

CÁLCULO DE DIREÇÕES DE BUSCA EXATAS PARA OTIMIZAÇÃO DE FILTROS DE LAGUERRE E DE KAUTZ

ALEX DA ROSA*, RICARDO J. G. B. CAMPELLO†, WAGNER C. AMARAL*

**Faculdade de Eng. Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), CEP 13083-852, Campinas-SP, Brasil.*

†*Depto. Ciências de Computação, Universidade de São Paulo (USP), CEP 13560-970, São Carlos-SP, Brasil.*

Emails: alex@dca.fee.unicamp.br, campello@icmc.usp.br, wagner@dca.fee.unicamp.br

Abstract— A novel technique for selecting the poles of orthonormal basis functions in the modeling of dynamic systems is proposed. To optimize the poles that parameterize the Kautz basis, the gradients of the outputs of the Kautz filters with respect to their parameters are computed analytically. The gradients for Laguerre filters follow straightforwardly as a particular case. The dynamic of the orthonormal filters is fully considered in the gradient computations. These gradients can be used to provide exact search directions for optimizing the poles of a given orthonormal basis. The search directions can, in turn, be used as part of an optimization procedure to locate the minimum of a quadratic cost-function that takes into consideration the error of estimation of the system output. In this paper, the Levenberg-Marquardt algorithm is adopted as the optimization procedure. A simulated example is also presented to illustrate the method proposed.

Keywords— System identification, Orthonormal basis functions, Optimization, Gradient-based methods.

Resumo— Este trabalho propõe uma nova técnica para a seleção dos pólos das funções de base ortonormal na modelagem de sistemas dinâmicos. Para otimizar os pólos que parametrizam a base de Kautz, calcula-se analiticamente os gradientes das saídas dos filtros de Kautz com relação aos seus parâmetros. Os gradientes para os filtros de Laguerre seguem diretamente como um caso particular. A dinâmica dos filtros ortonormais é completamente considerada no cálculo dos gradientes. Estes gradientes podem ser utilizados para fornecer direções de busca exatas a fim de otimizar os pólos de uma dada base ortonormal. As direções de busca, por sua vez, podem ser utilizadas como parte de um procedimento de otimização para obter o mínimo de uma função de custo quadrática que descreve o erro de estimação da saída do sistema. Neste trabalho, utiliza-se o algoritmo de Levenberg-Marquardt como o procedimento de otimização. Apresenta-se também um exemplo simulado para ilustrar o método proposto.

Keywords— Identificação de sistemas, Funções de base ortonormal, Otimização, Métodos do gradiente.

1 Introdução

Nas últimas décadas, o uso de funções de base ortonormal (*Orthonormal Basis Functions* – OBF) tem sido tema de inúmeros estudos nas áreas de identificação e controle de sistemas dinâmicos – e.g. veja (Heuberger et al., 2005; Campello et al., 2007) e referências inclusas. Esta abordagem consiste em representar o sistema em termos de uma base ortonormal no espaço de interesse. A principal motivação para o uso de OBF naquelas áreas é que os problemas aproximados correspondentes usualmente possuem soluções simplificadas, uma vez que a ortonormalidade dessas funções leva a modelos mais simples. Em particular, o uso de OBF na estrutura do modelo permite a incorporação de conhecimento *a priori* sobre a dinâmica do sistema (den Hof et al., 1995; Ninness and Gustafsson, 1997; Nelles, 2001), o que permite reduzir significativamente o número de parâmetros do modelo.

As bases ortonormais mais comumente utilizadas na modelagem de sistemas são as bases de Laguerre e de Kautz (Broome, 1965; Bokor and Schipp, 1998; Heuberger et al., 2005), que são adequadas para modelar sistemas com dinâmica dominante de primeira e segunda ordem, respectivamente. Para representar dinâmicas mais complexas, as bases de funções

ortonormais generalizadas (*Generalized Orthonormal Basis Functions* – GOBF) (den Hof et al., 1995; Ninness and Gustafsson, 1997) são mais apropriadas, embora elas envolvam uma parametrização mais complexa.

Modelos baseados em OBF podem em geral ser construídos com um número reduzido de termos para descrever o sistema. Quando parametrizada adequadamente, a base de funções ortonormais leva a um aumento na velocidade de convergência em problemas de identificação (den Hof et al., 1995; Heuberger et al., 2005), uma vez que o número de parâmetros dos modelos OBF depende fortemente da escolha dos pólos da base. Muitos trabalhos têm sido escritos sobre como esta seleção pode ser feita, especialmente por meio de abordagens analíticas. Para a base de Laguerre, otimizações analíticas que levaram à soluções fechadas já foram extensivamente abordadas (Fu and Dumont, 1993; Silva, 1994; Tanguy et al., 1995; Campello et al., 2004; Campello et al., 2006). Para a base de Kautz, investigou-se condições de otimalidade para a seleção dos pólos (Silva, 1995) e ainda abordou-se formulações de otimização para se obter parâmetros ótimos analiticamente (Tanguy et al., 2002; da Rosa et al., 2007; da Rosa et al., 2008). Apesar dos importantes resultados teóricos e avanços obtidos com respeito à abordagens analíticas, as

soluções apresentadas na literatura possuem pelo menos uma das seguintes desvantagens: (i) um modelo do sistema (e.g. FIR ou Volterra) precisa ser conhecido a priori; (ii) a solução é apenas sub-ótima, pois otimiza somente um sub-conjunto dos parâmetros da base.

Em (Bodin et al., 1997; Sarroukh et al., 2001; Patwardhan and Shah, 2005), são propostos métodos de seleção de pólos envolvendo procedimentos de otimização numérica. Neste contexto, sugeriu-se em (Patwardhan and Shah, 2005) uma formulação geral de otimização que envolve a decomposição do problema de identificação do modelo em um procedimento iterativo composto por dois passos: uma atualização nos parâmetros (pólos) da base ortonormal seguida por uma atualização nos parâmetros restantes do modelo (com os pólos mantidos fixos). Esta decomposição reduz a complexidade do problema de identificação, porém não se apresentou nenhuma estratégia para obter a direção de busca a ser seguida pelo algoritmo de otimização. O presente trabalho propõe uma metodologia para determinar as direções de busca baseado no cálculo analítico recursivo das derivadas da saída dos filtros de Kautz em relação aos seus pólos. As derivadas para os filtros de Laguerre seguem diretamente como um caso particular. Estas derivadas podem ser utilizadas como parte de um método de otimização para obter uma direção de busca para os pólos da base. Adota-se aqui o método de Levenberg-Marquardt (ver Seção 3.2), uma vez que ele contempla vantagens dos métodos de gradiente de maior descida e gradiente conjugado em um único algoritmo, sendo uma escolha usual em problemas de otimização não-linear de porte médio.

Este trabalho está organizado como segue. Na próxima seção, faz-se uma breve revisão dos modelos OBF lineares. Na Seção 3, formula-se um problema de otimização dos pólos das bases ortonormais, apresentando os principais fundamentos matemáticos para resolvê-lo iterativamente. Na Seção 4, obtêm-se as equações necessárias para implementar a metodologia proposta quando o modelo OBF é baseado em funções de Kautz. Na Seção 5, um exemplo simulado ilustra os resultados teóricos e, finalmente, a Seção 6 apresenta as conclusões.

2 Aproximação de modelos lineares usando funções ortonormais

No domínio do tempo discreto, o clássico modelo de resposta ao impulso relaciona a saída $y(k)$ de um processo físico com sua entrada $u(k)$ como (Aguirre, 2004; Heuberger et al., 2005):

$$y(k) = \sum_{\tau=0}^{\infty} h(\tau)u(k-\tau) \quad (1)$$

onde a função $h(\tau)$ é a resposta ao impulso unitário do sistema. Por razões computacionais, os elementos de ordem superior de $h(\tau)$ são usualmente conside-

rados nulos, permitindo-se assim truncar a soma infinita no modelo em (1), o que é possível se o sistema for estável. Entretanto, como este modelo não incorpora recursão (realimentação), ele geralmente requer um número considerável de termos, especialmente se o sistema possuir dinâmica lenta. Esta desvantagem pode ser superada ao se representar h utilizando um conjunto de funções ortonormais $\{\psi_m\}$, como segue (Nelles, 2001; Heuberger et al., 2005):

$$h(k) = \sum_{m=1}^{\infty} \alpha_m \psi_m(k) \quad (2)$$

Se o sistema for estável, a expansão em (2) converge, e portanto ela pode ser aproximada em um número finito (M) de termos. Tem-se assim:

$$\hat{h}(k) = \sum_{m=1}^M \alpha_m \psi_m(k) \quad (3)$$

Assumindo que $u(k) = 0$ para $k < 0$, então o modelo original em (1) pode ser reescrito usando (3) como:

$$\hat{y}(k) = \sum_{m=1}^M \alpha_m \sum_{\tau=0}^k \psi_m(\tau)u(k-\tau) \quad (4)$$

O número M de funções necessárias para se obter uma boa representação depende da complexidade do problema, e representa um compromisso entre a precisão da aproximação e o número de parâmetros do modelo. As funções ortonormais mais comumente utilizadas na representação de sinais e sistemas são apresentadas a seguir.

As bases de funções ortonormais generalizadas (GOBF) são definidas no domínio complexo z como (den Hof et al., 1995; Ninness and Gustafsson, 1997):

$$\Psi_m(z) = \frac{z\sqrt{1-|\beta_m|^2}}{z-\beta_m} \prod_{j=1}^{m-1} \left(\frac{1-\bar{\beta}_j z}{z-\beta_j} \right) \quad (5)$$

$m = 1, 2, \dots$

onde $\beta_m, \bar{\beta}_m \in \mathbb{C}$ são os pólos das GOBF. As realizações correspondentes no domínio do tempo, $\psi_m(k)$, são dadas pela transformada Z inversa de (5) e satisfazem a propriedade de ortonormalidade. O conjunto $\{\psi_m\}$ é completo em $\ell^2[0, \infty)$ se e somente se $\sum_{m=1}^{\infty} (1-|\beta_m|) < \infty$ (Heuberger et al., 2005). Nesse caso, qualquer sinal com energia finita pode ser aproximado com uma precisão pré-determinada fazendo-se o truncamento da expansão (infinita) do sinal nesta base de funções.

Quando todos os pólos de (5) são reais e iguais entre si para qualquer valor de m , i.e. $\beta_m = \bar{\beta}_m = c$, tem-se a base de Laguerre (Fu and Dumont, 1993):

$$\Psi_m(z) = \frac{z\sqrt{1-c^2}}{z-c} \left(\frac{1-cz}{z-c} \right)^{m-1} \quad (6)$$

onde c denota o pólo de Laguerre.

O caso particular da base GOBF quando os pólos em (5) são tais que $\beta_{m+1} = \bar{\beta}_m$ resulta nas funções de Kautz (Wahlberg, 1994; Tanguy et al., 2002):

$$\begin{aligned}\Psi_{2m}(z) &= \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2 + b(c-1)z - c} z \left[\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right]^{m-1} \\ \Psi_{2m-1}(z) &= \frac{\sqrt{1-c^2}}{z^2 + b(c-1)z - c} z(z-b) \left[\frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right]^{m-1}\end{aligned}\quad (7)$$

com $\Psi_{2m}(z)$ denotando os filtros pares de Kautz e $\Psi_{2m-1}(z)$ representando os filtros ímpares. Neste caso, b e c são parâmetros reais que estão relacionados com o par de pólos de Kautz $(\beta, \bar{\beta})$ como $b = (\beta + \bar{\beta})/(1 + \beta\bar{\beta})$ e $c = -\beta\bar{\beta}$.

3 Definição do problema

Esta seção formula o problema de descrever de forma exata e analítica a relação dinâmica entre os parâmetros de modelos OBF com a sua saída. Primeiramente, faz-se uma breve revisão da técnica *back-propagation-through-time* (Nelles, 2001). Depois, discute-se a otimização dos parâmetros de modelos OBF baseada na adaptação desta técnica.

3.1 Técnica *back-propagation-through-time*

Considere um sistema dinâmico linear representado por um mapeamento estático \mathcal{H} aplicado sobre os valores passados dos sinais de entrada e saída, ou seja, $\hat{y}(k) = \mathcal{H}[u(k-1), \dots, u(k-n_u), \hat{y}(k-1), \dots, \hat{y}(k-n_y)]$, onde $\hat{y}(k)$ denota a saída do modelo. Neste tipo de modelo, conhecido como *Output-Error* (OE) (Nelles, 2001), o mapeamento \mathcal{H} possui uma relação dinâmica com a saída do modelo, uma vez que ele está implicitamente e recursivamente incluso nos termos $\hat{y}(k-i)$. Logo, não é possível estimar os parâmetros deste mapeamento diretamente a partir de dados usando uma abordagem padrão de estimação, como um algoritmo de mínimos quadrados. Em vez disso, a abordagem utilizada consiste em representar a dinâmica do sistema por meio de vários modelos estáticos, o que pode ser feito através de recursões no tempo. Para ilustrar essa idéia, considere, por exemplo, o modelo acima com $n_u = n_y = 1$, i.e.:

$$\hat{y}(k) = \mathcal{H}[u(k-1), \hat{y}(k-1)] \quad (8)$$

Suponha que o mapeamento estático \mathcal{H} seja parametrizado por um vetor de parâmetros θ . Este vetor pode ser otimizado se o gradiente de uma função de custo em relação aos seus elementos (parâmetros do modelo) estiver disponível. Por sua vez, o gradiente depende das derivadas da saída do modelo com relação aos elementos de θ . Dado que o modelo é linear, pode-se inferir a partir de (8) que:

$$\frac{\partial \hat{y}(k)}{\partial \theta_i} = \frac{\partial \mathcal{H}_1[\cdot]}{\partial \theta_i} + \frac{\partial \mathcal{H}_2[\cdot]}{\partial \hat{y}(k-1)} \cdot \frac{\partial \hat{y}(k-1)}{\partial \theta_i} \quad (9)$$

onde \mathcal{H}_2 e \mathcal{H}_1 são as porções complementares de \mathcal{H} que dependem e não dependem de $\hat{y}(k-1)$, respectivamente, e θ_i são as componentes do vetor de

parâmetros θ . O primeiro termo de (9) representa a parte estática daquela equação. O segundo termo surge a partir da componente recursiva e representa a parte dinâmica da equação. Note que, se $\hat{y}(k-1)$ fosse um sinal externo, tal como a saída medida do processo, $y(k-1)$, não haveria nenhuma dependência deste sinal com os parâmetros do modelo, e assim o segundo termo de (9) seria nulo. Em modelos OE, entretanto, $\hat{y}(k-1)$ é a saída do próprio modelo em um instante de tempo anterior, e que depende de θ . Portanto, o segundo termo de (9) requer as derivadas da saída do modelo no instante de tempo anterior com relação aos parâmetros do modelo. Estas derivadas podem ser calculadas a partir de $\hat{y}(k-1) = \mathcal{H}[u(k-2), \hat{y}(k-2)]$ como:

$$\frac{\partial \hat{y}(k-1)}{\partial \theta_i} = \frac{\partial \mathcal{H}_1[\cdot]}{\partial \theta_i} + \frac{\partial \mathcal{H}_2[\cdot]}{\partial \hat{y}(k-2)} \cdot \frac{\partial \hat{y}(k-2)}{\partial \theta_i} \quad (10)$$

Novamente, o segundo termo em (10) requer as derivadas da saída no instante de tempo anterior. Este procedimento continua até se chegar às condições iniciais em $k=0$; neste instante, $\hat{y}(0) = y(0)$ e, conseqüentemente, $\partial \hat{y}(0)/\partial \theta_i = 0$.

Em resumo, a técnica *back-propagation-through-time*, ilustrada na Figura 1, decompõe a dinâmica do sistema em uma série de representações estáticas. Esta abordagem permite descrever as derivadas da saída do modelo em termos somente das condições iniciais e do sinal de entrada, fazendo-se uma recursão de k passos no tempo.

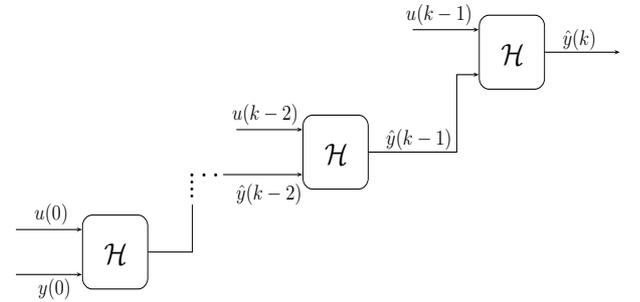


Figura 1: Técnica *back-propagation-through-time*.

3.2 Otimização dos pólos

Denotando $\check{\psi}_m(k)$ como a saída do filtro com resposta ao impulso $\psi_m(k)$ à entrada $u(k)$, ou seja

$$\check{\psi}_m(k) = \sum_{\tau=0}^k \psi_m(\tau) u(k-\tau) \quad (11)$$

$$\check{\Psi}_m(z) = \Psi_m(z) U(z) \quad (12)$$

o modelo (4) pode ser reescrito como:

$$\hat{y}(k) = \sum_{m=1}^M \alpha_m \check{\psi}_m(k) \quad (13)$$

A estratégia adotada aqui consiste em otimizar o vetor de parâmetros de Kautz $\mathbf{p} \triangleq [b \quad c]^T$ e o vetor de

coeficientes $\alpha \triangleq [\alpha_1 \ \alpha_2 \ \dots \ \alpha_M]^T$ que parametriza a base ortonormal $\{\psi_m\}$. Para isso, defina o seguinte problema de otimização:

$$\min_{\theta} J \triangleq \frac{1}{2} \sum_{k=1}^{N_d} [y(k) - \hat{y}(k)]^2 \quad (14)$$

onde $\theta = [\mathbf{p}^T \quad \alpha^T]^T$ e N_d é o número de dados entrada/saída disponíveis. O gradiente da função de custo J em (14) com relação ao vetor de pólos \mathbf{p} é obtido usando (13) como:

$$\begin{aligned} \nabla_{\mathbf{p}} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\mathbf{p}} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} \sum_{m=1}^M [\hat{y}(k) - y(k)] \alpha_m \nabla_{\mathbf{p}} \check{\psi}_m(k) \end{aligned} \quad (15)$$

O gradiente de J com relação a α pode ser calculado como segue:

$$\begin{aligned} \nabla_{\alpha} J &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \nabla_{\alpha} \hat{y}(k) \\ &= \sum_{k=1}^{N_d} [\hat{y}(k) - y(k)] \check{\psi}(k) \end{aligned} \quad (16)$$

onde $\check{\psi}(k) = [\check{\psi}_1(k) \ \check{\psi}_2(k) \ \dots \ \check{\psi}_M(k)]^T$.

A contribuição central deste trabalho é a obtenção de equações detalhadas para o cálculo de $\nabla_{\mathbf{p}} \check{\psi}_m(k) = [\partial \check{\psi}_m(k) / \partial b \quad \partial \check{\psi}_m(k) / \partial c]$ e a sua aplicação à otimização dos pólos do modelo (13) usando o bem conhecido algoritmo de Levenberg-Marquardt (Nelles, 2001). A base deste algoritmo é uma aproximação quadrática de J na vizinhança de seus argumentos (parâmetros do modelo). O algoritmo aproxima a matriz Hessiana de J com relação aos parâmetros usando somente informação das derivadas de primeira ordem correspondentes (gradientes). Especificamente, ele estima a matriz Hessiana usando a soma de produtos dos gradientes em (16) e (15). Portanto, o algoritmo de Levenberg-Marquardt é um método de pseudo-segunda ordem, pois ele utiliza apenas avaliações da função e de seu gradiente. Quando inicializado em um ponto θ^0 , o algoritmo produz uma série de pontos que convergem para uma solução ótima θ_{otimo} de $J(\theta)$. Ele pode ser visto como uma combinação do método do gradiente de maior descida e do método de Gauss-Newton (Nelles, 2001): quando o resultado está longe da solução ótima, o algoritmo se comporta como um método do gradiente (lento, mas com convergência garantida); próximo da solução ótima, o algoritmo se torna um método de Gauss-Newton, convergindo assim mais rápido.

A Seção 4 apresenta um estudo quando a base de Kautz $\{\psi_m\}$ é utilizada para implementar o modelo OBF em (11) e (13). A formulação baseada em funções de Laguerre segue como um caso particular.

4 Formulação baseada em funções de Kautz

Esta seção desenvolve um esquema como aquele na Figura 1 para obter a saída de um modelo baseado em funções de Kautz em termos das condições iniciais e do sinal de entrada. Estuda-se somente os filtros pares de Kautz; as equações relativas aos filtros ímpares são obtidas de maneira similar e foram omitidas por questão de espaço.

De acordo com as equações (12) e (7), pode-se escrever a saída do primeiro filtro par de Kautz como:

$$\check{\Psi}_2(z) = \frac{\sqrt{(1-c^2)(1-b^2)} z}{z^2 + b(c-1)z - c} U(z)$$

Equivalentemente, no domínio do tempo:

$$\begin{aligned} \check{\psi}_2(k+2) + b(c-1)\check{\psi}_2(k+1) - c\check{\psi}_2(k) &= \\ &= \sqrt{(1-c^2)(1-b^2)} u(k+1) \end{aligned} \quad (17)$$

Definindo $x_a(k) \triangleq \check{\psi}_2(k)$ e $x_b(k) \triangleq x_a(k+1) = \check{\psi}_2(k+1)$, reescreve-se a equação a diferenças (17) como a seguinte realização no espaço de estados:

$$\begin{aligned} \begin{bmatrix} x_a(k+1) \\ x_b(k+1) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ c & -b(c-1) \end{bmatrix} \begin{bmatrix} x_a(k) \\ x_b(k) \end{bmatrix} + \\ &+ \begin{bmatrix} 0 \\ \sqrt{(1-c^2)(1-b^2)} \end{bmatrix} u(k+1) \end{aligned} \quad (18)$$

$$\check{\psi}_2(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_a(k) \\ x_b(k) \end{bmatrix}$$

que corresponde à clássica representação:

$$\begin{cases} \mathbf{x}_2(k+1) = \mathcal{A}\mathbf{x}_2(k) + \mathcal{B}u(k+1) \\ \check{\psi}_2(k) = \mathcal{C}\mathbf{x}_2(k) \end{cases} \quad (19)$$

A solução de (19) para $\check{\psi}_2(k)$ é bem conhecida e dada por:

$$\check{\psi}_2(k) = \mathcal{C}\mathcal{A}^k \mathbf{x}_2(0) + \sum_{i=0}^{k-1} \mathcal{C}\mathcal{A}^i \mathcal{B}u(k-i) \quad (20)$$

onde $\mathbf{x}_2(0) = [\check{\psi}_2(0) \ \check{\psi}_2(1)]^T$ e $\mathcal{A}, \mathcal{B}, \mathcal{C}$ estão em (18).

O cálculo das derivadas de (20) com relação aos parâmetros b e c de Kautz resulta em:

$$\begin{aligned} \frac{\partial \check{\psi}_2(k)}{\partial b} &= \mathcal{C} \frac{\partial (\mathcal{A}^k)}{\partial b} \mathbf{x}_2(0) + \\ &+ \mathcal{C} \sum_{i=0}^{k-1} \left[\frac{\partial (\mathcal{A}^i)}{\partial b} \mathcal{B} + \mathcal{A}^i \frac{\partial \mathcal{B}}{\partial b} \right] u(k-i) \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{\partial \check{\psi}_2(k)}{\partial c} &= \mathcal{C} \frac{\partial (\mathcal{A}^k)}{\partial c} \mathbf{x}_2(0) + \\ &+ \mathcal{C} \sum_{i=0}^{k-1} \left[\frac{\partial (\mathcal{A}^i)}{\partial c} \mathcal{B} + \mathcal{A}^i \frac{\partial \mathcal{B}}{\partial c} \right] u(k-i) \end{aligned} \quad (22)$$

Expressões relativas aos filtros de Kautz com $m \geq 2$ podem ser obtidas de maneira semelhante, como mostrado a seguir. Voltando à definição (7), tem-se:

$$\check{\Psi}_{2m}(z) = \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \check{\Psi}_{2(m-1)}(z) \quad (23)$$

Usando um raciocínio análogo àquele usado para $\check{\psi}_2(k)$, é possível reescrever (23) como:

$$\begin{cases} \mathbf{x}_{2m}(k+1) = \mathbf{A}\mathbf{x}_{2m}(k) + \mathbf{B}_1\check{\psi}_{2(m-1)}(k+2) \\ \quad + \mathbf{B}_2\check{\psi}_{2(m-1)}(k+1) \\ \quad + \mathbf{B}_3\check{\psi}_{2(m-1)}(k) \\ \check{\psi}_{2m}(k) = \mathbf{C}\mathbf{x}_{2m}(k) \end{cases} \quad (24)$$

onde $\mathbf{x}_{2m}(0) = [\check{\psi}_{2m}(0) \quad \check{\psi}_{2m}(1)]^T$. As matrizes \mathbf{A}, \mathbf{C} em (24) são as mesmas de (18). Além disso:

$$\mathbf{B}_1 = \begin{bmatrix} 0 \\ -c \end{bmatrix} \quad \mathbf{B}_2 = \begin{bmatrix} 0 \\ b(c-1) \end{bmatrix} \quad \mathbf{B}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

A solução de (24) para $\check{\psi}_{2m}(k)$ é dada por:

$$\begin{aligned} \check{\psi}_{2m}(k) = & \mathbf{C}\mathbf{A}^k\mathbf{x}_{2m}(0) + \\ & \sum_{i=0}^{k-1} \mathbf{C}\mathbf{A}^i \left[\mathbf{B}_1\check{\psi}_{2(m-1)}(k+1-i) + \right. \\ & \left. \mathbf{B}_2\check{\psi}_{2(m-1)}(k-i) + \mathbf{B}_3\check{\psi}_{2(m-1)}(k-1-i) \right] \end{aligned} \quad (25)$$

A derivada de (25) com relação aos parâmetros b e c resulta em (26) e (27) (veja na próxima página).

5 Exemplo ilustrativo

Considere um sistema com resposta ao impulso unitário dada por:

$$H_s(z) = \frac{3z^3 + 1.2z^2 - 2.85z + 1.5}{5z^3 - 8z^2 + 4.5z - 0.6} \quad (28)$$

Suponha que este sistema seja excitado por um sinal de entrada aleatório $u(k)$ com média zero e variância unitária. Os dados de saída são gerados através da relação $y(k) = \sum_{\tau=0}^k h_s(\tau)u(k-\tau)$, onde $h_s(k) = \mathcal{Z}^{-1}[H_s(z)]$. O objetivo deste exemplo é obter os pólos ótimos de Kautz a partir apenas de dados entrada/saída do sistema, usando o método de otimização proposto.

Neste exemplo, os pólos de Kautz iniciais são $(\beta^0, \bar{\beta}^0) = 0,1 \pm i0,1$, o que implica em $\mathbf{p}^0 = [b^0 \quad c^0] = [0,196 \quad -0,02]^T$. Os resultados obtidos na simulação estão resumidos na Tabela 1, para diferentes níveis de precisão, usando $M = 6$ funções de Kautz. A Figura 2 ilustra a evolução de um dos pólos conjugados de Kautz.

Tabela 1: Pólos ótimos de Kautz para $H_s(z)$.

$(\beta_{\text{otimo}}, \bar{\beta}_{\text{otimo}})$	J_{otimo}	Iterações
$0,6698 \pm i0,3252$	$1,034 \times 10^{-3}$	18
$0,6706 \pm i0,3242$	$9,997 \times 10^{-4}$	19
$0,6709 \pm i0,3241$	$9,747 \times 10^{-5}$	20
$0,6710 \pm i0,3240$	$5,881 \times 10^{-6}$	21

O método convergiu para o par de pólos complexos conjugados $(\beta_{\text{otimo}}, \bar{\beta}_{\text{otimo}}) = 0,671 \pm i0,324$, com erro quadrático associado $J_{\text{otimo}} = 5,881 \times 10^{-6}$,

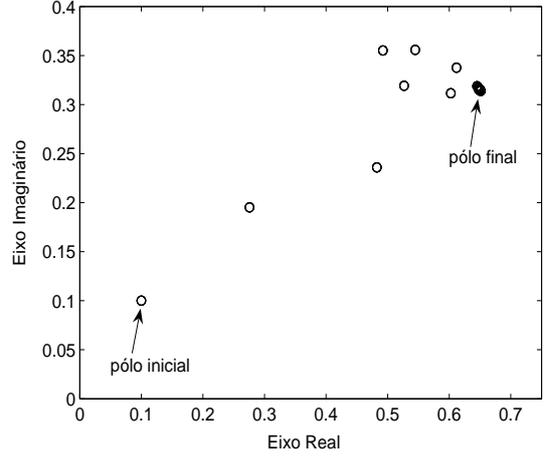


Figura 2: Evolução de um dos pólos de Kautz.

o que representa uma descrição quase perfeita da saída do sistema. Experimentos adicionais, omitidos por questão de espaço, também foram realizados usando uma variedade de sistemas distintos e com diferentes pólos iniciais, levando a resultados similares. Estes experimentos produziram modelos muito precisos quando avaliados frente a dados diferentes daqueles adotados na otimização (dados de validação).

6 Conclusões

Este trabalho apresentou uma nova estratégia de otimização numérica dos pólos de Laguerre e de Kautz em modelos OBF. Obteve-se expressões analíticas para as derivadas da saída dos filtros de Kautz com relação aos pólos de Kautz. Expressões para os filtros de Laguerre seguem diretamente como um caso particular. Estas expressões fornecem direções de busca exatas a serem utilizadas por um método iterativo para a otimização dos pólos da base ortonormal. Resultados de simulação mostraram que a metodologia proposta permitiu obter uma boa aproximação de modelos OBF, tornando assim um assunto promissor para estudos futuros. Atualmente sob investigação estão extensões dos resultados apresentados neste trabalho com respeito as bases ortonormais generalizadas e aos modelos não-lineares de Volterra (Nelles, 2001; Aguirre, 2004).

Agradecimentos

Os autores agradecem o apoio do CNPq e da Fapesp.

Referências

- Aguirre, L. A. (2004). *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*, Editora da UFMG.
- Bodin, P., Villemoes, L. F. and Wahlberg, B. (1997). An algorithm for selection of best orthonormal

$$\begin{aligned}
\frac{\partial \check{\psi}_{2m}(k)}{\partial b} &= c \frac{\partial (\mathcal{A}^k)}{\partial b} \mathbf{x}_{2m}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\mathcal{B}_1 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k+1-i) \right. \right. \\
&+ \left. \frac{\partial \mathcal{B}_2}{\partial b} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_3 \frac{\partial}{\partial b} \check{\psi}_{2(m-1)}(k-1-i) \right) \\
&+ \left. \frac{\partial (\mathcal{A}^i)}{\partial b} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_3 \check{\psi}_{2(m-1)}(k-1-i) \right) \right] \quad (26)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \check{\psi}_{2m}(k)}{\partial c} &= c \frac{\partial (\mathcal{A}^k)}{\partial c} \mathbf{x}_{2m}(0) + c \sum_{i=0}^{k-1} \left[\mathcal{A}^i \left(\frac{\partial \mathcal{B}_1}{\partial c} \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_1 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k+1-i) \right. \right. \\
&+ \left. \frac{\partial \mathcal{B}_2}{\partial c} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_2 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_3 \frac{\partial}{\partial c} \check{\psi}_{2(m-1)}(k-1-i) \right) \\
&+ \left. \frac{\partial (\mathcal{A}^i)}{\partial c} \left(\mathcal{B}_1 \check{\psi}_{2(m-1)}(k+1-i) + \mathcal{B}_2 \check{\psi}_{2(m-1)}(k-i) + \mathcal{B}_3 \check{\psi}_{2(m-1)}(k-1-i) \right) \right] \quad (27)
\end{aligned}$$

- rational basis, *Proc. 36th Conference on Decision and Control*, San Diego, California, USA.
- Bokor, J. and Schipp, F. (1998). Approximate identification in Laguerre and Kautz bases, *Automatica* **34**(4): 463–468.
- Broome, P. W. (1965). Discrete orthonormal sequences, *Journal of the Assoc. for Comp. Machinery* **12**(2): 151–168.
- Campello, R. J. G. B., Amaral, W. C. and Favier, G. (2006). A note on the optimal expansion of Volterra models using Laguerre functions, *Automatica* **42**(4): 689–693.
- Campello, R. J. G. B., Favier, G. and Amaral, W. C. (2004). Optimal expansions of discrete-time Volterra models using Laguerre functions, *Automatica* **40**(5): 815–822.
- Campello, R. J. G. B., Oliveira, G. H. C. and Amaral, W. C. (2007). Identificação e Controle de Processos via Desenvolvimentos em Séries Ortonormais: Partes A (Identificação) e B (Controle), *Controle & Automação* **18**(3): 301–336.
- da Rosa, A., Campello, R. J. G. B. and Amaral, W. C. (2007). Choice of free parameters in expansions of discrete-time Volterra models using Kautz functions, *Automatica* **43**(6): 1084–1091.
- da Rosa, A., Campello, R. J. G. B. and Amaral, W. C. (2008). An optimal expansion of Volterra models using independent Kautz bases for each kernel dimension, *International Journal of Control* **81**(6): 962–975.
- den Hof, P. M. J. V., Heuberger, P. S. C. and Bokor, J. (1995). System identification with generalized orthonormal basis functions, *Automatica* **31**(12): 1821–1834.
- Fu, Y. and Dumont, G. A. (1993). An optimum time scale for discrete Laguerre network, *IEEE Trans. Automatic Control* **38**(6): 934–938.
- Heuberger, P. S. C., den Hof, P. M. J. V. and Wahlberg, B. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions*, Springer.
- Nelles, O. (2001). *Nonlinear System Identification*, Springer.
- Ninness, B. and Gustafsson, F. (1997). A unifying construction of orthonormal bases for system identification, *IEEE Trans. Automatic Control* **42**(4): 515–521.
- Patwardhan, S. C. and Shah, S. L. (2005). From data to diagnosis and control using generalized orthonormal basis filters. Part I: Development of state observers, *Journal of Process Control* **15**(7): 819–835.
- Sarroukh, B. E., van Eijndhoven, S. J. L. and den Brinker, A. C. (2001). An iterative solution for the optimal poles in a Kautz series, *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Piscataway, New Jersey, USA.
- Silva, T. O. (1994). Optimality conditions for truncated Laguerre networks, *IEEE Trans. Signal Processing* **42**(9): 2528–2530.
- Silva, T. O. (1995). Optimality conditions for truncated Kautz networks with two periodically repeating complex conjugate poles, *IEEE Trans. Automatic Control* **40**(2): 342–346.
- Tanguy, N., Morvan, R., Vilbé, P. and Calvez, L. C. (1995). Optimum choice of free parameter in orthonormal approximations, *IEEE Trans. Automatic Control* **40**(10): 1811–1813.
- Tanguy, N., Morvan, R., Vilbé, P. and Calvez, L. C. (2002). Pertinent choice of parameters for discrete Kautz approximation, *IEEE Trans. Automatic Control* **47**(5): 783–787.
- Wahlberg, B. (1994). System identification using Kautz models, *IEEE Trans. Automatic Control* **39**(6): 1276–1282.