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:

\[ \text{max} \quad J_i(X,Y) \quad (1) \]

\[ \text{min} \quad J_i(X,Y) \quad (2) \]

subject to satisfying:

\[ F_i(X,Y) = 0, \quad i = 1,2,\ldots,n \quad (3) \]

\[ 0.8(X_i)_0 \leq X_i \leq 1.2(X_i)_0 \quad (4) \]

\[ Y_i^c \leq Y_i \leq Y_i^v, \quad k = 1,2,\ldots,m \quad (5) \]

where \( X = (X_1, X_2, \ldots, X_n)^T \in \mathbb{R}^n \), \( Y = (Y_1, Y_2, \ldots, Y_m)^T \in \mathbb{R}^m \). \( X_i \) represent the metabolite concentrations, and \( Y_i \) denote the enzyme activities; the objective function \( J_i \) is usually a flux, and \( J_i \) is the sum of metabolite concentrations; constraint (3) is the steady-state condition (i.e., \( \frac{dX_i}{dt} = 0 \)); \( (X_i)_0 \) 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:

\[ \frac{dX_i}{dt} = F_i(X,Y) \quad i = 1,2,\ldots,n \quad (6) \]

can be expressed as:

\[ \frac{dX_i}{dt} = F_i(X,Y) = V_i^c - V_i^v \]

\[ = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} \prod_{k=1}^{m} Y_k^{b_{ik}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \prod_{k=1}^{m} Y_k^{b_{ik}}, i = 1,2,\ldots,n \quad (7) \]

where the model parameters \( g_{ij}, h_{ij}, b_{ik}, \) and \( h_{ij} \) are the kinetic orders, and \( \alpha_i \) and \( \beta_i \) are the rate constants. Their definition can be found in [5]. The objective functions \( J_i(X,Y) \) and \( J_j(X,Y) \) can also be written as the following S-system forms:

\[ J_i(X,Y) = \gamma^{c_{ij}} \prod_{j=1}^{n} X_j^{f_{ij}} \prod_{k=1}^{m} Y_k^{b_{ik}} \quad (8) \]

\[ J_j(X,Y) = \gamma^{v_{ij}} \prod_{j=1}^{n} X_j^{f_{ij}} \prod_{k=1}^{m} Y_k^{b_{ik}} \quad (9) \]

where \( f_{ij}, f_{ij}^{}, f_{ij}^{} \), and \( f_{ij} \) terms stand for the kinetic orders, and \( \gamma^{c_{ij}}, \gamma^{v_{ij}} \) represent the corresponding rate constants.
At steady-state the S-system (7) can be represented as the following linear equations:
\[
\sum_{j=1}^{m}(g_{ij} - h_{ij})x_{j} + \sum_{k=1}^{n}(g_{ik} - h_{ik})y_{i} = \ln\left(\frac{\beta}{\alpha}\right), \quad i = 1, 2, \ldots, n
\]  
where \( x_{j} = \ln(x_{j}), \quad j = 1, 2, \ldots, n, \quad y_{i} = \ln(y_{i}), \quad k = 1, 2, \ldots, m. \)

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:

Objective function: \( J_{i}(x, y) \) and \( J_{j}(x, y) \) can be written as respectively:
\[
J_{i}(x, y) = \ln(J_{i}(x, y)) = \ln(y_{i}) + \sum_{j=1}^{m}f_{ij}x_{j} + \sum_{k=1}^{n}f_{ik}y_{i} = \ln(y_{i}) + \sum_{j=1}^{m}f_{ij}x_{j} + \sum_{k=1}^{n}f_{ik}y_{i} = J_{i}(x, y) = \ln(J_{i}(x, y)) = \ln(y_{i}) + \sum_{j=1}^{m}f_{ij}x_{j} + \sum_{k=1}^{n}f_{ik}y_{i} = \ln(y_{i}) + \sum_{j=1}^{m}f_{ij}x_{j} + \sum_{k=1}^{n}f_{ik}y_{i} = \ln(y_{i}) + \sum_{j=1}^{m}f_{ij}x_{j} + \sum_{k=1}^{n}f_{ik}y_{i}
\]

subject to satisfying:
\[
\sum_{j=1}^{m}(g_{ij} - h_{ij})x_{j} + \sum_{k=1}^{n}(g_{ik} - h_{ik})y_{i} = \ln\left(\frac{\beta}{\alpha}\right), \quad i = 1, 2, \ldots, n
\]

\[
\ln(0.8(X_{i}^{0})) \leq x_{i} \leq \ln(1.2(X_{i}^{0})), \quad i = 1, 2, \ldots, n
\]

\[
\ln(Y_{i}^{0}) \leq y_{i} \leq \ln(Y_{i}^{0}), \quad k = 1, 2, \ldots, m
\]

where the new objective functions \( J_{i}(x, y) \) and \( J_{j}(x, y) \) can be formally written as:

\[
J_{i}(x, y) = \ln(J_{i}(x, y))
\]

\[
\min J_{i}(x, y)
\]

\[
\max J_{i}(x, y)
\]

subject to satisfying:

\[
\ln(0.8(X_{i}^{0})) \leq x_{i} \leq \ln(1.2(X_{i}^{0})), \quad i = 1, 2, \ldots, n
\]

\[
\ln(Y_{i}^{0}) \leq y_{i} \leq \ln(Y_{i}^{0}), \quad k = 1, 2, \ldots, m
\]

where the objective function \( J_{i}(x, y) \) has the following expression:

\[
J_{i}(x, y) = \sum_{j=1}^{m}(g_{ij} - h_{ij})x_{j} + \sum_{k=1}^{n}(g_{ik} - h_{ik})y_{i} - \ln\left(\frac{\beta}{\alpha}\right)
\]

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 \( J_{i}(x, y) \).

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:

\[
\frac{dx_{1}}{dt} = \frac{x_{1} + 1}{1 + (1 + y_{1})x_{1}} - (y_{1} + y_{1})x_{1} = V_{1} - V_{1}^{*}
\]

\[
\frac{dx_{2}}{dt} = x_{2} - (y_{1} + y_{1})x_{1} = V_{2} - V_{2}^{*}
\]

\[
\frac{dx_{3}}{dt} = \frac{y_{1}^{2} + x_{3}^{2}}{y_{1}^{2} + x_{3}^{2} - y_{1}(1 - y_{1})y_{1}x_{1}} - \frac{x_{3}y_{1}}{x_{1} + y_{1}} = V_{3} - V_{3}^{*}
\]

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:

\[
\max J_{i}(X, Y) = \frac{Y_{i}(1 - Y_{i})Y_{i}X_{i}}{X_{i} + Y_{i}}
\]

\[
\min J_{j}(X, Y) = X_{i} + X_{3} + X_{3}
\]

subject to satisfying:

\[
X_{i}^{0} + 1 = (Y_{i} + Y_{i})X_{i}
\]

\[
X_{i} = (Y_{i} + Y_{i})X_{i}
\]

\[
\frac{X_{i}Y_{i}^{2}}{Y_{i}^{2} + X_{3}^{2}} = (Y_{i} + Y_{i})X_{i} + \frac{X_{i}Y_{i}^{2} + Y_{i}(1 - Y_{i})Y_{i}X_{i}}{1 + X_{3} + Y_{i}}
\]

\[
0.8(X_{i})_{n} \leq X_{i} \leq 1.2(X_{i})_{n}, \quad i = 1, 2, 3
\]

\[
0 < X_{j}^{0} \leq 0.00624
\]

\[
4 \leq X_{j} \leq 10
\]

\[
\frac{Y_{i}^{2} + X_{3}^{2}}{X_{i}^{2} + Y_{i}^{2}} = (Y_{i} + Y_{i})X_{i} + \frac{X_{i}Y_{i}^{2} + Y_{i}(1 - Y_{i})Y_{i}X_{i}}{1 + X_{3} + Y_{i}}
\]
500 \leq Y_i \leq 5000 \quad \text{(25)}

Y_i = 0.0022W_i \quad \text{(26)}

0 < Y_i \leq 1000 \quad \text{(27)}

(Y_{i_1}, Y_{i_2}, Y_{i_3}, Y_{i_4}) = (7.5, 0.005, 0.9, 0.02, 0) \quad \text{(28)}

Note that we replaced the variable \( W_i \) with the constraint \( Y_i = 0.0022W_i \) in our simulation experiments.

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

\[
\frac{dY_i}{dt} = 0.6403X_1^{0.8356}Y_i^{0.3332} - 1.0233X_i^{0.0033}Y_i^{0.3865} \quad \text{(29)}
\]

\[
\frac{dX_i}{dt} = X_i - 1.4854X_i^{0.3146}Y_i^{0.8561} \quad \text{(30)}
\]

\[
\frac{dY_i}{dt} = 0.5534X_1^{0.5573}Y_i^{0.5573} \quad \text{(31)}
\]

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

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

\[
\min Q_i(x, y) = -3.5 \times 10^{-06}x_i - 0.976y_i - y_i + 0.024y_i + 3.5 \times 10^{-06}y_i \quad \text{(32)}
\]

\[
\min Q_i(x) = 0.0001x_i + 0.005x_i + 0.9943x_i \quad \text{(33)}
\]

\[
\min Q_i(x, y) = (x_i + 5.87 \times 10^{-04}x_i + 0.0035y_i + 0.8332y_i + 0.9965y_i + 0.4689)^2 + (x_i - x_i - 0.1349y_i - 0.8651y_i - 0.3957)^2 + (x_i - 1.3257x_i - 0.9904y_i - 0.0042y_i + 0.5573y_i - 0.2274y_i + 5.45 \times 10^{-04}y_i + 0.8 \times 10^{-04}y_i - 1.1021)^2
\]

subject to satisfying:

\[
\ln(0.8(x_i)) \leq x_i \leq \ln(1.2(x_i)), i = 1, 2, 3
\]

\[
\ln(9) \leq y_i \leq \ln(0.00624)
\]

\[
(4) \leq y_i \leq (10)
\]

\[
(500) \leq y \leq (5000)
\]

\[
(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (\ln(7.5), \ln(0.005), \ln(0.9), \ln(0.02))
\]

where \( Q_i(x, y) \) is \( \bar{J}_i(x, y) + \ln(y_i) \), \( Q_i(x) \) is \( \bar{J}_i(x) - \ln(y_i) \), \( Q_i(x, y) \) is \( \bar{J}_i(x, y) \), and \( \varepsilon_1 \) and \( \varepsilon_2 \) 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 re-formulations of multi-objective nonlinear problem (17)-(28):

\[
\max J_i(x, y) = \frac{Y_i(1 - Y_i)}{Y_i + X_i}
\]

\[
\min J_i(x, Y) = X_i + X_i + X_i
\]

\[
\min J_i(x, Y) = \sum (V_i^* - V_i^*)^2
\]

subject to satisfying:

\[
0.8(x_i) \leq x_i \leq 1.2(x_i), i = 1, 2, 3
\]

\[
0 < Y_i \leq 0.00624
\]

\[
4 \leq y_i \leq 10
\]

\[
500 \leq Y_i \leq 5000
\]

\[
0 < y_i \leq 1000
\]

\[
(Y_{i_1}, Y_{i_2}, Y_{i_3}, Y_{i_4}) = (7.5, 0.005, 0.9, 0.02, 0)
\]

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>
<thead>
<tr>
<th>TABLE I. BASED STEADY-STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>( x_1 )</td>
</tr>
<tr>
<td>( x_3 )</td>
</tr>
<tr>
<td>( y_2 )</td>
</tr>
<tr>
<td>( y_3 )</td>
</tr>
<tr>
<td>( y_4 )</td>
</tr>
<tr>
<td>( y_5 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II. MOST EFFICIENT POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>( x_1 )</td>
</tr>
<tr>
<td>( x_2 )</td>
</tr>
<tr>
<td>( x_3 )</td>
</tr>
<tr>
<td>( y_1 )</td>
</tr>
<tr>
<td>( y_2 )</td>
</tr>
<tr>
<td>( y_3 )</td>
</tr>
<tr>
<td>( y_4 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III. COMPARISON OF PERFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>( J_1 )</td>
</tr>
<tr>
<td>( J_2 )</td>
</tr>
<tr>
<td>( J_1 )</td>
</tr>
<tr>
<td>( J_2 )</td>
</tr>
</tbody>
</table>

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).
that is, these points are close 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
Figure 2. Final Pareto front of problem (33).