Multi-objective Optimization of Biological Systems Represented by S-system Models

Gongxian Xu
Department of Mathematics
Bohai University
Jinzhou, China
E-mail: dutxugx@yahoo.com.cn

***Abstract*—This paper considers multi-objective optimization problems of biological systems. The biological system is represented by the S-system formalism. The advantage of this representation is that the steady-state equations are linear when the variables of the models are expressed in logarithmic coordinates. Profiting from this special property of S-system models, we transform the original nonlinear problem into a multi-objective linear programming. The obtained problem is then reformulated as a new multi-objective programming that has no equality or inequality constraints. The example of tryptophan biosynthesis is performed to the proposed framework and shown to the effectiveness of the approach. The simulation is also studied to give a performance comparison between the proposed and nonlinear approaches.**

Keywords-biological systems; multi-objective optimization; S-system

I. INTRODUCTION

Multi-objective optimization of biological systems has recently received interest by researchers [1-4]. In general an appropriate mathematical model describing the biotechnological process is first needed to do this task. Several such models including Michaelis-Menten, S-system, Generalized Mass Action and stoichiometric equations have been proposed in the literatures [5-6]. Among these formulations, one convenient tool to model a biological system is the S-system [5]. The advantage of this representation is that the steady-state equations are linear when the variables of the models are expressed in logarithmic coordinates. In this paper, we take advantage of this special property of S-system models and transform the original multi-objective nonlinear optimization problem of biological systems into a multi-objective linear programming. The obtained problem is then reformulated as a new multi-objective programming that has no equality or inequality constraints. The tryptophan biosynthesis in *Escherichia coli* is performed to verify the proposed framework of multi-objective optimization. The simulation is also studied to give a performance comparison between the proposed and nonlinear approaches.

II. OPTIMIZATION PROBLEM STATEMENT

Consider the following multi-objective problem of optimizing a biological system:

$$\max \quad J_1(X,Y) \tag{1}$$

$$\min \quad J_2(X,Y) \tag{2}$$

subject to satisfying:

$$F_i(X,Y)=0, \quad i=1,2,\ldots,n \tag{3}$$

$$0.8(X_i)_0 \le X_i \le 1.2(X_i)_0 \tag{4}$$

$$Y_k^L \le Y_k \le Y_k^U, \quad k=1,2,\ldots,m \tag{5}$$

where $X=(X_1,X_2,\ldots,X_n)^{\mathrm{T}} \in R_+^n$, $Y=(Y_1,Y_2,\ldots,Y_m)^{\mathrm{T}} \in R_+^m$; X_i represent the metabolite concentrations, and Y_k denote the enzyme activities; the objective function J_1 is usually a flux, and J_2 is the sum of metabolite concentrations; constraint (3) is the steady-state condition (i.e., $\mathrm{d}X_i/\mathrm{d}t=0$); $(X_i)_0$ is the basal steady-state of a biological system.

III. OPTIMIZATION METHOD

A. Multi-objective Linear Formalism

The S-system formalism is based on the Biochemical System Theory which proposes the use of power law functions to describe the nonlinear nature of biological processes [7]. Under this representation, the original model:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t}=F_i(X,Y) \qquad i=1,2,\ldots,n \tag{6}$$

can be expressed as:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t}=F_i(X,Y)=V_i^+-V_i^- =\alpha_i\prod_{j=1}^{n}X_j^{g_{ij}}\prod_{k=1}^{m}Y_k^{g'_{ik}}-\beta_i\prod_{j=1}^{n}X_j^{h_{ij}}\prod_{k=1}^{m}Y_k^{h'_{ik}}, \; i=1,2,\ldots,n \tag{7}$$

where the model parameters g_{ij}, g'_{ik}, h_{ij} and h'_{ik} are the kinetic orders, and α_i and β_i are the rate constants. Their definition can be found in [5]. The objective functions $J_1(X,Y)$ and $J_2(X,Y)$ can also be written as the following S-system forms:

$$J_1'(X,Y)=\gamma_1\prod_{i=1}^{n}X_i^{f_{1i}}\prod_{k=1}^{m}Y_k^{f'_{1k}} \tag{8}$$

$$J_2'(X,Y)=\gamma_2\prod_{i=1}^{n}X_i^{f_{2i}}\prod_{k=1}^{m}Y_k^{f'_{2k}} \tag{9}$$

where f_{1i}, f'_{1k}, f_{2i} and f'_{2k} terms stand for the kinetic orders, and γ_1 and γ_2 represent the corresponding rate constants.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

At steady-state the S-system (7) can be represented as the following linear equations:

$$\sum_{j=1}^{n}(g_{ij}-h_{ij})x_j+\sum_{k=1}^{m}(g_{ik}^{'}-h_{ik}^{'})y_k=\ln\left(\frac{\beta_i}{\alpha_i}\right),\ i=1,2,\ldots,n \quad (10)$$

where $x_j=\ln(X_j)$, $j=1,2,\ldots,n$, $y_k=\ln(Y_k)$, $k=1,2,\ldots,m$.

Due to the fact that the logarithmic transformation does not change the locations of maximum/minimum of a function, the multi-objective nonlinear optimization problem (1)-(5) can be transformed into the following linear formulations:

$$\begin{cases}\max \quad \bar{J}_1(x,y)\\ \min \quad \bar{J}_2(x,y)\\ \text{subject to satisfying :}\\ \sum_{j=1}^{n}(g_{ij}-h_{ij})x_j+\sum_{k=1}^{m}(g_{ik}^{'}-h_{ik}^{'})y_k=\ln\left(\frac{\beta_i}{\alpha_i}\right), i=1,2,\ldots,n\\ \ln(0.8(X_i)_0)\le x_i\le\ln(1.2(X_i)_0), i=1,2,\ldots,n\\ \ln(Y_k^L)\le y_k\le\ln(Y_k^U), k=1,2,\ldots,m\end{cases} \quad (11)$$

where the new objective functions $\bar{J}_1(x,y)$ and $\bar{J}_2(x,y)$ can be written respectively as:

$$\begin{aligned}\bar{J}_1(x,y)&=\ln(J_1(X,Y))\\&=\ln(J_1^{'}(X,Y))\\&=\ln(\gamma_1)+\sum_{i=1}^{n}f_{1i}\ln(X_i)+\sum_{k=1}^{m}f_{1k}^{'}\ln(Y_k)\\&=\ln(\gamma_1)+\sum_{i=1}^{n}f_{1i}x_i+\sum_{k=1}^{m}f_{1k}^{'}y_k\\&=\ln(\gamma_1)+f_1^{\mathrm{T}}x+f_1^{'\mathrm{T}}y\end{aligned}$$

$$\begin{aligned}\bar{J}_2(x,y)&=\ln(J_2(X,Y))\\&=\ln(J_2^{'}(X,Y))\\&=\ln(\gamma_2)+\sum_{i=1}^{n}f_{2i}\ln(X_i)+\sum_{k=1}^{m}f_{2k}^{'}\ln(Y_k)\\&=\ln(\gamma_2)+\sum_{i=1}^{n}f_{2i}x_i+\sum_{k=1}^{m}f_{2k}^{'}y_k\\&=\ln(\gamma_2)+f_2^{\mathrm{T}}x+f_2^{'\mathrm{T}}y\end{aligned}$$

where $f_1=(f_{11},f_{12},\cdots,f_{1n})^{\mathrm{T}}$,
$f_1^{'}=(f_{11}^{'},f_{12}^{'},\cdots,f_{1m}^{'})^{\mathrm{T}}$,
$f_2=(f_{21},f_{22},\cdots,f_{2n})^{\mathrm{T}}$,
$f_2^{'}=(f_{21}^{'},f_{22}^{'},\cdots,f_{2m}^{'})^{\mathrm{T}}$.

B. Solution of Multi-objective Linear Problem

Many strategies have been presented to handle a multi-objective optimization problem in recent years. In this work, we propose the following reformulations of multi-objective optimization problem (11):

$$\begin{cases}\min \quad \{-\bar{J}_1(x,y)\}\\ \min \quad \bar{J}_2(x,y)\\ \min \quad \bar{J}_3(x,y)\\ \text{subject to satisfying :}\\ \ln(0.8(X_i)_0)\le x_i\le\ln(1.2(X_i)_0), i=1,2,\ldots,n\\ \ln(Y_k^L)\le y_k\le\ln(Y_k^U), k=1,2,\ldots,m\end{cases} \quad (12)$$

where the objective function $\bar{J}_3(x,y)$ has the following expression:

$$\bar{J}_3(x,y)=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(g_{ij}-h_{ij})x_j+\sum_{k=1}^{m}(g_{ik}^{'}-h_{ik}^{'})y_k-\ln\left(\frac{\beta_i}{\alpha_i}\right)\right)^2, \quad (13)$$

Compared with optimization problem (11), multi-objective problem (12) has no equality constraints because of these constraints having been integrated into the third objective function $\bar{J}_3(x,y)$.

IV. Multi-objective Optimization of Tryptophan Biosynthesis in *Escherichia coli*

In this section, to illustrate the calculation algorithm, we will apply the proposed optimization method to tryptophan biosynthesis in *Escherichia coli*. This metabolic pathway is an appealing benchmark system that has already been optimized with other methods [8-9]. A complete description of the biological system can be found in [10]. The differential equations in dimensionless variables are given as:

$$\frac{\mathrm{d}X_1}{\mathrm{d}t}=\frac{X_3+1}{1+(1+Y_2)X_3}-(Y_8+Y_1)X_1=V_1^{+}-V_1^{-} \quad (14)$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t}=X_1-(Y_9+Y_1)X_2=V_2^{+}-V_2^{-} \quad (15)$$

$$\begin{aligned}\frac{\mathrm{d}X_3}{\mathrm{d}t}&=\frac{X_2Y_3^2}{Y_3^2+X_3^2}-(Y_{10}+Y_1)X_3-\frac{X_3Y_4}{1+X_3}-\frac{Y_5(1-Y_6Y_1)Y_1X_3}{X_3+Y_7}\\&=V_3^{+}-V_3^{-}\end{aligned} \quad (16)$$

Here, X_1 is used for mRNA concentration, X_2 is used for enzyme concentration and X_3 is used for tryptophan concentration.

Consider the following multi-objective steady-state optimization problem:

$$\max \quad J_1(X,Y)=\frac{Y_5(1-Y_6Y_1)Y_1X_3}{X_3+Y_7} \quad (17)$$

$$\min \quad J_2(X,Y)=X_1+X_2+X_3 \quad (18)$$

subject to satisfying:

$$\frac{X_3+1}{1+(1+Y_2)X_3}=(Y_8+Y_1)X_1 \quad (19)$$

$$X_1=(Y_9+Y_1)X_2 \quad (20)$$

$$\frac{X_2Y_3^2}{Y_3^2+X_3^2}=(Y_{10}+Y_1)X_3+\frac{X_3Y_4}{1+X_3}+\frac{Y_5(1-Y_6Y_1)Y_1X_3}{X_3+Y_7} \quad (21)$$

$$0.8(X_i)_0\le X_i\le 1.2(X_i)_0,\quad i=1,2,3 \quad (22)$$

$$0<Y_1\le 0.00624 \quad (23)$$

$$4\le Y_2\le 10 \quad (24)$$

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

$$500 \le Y_3 \le 5000 \tag{25}$$

$$Y_4 = 0.0022 Y_2 \tag{26}$$

$$0 < Y_5 \le 1000 \tag{27}$$

$$(Y_6, Y_7, Y_8, Y_9, Y_{10}) = (7.5, 0.005, 0.9, 0.02, 0) \tag{28}$$

Note that we replaced the variable Y_4 with the constraint $Y_4 = 0.0022 Y_2$ in our simulation experiments.

At the basal steady-state (see Table 1), the S-system representation of biological model (14)-(16) is written as:

$$\frac{dX_1}{dt} = 0.6403 X_3^{-5.87\times10^{-4}} Y_2^{-0.8332} - 1.0233 X_1 Y_1^{0.0035} Y_8^{0.9965} \tag{29}$$

$$\frac{dX_2}{dt} = X_1 - 1.4854 X_2 Y_1^{0.1349} Y_9^{0.8651} \tag{30}$$

$$\frac{dX_3}{dt} = 0.5534 X_2 X_3^{-0.5573} Y_3^{0.5573} - 1.6660 X_3^{0.7684} Y_1^{0.9904} Y_2^{0.0042} Y_5^{0.2274} Y_6^{-5.45\times10^{-3}} Y_7^{-0.8\times10^{-6}} \tag{31}$$

This S-system representation is modified slightly from [9].

Based on problem (12), we have the following formulations of problem (17)-(28):

$$\begin{cases} \min \quad Q_1(x,y) = -3.5\times10^{(-6)} x_3 - 0.976 y_1 - y_5 + 0.024 y_6 + 3.5\times10^{(-6)} y_7 \\ \min \quad Q_2(x) = 0.0001 x_1 + 0.0056 x_2 + 0.9943 x_3 \\ \min \quad Q_3(x,y) = (x_1 + 5.87\times10^{-4} x_3 + 0.0035 y_1 + 0.8332 y_2 + 0.9965 y_8 + 0.4689)^2 + (x_1 - x_2 - 0.1349 y_1 - 0.8651 y_9 - 0.3957)^2 + (x_2 - 1.3257 x_3 - 0.9904 y_1 - 0.0042 y_2 + 0.5573 y_3 - 0.2274 y_5 + 5.45\times10^{-3} y_6 + 0.8\times10^{-6} y_7 - 1.1021)^2 \\ \text{subject to satisfying :} \\ \ln(0.8(X_i)_0) \le x_i \le \ln(1.2(X_i)_0), i = 1,2,3 \\ \ln(\varepsilon_1) \le y_1 \le \ln(0.00624) \\ \ln(4) \le y_2 \le \ln(10) \\ \ln(500) \le y_3 \le \ln(5000) \\ \ln(\varepsilon_2) \le y_5 \le \ln(1000) \\ (y_6, y_7, y_8, y_9) = (\ln(7.5), \ln(0.005), \ln(0.9), \ln(0.02)) \end{cases} \tag{32}$$

where $Q_1(x,y) = -\bar{J}_1(x,y) + \ln(\gamma_1)$, $Q_2(x) = \bar{J}_2(x) - \ln(\gamma_2)$, $Q_3(x,y) = \bar{J}_3(x,y)$, and ε_1 and ε_2 are very small positive numbers.

In order to investigate the performance comparisons of linearization and nonlinear approaches in dealing with multi-objective optimization of biological systems, we also performed the following reformulations of multi-objective nonlinear problem (17)-(28):

$$\begin{cases} \max \quad J_1(X,Y) = \dfrac{Y_5(1 - Y_6 Y_1) Y_1 X_3}{X_3 + Y_7} \\ \min \quad J_2(X,Y) = X_1 + X_2 + X_3 \\ \min \quad J_3(X,Y) = \sum_{i=1}^{3} \left(V_i^+ - V_i^-\right)^2 \\ \text{subject to satisfying :} \\ 0.8(X_i)_0 \le X_i \le 1.2(X_i)_0, i = 1,2,3 \\ 0 < Y_1 \le 0.00624 \\ 4 \le Y_2 \le 10 \\ 500 \le Y_3 \le 5000 \\ 0 < Y_5 \le 1000 \\ (Y_6, Y_7, Y_8, Y_9, Y_{10}) = (7.5, 0.005, 0.9, 0.02, 0) \end{cases} \tag{33}$$

In our simulation experiments, both problems (32) and (33) were optimized using the MATLAB based solver GODLIKE [11]. GODLIKE stands for Global Optimum Determination by Linking and Interchanging Kindred Evaluators. It is a global optimizer that combines the power of GA (Genetic Algorithm), DE (Differential Evolution), PSO (Particle Swarm Optimization) and ASA (Adaptive Simulated Annealing) algorithms. The default settings of algorithm parameters were assumed in the run of GODLIKE solver.

TABLE I. BASAL STEADY-STATE

Variables	Basal steady-state	Variables	Basal steady-state
X_1	0.184654	X_2	7.986756
X_3	1418.931944	Y_1	0.00312
Y_2	5	Y_3	2283
Y_5	430	Y_6	7.5
Y_7	0.005	Y_8	0.9
Y_9	0.02	Y_{10}	0

TABLE II. MOST EFFICIENT POINTS

Variables	Basal steady-state	Linearization approach	Nonlinear approach
X_1	0.184654	0.191865	0.150365
X_2	7.986756	8.367718	7.256289
X_3	1418.931944	1146.204808	1135.441318
Y_1	0.00312	0.006107	0.003746
Y_2	5	5.288440	6.608242
Y_3	2283	4730.568656	4647.472547
Y_5	430	991.584907	996.145655

TABLE III. COMPARISON OF PERFORMANCES

Variables	Basal steady-state	Linearization approach	Nonlinear approach
J_1	1.310202	5.778472	3.626488
J_2	1427.103354	1154.764392	1142.8479728
$J_1/(J_1)_0$	-	4.410367	2.767885
$J_2/(J_2)_0$	-	0.809167	0.800817

Figs 1 and 2 show the final Pareto fronts of problems (32) and (33). The green dots in these figures are the corresponding most efficient points of problems (32) and (33),

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

that is, these points are closet to their origins and thus implies the "best efficient" compromise. Table 2 presents the optimization results of both problems in terms of their most efficient solutions. Table 3 shows a comparative study between the linearization and nonlinear approaches. As can be seen for the J_2 value, both methods yield an almost same sum of metabolite concentrations. However, for the J_1 value, the result obtained by the proposed linearization method is better than the nonlinear approach, with an improvement in about 59.34%. These conclusions clearly show the effectiveness of the presented linearization strategy in handling multi-objective nonlinear optimization of biological systems.

V. CONCLUSIONS

In this paper, we have presented a framework of multi-objective optimization of biological systems. The S-system representation of a biological process is easily transformed into a linear model in logarithmic coordinates. Therefore, the optimization problem can be transformed into a multi-objective linear programming. The proposed framework has been applied to tryptophan biosynthesis in *Escherichia coli*. Compared with the optimization results attained by using a nonlinear approach, a significant improvement in the rate of tryptophan production can be obtained through the use of the proposed linearization approach. This illustrates the effectiveness of the proposed linearization strategy in handling the multi-objective nonlinear optimization of biological systems.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 11101051), Liaoning Province Doctor Startup Fund of China (No. 20101001).

REFERENCES

[1] H. Link, J. Vera, D. Weuster-Botz, N. T. Darias and E. Franco-Lara, "Multi-objective steady state optimization of biochemical reaction networks using a constrained genetic algorithm", *Computers and Chemical Engineering*, vol. 32, pp. 1707-1713, 2008.

[2] G. Xu, "Bi-objective optimization of biochemical systems by linear programming", *Applied Mathematics and Computation*, vol. 218, pp. 7562-7572, 2012.

[3] O. H. Sendin, J. Vera, N. V. Torres and J. R. Banga, "Model based optimization of biochemical systems using multiple objectives: A comparison of several solution strategies", *Mathematical and Computer Modelling of Dynamical Systems*, vol. 12, pp. 469-487, 2006.

[4] W.-H. Wu, F.-S. Wang and M.-S. Chang, "Multi-objective optimization of enzyme manipulations in metabolic networks considering resilience effects", *BMC Systems Biology*, vol. 5: 145, 2011.

[5] E. O. Voit, *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge, U.K., 2000.

[6] N. V. Torres and E. O. Voit, *Pathway Analysis and Optimization in Metabolic Engineering*, Cambridge University Press, Cambridge, U.K., 2002.

[7] M. A. Savageau, *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology*, Addison-Wesley, Reading, MA, 1976.

[8] A. Marín-Sanguino and N. V. Torres, "Optimization of tryptophan production in bacteria design of a strategy for genetic manipulation of the tryptophan operon for tryptophan flux maximization", *Biotechnology Progress*, vol. 16, pp. 133-145, 2000.

[9] G. Xu, C. Shao and Z. Xiu, "A modified iterative IOM approach for optimization of biochemical systems", *Computers and Chemical Engineering*, vol. 32, pp. 1546-1568, 2008.

[10] Z. L. Xiu, A.-P. Zeng and W.-D. Deckwer, "Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria", *Journal of Biotechnology*, vol. 58, pp. 125-140, 1997.

[11] R. Oldenhuis, "GODLIKE- A robust single- & multi-objective optimizer", http://www.mathworks.com/matlabcentral/fileexchange, 2009.

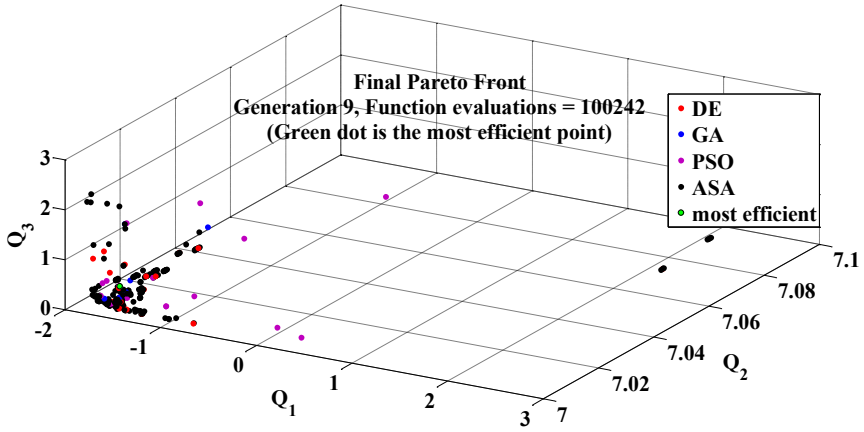

Figure 1. Final Pareto front of problem (32).

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

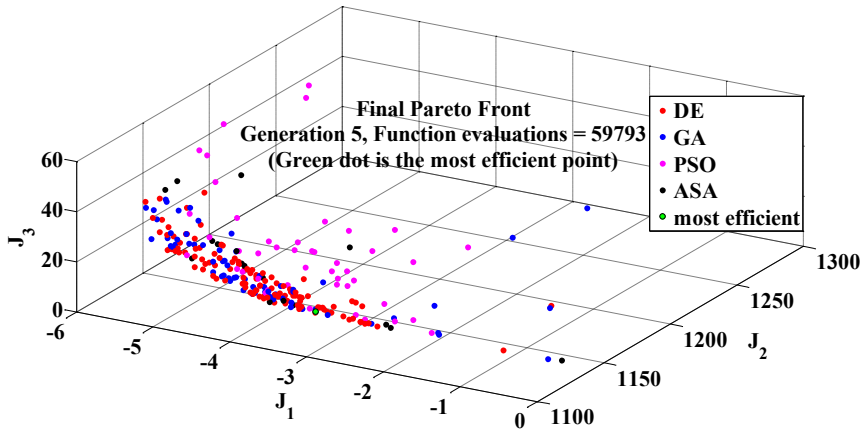

Figure 2. Final Pareto front of problem (33).

978-1-4673-4398-5/12/$31.00 ©2012 IEEE