

Estimação de Parâmetros de Modelos S-Systems Utilizando B-Spline

Carolina Borges, Esdras Penêdo de Carvalho, Mauro A. da Silva Sá Ravagnani
Programa de Pós-Graduação em Engenharia Química
Universidade Estadual de Maringá (UEM)
Maringá, PR, Brasil
carolinaborges.eq@gmail.com

Resumo—Em problemas de modelagem bioquímica, frequentemente é necessário estimar parâmetros de modelos oriundos de equações diferenciais não-lineares que não possuem, em sua maioria, solução analítica. A modelagem de perfis metabólicos, por exemplo, exige modelos que representem séries de dados temporais adequadas para o processo estudado. Um dos principais modelos usados é o chamado *S-system*, que surgiu baseado na Lei de Ação das Massas e descreve o comportamento dinâmico do fenômeno de modo satisfatório. Desenvolvimento de algoritmos matemáticos baseando-se em métodos gradientes ou técnicas heurísticas para estimação de parâmetros de tais modelos têm sido largamente estudados. Porém, problemas de convergência ou alto custo computacional destes métodos motivaram a busca por métodos alternativos. No presente trabalho, é considerada uma representação da função aproximante, por meio de B-splines e minimiza-se os desvios entre as derivadas dessa função e o lado direito do sistema de EDOs, aplicando-se o método sem derivadas *Direct MultiSearch* (DMS). Resultados preliminares para sistemas com quatro, cinco e seis metabólitos apresentaram bons resultados, mostrando que a metodologia é promissora.

Palavras-chave—S-Systems; B-Spline; Estimação de Parâmetros; Otimização

I. INTRODUÇÃO

O papel da modelagem biomatemática compreende a representação de um sistema biológico por um modelo simplificado que seja mais fácil de analisar, inferir, prever, extrapolar, manipular e otimizar que o sistema em estudo. Em pesquisas recentes, uma das principais atividades na Biologia Molecular tem sido o estudo e desenvolvimento de modelos matemáticos que possam representar de modo satisfatório a estrutura e/ou dinâmica de processos e caminhos metabólicos ou redes de regulação gênica. Os avanços nas tecnologias de medições, tais como, espectrometria de massa, ressonância magnética nuclear etc., têm fornecido uma melhor compreensão de perfis temporais de metabólitos ou proteínas que podem ser usados, por exemplo, para mapear caminhos metabólicos e/ou proteômicos que produzam novas informações acerca da natureza de certas doenças e possibilidades de tratamento. Esses perfis metabólicos são medições simultâneas de elementos que podem ser obtidas num instante de tempo ou numa série de dados temporais. A fim de representar matematicamente essas informações, um padrão de modelo matemático é estabelecido e assim torna-se necessário aplicar algum algoritmo computacional para determinar o conjunto de parâmetros que possibilite o ajuste adequado da estrutura de modelagem especificada às

informações medidas. Um dos principais modelos utilizados para representar tais fenômenos biológicos são os conhecidos modelos *S-systems*, que se caracterizam por descrever o comportamento dinâmico do sistema de modo satisfatório com precisão e flexibilidade matemática. Esses modelos surgem da aplicação da Lei de Ação das Massas, sendo representados por um conjunto de equações diferenciais ordinárias (EDO) não-lineares que dificilmente possuem solução analítica. Tais modelos apresentam um grande número de parâmetros a serem estimados e, em geral, essa estimação é um problema de otimização não-linear em que os mesmos são obtidos minimizando-se a soma ponderada das diferenças entre os dados experimentais e os valores calculados teoricamente. Diferentes abordagens foram aplicadas para estimar parâmetros de modelos *S-systems*. Muitos autores estudaram a modelagem e estimação de parâmetros de sistemas bioquímicos e aplicaram algoritmos de otimização para tais tipos de problemas ([3], [6], [7], [8], [11]). Métodos gradientes são alguns dos principais algoritmos utilizados para estimativa de parâmetros, mas em geral convergem para soluções locais devido à falta de uma boa estimativa inicial para a solução. Por outro lado, métodos heurísticos, tais como Algoritmos Genéticos (GA), são mais robustos na busca de um minimizador global, mas podem exigir um custo computacional substancial e não garantem convergência global. Uma alternativa consiste na utilização de algoritmos híbridos, combinando estratégias heurísticas com métodos gradientes ([2], [5], [9]). Uma outra alternativa para otimização dos parâmetros consiste em se empregar métodos de colocação que utilizam uma representação global da solução através de um conjunto conveniente de funções de base, geralmente polinômios de baixa ordem [10]. Neste trabalho utilizamos B-splines. Em seguida, aproxima-se as derivadas dessa e define-se uma função objetivo que consiste em minimizar a diferença entre essas derivadas e o lado direito do sistema de EDOs. Neste trabalho, apresentamos um algoritmo para estimar parâmetros de modelos *S-systems* com dados de séries temporais sem ruído. Inicialmente, esses dados são ajustados considerando-se uma função de ajuste como sendo uma combinação linear de B-splines e, em seguida, os parâmetros do modelo *S-systems* são determinados aplicando-se o método sem derivadas *Direct MultiSearch* (DMS) [1].

II. FORMULAÇÃO DO PROBLEMA

A. Modelo S-system

Um modelo *S-system* pode ser definido como:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^N X_j^{g_{ij}} - \beta_i \prod_{j=1}^N X_j^{h_{ij}}, \quad i = 1, \dots, N \quad (1)$$

em que N é o número de componentes da rede metabólica, α_i e β_i são parâmetros não-negativos denominados constantes de taxa, os quais referem-se às taxas de produção e consumo, respectivamente. Os parâmetros g_{ij} e h_{ij} são parâmetros denominados coeficientes cinéticos, que refletem a ação que determinado metabólito X_j possui no termo de produção ou consumo. Valores positivos significam um efeito catalisador exercido por X_j , enquanto que valores negativos significam efeito inibitório. Valores iguais a zero significam que X_j não exerce nenhum efeito [4].

III. METODOLOGIA

A metodologia para otimização dos parâmetros, com consequente resolução do sistema de EDOs, consiste em empregar-se métodos de colocação que utilizam uma representação global da solução através de um conjunto conveniente de funções de base, tais como, B-splines [10]. Assim, supõe-se que uma determinada variável de estado $X_i(t)$ possa ser representada por uma combinação linear da seguinte forma:

$$X_i(t) = \sum_{j=0}^p b_j B_j(t), \quad (2)$$

em que b_j são os coeficientes a serem determinados e B_j , polinômios denominados B-splines. Em seguida, aproxima-se as derivadas de (2) e define-se uma função objetivo que consiste em minimizar a diferença entre essas derivadas e o lado direito de (1).

IV. RESULTADOS PARCIAIS

Dados experimentais foram simulados para testes em sistemas contendo quatro, cinco e seis elementos, e os resultados são mostrados nas Figuras 1 a 3.

A. Exemplo Numérico 1

O primeiro modelo *S-system* considerado neste trabalho possui quatro metabólitos [4]:

$$\begin{aligned} \frac{dX_1}{dt} &= 12X_3^{-0.8} - X_1^{0.5} \\ \frac{dX_2}{dt} &= 8X_1^{0.5} - 3X_2^{0.75} \\ \frac{dX_3}{dt} &= 3X_2^{0.75} - 5X_3^{0.5}X_4^{0.2} \\ \frac{dX_4}{dt} &= 2X_1^{0.5} - 6X_4^{0.8}. \end{aligned} \quad (3)$$

O modelo possui 17 parâmetros a serem determinados. Os dados foram gerados resolvendo-se o sistema acima com condições iniciais $X(0) = [10, 1, 2, 3]^T$ no intervalo $[0, 5]$ com $\Delta t = 0.1$. A Tabela I mostra os valores dos parâmetros estimados, os quais estão muito próximos do respectivos valores exatos.

Tabela I

PARÂMETROS ESTIMADOS PARA O MODELO DE QUATRO ELEMENTOS.

Parâmetro	Valor Exato	Valor Estimado	Erro Relativo (%)
α_1	12.0000	11.8906	0.9115
β_1	10.0000	9.9062	0.9375
g_{13}	-0.8000	-0.8000	0
h_{11}	0.5000	0.5000	0
α_2	8.0000	7.7969	2.5391
β_2	3.0000	2.9375	2.0833
g_{21}	0.5000	0.5000	0
h_{22}	0.7500	0.7500	0
α_3	3.0000	2.9375	2.0833
β_3	5.0000	4.8750	2.5000
g_{32}	0.7500	0.7500	0
h_{33}	0.5000	0.5000	0
h_{34}	0.2000	0.2000	0
α_4	2.0000	2.0000	0
β_4	6.0000	5.9688	0.5208
g_{41}	0.5000	0.5000	0
h_{44}	0.8000	0.8000	0

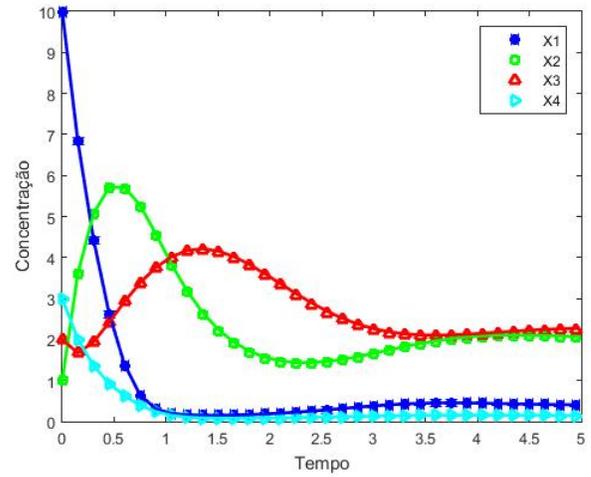


Figura 1. Dados de Tempo \times Concentração para o modelo de quatro metabólitos.

B. Exemplo Numérico 2

O segundo modelo *S-system* considerado possui cinco metabólitos [11]:

$$\begin{aligned} \frac{dX_1}{dt} &= 5X_3X_5^{-1} - 10X_1^2 \\ \frac{dX_2}{dt} &= 10X_1^2 - 10X_2^2 \\ \frac{dX_3}{dt} &= 10X_2^{-1} - 10X_2^{-1}X_3^2 \\ \frac{dX_4}{dt} &= 8X_3^2X_5^{-1} - 10X_4^2 \\ \frac{dX_5}{dt} &= 10X_4^2 - 10X_5^2. \end{aligned} \quad (4)$$

O modelo possui 23 parâmetros a serem determinados.

Tabela II
PARÂMETROS ESTIMADOS PARA O MODELO DE CINCO METABÓLITOS.

Parâmetro	Valor Exato	Valor Estimado	Erro Relativo (%)
α_1	5	5.0000	0
β_1	10	10.0625	0.6250
g_{13}	1	1.0000	0
g_{15}	-1	-1.0000	0
h_{11}	2	2.0000	0
α_2	10	9.9062	0.9375
β_2	10	9.9375	0.6250
g_{21}	2	2.0000	0
h_{22}	2	2.0000	0
α_3	10	10.0000	0
β_3	10	10.0000	0
g_{32}	-1	-1.0000	0
h_{32}	-1	-1.0000	0
h_{33}	2	2.0000	0
α_4	8	8.0000	0
β_4	10	10.0312	0.3125
g_{43}	2	2.0000	0
g_{45}	-1	-1.0000	0
h_{44}	2	1.9375	3.1250
α_5	10	10.1250	1.2500
β_5	10	9.9375	0.6250
g_{54}	2	2.1250	6.2500
h_{55}	2	1.8125	9.3750

Tabela III
PARÂMETROS ESTIMADOS PARA O MODELO DE SEIS METABÓLITOS.

Parâmetro	Valor Exato	Valor Estimado	Erro Relativo (%)
α_1	10	9.9980	0.0195
β_1	5	5.0000	0
g_{13}	-2	-1.9980	0.0977
g_{15}	1	1.0000	0
h_{11}	0.5	0.5000	0
α_2	5	4.9961	0.0781
β_2	10	9.9883	0.1172
g_{21}	0.5	0.5000	0
h_{22}	0.5	0.5000	0
α_3	2	2.0000	0
β_3	1.25	1.2500	0
g_{32}	0.5	0.5000	0
h_{33}	0.5	0.5000	0
α_4	8	7.9941	0.0732
β_4	5	5.0000	0
g_{42}	0.5	0.5000	0
h_{44}	0.5	0.5000	0
α_5	0.5	0.5000	0
β_5	1	1.0000	0
h_{56}	1	1.0000	0
α_6	1	1.0000	0
β_6	0.5	0.5000	0
g_{65}	1	1.0000	0

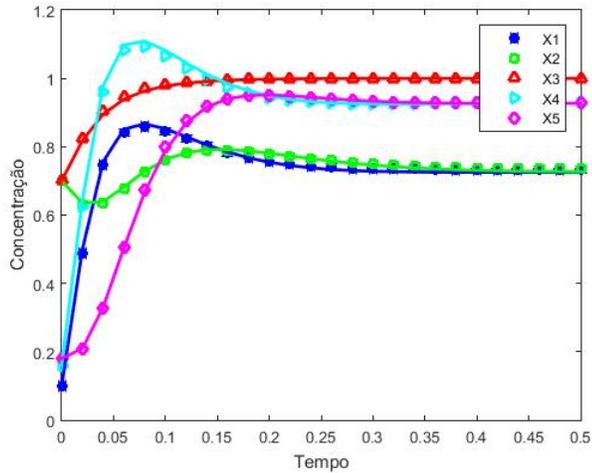


Figura 2. Dados de Tempo \times Concentração para o modelo de cinco metabólitos.

Os dados foram gerados resolvendo-se o sistema acima com condições iniciais $X(0) = [0.1, 0.7, 0.7, 0.16, 0.18]^T$ no intervalo $[0, 0.5]$ com $\Delta t = 0.01$. A Tabela II mostra os valores dos parâmetros estimados, os quais estão muito próximos dos respectivos valores exatos.

C. Exemplo Numérico 3

O terceiro modelo S -system considerado possui seis metabólitos [11]:

$$\begin{aligned}
 \frac{dX_1}{dt} &= 10X_3^{-2}X_5 - 5X_1^{0.5} \\
 \frac{dX_2}{dt} &= 5X_1^{0.5} - 10X_2^{0.5} \\
 \frac{dX_3}{dt} &= 2X_2^{0.5} - 1.25X_3^{0.5} \\
 \frac{dX_4}{dt} &= 8X_2^{0.5} - 5X_4^{0.5} \\
 \frac{dX_5}{dt} &= 0.5 - X_6 \\
 \frac{dX_6}{dt} &= X_5 - 0.5.
 \end{aligned} \tag{5}$$

O modelo possui 23 parâmetros a serem determinados. Os dados foram gerados resolvendo-se o sistema acima com condições iniciais $X(0) = [1.1, 0.5, 0.9, 0.75, 0.5, 0.75]^T$ no intervalo $[0, 10]$ com $\Delta t = 0.1$. A Tabela III mostra os valores dos parâmetros estimados, os quais estão muito próximos dos respectivos valores exatos.

V. CONCLUSÃO

A estimação de parâmetros em modelos que envolvem processos bioquímicos é usualmente um problema de otimização não-linear. Neste trabalho, apresentamos um procedimento para estimar parâmetros de modelos S -systems com dados de séries temporais sem ruído. Foi considerada uma representação

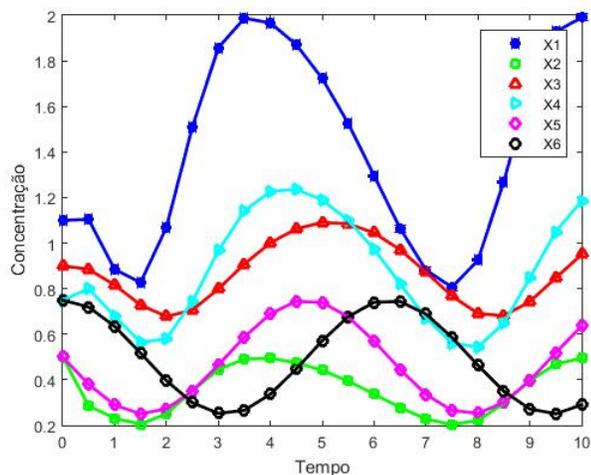


Figura 3. Dados de Tempo \times Concentração para o modelo de seis metabólitos.

da função aproximante, por meio de B-splines, e o problema de otimização resumiu-se a minimizar os desvios entre as derivadas dessa função e o lado direito do sistema de EDOs nos dados experimentais. Foi aplicado o algoritmo Direct MultiSearch (DMS) para estimar os parâmetros dos modelos *S-systems* e os resultados computacionais demonstram a aplicabilidade e eficiência da abordagem proposta.

AGRADECIMENTOS

Este trabalho é financiado pela CAPES.

REFERÊNCIAS

- [1] Custódio, A. L., Madeira, J. F. A., Vaz, A. I. F., & Vicente, L. N. (2011). Direct multisearch for multiobjective optimization. *SIAM Journal on Optimization*, 21, 1109-1140.
- [2] Da Ros, S., Colusso, G., Weschenfelder, T. A., de Marsillac Terra, L., De Castilhos, F., Corazza, M. L., & Schwaab, M. (2013). A comparison among stochastic optimization algorithms for parameter estimation of biochemical kinetic models. *Applied Soft Computing*, 13, 2205-2214.
- [3] L.-Z. Liu, F.-X. Wu, e W. Zhang, *Alternating Weighted Least Squares Parameter Estimation for Biological S-Systems*. IEEE 6th International Conference on Systems Biology (ISB), Xian, China, August 18-20, 2012.
- [4] Voit, E. O., *Computational analysis of biochemical systems: A practical guide for biochemists and molecular biologists*. Cambridge, UK: Cambridge University, 2000.
- [5] Katare, S., Bhan, A., Caruthers, J. M., Delgass, W. N., & Venkatasubramanian, V. (2004). A hybrid genetic algorithm for efficient parameter estimation of large kinetic models. *Computers & Chemical Engineering*, 28, 2569-2581.
- [6] Liu, L.-Z. Wu, F. -X. e Zhang, W.-J. *Estimating parameters of S-systems by an auxiliary function guided coordinate descent method.*, *Systems Science & Control Engineering*, 2:1, 125-134, 2014
- [7] Matsuoka, Y., Jahan, N. e Kurata, H. *S-system-based analysis of the robust properties common to many biochemical network models.*, *Bioprocess Biosystem Engineering*, 39, 735-746, 2016
- [8] Mekin, N., Nounou, H., Nounou, M., Datta, A. e Dougherty, E. R. *S-system-based analysis of the robust properties common to many biochemical network models.*, 50th IEEE Conference on Decision and Control and European Control Conference, 4424 - 4429, 2011
- [9] Shacham, M., & Brauner, N. (2014). Application of stepwise regression for dynamic parameter estimation. *Computers & Chemical Engineering*, 69, 26-38.
- [10] Brewer, D., Barenco, M., Callard, R., Hubank, M., Stark, J. *Fitting ordinary differential equations to short time course data*. *Philosophical Transactions of Royal Society A*, 366, 519-544, 2008.
- [11] Yang, X., Dent, J. E., Nardini, C. *An S-System Parameter Estimation Method (SPEM) for Biological Networks*. *Journal of Computational Biology*, 19(2), 175-187. 2012.