

Research Article

Complexity Analysis and Parameter Estimation of Dynamic Metabolic Systems

Li-Ping Tian,¹ Zhong-Ke Shi,² and Fang-Xiang Wu^{3,4}

¹ School of Information, Beijing Wuzi University, Beijing 101149, China

² School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

³ Department of Mechanical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada S7N 5A9

⁴ Division of Biomedical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada S7N 5A9

Correspondence should be addressed to Fang-Xiang Wu; faw341@mail.usask.ca

Received 24 April 2013; Revised 18 August 2013; Accepted 5 September 2013

Academic Editor: Shengyong Chen

Copyright © 2013 Li-Ping Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A metabolic system consists of a number of reactions transforming molecules of one kind into another to provide the energy that living cells need. Based on the biochemical reaction principles, dynamic metabolic systems can be modeled by a group of coupled differential equations which consists of parameters, states (concentration of molecules involved), and reaction rates. Reaction rates are typically either polynomials or rational functions in states and constant parameters. As a result, dynamic metabolic systems are a group of differential equations nonlinear and coupled in both parameters and states. Therefore, it is challenging to estimate parameters in complex dynamic metabolic systems. In this paper, we propose a method to analyze the complexity of dynamic metabolic systems for parameter estimation. As a result, the estimation of parameters in dynamic metabolic systems is reduced to the estimation of parameters in a group of decoupled rational functions plus polynomials (which we call improper rational functions) or in polynomials. Furthermore, by taking its special structure of improper rational functions, we develop an efficient algorithm to estimate parameters in improper rational functions. The proposed method is applied to the estimation of parameters in a dynamic metabolic system. The simulation results show the superior performance of the proposed method.

1. Introduction

Living cells require energy and material for maintaining their essential biological processes through metabolism, which is a highly organized process. Metabolic systems are defined by the enzymes dynamically converting molecules of one type into molecules of another type in a reversible or irreversible manner. Modeling and parameter estimation in dynamic metabolic systems provide new approaches towards the analysis of experimental data and properties of the systems, ultimately leading to a great understanding of the language of living cells and organisms. Moreover, these approaches can also provide systematic strategies for key issues in medicine, pharmaceutical, and biotechnological industries [1]. The formulation and identification of metabolic systems generally includes the building of the mathematical model of biological process and the estimating of system parameters. Because the components of a pathway interact not only with each other

in the same pathway but also with those in different pathways, most (if not all) of mathematical models of metabolic systems are highly complex and nonlinear. The widely used approaches for modeling inter- and intracellular dynamic processes are based on mass action law [1–4]. By mass action law, the reaction rates are generally polynomials in concentrations of metabolites with reaction constants or rational functions which are a fraction and whose denominator and numerators are polynomials in concentrations of metabolites with reaction constants [1–4]. As a result, the mathematical model is nonlinear not only in the states but also in the parameters. Estimation of these parameters is crucial to construct a whole metabolic system [5–7].

In general, all algorithms for nonlinear parameter estimation can be used to estimate parameters in metabolic systems, for example, Gauss-Newton iteration method, and its variants such as Box-Kanemasu interpolation method, Levenberg damped least squares methods and Marquardt's

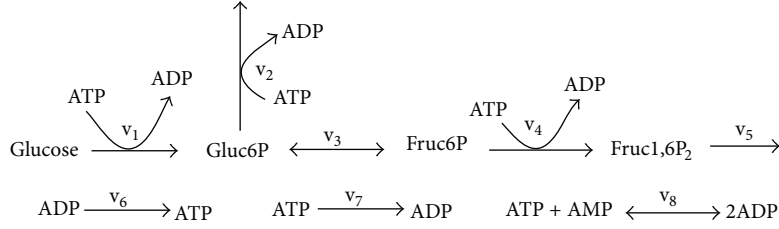


FIGURE 1: Schematic representation of the upper part of glycolysis [4].

method [8, 9]. However, these iteration methods are initial-sensitive. Another main shortcoming is that these methods may converge to the local minimum of the least squares cost function and thus cannot find the real values of parameters. Furthermore, because of their highly complexity and nonlinearity, Gauss-Newton iteration method and its variants cannot efficiently and accurately estimate the parameters in metabolic systems [5–7, 10, 11].

In this paper, we propose a systematic method for estimating parameters in dynamic metabolic systems. Typically mathematical model of dynamic metabolic systems consists of a group of nonlinear differential equations, some of which contains several rational functions in which parameters are nonlinear. In Section 2, we propose a method for model complexity analysis via the stoichiometric matrix. As a result, we obtain a group of equations, each of which contains only one-rational function plus polynomial functions which we called an improper rational function. Then, based on the observation that in the improper rational functions both the denominator and numerator are linear in parameters while polynomials are also linear in parameters, we develop an iterative linear least squares method for estimating parameters in dynamic metabolic systems in Section 3. The basic idea is to transfer optimizing a nonlinear least squares objective function into iteratively solving a sequence of linear least squares problems. In Section 4, we apply our developed method to estimate parameters in a metabolism system. Finally we give conclusions and some directions of future work along with this study in Section 5.

2. Model Complexity Analysis for Parameter Estimation

A dynamic metabolic system consists of k substances (molecules), and m reactions can be described by a system of differential equations as follows:

$$\frac{dx_i}{dt} = \sum_{j=1}^m c_{ij} r_j, \quad \text{for } i = 1, \dots, k, \quad (1)$$

where x_i represents the concentrations of molecule i , r_j represents the reaction rate j , and c_{ij} represents the stoichiometric coefficient of molecule i in reaction j . The mass action law in biochemical kinetics [2–4, 12] states that the reaction rate is proportional to the probability of a collision of the reactants. This probability is in turn proportional to the concentration of reactants. Therefore, reaction rate r_j is a function of the concentrations of molecules involved in reaction j and proportion constants.

The stoichiometric coefficient c_{ij} assigned to molecule i and reaction j can be put into a so-called stoichiometric matrix $\mathbf{C} = [c_{ij}]_{k \times m}$. Let $\mathbf{X} = [x_1, x_2, \dots, x_k]^T$ and $\mathbf{r} = [r_1, r_2, \dots, r_m]^T$, and let $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$ represent the vector consisting of all independent proportion constants, and then (1) can be rewritten in the following vector-matrix format:

$$\frac{d\mathbf{X}}{dt} = \mathbf{C}\mathbf{r}(\mathbf{X}, \boldsymbol{\beta}). \quad (2)$$

In principle, the stoichiometric coefficient c_{ij} in matrix \mathbf{C} is a constant integer and can be decided according to how molecule i is involved in reaction j . According to mass action law, the expression of reaction rates can be determined to be polynomials or rational functions with reaction constants [2–4, 12]. The challenge to build up the mathematic model of dynamic metabolic system (2) is to estimate the parameter vector $\boldsymbol{\beta}$, especially when some reaction rates are in the form of rational functions in which parameters are nonlinear.

If each differential equation in (2) contains one-rational function without or with polynomial functions, the parameters in model (2) can be estimated by algorithms in [13, 14] or a new algorithm proposed in the next section of this paper. Unfortunately, each differential equation contains a linear combination of several rational functions, which makes the parameter estimation in those coupled differential equations more difficult. The stoichiometric matrix contains very important information about the structure of the metabolic systems and is widely used to analyze the steady state and flux balance of metabolic systems [2–4]. In this paper, via the stoichiometric matrix, we propose a systematic method to transfer a system of differential equations (2) into another system of differential equations, in which each differential equation contains at most one-rational function.

Running Example. To illustrate the proposed method, we use the upper part of glycolysis system as a running example, showing how the method is applied to this system step after step. The schematic representation of this system is shown in Figure 1. The model for this metabolic system is described by the system of differential equations (2) as follows:

$$\begin{aligned} \frac{d}{dt} \text{Gluc6P} &= r_1 - r_2 - r_3, \\ \frac{d}{dt} \text{Fruc6P} &= r_3 - r_4, \\ \frac{d}{dt} \text{Fruc1,6P}_2 &= r_4 - r_5, \end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{ATP} &= -r_1 - r_2 - r_4 + r_6 - r_7 - r_8, \\
\frac{d}{dt} \text{ADP} &= r_1 + r_2 + r_4 - r_6 + r_7 + 2r_8, \\
\frac{d}{dt} \text{AMP} &= -r_8.
\end{aligned} \tag{3}$$

Based on the mass action law, the individual reaction rates can be expressed as

$$\begin{aligned}
r_1 &= \frac{V_{\max,2} \text{ATP}(t)}{K_{\text{ATP},1} + \text{ATP}(t)}, \\
r_2 &= k_2 \text{ATP}(t) \cdot \text{Gluc6P}(t), \\
r_3 &= \left(\frac{V_{\max,3}^f}{K_{\text{Gluc6P},3}} \text{Gluc6P}(t) \right. \\
&\quad \left. - \frac{V_{\max,3}^r}{K_{\text{Fruc6P},3}} \text{Fruc6P}(t) \right) \\
&\quad \times \left(1 + \left(\frac{\text{Gluc6P}(t)}{K_{\text{Gluc6P},3}} \right) \right. \\
&\quad \left. + \frac{\text{Fruc6P}(t)}{K_{\text{Fruc6P},3}} \right)^{-1}, \\
r_4 &= \frac{V_{\max,4} (\text{Fruc6P}(t))^2}{K_{\text{Fruc6P},4} \left(1 + \kappa (\text{ATP}(t) / \text{AMP}(t))^2 \right) + (\text{Fruc6P}(t))^2}, \\
r_5 &= k_5 \text{Fruc1,6P}_2(t), \\
r_6 &= k_6 \text{ADP}(t), \\
r_7 &= k_7 \text{ATP}(t), \\
r_8 &= k_{8f} \text{ATP}(t) \cdot \text{AMP}(t) - k_{8r} (\text{ADP}(t))^2.
\end{aligned} \tag{4}$$

Model (3) has six ordinary differential equations (ODEs) and 15 parameters contained in eight reaction rates, three out of which are rational functions. Some ODEs contain more than one rational reaction rates, which makes the parameter more difficult.

Comparing (3) to (2) we have the state vector: $\mathbf{X} = [\text{Gluc6P}; \text{Fruc6P}; \text{Fruc1,6P}_2; \text{ATP}, \text{ADP}, \text{AMP}]$ and stoichiometric matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & -1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \tag{5}$$

In the following, we describe our proposed method to analyze the complexity of model (2) through the running example.

Step 1. Collect the columns in the stoichiometric matrix corresponding to the rational reaction rates in model (2) to construct a submatrix \mathbf{C}_r and collect other columns (corresponding to polynomial reaction rates) to construct a submatrix \mathbf{C}_p . Therefore, we have

$$\frac{d\mathbf{X}}{dt} = \mathbf{C}\mathbf{r}(X, \boldsymbol{\beta}) = \mathbf{C}_r \mathbf{r}_r(X, \boldsymbol{\beta}) + \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}), \tag{6}$$

where \mathbf{r}_r is the subvector of \mathbf{r} and consists of all rational reaction rates while \mathbf{r}_p is another subvector of \mathbf{r} and consists of all polynomial reaction rates. In this step, we should make sure that the rank of matrix \mathbf{C}_r equals the number of rational reaction rates. If the rank of matrix \mathbf{C}_r does not equal the number of rational reaction rates, it means that some rational reaction rates are not independent. Then we combine dependent rational reaction rates together to create a new reaction rate such that all resulted rational reaction rates should be linearly independent [14]. As a result, the rank of matrix \mathbf{C}_r will equal the number of rational reaction rates.

For the running example, we have

$$\mathbf{C}_r = [c_1, c_3, c_4] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \\ -1 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{7}$$

$$\mathbf{C}_p = [c_2, c_5, c_6, c_7, c_8] = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

and $\mathbf{r}_r = [r_1, r_3, r_4]$ and $\mathbf{r}_p = [r_2, r_5, r_6, r_7, r_8]$. The rank of matrix \mathbf{C}_r equals 3, which is the number of rational reaction rates.

Step 2. Calculate the left inverse matrix of \mathbf{C}_r . That is, calculate \mathbf{C}_r^- such that

$$\mathbf{C}_r^- \mathbf{C}_r = \mathbf{I}. \tag{8}$$

As matrix \mathbf{C}_r has the column full rank, matrix \mathbf{C}_r^- satisfying (8) exists although it is typically not unique. For a given matrix \mathbf{C}_r , \mathbf{C}_r^- can be easily found by solving (8) which is a linear algebraic system. If it is not unique, any matrix satisfying (8) works for our proposed method.

For the running example, we can have

$$\mathbf{C}_r^- = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{9}$$

Step 3. Multiply (6) by matrix \mathbf{C}_r^- from the left to obtain

$$\begin{aligned}
\mathbf{C}_r^- \frac{d\mathbf{X}}{dt} &= \mathbf{C}_r^- \mathbf{C}_r \mathbf{r}_r(X, \boldsymbol{\beta}) + \mathbf{C}_r^- \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) \\
&= \mathbf{r}_r(X, \boldsymbol{\beta}) + \mathbf{C}_r^- \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta})
\end{aligned} \tag{10}$$

or

$$\mathbf{r}_r(X, \boldsymbol{\beta}) + \mathbf{C}_r^- \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) = \mathbf{C}_r^- \frac{dX}{dt}. \quad (11)$$

From its expression, each differential equation in the system (11) contains only one-rational reaction rates plus a linear combination of polynomial reaction rates.

For the running example, we have

$$\begin{aligned} r_1 - r_2 - r_5 &= \frac{d}{dt} (\text{Gluc6P} + \text{Fruc6P} + \text{Fruc1}, 6P_2), \\ r_3 - r_5 &= \frac{d}{dt} (\text{Fruc6P} + \text{Fruc1}, 6P_2), \\ r_4 - r_5 &= \frac{d}{dt} \text{Fruc1}, 6P_2. \end{aligned} \quad (12)$$

Step 4. Calculate matrix \mathbf{C}_r^\perp such that

$$\mathbf{C}_r^\perp \mathbf{C}_r = \mathbf{0}, \quad (13)$$

where \mathbf{C}_r^\perp has the full row rank and $\text{rank}(\mathbf{C}_r^\perp) + \text{rank}(\mathbf{C}_r^-) =$ the number of rows in \mathbf{C}_r . Note that \mathbf{C}_r^\perp can be easily found by solving (13), which is a homogenous linear algebraic system. Again if it is not unique, any matrix satisfying (13) works for our proposed method.

Then multiply (6) by matrix \mathbf{C}_r^\perp from the left to obtain

$$\mathbf{C}_r^\perp \frac{dX}{dt} = \mathbf{C}_r^\perp \mathbf{C}_r \mathbf{r}_r(X, \boldsymbol{\beta}) + \mathbf{C}_r^\perp \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) = \mathbf{C}_r^\perp \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) \quad (14)$$

or

$$\mathbf{C}_r^\perp \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) = \mathbf{C}_r^\perp \frac{dX}{dt}. \quad (15)$$

For the running example, we can have

$$\begin{aligned} \mathbf{C}_r^\perp &= \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{C}_r^\perp \mathbf{C}_p &= \begin{bmatrix} -2 & -2 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \end{aligned} \quad (16)$$

Step 5. Let $D = \mathbf{C}_r^\perp \mathbf{C}_p$. If $\text{rank}(D) \geq$ the number of columns, then solving (15) yields

$$\mathbf{r}_p(X, \boldsymbol{\beta}) = (D^T D)^{-1} D^T \mathbf{C}_r^\perp \frac{dX}{dt}. \quad (17)$$

If $\text{rank}(D) <$ the number of columns, it means that some polynomial reaction rates in (15) are linearly dependent. Then combine the linearly dependent rates and construct a new reaction rate vector $\bar{\mathbf{r}}_p(X, \boldsymbol{\beta})$ and full column rank matrix \bar{D} such that

$$\bar{D} \bar{\mathbf{r}}_p(X, \boldsymbol{\beta}) = D \mathbf{r}_p(X, \boldsymbol{\beta}) = \mathbf{C}_r^\perp \mathbf{C}_p \mathbf{r}_p(X, \boldsymbol{\beta}) = \mathbf{C}_r^\perp \frac{dX}{dt}, \quad (18)$$

and then solving (18) yields

$$\bar{\mathbf{r}}_p(X, \boldsymbol{\beta}) = (\bar{D}^T \bar{D})^{-1} \bar{D}^T \mathbf{C}_r^\perp \frac{dX}{dt}. \quad (19)$$

For the running example, we have $\text{rank}(D) <$ the number of columns. As the first four columns are linearly dependent, we can have a new reaction rates $-2r_2 - 2r_5 + r_6 - r_7$. Therefore, we have

$$\bar{D} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}, \quad \bar{D}^T \mathbf{C}_r^\perp = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ -1 & -1 & -2 & 0 & 1 & -1 \end{bmatrix}, \quad (20)$$

and furthermore, noting that $(d/dt)(\text{ATP} + \text{ADP} + \text{AMP}) = 0$, from (19) we have

$$\begin{aligned} r_6 - r_7 - 2r_2 - 2r_5 &= \frac{d}{dt} (\text{Gluc6P} + \text{Fruc6P} \\ &\quad + 2\text{Fruc1}, 6P_2 + \text{ATP} - \text{AMP}), \\ r_8 &= -\frac{d}{dt} \text{AMP}. \end{aligned} \quad (21)$$

After these five steps, dynamic metabolic system (2) is transferred into a system of differential equations, in which each differential equation contains one-rational function plus polynomial functions ((11) or (12)) or only polynomial function ((19) or (21)). Parameters in (19) can be analytically estimated by well-known least squares methods. In the next section, we describe an algorithm to estimate parameters in (11).

3. Parameter Estimation Algorithm

After its complexity analysis, estimating parameters in dynamic metabolic system is reduced to mainly estimating parameters in a rational function plus polynomial, which we call the improper rational function. These functions are nonlinear in both parameters and state variables. Therefore, estimation of parameters in these models is a nonlinear estimation problem. In general, all algorithms for nonlinear parameter estimation can be used to estimate parameters in the improper rational functions, for example, Gauss-Newton iteration method and its variants such as Box-Kanemasu interpolation method, Levenberg damped least squares methods, Marquardt's method [9–12, 15], and more sophisticated methods [16]. However, these iteration methods are initial sensitive. Another main shortcoming is that most of these methods may converge to the local minimum of the least squares cost function and thus cannot find the real values of parameters. In the following, we describe an iterative linear least squares method to estimate parameters in the improper rational functions. The basic idea is to transfer optimizing a nonlinear least squares objective function into iteratively solving a sequence of linear least squares problems.

Consider the general form of the following improper rational functions:

$$\eta(\mathbf{X}, \boldsymbol{\beta}) = \frac{N_0(\mathbf{X}) + \sum_{i=1}^{P_N} N_i(\mathbf{X}) \beta_{N_i}}{D_0(\mathbf{X}) + \sum_{j=1}^{P_D} D_j(\mathbf{X}) \beta_{D_j}} + \sum_{k=1}^{P_P} P_k(\mathbf{X}) \beta_{P_k}, \quad (22)$$

where the vector \mathbf{X} consists of the state variables and the p -dimensional vector $\boldsymbol{\beta}$ consists of all parameters in the improper rational function (22), which can naturally be divided into three groups: those in the numerator of the rational functions β_{N_i} ($i = 1, \dots, p_N$), those in the denominator of the rational function β_{D_j} ($j = 1, \dots, p_D$), and those in the polynomial β_{P_k} ($k = 1, \dots, p_P$), where we have that $p_D + p_N + p_P = p$. $N_i(\mathbf{X})$ ($i = 0, 1, \dots, p_N$), $D_j(\mathbf{X})$ ($j = 0, 1, \dots, p_D$), and $P_k(\mathbf{X})$ ($k = 1, \dots, p_P$) are the known functions nonlinear in the state variable \mathbf{X} and do not contain any unknown parameters. Either $N_0(\mathbf{X})$ or $D_0(\mathbf{X})$ must be nonzero, and otherwise from sensitivity analysis [9, 16] the parameters in model (22) cannot be uniquely identified.

If there is no polynomial part, model (22) is reduced to a rational function. Recently, several methods have been proposed for estimating parameters in rational functions [5, 6, 13, 14]. The authors in [5, 6] have employed general nonlinear parameter estimation methods to estimate parameters in rational functions. As shown in their results, the estimation error is fairly large. We have observed that in rational functions both the denominator and numerator are linear in the parameters. Based on this observation, we have developed iterative linear least squares methods in [13, 14] for estimating parameters in rational functions. Mathematically, improper rational function (22) can be rewritten as the following rational function:

$$\begin{aligned} \eta(\mathbf{X}, \boldsymbol{\beta}) = & \left(N_0(\mathbf{X}) + \sum_{i=1}^{p_N} N_i(\mathbf{X}) \beta_{N_i} + \left(\sum_{k=1}^{p_P} P_k(\mathbf{X}) \beta_{P_k} \right) \right. \\ & \times \left(D_0(\mathbf{X}) + \sum_{j=1}^{p_D} D_j(\mathbf{X}) \beta_{D_j} \right) \Big) \\ & \times \left(D_0(\mathbf{X}) + \sum_{j=1}^{p_D} D_j(\mathbf{X}) \beta_{D_j} \right)^{-1}. \end{aligned} \quad (23)$$

However, in the numerator of the model above, there are $p_D p_P + p_N + p_P$ coefficients while there are $p_D + p_N + p_P$ unknown parameters. When $p_P = 1$, the number of parameters is equal to the numbers of coefficients, and the methods developed in [13, 14] can be applied. However, when $p_P > 1$, those methods are not applicable as the number of parameters is less than the number of coefficients in the numerator.

In order to describe an algorithm to estimate parameters in the improper rational function (22) for n given groups of observation data y_t and \mathbf{X}_t ($t = 1, 2, \dots, n$), we introduce the following notation:

$$\begin{aligned} \boldsymbol{\beta}_N &= [\beta_{N_1}, \beta_{N_2}, \dots, \beta_{N_{p_N}}]^T \in R^{p_N}, \\ \boldsymbol{\beta}_D &= [\beta_{D_1}, \beta_{D_2}, \dots, \beta_{D_{p_D}}]^T \in R^{p_D}, \\ \boldsymbol{\beta}_P &= [\beta_{P_1}, \beta_{P_2}, \dots, \beta_{P_{p_P}}]^T \in R^{p_P}, \\ \boldsymbol{\beta} &= [\boldsymbol{\beta}_P^T \ \boldsymbol{\beta}_N^T \ \boldsymbol{\beta}_D^T]^T, \end{aligned}$$

$$\varphi_N(\mathbf{X}_t) = [N_1(\mathbf{X}_t), N_2(\mathbf{X}_t), \dots, N_{p_N}(\mathbf{X}_t)] \in R^{p_N},$$

$$\varphi_D(\mathbf{X}_t) = [D_1(\mathbf{X}_t), D_2(\mathbf{X}_t), \dots, D_{p_D}(\mathbf{X}_t)] \in R^{p_D},$$

$$\varphi_P(\mathbf{X}_t) = [P_1(\mathbf{X}_t), P_2(\mathbf{X}_t), \dots, P_{p_P}(\mathbf{X}_t)] \in R^{p_P},$$

$$\mathbf{Y} = [y(1), y(2), \dots, y(n)]^T \in R^n,$$

$$\Phi_{N_0} = [N_0(\mathbf{X}_1), N_0(\mathbf{X}_2), \dots, N_0(\mathbf{X}_n)]^T \in R^n,$$

$$\Phi_{D_0} = [D_0(\mathbf{X}_1), D_0(\mathbf{X}_2), \dots, D_0(\mathbf{X}_n)]^T \in R^n,$$

$$\Phi_N = \begin{bmatrix} \varphi_N(\mathbf{X}_1) \\ \varphi_N(\mathbf{X}_2) \\ \vdots \\ \varphi_N(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_N},$$

$$\Phi_D = \begin{bmatrix} \varphi_D(\mathbf{X}_1) \\ \varphi_D(\mathbf{X}_2) \\ \vdots \\ \varphi_D(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_D},$$

$$\Phi_P = \begin{bmatrix} \varphi_P(\mathbf{X}_1) \\ \varphi_P(\mathbf{X}_2) \\ \vdots \\ \varphi_P(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_P},$$

$$\Psi(\boldsymbol{\beta}_D) = \text{diag} \begin{bmatrix} D_0(\mathbf{X}_1) + \varphi_D(\mathbf{X}_1) \boldsymbol{\beta}_D \\ D_0(\mathbf{X}_2) + \varphi_D(\mathbf{X}_2) \boldsymbol{\beta}_D \\ \vdots \\ D_0(\mathbf{X}_n) + \varphi_D(\mathbf{X}_n) \boldsymbol{\beta}_D \end{bmatrix} \in R^{n \times n}. \quad (24)$$

To estimate parameters in the improper rational function (22), as in [11], we form a sum of the weighted squared errors (the cost function) with the notions above as follows:

$$\begin{aligned} J(\boldsymbol{\beta}) &= J(\boldsymbol{\beta}_P, \boldsymbol{\beta}_N, \boldsymbol{\beta}_D) \\ &= \sum (D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D)^2 \\ &\quad \times \left(\frac{N_0(\mathbf{X}_t) + \varphi_N(\mathbf{X}_t) \boldsymbol{\beta}_N}{D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D} + \Phi_P \boldsymbol{\beta}_P - y_t \right)^2. \end{aligned} \quad (25)$$

Minimizing $J(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta} = [\boldsymbol{\beta}_P^T, \boldsymbol{\beta}_N^T, \boldsymbol{\beta}_D^T]^T$ can give the nonlinear least squares estimation of parameters $\boldsymbol{\beta}_P$, $\boldsymbol{\beta}_N$, and $\boldsymbol{\beta}_D$. We rewrite the objective function (22) as follows:

$$\begin{aligned} J(\boldsymbol{\beta}) &= \sum [(D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D) \Phi_P \boldsymbol{\beta}_P + \varphi_N(\mathbf{X}_t) \boldsymbol{\beta}_N \\ &\quad - \varphi_D(\mathbf{X}_t) y_t \boldsymbol{\beta}_D - D_0(\mathbf{X}_t) y_t + N_0(\mathbf{X}_t)]^2. \end{aligned} \quad (26)$$

TABLE 1: The true value (from [4]), estimated value, and relative estimation errors.

Parameter name	True value	Estimated value	REE (%)
$V_{\max,2}$ (mM·min ⁻¹)	50.2747	50.2447	0.0001
$K_{\text{ATP}1}$ (mM)	0.10	0.10000	0.0399
k_2 (mM ⁻¹ ·min ⁻¹)	2.26	2.2599	0.0049
$V_{\max,3}^f$ (mM·min ⁻¹)	140.282	139.4917	0.5633
$V_{\max,3}^r$ (mM·min ⁻¹)	140.282	141.3623	0.7701
$K_{\text{Gluc6P}3}$ (mM)	0.80	0.7999	1.3884
$K_{\text{Fruc6P}3}$ (mM)	0.15	0.1499	0.0930
$V_{\max,4}$ (mM·min ⁻¹)	44.7287	44.6664	0.1372
$K_{\text{Fruc6P}4}$ (mM ²)	0.021	0.0206	1.8457
k	0.15	0.1526	1.7447
k_5 (min ⁻¹)	6.04662	6.0466	0.0007
k_6 (min ⁻¹)	68.48	68.4837	0.0054
k_7 (min ⁻¹)	3.21	3.20797	0.0078
k_{8f} (min ⁻¹)	432.9	432.8408	0.0137
k_{8r} (min ⁻¹)	133.33	133.314	0.0120

In the objective function (26), for a given parameters $\bar{\beta}_D$ in the first term, we have

$$\begin{aligned} J(\beta) &= J(\beta_P, \beta_N, \beta_D, \bar{\beta}_D) \\ &= [\mathbf{A}(\bar{\beta}_D)\beta - \mathbf{b}]^T [\mathbf{A}(\bar{\beta}_D)\beta - \mathbf{b}], \end{aligned} \quad (27)$$

where

$$\mathbf{A}(\bar{\beta}_D) = \begin{bmatrix} \Psi(\bar{\beta}_D)\Phi_P^T \\ \Phi_N^T \\ -\text{diag}(Y)\Phi_D^T \end{bmatrix} \in R^{n \times p}, \quad (28)$$

$$\mathbf{b} = (\Phi_{D_0} \text{diag}(Y) - \Phi_{N_0}) \in R^n. \quad (29)$$

Then for given parameters $\bar{\beta}_D$, we can estimate the parameters $\beta = [\beta_P^T, \beta_N^T, \beta_D^T]^T$ by linear least squares method as follows:

$$\beta = [\mathbf{A}^T(\bar{\beta}_D)\mathbf{A}(\bar{\beta}_D)]^{-1}\mathbf{A}^T(\bar{\beta}_D)\mathbf{b}. \quad (30)$$

Based on the above discussion, we propose the following iterative linear least squares method.

Step 1. Choose the initial guess for β_D^0 .

Step 2. Iteratively construct matrix $\mathbf{A}(\beta_D^s)$ and vector \mathbf{b} by (28) and (29), respectively, and then solve the linear least squares problem:

$$J(\beta^{s+1}) = [\mathbf{A}(\beta_D^s)\beta^{s+1} - \mathbf{b}]^T [\mathbf{A}(\beta_D^s)\beta^{s+1} - \mathbf{b}], \quad (31)$$

which gives the solution

$$\beta^{s+1} = [\mathbf{A}^T(\beta_D^s)\mathbf{A}(\beta_D^s)]^{-1}\mathbf{A}^T(\beta_D^s)\mathbf{b} \quad (32)$$

until the stopping criterion is met, where $\beta^s = [\beta_P^{sT}, \beta_N^{sT}, \beta_D^{sT}]^T$ is the estimation of parameters β at step s .

From (31), if the estimation sequence β^1, β^2, \dots is converged to β^* , the objective function (26) reaches its minimum value at β^* . That is, β^* is the estimation of parameters in model (22).

There are several ways to set up a stopping criterion. In this paper the stopping criteria are chosen as

$$\frac{\|\beta^k - \beta^{k-1}\|}{\|\beta^{k-1}\| + 1} \leq \varepsilon, \quad (33)$$

where $\|\cdot\|$ is the Euclidean norm of the vector and ε is a preset small positive number, for example, 10^{-5} .

4. Application

To investigate the method developed in previous sections, this study generates artificial data from the dynamic metabolic system in the running example with the biochemically plausible parameter values [4] listed in column 2 of Table 1 and initial values: Gluc6P(0) = 1 mM, Fruc6P(0) = 0 mM, Fruc1,6P₂(0) = 0 mM, ATP(0) = 2.1 mM, ADP(0) = 1.4 mM, and AMP(0) = 0.1 mM. The trajectory of this system is depicted in Figure 2. From Figure 2, the concentrations of all molecules except for Fructose-1,6-biphosphate reach their steady states after about 0.1 minutes while Fructose-1,6-biphosphate after 0.5 minutes. Therefore, we do not use the data simulated after 0.5 minutes.

Although no noise is added to the artificial data in the simulation, noises are introduced in numerically calculating the derivatives by finite difference formulas. In general, the higher the sampling frequency and more data points are used, the more accurate the numerical derivatives are. On the other hand, we may not obtain data with the high frequency because of experimental limitations in practice. In this study, the sampling frequency is 100 data points per minute. In numerically calculating the concentration change rate at each

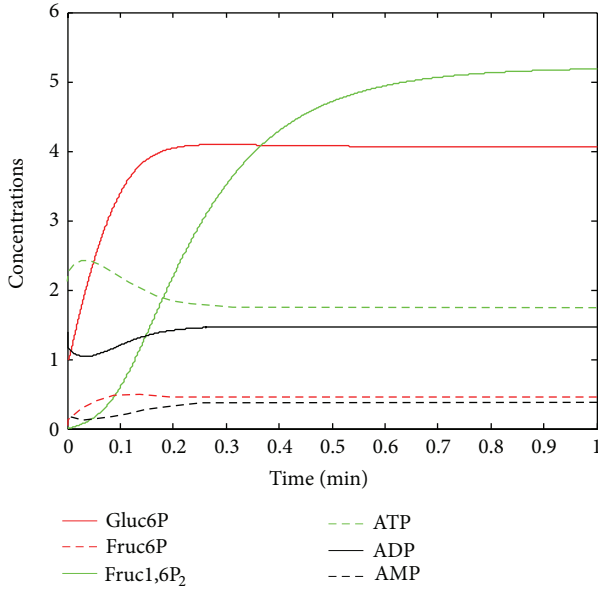


FIGURE 2: Trajectory of system (3).

time point from concentration x , we adopt the five-point central finite difference formula as follows:

$$\dot{x}(t_n) = \frac{1}{12\Delta t} [x(t_{n-2}) - 8x(t_{n-1}) + 8x(t_{n+1}) - x(t_{n+2})]. \quad (34)$$

The estimation accuracy of the proposed method is investigated in terms of relative estimation error which is defined as

$$\text{REE} = \frac{\|\text{estimate_value} - \text{true_value}\|}{\|\text{true_value}\|}. \quad (35)$$

As all parameters to be estimated are nonnegative, initial values are chosen as 0 or 1 in this study. The experimental results are listed in columns 3 and 4 in Table 1. From column 3 in Table 1, the estimated parameter values are very close to the corresponding true values. Actually the relative estimation errors calculated from (29) for all estimated parameters except for two are less than 1%. This indicates that the proposed method can accurately estimate the parameters in this system.

5. Conclusions and Future Work

In this study, we have first described a method to analyze the complexity of metabolic systems for parameter estimation, based on the stoichiometric matrix of the metabolic systems. As a result, the estimation of parameters in the metabolic systems has been reduced to the estimation of parameters in the improper rational functions or polynomial functions. Then we have developed an iterative linear least squares method for estimating parameters in the improper rational models. The results from its application to a metabolism system have shown that the proposed method can accurately estimate the parameters in metabolic systems.

We do not consider the noises in the data except those introduced by numerical derivatives in this study. One direction of future work is to investigate the influence of noises in the data to the estimation accuracy. In addition, low sampling frequency is expected, particularly for molecular biological systems as in practice measurements from them may be very expensive or it is impossible to sample measurements with high frequencies. Another direction of future work is to improve the estimation accuracy of the proposed method with low sampling frequencies.

Acknowledgments

This work was supported by the Special Fund of Ministry of Education of Beijing for Distinguishing Professors and Science and Technology Funds of Beijing Ministry of Education (SQKM201210037001) to Li-Ping Tian, by National Natural Science Foundation of China (NSFC 61134004) to Zhong-Ke Shi, and by Natural Sciences and Engineering Research Council of Canada (NSERC) to Fang-Xiang Wu.

References

- [1] M. Fussenegger, J. E. Bailey, and J. Varner, "A mathematical model of caspase function in apoptosis," *Nature Biotechnology*, vol. 18, no. 7, pp. 768–774, 2000.
- [2] J. Nielsen, J. Villadsen, and G. Liden, *Bioreaction Engineering Principles*, Kluwer Academic Publishers, New York, NY, USA, 2nd edition, 2003.
- [3] G. N. Stephanopoulos, A. A. Aritidou, and J. Nielsen, *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego, Calif, USA, 1998.
- [4] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems Biology in Practice: Concepts, Implementation and Application*, Wiley-VCH and KGaA, Weinheim, Germany, 2005.
- [5] K. G. Gadkar, J. Varner, and F. J. Doyle III, "Model identification of signal transduction networks from data using a state regulator problem," *Systems Biology*, vol. 2, no. 1, pp. 17–29, 2005.
- [6] K. G. Gadkar, R. Gunawan, and F. J. Doyle III, "Iterative approach to model identification of biological networks," *BMC Bioinformatics*, vol. 6, article 155, 2005.
- [7] I.-C. Chou and E. O. Voit, "Recent developments in parameter estimation and structure identification of biochemical and genomic systems," *Mathematical Biosciences*, vol. 219, no. 2, pp. 57–83, 2009.
- [8] J. V. Beck and K. J. Arnold, *Parameter Estimation in Engineering and Science*, John Wiley & Sons, New York, NY, USA, 1977.
- [9] A. van den Bos, *Parameter Estimation for Scientists and Engineers*, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- [10] P. Mendes and D. B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.
- [11] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods," *Genome Research*, vol. 13, no. 11, pp. 2467–2474, 2003.
- [12] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig, *Systems Biology: A Textbook*, Wiley-VCH and KGaA, Weinheim, Germany, 2009.

- [13] F. X. Wu, L. Mu, and Z. K. Shi, "Estimation of parameters in rational reaction rates of molecular biological systems via weighted least squares," *International Journal of Systems Science*, vol. 41, no. 1, pp. 73–80, 2010.
- [14] F. X. Wu, Z. K. Shi, and L. Mu, "Estimating parameters in the caspase activated apoptosis system," *Journal of Biomedical Engineering and Technology*, vol. 4, no. 4, pp. 338–354.
- [15] L. Marucci, S. Santini, M. di Bernardo, and D. di Bernardo, "Derivation, identification and validation of a computational model of a novel synthetic regulatory network in yeast," *Journal of Mathematical Biology*, vol. 62, no. 5, pp. 685–706, 2011.
- [16] L. Cheng, Z. G. Hou, Y. Lin, M. Tan, W. C. Zhang, and F.-X. Wu, "Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 714–726, 2011.