 1970; Voit 2000). Este trabalho propõe uma nova técnica de parametrização do filtro Whittaker acoplado à um algoritmo de segmentação aplicado a suavização e a extração da derivada de perfis metabólicos experimentais.

O método proposto neste trabalho para a solução do problema da suavização de perfis metabólicos é composto por dois procedimentos: otimização dos parâmetros do filtro Whittaker e segmentação da série temporal em janelas com a variância do ruído constante. Primeiramente, a otimização dos parâmetros do filtro Whittaker foi obtida usando o método do gradiente descendente na minimização da entropia quadrática de Rényi do vetor do erro de validação cruzada - Evc. Essa entropia foi calculada usando o método das Janelas de

Parzen para a estimação de função de densidade de probabilidade do Evc. A equação do gradiente foi obtida escrevendo o erro de validação cruzada como o produto de Hadamard entre dois vetores dependentes da matriz de mapeamento do filtro Whittaker, chamada de matriz chapéu. Essa solução apresentou bons resultados na suavização dos perfis metabólicos testados, porém com algumas fracas suavizações em sinais sintéticos. Após vários testes, concluiu-se que o problema das fracas suavizações era devido a uma má estimação da largura do kernel gaussiano usado nas Janelas de Parzen, detendo o gradiente em um ponto de mínimo local. Para superar esse problema, outra função custo composta pelo Evc e uma medida de suavização dada pelo operador de diferenças do próprio filtro foi proposta. Um novo gradiente foi obtido a partir dessa função custo e todos os testes foram refeitos, apresentando bons resultados e robustez com as diferentes séries temporais.

Outra parcela da solução do suavizador é o algoritmo de segmentação baseado na entropia quadrática de Rényi do vetor do Evc das janelas suavizadas. Combinado com o algoritmo de otimização dos parâmetros do filtro Whittaker, o método de segmentação mostrou resultados consistentes quando aplicado aos perfis metabólicos da bactéria *Lactococcus lactis* assim como em séries sintéticas com diferentes variâncias no ruído. Foi observado que os melhores resultados da suavização foram obtidos em sinais com janelas de discrepância significativa na variância do ruído, o que facilita para o algoritmo no reconhecimento dos diferentes segmentos. Essa é uma característica marcante dos perfis metabólicos, o que faz do método desenvolvido uma ferramenta útil no processamento de perfis metabólicos numa etapa anterior a parametrização.

5.1 Perspectivas Futuras

Um método robusto para a suavização e extração da derivada temporal dos perfis metabólicos abre a porta para testes e aprimoramento das técnicas de parametrização de Sistemas-S a partir de dados reais. Estudos com simulações de modelos de vias metabólicas mostraram que mais de um grupo de parâmetros pode ser encontrado tal que o Sistema-S tenha o mesmo comportamento dinâmico, indicando a existência de mais de uma possível solução como resultado do processo de parametrização. Múltiplas soluções dos valores dos parâmetros do Sistema-S não implicam necessariamente em diferentes redes de regulação, já que no processo de construção da rede de regulação, a partir de um Sistema-S, somente a informação do sinal dos parâmetros é usada. Uma unidade computacional automática que gera as possíveis soluções para os parâmetros do Sistema-S, e conseqüentemente as possíveis topologias da rede de regulação metabólica a partir dos perfis metabólicos surgem como sendo o próximo passo no desenvolvimento de ferramentas para o auxílio no estudo das atividades metabólicas celulares.

Apêndice A – Operações de Hadamard

Neste apêndice são introduzidas breves definições de algumas operações de Hadamard para o melhor entendimento da notação e operações usadas no desenvolvimento das equações do gradiente apresentado neste trabalho (Magnus and Neudecker 1988; Chen and Zhing 1997).

Definição a.1 – Sejam as matrizes de mesma ordem $A = [a_{ij}]$, $B = [b_{ij}] \in C^{N \times M}$. O produto $A \circ B = [a_{ij} b_{ij}] \in C^{N \times M}$ é definido como o produto de Hadamard.

Definição a.2 – Dada a matriz $A = [a_{ij}] \in C^{N \times M}$, a operação $A^{\circ q} = [a_{ij}^q] \in C^{N \times M}$ é definida como a potência q de Hadamard da matriz A.

Definição a.3 – Dada a Definição a.2 e uma matriz com elementos não nulos

$A = [a_{ij}] \in C^{N \times M}$, $a_{ij} \neq 0$ define-se a matriz $A^{\circ -1} = \left[\frac{1}{a_{ij}} \right] \in C^{N \times M}$ como a matriz inversa de

Hadamard. A matriz $A^{\circ 0} = [1_{ij}] \in C^{N \times M}$ é definida como a matriz unidade de Hadamard.

A.1 Propriedades do produto de Hadamard

Dada as matrizes $A = [a_{ij}]$, $B = [b_{ij}] \in C^{N \times M}$ e a função matricial $dg(.)$ definida tal que aplicada na matriz quadrada $A = [a_{ij}]$, $i, j = 1, 2, \dots, n$, $dg(A) = [a_{11} \ a_{22} \ \dots \ a_{nn}]^t$, tem-se que:

$$\begin{aligned}
PA.1 \quad & A \circ B = B \circ A \\
PA.2 \quad & (A \circ B)^t = A^t \circ B^t \\
PA.3 \quad & (A \circ B) \circ C = A \circ (B \circ C) \\
PA.4 \quad & (A + B) \circ (C + D) = A \circ C + A \circ D + B \circ C + B \circ D \\
PA.5 \quad & A \circ 0 = 0 \\
PA.6 \quad & A \circ A^{\circ-1} = A^{\circ 0} \\
PA.7 \quad & A^p \circ A^q = A^{\circ(p+q)} \\
PA.8 \quad & A \circ I = \text{diag}(\text{dg}(A))
\end{aligned}$$

PA.9 - Dada as matrizes $A = [a_{ij}]$, $B = [b_{ij}] \in C^{N \times M}$ e o produto de Hadamard definido acima, temos que derivada de $A \circ B = [a_{ij} b_{ij}] \in C^{N \times M}$ em relação ao escalar α :

$$\frac{\partial(A \circ B)}{\partial \alpha} = \frac{\partial A}{\partial \alpha} \circ B + \frac{\partial B}{\partial \alpha} \circ A \tag{a.1}$$

Outra importante propriedade é a derivada da matriz inversa, dada por:

$$PA.10 \quad \frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}$$

Apêndice B

Listas de Congressos e Simpósios onde o trabalho da dissertação foi apresentado:

Marco Vilela, Carlos C.H. Borges, Ana Tereza Vasconcelos, Helena Santos, Eberhard Voit and Jonas S Almeida. .Automated Smoother for Decoupling of the S-system Models. In: Meeting of Computational Cell Biology, Cold Spring Harbor Laboratory, 2007, Cold Spring Harbor – NY.

Marco Vilela, Carlos C.H. Borges, Ana Tereza Vasconcelos, Helena Santos, Eberhard Voit and Jonas S Almeida. .Automated Smoother for Decoupling of the S-system Models. In: Symposium of Collaborative Research 2006, Rice University, Houston – TX.

Marco Vilela, Carlos C.H. Borges, Ana Tereza Vasconcelos, Helena Santos, Eberhard Voit and Jonas S Almeida. .Automated Smoother for Biological Time Series. In: Cancer Systems Biology Symposium, University of Texas MDAnderson Cancer Center, 2006, Houston – TX.

Artigo submetido

Marco Vilela, Carlos C.H. Borges, Susana Vinga, Ana Tereza Vasconcelos, Helena Santos, Eberhard O. Voit and Jonas S. Almeida. Automated Smoother for the Numerical Decoupling of Dynamics Models. Artigo submetido na revista BMCBioinformatics.

Referências Bibliográficas

Almeida, J. S. (2002). "Predictive non-linear modeling of complex data by artificial neural networks." *Curr Opin Biotechnol* 13(1): 72-6.

Almeida, J. S. and E. O. Voit (2003). "Neural-network-based parameter estimation in S-system models of biological networks." *Genome Inform* 14: 114-23.

Borges, C. C. H., M. A. M. Vilela, et al. (2006). A Support Vector Machine Regression Applied in Signal Extraction for Numerical Estimation. Iberian Latin-American Congress on Computational Methods, Belem, Para, Brazil.

Borges, C. C. H., E. O. Voit, et al. (2004). Signal Extraction for Numerical Decoupling of S-systems International Conference on Molecular Systems Biology, Tahoe City, California, USA.

Chen, W. and t. Zhing (1997). "The study on the nonlinear computations of the DQ and DC methods." *Numerical Methods for Partial Differential Equations* 13: 57-75.

Chou, I. C., H. Martens, et al. (2006). "Parameter estimation in biochemical systems models with alternating regression." *Theor Biol Med Model* 3(1): 25.

Comon, P. (1992). "Independent component analysis, A new concept?" *Signal Processing* 36: 287-314.

Cover, T. M. and J. A. Thomas (1991). *Elements of information theory*. New York, Wiley.

Csiszár, I. and J. Körner (1981). *Information theory : coding theorems for discrete memoryless systems*. New York, Academic Press.

Eilers, P. H. (2003). "A perfect smoother." *Anal Chem* 75(14): 3631-6.

Erdogmus, D. and J. C. Principe (2000). Comparison of Entropy and Mean Square Error Criteria in Adaptive System Training Using Higher Order Statistics. Workshop on Independent Component Analysis and Signal Separation, Helsinki, Finland.

Erdogmus, D. and J. C. Principe (2002). "An Error-Entropy Minimization Algorithm for Supervised Training of Nonlinear Adaptive Systems." *IEEE Transactions on Signal Processing* 50(7): 1780-1786.

Fan, J. and Q. Yao (2005). *Nonlinear time series : nonparametric and parametric methods*. New York, Springer.

Gonzalez, O. R., C. Kuper, et al. (2006). "Parameter estimation using Simulated Annealing for S-system models of biochemical networks." *Bioinformatics*.

- Hoaglin, D. C. and R. E. Welsch (1977). *The hat matrix in regression and anova*. Cambridge, Mass., M.I.T. Alfred P. Sloan School of Management.
- Hodrick, R. J. and E. C. Prescott (1997). "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, 29(1): 1-16.
- Hornik, K., M. Stinchcombe, et al. (1989). "Multilayer feedforward networks are universal approximators." *Neural Networks* 2: 359–366.
- Imoto, S. and S. Konishi (2003). "Selection of Smoothing Parameters in B-Spline Nonparametric Regression Models Using Information Criteria." *Ann. Ins. Statist. Math* 55(4): 671-687.
- Jenssen, R., D. Erdogmus, et al. (2005). "Some Equivalences Between Kernel Methods and Information Theoretic Methods." *Journal of VLSI Signal Processing Special Issue on Machine Learning for Signal Processing*.
- Kikuchi, S., D. Tominaga, et al. (2003). "Dynamic modeling of genetic networks using genetic algorithm and S-system." *Bioinformatics* 19(5): 643-50.
- Kullback, S. and R. A. Leibler (1951). "On information and sufficiency." *Annals of Mathematical Statistics* 22: 79-86.
- Liano, K. (1996). "Robust Error Measure for Supervised Neural Network Learning with Outliers." *IEEE Transactions on Neural Networks* 7(1): 246-250.
- Magnus, J. R. and H. Neudecker (1988). *Matrix differential calculus with applications in statistics and econometrics*. Chichester [England] ; New York, Wiley.
- Marino, S. and E. O. Voit (2006). "An automated procedure for the extraction of metabolic network information from time series data." *J Bioinform Comput Biol* 4(3): 665-91.
- Martins, S. V. (2005). *Biological sequence analysis by vector maps: alignment-free comparison of DNA and proteins*. Instituto de Tecnologia Quimica e Biologica Lisboa, Universidade de Lisboa. Ph.D: 1-159.
- Mosheiov, G. and A. Raveh (1997). "On Trend Estimation of Time-Series: A Simpler Linear Programming Approach." *Journal of Operational Research Society* 48: 90-96.
- Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. New York, McGraw-Hill.
- Parzen, E. (1962). "On the estimation of probability density function and the mode." *Annals of Mathematical Statistics* 33: 1065.
- Principe, J., D. Xu, et al. (1999). *Information-Theoretic Learning. Advances in unsupervised adaptive filtering*, Wiley.
- Ramos, A., A. R. Neves, et al. (2002). "Metabolism of lactic acid bacteria studied by nuclear magnetic resonance." *Antonie Van Leeuwenhoek* 82(1-4): 249-61.
- Rényi, A. (1976). *Selected papers of Alfréd Rényi*. Budapest, Akadémiai Kiadó.
- Rhodes, I. B. (1971). "A Tutorial Introduction to Estimation and Filtering." *IEEE Transactions on Automatic Control* 16(6): 688-706.

Santos, J. M., J. M. d. Sa, et al. (2005). Neural Networks Trained with the EEM Algorithm: Tuning the Smoothing Parameter. In 6th WSEAS Int. Conf. on Neural Networks.

Savageau, M. A. (1969a). "Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions." J Theor Biol 25(3): 365-9.

Savageau, M. A. (1969b). "Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation." J Theor Biol 25(3): 370-9.

Savageau, M. A. (1970). "Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation." J Theor Biol 26(2): 215-26.

Shannon, C. E. (1948). "A mathematical theory of communication." Bell System Technical Journal 27: 379-423 and 623-656.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman & Hall.

Smola, A. J. and B. Scholkopf (2004). "A tutorial on support vector regression." Statistics and Computing 14: 199-222.

Voit, E. O. (2000). Computational analysis of biochemical systems : a practical guide for biochemists and molecular biologists. Cambridge ; New York, Cambridge University Press.

Voit, E. O. and J. Almeida (2004). "Decoupling dynamical systems for pathway identification from metabolic profiles." Bioinformatics 20(11): 1670-81.

Voit, E. O., J. Almeida, et al. (2006). "Regulation of glycolysis in Lactococcus lactis: an unfinished systems biological case study." Syst Biol (Stevenage) 153(4): 286-98.

Whittaker, E. T. (1923). "On a new method of graduation." Proceedings of the Edinburgh Mathematical Society 41: 63-75.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)