Parameter Estimation Methods for Biological Systems

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Mechanical Engineering
University of Saskatchewan
Saskatoon

By
Lei MU

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mechanical Engineering
University of Saskatchewan
57 Campus Drive
Saskatoon, Saskatchewan S7N 5A9
Canada

Abstract

The inverse problem of modeling biochemical processes mathematically from measured time course data falls into the category of system identification and parameter estimation. Analyzing the time course data would provide valuable insights into the model structure and dynamics of the biochemical system. Based on the types of biochemical reactions, such as metabolic networks and genetic networks, several modeling frameworks have been proposed, developed and proved effective, including the Michaelis-Menten equation, the Biochemical System Theory (BST), etc. One bottleneck in analyzing the obtained data is the estimation of parameter values within the system model.

As most models for molecular biological systems are nonlinear with respect to both parameters and system state variables, estimation of parameters in these models from experimental measurement data is thus a nonlinear estimation problem. In principle, all algorithms for nonlinear optimization can be used to deal with this problem, for example, the Gauss-Newton iteration method and its variants. However, these methods do not take the special structures of biological system models into account. When the number of parameters to be determined increases, it will be challenging and computationally expensive to apply these conventional methods.

In this research, several methods are proposed for estimating parameters in two classes of widely used biological system models: the S-system model and the linear fractional model (LFM), by utilizing their structure specialties. For the S-system, two estimation methods are designed. 1) Based on the two-term structure (production and degradation) of the model, an alternating iterative least squares method is proposed. 2) A separation nonlinear least squares method is proposed to deal with the partially linear structure of the model. For the LFM, two estimation methods are provided. 1) The separation nonlinear least squares method can also be adopted to treat the partially linear structure of the LFM, and moreover a modified iterative version is included. 2) A special strategy using the separation principle and the weighted least squares method is implemented to turn the cost function into a

quadratic form and thus the estimates for parameters can be analytically solved. Simulation results have demonstrated the effectiveness of the proposed methods, which have shown better performance in terms of estimation accuracy and computation time, compared with those conventional nonlinear estimation methods.

Keywords: parameter estimation, nonlinear biological system, S-system, linear fractional model (LFM), time course data, least squares, optimization, separation method

Acknowledgements

I am deeply indebted to my supervisor Prof. FangXiang Wu, for his help and advice both in study and life throughout my staying in university. From him I learned parameter estimation techniques and chose this area to explore during my graduate studies. I would like to thank Prof. Wu for his encouragement and patience in my thesis writing stage, also his comments and recommendations for my draft. It is hard to imagine this work could be completed without his guidance and modification.

I would like to give my sincere thanks to my committee members: Prof. Allan Dolovich and Prof. Yang Shi, for their evaluations and suggestions to improve my research work during the proposal session and after. I highly appreciate that my external examiner, Prof. Li Chen, is willing to review and examine my thesis work during the defense session. A special thanks goes to Prof. James Bugg for kindly offering much help in this last stage. My gratitude is also expressed to professors who acted as my course instructors. They are Eric Salt, Nurul Chowdhury, Yang Shi, Longhai Li and Fangxiang Wu.

My gratefulness also goes to my group members. They are JiaRui Ding, JinHong Shi, AnMin Zou, RuiZhi Luo, and Jian Sun. Being friends, studying together and sharing joys with them will be a great memory for me.

I would say thank you to faculty and staff in the Department of Mechanical Engineering for the friendly environment and help you offered. I'm grateful to my supervisor and department for providing me financial support to complete my study and research.

Finally, my appreciation and missing fly to my parents. Their love and care support me a lot and their encouragement helps throughout the long way.

Contents

List of Tables

List of Figures

List of Abbreviations

5CD	5-point Central Difference
ALS	Alternating Least Squares
ANN	Artifical Neural Network
AS	AutoSmooth
BST	Biochemical System Theory
CPU	Central Processing Unit
DNA	Deoxyribonucleic Acid
EA	Evolutionary Algorithm
EC	Evolutionary Computation
GA	Genetic Algorithm
GMA	Generalized Mass Action
GNA	Gauss-Newton Algorithm
GNA_C	Gauss-Newton Algorithm with the Conventional routine
GNA_S	Gauss-Newton Algorithm with the Separation method
LFM	Linear Fractional Model
LHS	Left-Hand Side
LMA	Levenberg-Marquart Algorithm
LMA_C	Levenberg-Marquart Algorithm with the Conventional routine
LMA_S	Levenberg-Marquart Algorithm with the Separation method
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
ODE	Ordinary Differential Equation
REE	Relative Estimation Error
RHS	Right-Hand Side
RNA	Ribonucleic Acid
RNAP	RNA Polymerase
SSE	Sum of Squared Errors

Chapter 1

Introduction

1.1 Background

Over the past decade, advances in high-throughput experiment tools and methods have helped generate huge amount of biological data and information at the microscopic levels, such as molecular and cellular levels. One example of such achievements is that the sequencing processes of genomes have been accomplished for several species or organisms [2]. Nowadays, researchers face a new challenge regarding how these available biological data and information can be integrated and utilized in order to quantitatively understand the dynamic behaviors of biochemical reactions at the system level. Such ideas give rise to a new emerging research area, named 'systems biology' [3], and it exerts power in modeling and quantitatively analyzing the details and principles of biological systems.

Conventionally the biological methods and efforts have focused primarily upon separated fundamental parts, such as genes, proteins and cells, to dig deep into those individual functions and mechanisms. However, as these components usually work together and have complex interactions, limited understanding of the system's operation has been provided by the traditional research approach. Systems biology tries to seek the patterns and reasons of these interactions, in order to better understand the entire processes that happen in a biological system [4]. Many ideas and techniques from mathematical and engineering disciplines, such as the dynamical systems theory and the system engineering approach, are incorporated in this research field. The ultimate goal of systems biology research is to design and manipulate cell functions after analyzing and understanding the characteristics and mechanisms of complex

biological systems.

One objective of systems biology is to construct a model to represent the dynamics and interactions within a system. It has to be decided what kind of model is appropriate for the research target and the type of experimental data. Several approaches have been proposed in order to quantitatively model the dynamics of biological system networks, including continuous deterministic methods such as differential equation models, or discrete probabilistic models based on stochastic dynamics [20]. In most situations, researchers are mainly interested in average responses of the model rather than extreme or rare cases, and a deterministic model would be sufficient for the modeling process, if the stochastic aspect can be ignored.

1.2 Problem Statement

The first three phases for the construction of a biological system model include: 1) collecting observation data from experiments; 2) identifying the structure of the reaction model; and 3) estimating values of parameters within the model. At present, modern high-throughput experimental techniques, e.g., microarray, mass spectrometry (MS) and nuclear magnetic resonance (NMR), are used to collect diverse quantitative measurements [3, 44, 23]. For example, time series data of the metabolic concentrations involved in a certain biochemical reaction pathway are especially useful for the construction of a metabolic model. *In vivo* NMR measurements are able to produce this kind of data, which contain information about both the material flows and the regulations within the reaction network [26]. The final constructed model, whose predictions are consistent with the experimental data, will be considered capable of quantitatively describing the underlying biological systems.

The subsequent phase of modeling is to determine the network structure of biochemical reactions and select a modeling framework. From available measurement data and related biological knowledge, the modeler will consider which components, interactions and regulations are relevant and should be incorporated in the model. Then a network map is formulated to visualize the process of reactions (see Fig.

2.1 as an illustration). Now by selecting a proper modeling framework, the network map can be translated into corresponding equations (see Eq. (2.8) as an example), which are usually in the form of ordinary differential equations (ODE). Such a set of ODEs represent the velocities or fluxes of reactions within the system model. As the modeling equations have been set up in symbolic forms, now we face the problem of determining suitable values for parameters within these equations. The way is to make the model-based predictions comply with those experimental measurements. Parameter estimation, which is the third phase, is chosen as the subject for this research.

Parameter estimation is a key issue in the construction of a biological model. Once the reaction network structure is known, the corresponding equations are relatively easy to list using the selected modeling framework, such as the Michaelis-Menten rate law and the Biochemical System Theory (BST) [32, 43]. Within these equations, there are a group of unknown parameters, which determine the system's characteristics. In most cases it is very difficult or even impossible to measure the parameters experimentally. However, we usually have the chance to measure some of the variables involved in the model, such as the concentrations of reaction components [23]. The development of experimental tools of biology helps us in accumulating such desired biological information. Some modern high-throughput tools are able to collect time series data of reactants, under different experimental conditions. Our task would be to seek the optimal parameter values from these experimental measurement data.

Models in the form of nonlinear equations are more ubiquitous and the estimation problem becomes more complex with nonlinear factors involved. Thus, such nonlinear models are chosen to be considered in this research. Here we select the S-system model [43], one of the 'testbed' models in this research, to exemplify one form of modeling framework, which shows nonlinearity with respect to parameters:

$$\dot{x}_i = \alpha_i \prod_{k=1}^{d} x_k^{g_{ik}} - \beta_i \prod_{k=1}^{d} x_k^{h_{ik}}, \quad (i = 1, \cdots, d). \tag{1.1}$$

Here, x_i are called state variables, which represent concentrations of reaction com-

ponents. In the right-hand side (RHS), α_i, β_i, g_{ik} and h_{ik} are system parameters. Of all the parameters, α_i and β_i are known as rate constants, whereas g_{ij} and h_{ij} are referred to as kinetic orders.

The parameter estimation problem, i.e., the topic of this research, can be dealt with through finding the optimal values of parameters that make the model-based predictions mostly consistent with the experimental data. The agreement/consistency between the model-based prediction and the experimental data can be measured in different ways [36]. The following two functions, J_A and J_B, are objective functions usually adopted in the estimation process (taking the S-system as an example):

$$\begin{aligned} J_A &= \sum_{i=1}^{d}\sum_{j=1}^{n} w_{ij}^2[\dot{x}_{ij} - (\alpha_i \prod_{k=1}^{d} x_{kj}^{g_{ik}} - \beta_i \prod_{k=1}^{d} x_{kj}^{h_{ik}})]^2, \\ J_B &= \sum_{i=1}^{d}\sum_{j=1}^{n} w_{ij}^2(x_{ij} - x_{ij}^{\star})^2. \end{aligned} \tag{1.2}$$

The subscript i distinguishes different state variables, and the subscript j indexes time points of measurements. Therefore x_{ij} denotes the j-th measurement of the variable i. w_{ij} is the weighting factor and $x_{ij}^{\star}$ is the value of state variable calculated by integrating ODEs numerically.

As it costs considerable computational resources to perform numerical integrations (as high as 95% of the total computation time during the optimization process [44]), adoption of J_B is not appropriate for fast parameter estimation. In this research, J_A is chosen as the objective function.

1.3 Research Objectives and Basic Ideas

A general comparison of parameter estimation algorithms for biological systems has been reported by Moles *et al.* in [25]. All the examined algorithms can be categorized into two classes. One class is deterministic methods, usually gradient-based, such as the Gauss-Newton algorithm (GNA) and the Levenberg-Marquardt algorithm (LMA). The advantage of these gradient-based methods is their ease of implementation and fast computation to deal with small-scale or middle-scale problems. The

weakness lies in that the gradient-based method may get trapped in the local optimum depending on the initial starting point. Such a problem can be partly resolved by running the algorithm several times with different initial guess values. The other class is stochastic methods, including adaptive stochastic search methods, evolutionary computation (EC) methods, and so on. These algorithms are applied with the purpose of searching for a global optimal solution.

In several studies, stochastic optimization procedures like evolutionary algorithms (EA) have been applied successfully to biochemical systems under reasonable model forms [34]. As most of the time there is no standard in the settings for the optimization procedures, the choices of values for these settings greatly affect the performance of the optimization process. Moreover, it usually requires a long computation time (lasting several hours) for running such stochastic search algorithms. In many cases, a lack of time prevents researchers from systematically benchmarking these settings. For scientists and experimentalists from biology-related domains, the complexity of the above-mentioned procedures and the long computation time hinder the application of these algorithms. Therefore, a fast and efficient parameter estimation strategy is needed for the improvement of model quality.

This research focuses on some specially structured nonlinear models, specifically the S-system model (detailed in Chapter 2) and the linear fractional model (LFM) (detailed in Chapter 3), and develops parameter estimation strategies based on their structure specialties. Once these parameters are identified, we can further validate the mathematical model, by comparing the predicted system dynamics in simulation with some new experimental observations under different conditions. If the model accuracy is satisfactory (consistent with the measurement data), then the estimation process is finished. Otherwise the estimation results can be considered as a pre-step for providing possible initial values before more complex algorithms are applied.

As shall be seen from the detailed structure of the S-system model, the RHS of each ODE is composed of two terms. By utilizing the property that power-law functions can be considered linear under the logarithmic coordinates, the parameters could be estimated by linear regression iteratively. Another perspective is that the

parameter set can be divided into two parts: the linear set and the nonlinear set. The model shows up to have the character of partial linearity. Rate constants appear in linear form and kinetic orders fall into the nonlinear set. In this research, estimation strategies will be implemented by utilizing this type of separable property of the parameters in the system model.

Such a separation approach can also be applied to the LFM, as the model can be regarded as linear with respect to parameters in the numerator. Therefore, by using the above-mentioned separation method, the optimization problem is reduced by dimension, i.e., only with respect to parameters in the denominator.

The general difficulty residing in nonlinear optimization problems is that there is no analytical solution for minimizing the objective function. Thus, various numerical strategies, such as the gradient-based methods, are applied in order to find the minimum. For the parameter estimation problem in the LFM, it is interesting to note that, by using the separation strategy and the weighted least squares method with a properly chosen weight matrix (in specific form), the objective function becomes quadratic with respect to parameters in the denominator. Therefore, estimates for all parameters can be analytically expressed. The estimation process will be greatly simplified.

In a nutshell, the objectives of this research reside on two aspects. One is to adopt strategies to reduce the complexity of the estimation problem. The other is to develop easy-to-use and fast methods for estimating parameters in two classes of nonlinear biological systems, i.e., the S-system and the LFM.

1.4 Thesis Organization and Notations

This thesis is organized as five parts. Chapter 1 is composed of background and introduction of the research, and an outline of the whole work. Chapter 2 explains parameter estimation methods for the S-system, based on the alternating least squares method and the separation principle. The effectiveness of these methods is demonstrated through two numerical examples, with the methods applied on S-system

models. Chapter 3 introduces parameter estimation methods for the LFM. The separation approach is also applied here and moreover a specially designed algorithm for estimating parameters in the LFM is presented. The estimates can be analytically expressed using a weighted least squares approach. In Chapter 4, a framework is proposed for a more complex estimation problem with state and reaction rate estimation. Such a method will simplify the experimental design for biological reaction systems, and provide estimates for variables which are unnecessary to measure or cannot be measured directly. An example is provided out to illustrate the proposed method. In Chapter 5, the research work is summarized and additional comments are made. Also, some future work that could be considered concludes this last section.

The notations used in this thesis are explained as follows. $\mathbb{R}^n$ denotes the set of all n-dimensional column (or row) vectors; $\mathbb{R}^{n\times m}$ denotes the set of all $n \times m$ matrices. The superscript $^{\mathrm{T}}$ denotes the matrix transpose. The Euclidean norm of an n-dimensional vector $a = [a_1, a_2, \ldots, a_n] \in \mathbb{R}^n$ is defined by $\|a\| = \sqrt{\sum_{i=1}^{n} a_i^2}$. $\mathrm{diag}[a_1, a_2, \ldots, a_n]$ is a diagonal matrix whose diagonal elements are a_i $(i = 1, 2, \ldots, n)$. In ODE $\dot{y} = f(t, y)$, y, $\dot{y}$ and t are usually vector-valued.

Chapter 2

Parameter Estimation Methods for S-system

2.1 Introduction

2.1.1 Explanation of S-system

For time course data analysis, the Biochemical Systems Theory (BST) [30, 31, 32, 43, 38] is an appropriate framework for modeling reaction networks, such as metabolic networks. BST is based on the approximation of Taylor's series expansion. Under this framework, the products of power-law functions are used to approximate reaction rates or fluxes. The most commonly encountered nonlinear models within the BST are the S-system model and the Generalized Mass Action (GMA) model [43]. In this chapter, the S-system model is focused on as a testbed for the proposed estimation methods.

The general form of an S-system is shown in the following:

$$\dot{x}_i = \alpha_i \prod_{k=1}^{d} x_k^{g_{ik}} - \beta_i \prod_{k=1}^{d} x_k^{h_{ik}}, \quad (i = 1, \cdots, d), \tag{2.1}$$

where x represents the state variable (e.g., the concentration of metabolite), and d denotes the total number of variables within the system. The non-negative parameters α_i and β_i are called rate constants, and the real-valued parameters g_{ik} and h_{ik} are kinetic orders.

In the S-system model, the structure of the RHS of each ODE is composed of two terms, in the form of product of power-law functions. The positive term contributes to the production of reactant, and the term with negative sign ahead corresponds to the degradation of reactant [43]. The difference between these two terms represents

the change of state variables over time. For the kinetic order parameter g or h, a positive value means an activating effect, while a negative value represents an inhibitory effect, either to the production or the degradation process. A zero-valued kinetic order indicates that the corresponding variable x_k has no effect on the change of variable x_i.

The primary reason to adopt such an S-system model within the BST framework is that the form with power-law functions has the capability to capture various nonlinearity appearing in system dynamics. Another benefit is that such an S-system model provides a simple mapping relationship between the network structure and the symbolic equations. Once the structure features of the reaction system are obtained (from biological knowledge and information about the underlying system), the model equations are relatively easy to write down in a simplified form, with some kinetic order values set to zero and some deemed to be negative or positive. Moreover, the power-law structure of the model is amenable to numerical analysis and simulation. To sum up, the S-system model is a representative dynamic model suitable for several kinds of biological systems, such as metabolic networks, genetic regulation networks, and signaling pathways. It finds a good compromise which can capture the dynamics while keeping the mathematical form simple and unified.

2.1.2 Related Work

The development of efficient algorithms or methods is required for determining the optimal values of parameters, after the time series measurement of state variables is obtained. The parameter estimation task can be formulated as an optimization problem, to minimize the cost/objective function measuring the difference between the model-based prediction and the experimental data. The Euclidean distance and the least squared error criterion are usually adopted for optimizing the cost function.

There are two ways typically adopted to construct the cost function: one is based on concentration error and the other is formulated as slope error, see Eq. (1.2). The concentration error based cost function is a straightforward representation to measure the prediction consistency, and therefore was widely adopted in the past.

Many estimation algorithms have been developed to minimize such a form of cost function. For example, Kikuchi *et al.* proposed a genetic algorithm (GA) method to infer the dynamics of a small genetic network with five variables in an S-system model [21]. One weakness for such concentration error based cost functions is that the numerical integration is required to solve the ODE on each step during the optimization process. The numerical integration process will cost large amounts of computation time, as high as 95% of the total time for optimization [44].

The slope error based cost function is an alternative way to formulate the cost function. Such a formulation requires the slope information from measurement data, and the consistency is evaluated on basis of slope values. As this approach circumvents the time-consuming numerical integration process, it shows superiority in speed for optimization. Recently, several methods have been proposed to estimate parameters in the S-system model and the like, adopting such a cost function. Voit and Almeida developed an ANN-based method to decouple the dynamic systems to identify the structure of the S-system model and estimate parameters within the model [44]. Tucker *et al.* used the interval analysis technique to estimate parameters for the generalized mass action (GMA) models [39]. Ho *et al.* and Wang *et al.* respectively proposed an intelligent two-stage evolutionary algorithm [18] and a so-called unified approach [45] to estimate parameters in the S-system models.

Generally speaking, these proposed methods usually require large amounts of computation resources and time, and are difficult to implement. One aspect of concern is that these methods do not sufficiently take the special structures of biological system models into consideration. Previous research has shown that consideration of the model specialties may simplify the parameter estimation problem, e.g., reducing a nonlinear problem to a linear one [47].

2.1.3 New Ideas

The first idea is to utilize the two-term structure of an S-system model. By taking a logarithmic transformation and applying the linear least squares method iteratively, the final estimates for parameters can be obtained until the stop criterion is met.

This idea is quite intuitive and straight-forward.

Another idea is that parameters in the S-system model can be separated into two groups: one group of parameters shows linearity in the model while another group of parameters is nonlinear. From this viewpoint, we employ a separation parameter estimation strategy to estimate parameters in the S-system models. Some early studies on the separation estimation methods were based on matrix factorization [14, 19, 8]. Our strategy consists of three steps. In the first step, parameters linear in a model are estimated by optimizing the objective function using linear least squares method [5], assuming all parameters nonlinear in the model are known. In the second step, substituting the estimated parameters in the first step into the objective function yields a new objective function, which is only related to parameters nonlinear in the model. Then parameters nonlinear in the model are estimated by proper nonlinear estimation methods. In the last step, the estimates of parameters linear in the model are calculated using the estimates of parameters in the second step. The foreseeable advantages of this type of separation method include: 1) requiring less initial guesses as linear parameters have been eliminated; 2) reducing the dimension of parameter space to search; 3) reducing the computational effort to calculate the refinement term, when using nonlinear parameter estimation methods such as the Gauss-Newton method and its variants.

2.1.4 Contents of Chapter

Briefly, this chapter is organized as follows. Section 2.2 states some pre-processing steps, including the decoupling of the system model, the smoothing of measurement data and the slope approximation. In Section 2.3, we first give an intuitive method called the alternating least squares (ALS) method, and an illustrative example follows. In Section 2.4, a separation parameter estimation approach is introduced and derived. The separation parameter estimation approach is applied to estimating parameters in an S-system example. In this section, performance of the separation parameter estimation method is compared with that of conventional methods, the Gauss-Newton algorithm and the Levenberg-Marquardt algorithm. Finally in

Section 2.5, we summarize these two proposed methods for the S-system.

2.2 Pre-processing Steps

2.2.1 Decoupling

Among the challenges for parameter estimation problems, the development of efficient optimization methods is crucial but also very tough. As we know, numerical integration tools are needed to solve ODEs in the optimization process, and the fact is inevitable that the numerical integration is quite time-consuming and can fail sometimes. It is observed that the numerical integration requires even more than 95% of the total time in the optimization process [44].

To resolve such a problem, Voit and Almeida presented a decoupling method, splitting the ODEs into sets of separate algebraic equations and approximating the derivatives by calculated slopes from measurement data [44]. This method does not require the numerical integration of ODEs and thus the subsequent optimization process can be much faster. Here is an example. If one has an S-system model with d components, and with the measurement of each component at n time points, the decoupling of original S-system can be made as:

$$s_i(t_k) = \alpha_i \prod_{j=1}^{d} x_j^{g_{ij}}(t_k) - \beta_i \prod_{j=1}^{d} x_j^{h_{ij}}(t_k), \qquad (2.2)$$
$$(i = 1, \cdots, d \quad \text{and} \quad k = 1, \cdots, n),$$

where $s_i(t_k)$ is the calculated slope of the concentration of component x_i at the time point t_k. The model framework of parameter estimation problem is thus reformulated from a set of d ODEs to $d \times n$ algebraic equations. After such reformulation, the analysis of algebraic equations becomes an easy problem, and the computation time for parameter estimation will be greatly reduced.

2.2.2 Smoothing and Slope Approximation

Before the procedures of parameter estimation start, the input/measurement data need to be pre-processed. The two goals ahead include: 1) to use a filter/smoother for the measurement which will extract the original signal from the noise-contaminated data; 2) to calculate/approximate the slopes s_i with great precision. It is noteworthy that the accuracy of slope calculation, and furthermore of parameter estimation, are greatly dependent upon the efficiency of smoothing algorithm, because the calculation of slopes would amplify the effect of errors from input data. Therefore, smoothing algorithms should be used for pre-processing.

Vilela *et al.* proposed a smoothing algorithm, named 'autosmooth' (AS), to deal with the noise in measurement and calculate the slopes from smoothed data [42]. Advantages of this algorithm include the capability of handling varying/uneven noise structures and that the authors provided a closed-form solution for computing derivatives of the smoothed signal.

This AS algorithm is based on the Whittaker filter and the Eilers' extension [9]. The basic ideas of the Whittaker filter include two aspects. On one hand, the newly smoothed time series are required to give a close fit to the original/raw noisy data points. On the other hand, the connection curve of the smoothed output could not be too rough. That means the values of neighborhood points for any smoothed output are relatively similar to each other. This filter undergoes an optimization process, relying on two control parameters: one determining the width of neighborhood and the other weighting the two aspects (closeness and smoothness). These two control parameters are optimized by the Eilers' approach using cross-validation. Vilela *et al.* improved this process by minimizing an entropy function [42].

In this research we adopt the AS algorithm package [1, 42] developed by Vilela *et al.* and implement it in the Matlab environment. This package will be used for the smoothing of the input data and also the estimation of the slopes in Section 2.3.

Another slope approximation strategy we adopt is the 5-point central difference

(5CD) method [52], as shown in Eq. (2.3):

$$\dot{x}_i(t) = \frac{x_i(t-2\Delta t) - 8x_i(t-\Delta t) + 8x_i(t+\Delta t) - x_i(t+2\Delta t)}{12\Delta t}. \tag{2.3}$$

Here, Δt represents the sampling interval. This method will be applied to the smoothed or noise-free data.

2.3 Alternating Least Squares Method

2.3.1 Algorithm Description

Consider the ith ODE in the general form of the S-system model (the derivative $\dot{x}_i$ replaced with the slope s_i):

$$s_i = \alpha_i \prod_{j=1}^{d} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{d} x_j^{h_{ij}}. \tag{2.4}$$

The parameters within the equation can be intuitively divided into two groups – those in the first term of the right-hand side (RHS) and the ones in the second term. By using some transformation (taking the logorithm in this case), the estimation problem can be solved by simple linear regression. The detailed procedures are stated below.

Step 1. Given the initial values of parameters in the second term of the RHS in Eq. (2.4), i.e., β_i and h_{ij}, we can move this second term to the left-hand side (LHS), placed together with the calculated slope s_i, and then take the logarithm on both sides of the equation to have

$$\log(s_i + \beta_i \prod_{j=1}^{d} x_j^{h_{ij}}) = \log \alpha_i + \sum_{j=1}^{d} g_{ij} \log(x_j). \tag{2.5}$$

Therefore, the RHS in Eq. (2.5) can be regarded as linear with respect to parameters $\log \alpha_i$ and g_{ij}. By the standard least squares method, these parameters can be estimated.

Step 2. Given the estimated values of parameters in the first term of the RHS in Eq. (2.4), which is accomplished in Step 1, the first term can be moved to the LHS,

together with s_i. Take the logarithm on both sides to have

$$\log(\alpha_i \prod_{j=1}^{d} x_j^{g_{ij}} - s_i) = \log \beta_i + \sum_{j=1}^{d} h_{ij} \log(x_j). \tag{2.6}$$

Similarly as in Step 1, parameters in the second term of the RHS in the general form can be estimated.

Then Step 1 will be repeated with the newly acquired estimates for β_i and h_{ij}. By performing Step 1 and Step 2 in an alternating way iteratively, until any stop criterion is met, the estimated values of parameters can be obtained. The termination criteria we adopt here include two factors: the first is the sum of squared errors (SSE) representing the conformity result of linear regression, and the second is the number of iterations regulated by a specific value. The SSE is defined as follows:

$$\text{SSE} = \sum_{k=1}^{n} [y_d(k) - L_p(k) \cdot b_p]^2 + \sum_{k=1}^{n} [y_p(k) - L_d(k) \cdot b_d]^2 \tag{2.7}$$

where 'd' represents degradation and 'p' means production for the subscripts, and the time points of the measurement are indexed by k. In Eq. (2.7), $y_d = \log(s_i + \beta_i \prod_{j=1}^{d} x_j^{h_{ij}})$, $L_p = [1 \ \log x_1 \ \cdots \ \log x_d]$, $b_p = [\log \alpha_i \ g_{i1} \ \cdots \ g_{id}]^{\mathrm{T}}$, and similarly $y_p = \log(\alpha_i \prod_{j=1}^{d} x_j^{g_{ij}} - s_i)$, $L_d = [1 \ \log x_1 \ \cdots \ \log x_d]$, $b_d = [\log \beta_i \ h_{i1} \ \cdots \ h_{id}]^{\mathrm{T}}$. Here in L_p, only those x_j that are known to have effect on the production term are included in the expression. L_d is defined in the same way, incorporating those x_j with a non-zero kinetic order shown in the degradation term.

As the S-system has been decoupled, the estimates for parameters in one ODE can be obtained at one time. After applying the ALS method for all d equations in an S-system, we will have all parameter values within the whole S-system model.

2.3.2 Numerical Example

Here we adopt a small-scale biochemical network as a test example, and use it to generate the synthetic time series data. This network has been discussed by other researchers for parameter identification and estimation [44, 40]. There are four metabolites within this example network and therefore the S-system model has four ODEs,

as shown in Eq. (2.8). It is important to note that such an example has similar and relevant features to a more complex network, and as a result it is appropriate as a testbed to show the performance of parameter estimation methods.

$$\begin{aligned}
\dot{x_1} &= \alpha_1 x_3^{g_{13}} - \beta_1 x_1^{h_{11}}, \\
\dot{x_2} &= \alpha_2 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x_3} &= \alpha_3 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x_4} &= \alpha_4 x_1^{g_{41}} - \beta_4 x_4^{h_{44}}.
\end{aligned} \tag{2.8}$$

The corresponding network map is shown in Fig. 2.1. Straight arrows with solid lines represent material flows in the network, while regulatory influences are displayed as curved arrows with dashed lines. The network map clearly shows that the production of component X_1 is affected by the inhibition exerted by component X_3, which is generated from X_1 via the intermediate reactant X_2. Component X_1 also influences the production of X_4, and this product promotes the degradation of X_3. In addition, each component promotes its own degradation.

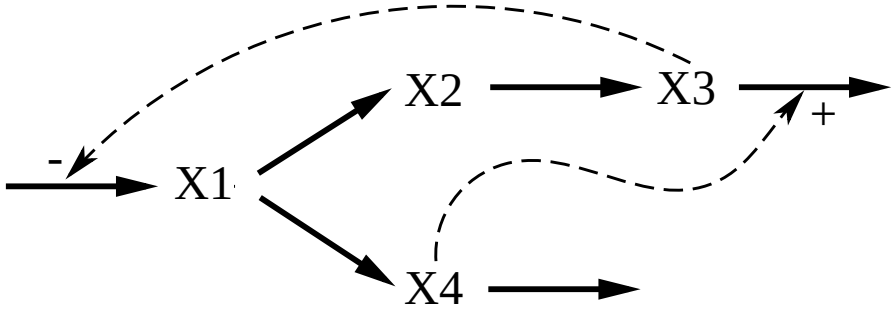

Figure 2.1: A metabolic pathway with four components and two regulatory signals (redrawn from [44])

The synthetic time-course data are obtained by simulating the model system with `ode45()` function in Matlab, which solves ODEs numerically. The initial values of x_1 to x_4 are set as appeared in reference [44], specifically 1.4, 2.7, 1.2 and 0.4. The true values of parameters in Eq. (2.8) are listed as in Table 2.1. The time-course profiles of metabolic concentrations are generated upon 51 time points, equally sampled over the time interval $0 \sim 5$ min. Fig. 2.2 shows the time series of metabolic concentrations to represent measurement data.

Table 2.1: True values of parameters in Eq. (2.8)

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
α_1	12	β_1	10	g_{13}	-0.8	h_{11}	0.5
α_2	8	β_2	3	g_{21}	0.5	h_{22}	0.75
α_3	3	β_3	5	g_{32}	0.75	h_{33} & h_{34}	0.5 & 0.2
α_4	2	β_4	6	g_{41}	0.5	h_{41}	0.8

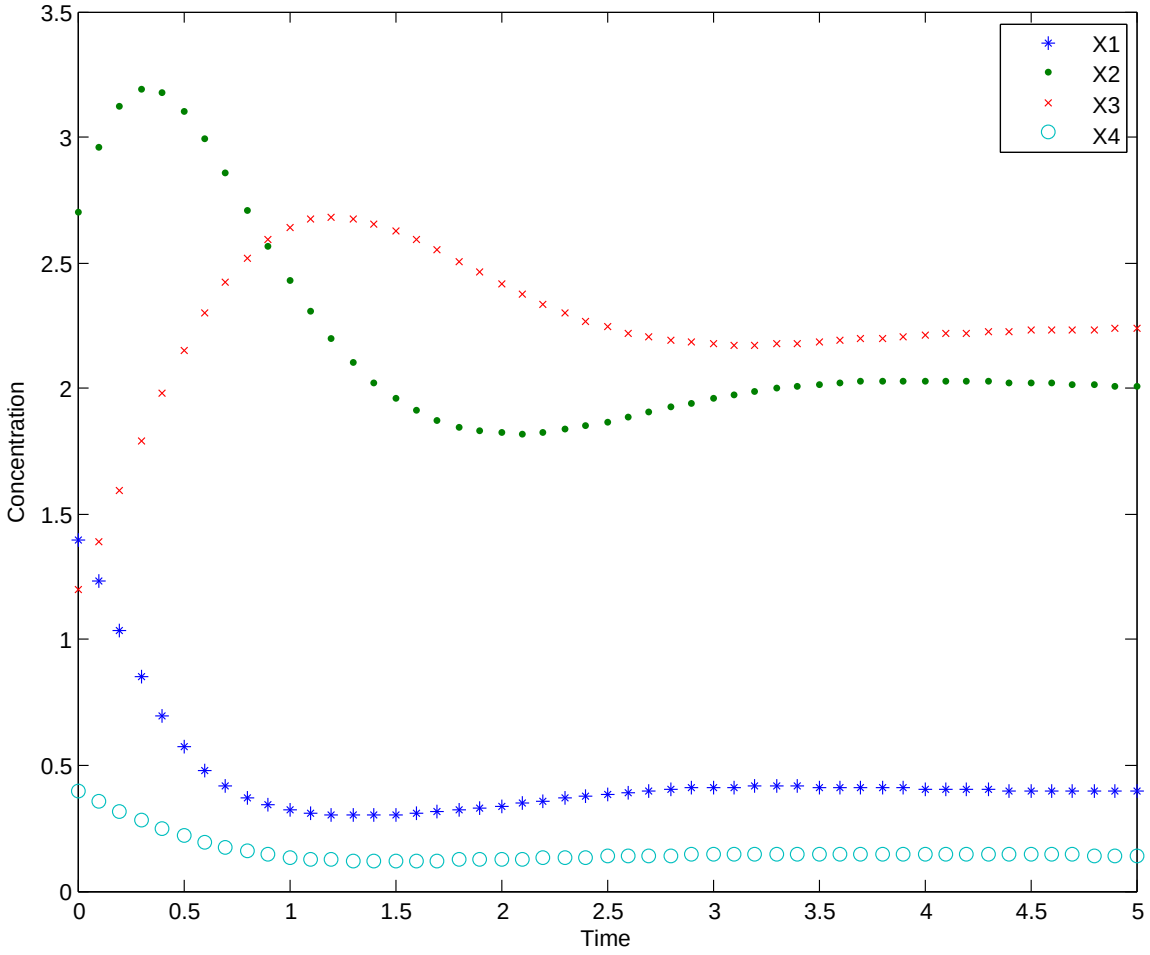

Figure 2.2: Time courses of model (2.8)

First let us make a trial on the noise-free time series data. Here we adopt two methods to calculate the first-order derivatives of state variables. One approach is that the slopes are calculated using the above-mentioned AS package. The other way is to use the 5-point central differences (5CD) [52]. Figure 2.3 shows the approximated slopes over time using these two methods. The slopes represented in the form of lines are the results using the AS package, while those shown in the form of symbols are given by the 5CD method. As we can see, for noise-free time series, these two methods give almost the same results of slope calculation.

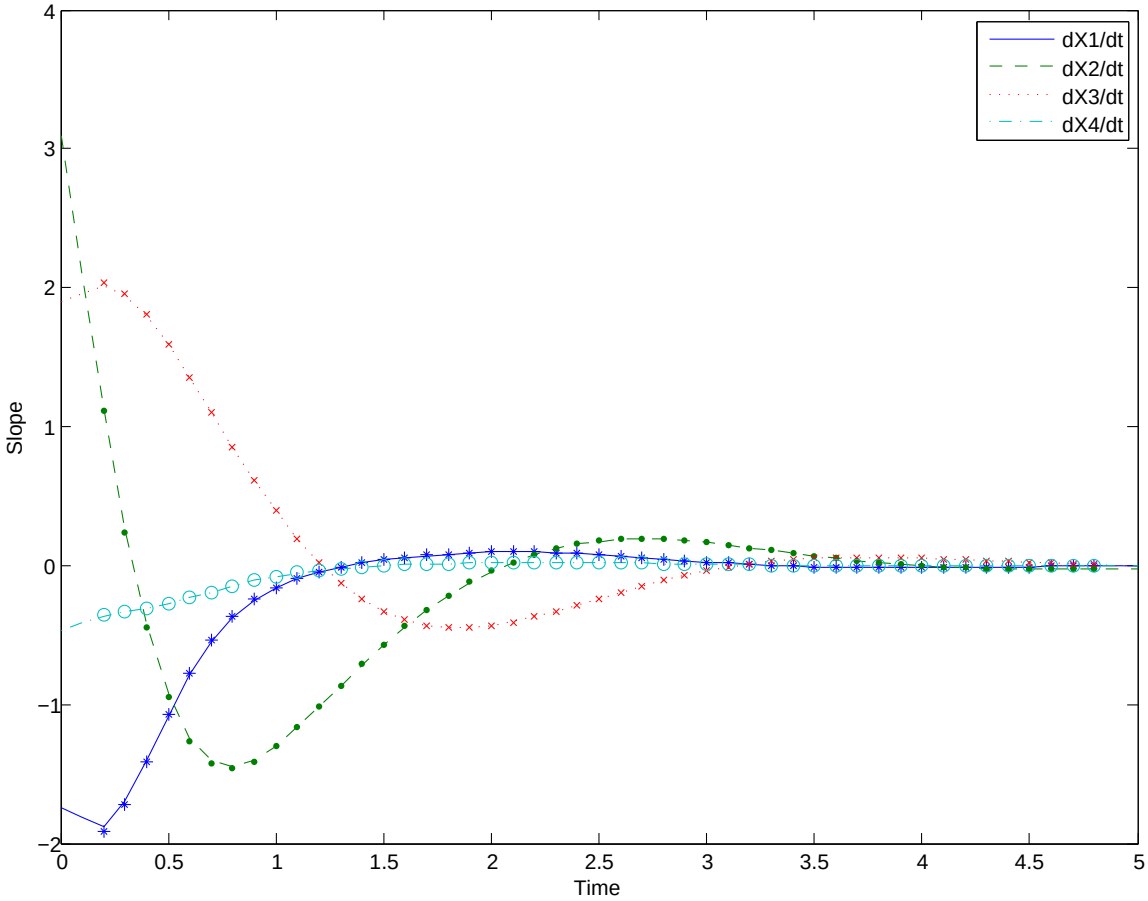

Figure 2.3: Calculated slopes of model (2.8)

Then the alternating least squares method is applied on the input data, including the time series and the approximated slopes. The termination criteria for parameter estimation are set as two conditions here: the lower threshold for ln(SSE) is -15 and the maximum number of iterations is 10^5. The initial values of rate constants β are

set as 17 [1]; the initial guesses for kinetic orders h_k are chosen to be 1.

Using the estimated slopes from the AS package, simulation results of the alternating least squares method are shown in Table 2.2. In the table the relative estimation error (REE) is defined as follows:

$$\text{REE} = \frac{\|\gamma_{\text{est}} - \gamma_{\text{true}}\|}{\|\gamma_{\text{true}}\|}, \tag{2.9}$$

where γ represents the target parameter, γ_{est} is the estimated value while γ_{true} is the true value.

Table 2.2: Estimation results using the ALS on the S-system (slopes calculated via AS)

True Value	Estimate	REE	True Value	Estimate	REE
12.00	10.9233	8.97%	10.00	8.6525	13.48%
8.00	8.2264	2.83%	3.00	3.2365	7.88%
3.00	2.9275	2.42%	5.00	4.8463	3.07%
2.00	2.0092	0.46%	6.00	6.0609	1.02%
−0.80	−0.9947	24.34%	0.50	0.6149	22.98%
0.50	0.4783	4.34%	0.75	0.7095	5.40%
0.75	0.7587	1.16%	0.50	0.5222	4.44%
			0.20	0.2029	1.45%
0.50	0.5104	2.08%	0.80	0.8078	0.97%

We can see that the estimates conform to the nominal values satisfactorily in general for most parameters, with REE less than 5%. Only the estimates for those parameters in the first ODE somewhat deviate from the true values. With the estimated parameter values, the state profiles over time can be generated by `ode45()` function. As shown in Fig. 2.4, they (in the form of lines) conform well to the original time series (in the form of symbols).

[1] Initial numbers were firstly chosen as 15 or 20, but incured the problem of taking the logarithm on negative data.

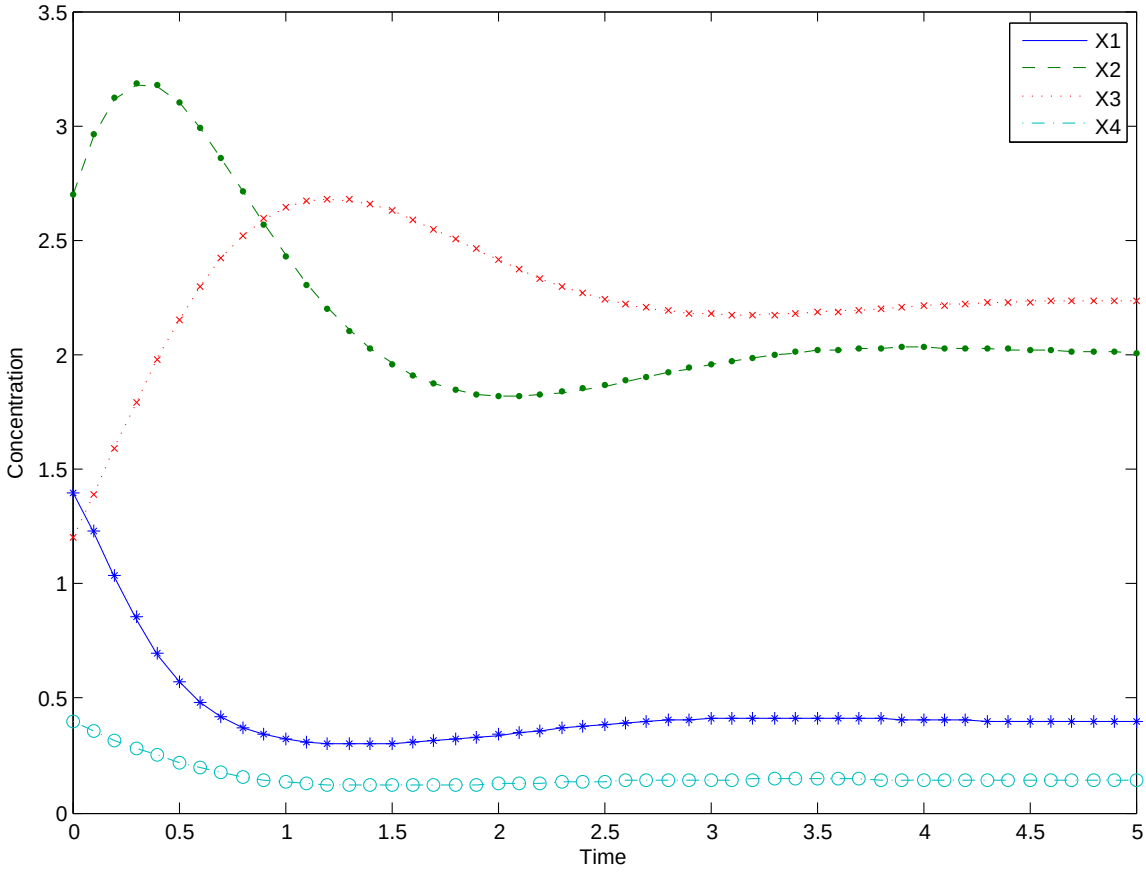

Figure 2.4: Comparisons of the estimated state profiles (slopes calculated via AS) with the original time courses of model (2.8)

The most obvious advantage of this alternating least squares method is that it performs much faster than those algorithms estimating whole parameters simultaneously. One shortcoming of this ALS method comes from that the logarithm may be applied on negative data during the iteration process, especially when the initial guess is not chosen properly. This problem sometimes can be partly resolved by setting a relatively large initial value for rate constants β. Another idea is to eliminate those time points whenever the logarithm of negative data would occur, and only apply the ALS method upon those appropriate time points step by step. Unfortunately, such an intuitive idea has been tested but cannot always give good performance.

Using the estimated slopes from the 5-point central differences method, simulation results of the alternating least squares method are shown in Table 2.3.

Table 2.3: Estimation results using the ALS on the S-system (slopes calculated via 5CD)

True Value	Estimate	REE	True Value	Estimate	REE
12.00	12.0188	0.16%	10.00	10.0202	0.20%
8.00	24.6456	208.07%	3.00	19.4116	547.05%
3.00	3.4822	16.07%	5.00	5.2450	4.90%
2.00	16.8248	741.24%	6.00	18.6121	210.20%
−0.80	−0.7983	0.21%	0.50	0.4990	0.20%
0.50	0.1212	75.76%	0.75	0.1826	75.65%
0.75	0.6659	11.21%	0.50	0.4185	16.30%
			0.20	0.1446	27.70%
0.50	0.0385	92.30%	0.80	0.0698	91.28%

The results of estimated parameters in the second and the fourth ODE are far away from the nominal values. This indicates that the performance of ALS is not steady, and will change with different slope approximation methods [2]. With these

[2]This does not mean the AS package is superior to the 5CD method for noise-free time series.

estimated parameter values, the state profiles over time can be generated by `ode45()` function in Matlab. As shown in Fig. 2.5, they (in the form of lines) basically conform to the original time series (in the form of symbols), but not as well as the previous result in Fig. 2.4. We consider this group of estimation results in Table 2.3 not meaningful on the whole, because under other experimental conditions (different initial concentrations of reaction components), the model-based prediction might not be in agreement with the time series measurement.

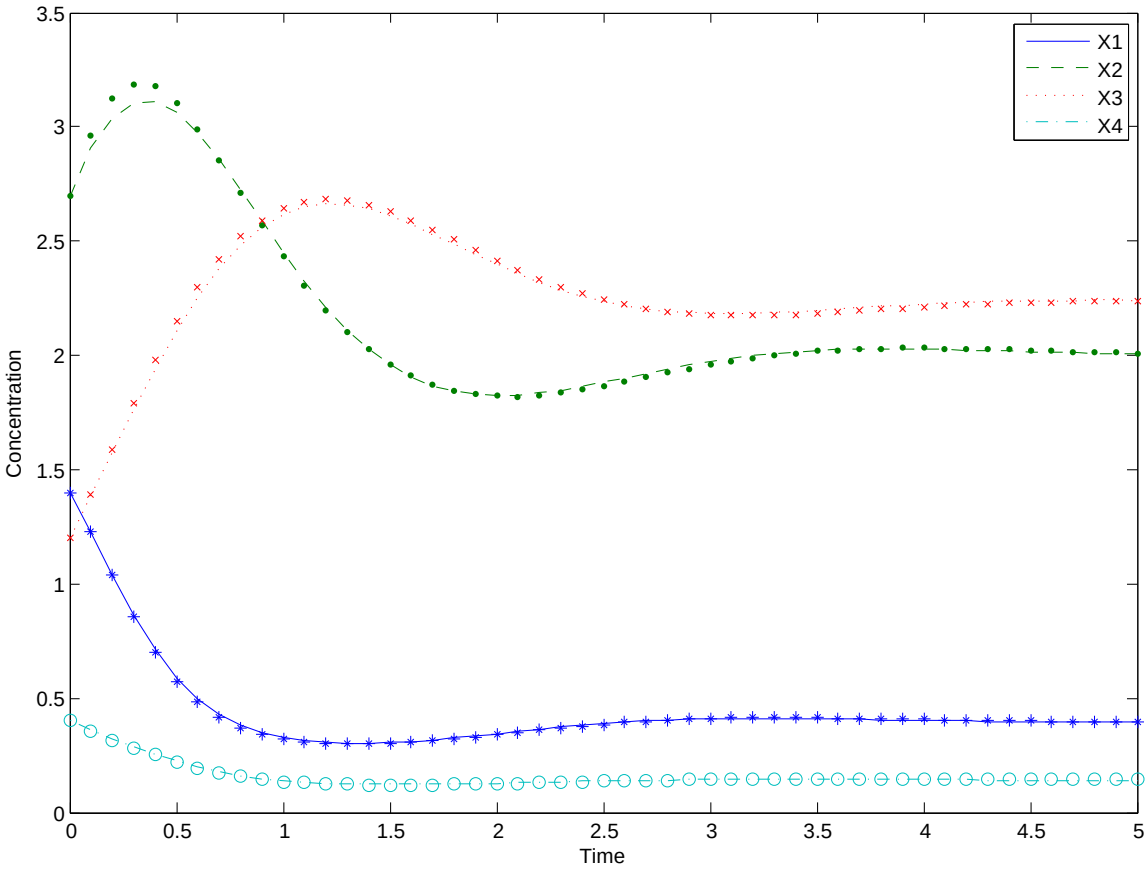

Figure 2.5: Comparisons of the estimated state profiles (slopes calculated via 5CD) with the original time courses of model (2.8)

Besides, it is quite informative to notice that simulating the system with these two groups of different estimated values will produce time profiles almost indistinguishable from the original synthetic time courses. This discovery gives rise to the insight that, for the S-systems, more than one group of parameter values may cause the identical system behavior to happen [16, 15, 29]. One possible explanation for this case

is that there are much more parameters than reaction variables within the network, and a proper combination of parameters could give similar dynamics. Therefore, to gain more accurate estimates, it is better to know the range for some parameters beforehand, or enforce constraints from prior knowledge of the system model. For example, the values of kinetic orders g and h should be within a reasonable range from -2 to 3, from the biological meaning in [43].

2.4 Separation Estimation Method

2.4.1 Algorithm Description

Consider the problem of estimating parameters $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ in a general model $\eta(t, \alpha, \beta)$ from observations

$$y_i = \eta(t_i, \alpha, \beta) + \varepsilon_i, \tag{2.10}$$

obtained at a sequence of distinct time points $t_1, \ldots, t_n$, where $\varepsilon_i (i = 1, 2, \ldots, n)$ is random measurement noise. Let $\gamma = [\alpha;\ \beta] \in \mathbb{R}^{(p+q)}$ be a vector consisting of all parameters.

The least squares method transfers the problem of estimating parameters in model $\eta(t, \alpha, \beta)$ into finding the optimal parameters $\hat{\gamma} \in \mathbb{R}^{(p+q)}$ that minimize the objective function

$$J(\gamma) = \sum_{i=1}^{n} [y_i - \eta(t_i, \gamma)]^2 = [Y - \eta(\gamma)]^{\mathrm{T}} [Y - \eta(\gamma)], \tag{2.11}$$

where $Y = [y_1, \ldots, y_n]^{\mathrm{T}}$ and $\eta(\gamma) = [\eta(t_1, \gamma), \ldots, \eta(t_n, \gamma)]^{\mathrm{T}}$.

Starting with an initial guess $\hat{\gamma}_0$, the optimization of objective function (2.11) can be solved by iteratively computing:

$$\hat{\gamma}_{i+1} = \hat{\gamma}_i + \Delta\hat{\gamma}_i, \tag{2.12}$$

where the increment $\Delta\hat{\gamma}_i$ is a correction/refinement term, until a convergence condition is met, such as that $\|\hat{\gamma}_{i+1} - \hat{\gamma}_i\|$ or $\|\hat{\gamma}_{i+1} - \hat{\gamma}_i\| / (\|\hat{\gamma}_i\| + \delta)$ is less than a preset small

value, where δ is a small positive number to avoid division by zero. For different algorithms, $\Delta\hat{\gamma}_i$ is calculated in various ways.

The Gauss-Newton algorithm (GNA) [6] is a common method to solve nonlinear least squares problems. This algorithm is derived from Newton's method for solving nonlinear equations. The increment term $\Delta\hat{\gamma}_i$ in Eq. (2.12) for GNA is calculated by

$$\Delta\hat{\gamma}_i = [P_i(\hat{\gamma}_i)]^{-1} f(\hat{\gamma}_i), \tag{2.13}$$

where $P_i(\hat{\gamma}_i) = X(\hat{\gamma}_i)^{\mathrm{T}} X(\hat{\gamma}_i)$, $f(\hat{\gamma}_i) = X(\hat{\gamma}_i)^{\mathrm{T}}[Y - \eta(\hat{\gamma}_i)]$, and $X(\gamma) = [\frac{\partial \eta(\gamma)^{\mathrm{T}}}{\partial \gamma}]^{\mathrm{T}}$.

The Levenberg-Marquardt algorithm (LMA) [6] is also a popular numerical solution to optimization problems. It is an improved version of the GNA and the increment term $\Delta\hat{\gamma}_i$ in Eq. (2.12) for the LMA is calculated by

$$\Delta\hat{\gamma}_i = [P_i(\hat{\gamma}_i) + \lambda D]^{-1} f(\hat{\gamma}_i), \tag{2.14}$$

where D is a positive diagonal matrix, and the adjustment parameter λ is changed at each step. The LMA is more robust than the GNA, but a little bit slower. More detailed procedures of the LMA can be found in [6].

In the GNA and the LMA methods mentioned above, at each step the value of $\Delta\hat{\gamma}_i$ has to be evaluated. In implementation of these algorithms, the inverse of matrices like P_i is not directly computed, but the elimination or decomposition approaches are adopted. When the dimension of parameter space is higher, namely, with a bigger sized $(p+q) \times (p+q)$ matrix P_i, it will consume more time in calculation. In this study, we focus on a class of models which have the property of partial linearity with respect to some parameters , i.e., $\eta(t, \alpha, \beta) = b(t, \beta)\alpha$. Therefore, model (2.10) becomes

$$y(t) = b(t, \beta)\alpha + \varepsilon. \tag{2.15}$$

To be multiplication-compatible, in model (2.15) $b(t, \beta)$ is assumed to be a p-dimensional row vector.

Many biological systems can be described in structure as in the parameter-separable model (2.15), especially models derived from the BST and the Michaelis-Menten rate law [22, 43]. To estimate parameters in such models, we need to find

the optimal values of parameters $\hat{\alpha} \in \mathbb{R}^p$ and $\hat{\beta} \in \mathbb{R}^q$ that minimize the following objective function

$$J(\alpha, \beta) = \sum_{i=1}^{n} [y_i - b(t_i, \beta)\alpha]^2. \tag{2.16}$$

Let $Y = [y_1, \ldots, y_n]^{\mathrm{T}}$ and $B(\beta) = [b(t_1, \beta)^{\mathrm{T}}, \ldots, b(t_n, \beta)^{\mathrm{T}}]^{\mathrm{T}}$, and then Eq. (2.16) can be rewritten as

$$J(\alpha, \beta) = [Y - B(\beta)\alpha]^{\mathrm{T}}[Y - B(\beta)\alpha]. \tag{2.17}$$

By considering the parameter-separable structure, this study presents a three-step estimation method for parameters in model (2.15). In the first step, for a given value of $\beta = \bar{\beta}$, using the linear least squares method, we obtain the estimates of α as follows

$$\hat{\alpha} = [B(\bar{\beta})^{\mathrm{T}} B(\bar{\beta})]^{-1} B(\bar{\beta})^{\mathrm{T}} Y. \tag{2.18}$$

Substitute Eq. (2.18) into Eq. (2.17) to get

$$J(\hat{\alpha}, \bar{\beta}) = \widetilde{J}(\bar{\beta}) = Y^{\mathrm{T}}\{I - B(\bar{\beta})[B(\bar{\beta})^{\mathrm{T}} B(\bar{\beta})]^{-1} B(\bar{\beta})^{\mathrm{T}}\}Y. \tag{2.19}$$

In the second step, find the optimal parameters $\hat{\beta}$ that minimize the objective function (2.19). In principle, the optimization of objective function (2.19) can be solved by any nonlinear optimization method such as the Gauss-Newton method and its variants. Note that the objective function (2.19) has only q parameters and thus the size of the corresponding matrix P_i is $q \times q$ in the process of applying the Gauss-Newton method. Once the optimal parameters $\hat{\beta}$ are obtained, the optimal parameters α can be computed from Eq. (2.18) by letting $\bar{\beta} = \hat{\beta}$ in the last step. It is expected that the introduced separation method can reduce the computational time, as the matrix P_i for calculating the refinement term $\Delta\hat{\gamma}_i$ has a smaller size than that using the conventional methods.

2.4.2 Numerical Example

In this study we use the same S-system example with the associated network as shown in Fig. 2.1 [43, 44], which consists of 4 different reaction components. To highlight

the separable strucuture of the system model, the set of ODEs are rewritten here using α and β to represent parameters, as shown in Eq. (2.20).

$$\begin{aligned}
\dot{x}_1 &= \alpha_{11} x_3^{\beta_{13}} - \alpha_{12} x_1^{\beta_{11}}, \\
\dot{x}_2 &= \alpha_{21} x_1^{\beta_{21}} - \alpha_{22} x_2^{\beta_{22}}, \\
\dot{x}_3 &= \alpha_{31} x_2^{\beta_{32}} - \alpha_{32} x_3^{\beta_{33}} x_4^{\beta_{34}}, \\
\dot{x}_4 &= \alpha_{41} x_1^{\beta_{41}} - \alpha_{42} x_4^{\beta_{44}}.
\end{aligned} \tag{2.20}$$

In this model there are totally 17 parameters: 8 positive rate constants α_{ij} and 9 kinetic order parameters β_{ij}.

Table 2.4: Nominal values for parameters in model (2.20)

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
α_{11}	12	α_{12}	10	β_{13}	-0.8	β_{11}	0.5
α_{21}	8	α_{22}	3	β_{21}	0.5	β_{22}	0.75
α_{31}	3	α_{32}	5	β_{32}	0.75	β_{33} & β_{34}	0.5 & 0.2
α_{41}	2	α_{42}	6	β_{41}	0.5	β_{41}	0.8

To investigate the performance of the separation estimation method, a group of artificial/synthetic data is generated from model (2.20) with parameter values listed in Table 2.4 and initial states as $x_1 = 1.4$, $x_2 = 2.7$, $x_3 = 1.2$, $x_4 = 0.4$. The time courses are plotted in Fig. 2.2. The time variable t starts from 0, and ends at 5 min. The data are evenly sampled on 51 time points and the sampling interval $\Delta t = 0.1$ min. In this study, synthetic data are adopted noise-free, in order to evaluate the proposed estimation method. Actually noisy data challenge any good method in parameter estimation problems. For noisy data, it is necessary to adopt some procedures to smooth them before the estimation process [44].

This study adopts the 5-point central difference to approximate derivatives of state variables, as in Eq. (2.3). The estimation error, the minimum value of objective function, the number of iterations and the consumed CPU time are employed to

evaluate the performance of estimation methods. Here the relative estimation error (REE) is defined as in Eq. (2.9).

Comparisons between the conventional method and the separation method both using the GNA and the LMA approaches have been conducted. In addition, to investigate the robustness of the methods, namely the insensitivity to various initial guesses, the initial values of α and β are chosen from the range with true values plus a relative Gaussian distribution, i.e., $[\alpha_0; \beta_0] = [\alpha_{\text{true}}; \beta_{\text{ture}}] \cdot (1 + \sigma \cdot \varepsilon)$, where ε follows the standard normal distribution and σ is a positive constant, taking various values in this study. The coded Matlab script runs 100 times with 100 initial guesses randomly selected for each value of σ. It is interesting to note that when converged, the minimum values of the objective function have the same number (1.9296×10^{-4} in this case), and the estimation errors have the same values as listed in Table 2.5 (the REE for the whole parameter vector is 0.43%), for all tested methods and all runs. This may indicate that the value 1.9296×10^{-4} is the global minimum of the objective function and thus all methods reach the best estimates of parameters. We can see that the REE for each estimate is much smaller than that using the ALS method. The estimation error is supposed to stem mainly from the slope approximation formula (2.3). It is expected that the estimation error will get smaller with a higher sampling frequency.

Table 2.6 lists the average results of the iteration count and the CPU time over 100 runs for all tested methods. GNA_C (LMA_C) stands for the conventional method with GNA (LMA), while GNA_S (LMA_S) stands for the separation method with GNA (LMA). The software environment for this simulation is: Matlab Version 6.5 Release 13, Windows XP Professional SP2, 1.81GHz AMD Turion 64bit $\times 2$ CPU, 960MB RAM.

From Table 2.6, the count of iterations and the consumed CPU time for the GNA_S are smaller than those for the GNA_C. Moreover, the number of iterations and the consumed CPU time for the LMA_S are less than half of those for the LMA_C. In addition, when initial guesses are far away from the true values (i.e., the values of σ are large), the separation method more obviously outperforms the

Table 2.5: Estimation results using the separation method on the S-system (slopes calculated via 5CD)

True Value	Estimate	REE	True Value	Estimate	REE
12.00	11.9945	0.0458%	10.00	9.9902	0.0980%
8.00	7.9952	0.0600%	3.00	2.9933	0.22%
3.00	3.0473	1.58%	5.00	5.0121	0.24%
2.00	2.0321	1.60%	6.00	5.9452	0.91%
−0.80	−0.8021	0.26%	0.50	0.5012	0.24%
0.50	0.5009	0.18%	0.75	0.7511	0.15%
0.75	0.7403	1.29%	0.50	0.4897	2.06%
			0.20	0.1924	3.80%
0.50	0.4874	2.52%	0.80	0.7811	2.36%

Table 2.6: Comparisons of the conventional method and the separation method applied on the S-system (results show the average of 100 runs with various initial guesses).

	GNA_C		GNA_S	
σ	iterations	CPU time (s)	iterations	CPU time (s)
0.1	∼19	0.0676	∼15	0.0519
0.3	∼29	0.0933	∼21	0.0653
0.5	∼47	0.1341	∼25	0.0773
	LMA_C		LMA_S	
σ	iterations	CPU time (s)	iterations	CPU time (s)
0.1	∼36	0.0800	∼16	0.0454
0.3	∼75	0.1479	∼24	0.0596
0.5	∼102	0.1944	∼29	0.0670

conventional method. This can be interpreted as follows: the separation method does not need to calculate the refined values of linear parameters at each step during the optimization process.

The most obvious advantage of this separation strategy is that the parameter space is divided into two sub-spaces, thus transferring the original optimization problem to a lower-dimensional problem, and therefore reducing the computational effort. For example, suppose there are p linear parameters and q items in the nonlinear set, the conventional estimation process will deal with a $(p + q)$ dimensional nonlinear estimation. When using the separation procedure, as the optimal linear parameters can be obtained by a simple linear regression if estimates of those nonlinear items are provided, the original least squares estimation problem can be rewritten with respect to q nonlinear parameters only. A common nonlinear estimation method, such as the GNA or the LMA, can then be applied to this dimension-reduced problem.

2.5 Summary

Two parameter estimation methods are introduced in this chapter, specially designed for the S-system. The first is called the alternating least squares (ALS) method. It is primarily based on the two-term structure of the S-system model (production term and degradation term). When parameters in either term are assumed known already, the estimation problem can be easily solved by a logarithmic transformation and the linear regression. Applying this process iteratively to the production term and the degradation term will lead to estimates of all parameters within the S-system model.

The rest of this chapter mainly presents a separation method to estimate the parameters in biological system models which can be characterized by partial linearity with respect to some parameters, exemplified by the S-system. The simulation results have shown that with use of the proposed separation method, the parameter estimation process can be improved in the following aspects:

- There is no need to provide initial guess values for parameters of the linear portion in the system model, and thus the proposed method will reduce the

risk of estimation divergence when improper initial values are selected.

- The proposed method can reduce the dimension of the search space of parameters, resulting in reduction of the computation cost when using algorithms such as the GNA and the LMA.

- As the partially linear structure appears in many biological system models [22, 47], the proposed separation method can be applied to those cases.

Chapter 3

Parameter Estimation Methods for LFM

3.1 Introduction

3.1.1 Explanation of LFM

The dynamics of molecular biological systems are commonly modeled in the form of a set of ordinary differential equations that involve parameters corresponding to kinetic constants. As some models are often derived on the basis of statistical thermodynamics [46, 10] or Michaelis-Menten kinetics [27, 35], nonlinear functions in the resultant models are rational fractional functions whose numerator and denominator are linear in parameters. Parameters in the linear fractional functions in a molecular biological system are typically reaction constants of interest. Estimation of these parameters is crucial to constructing the whole molecular biological system model [12, 11].

In models of complex biological systems, one differential equation can contain several rational fractional rates and one reaction rate can be contained in several differential equations [10, 22, 43]. Such couplings among reaction rates and differential equations make it very difficult to estimate parameters in reaction rates directly from differential equations. Recently, we have proposed a methodology to independently estimate the rational fractional reaction rates [50, 49]. As a result, estimating the parameters in coupled models is reduced to estimating the parameters in independent linear fractional reaction rates. In general, the form of a linear fractional model

(LFM) is as follows:

$$\eta(X,\beta) = \frac{N_0(X) + \sum_{i=1}^{p_N} N_i(X)\beta_{N_i}}{D_0(X) + \sum_{j=1}^{p_D} D_j(X)\beta_{D_j}}, \tag{3.1}$$

where the vector X consists of observation variables, and the p-dimensional vector β consists of all parameters in the linear fractional function, which can naturally be divided into two groups: those in the numerator β_{N_i} $(i = 1, \ldots, p_N)$ and those in the denominator β_{D_j} $(j = 1, \ldots, p_D)$, where we have that $p_D + p_N = p$. The coefficient functions $N_i(X)$ $(i = 0, 1, \ldots, p_N)$ and $D_j(X)$ $(j = 0, 1, \ldots, p_D)$ are the known functions of the variables X and do not contain any unknown parameters. Either $N_0(X)$ or $D_0(X)$ must be nonzero; otherwise from sensitivity analysis [6] the parameters cannot be uniquely identified.

3.1.2 Related Work

From a literature search, there are seldom special estimation methods designed for such linear fractional models, to utilize some structure properties of the model. Conventional nonlinear estimation algorithms, such as the GNA and the LMA, were the main deterministic approaches applied for seeking the values of parameters within the LFM in the past. Matsubara *et al.* proposed two schemes for parameter estimation in metabolic pathways (model equations in the form of LFM derived from the Michaelis-Menten rate law) [24]. One scheme is using a genetic algorithm (GA), and the other is using a designed artifical neural network (ANN). Some stochastic optimization methods were also compared by Moles *et al.*, using a model formulated as Michaelis-Menten type equations [25].

3.1.3 New Ideas

As in the S-system, parameters in the LFM can also be separated into two groups: parameters in the numerator and those in the denominator. The LFM can be considered linear with respect to the group of parameters in the numerator and nonlinear with respect to those in the denominator. Therefore, the separation parameter estimation strategy can also be applied to the LFM. The procedure of implementing

this separation approach is almost the same as that for the S-system. Also, we try to modify the procedure a little bit and derive an iterative approach. Moreover, specifically for the LFM, we try to make the optimization cost function quadratic with respect to parameters to be sought, and therefore design a weighted least squares method. Through this way, the separable structure of the LFM is utilized and the estimation gets a closed-form analytical solution. The weighted least squares method was studied and adopted in the system identification community [41, 13], and we will try to apply it to parameter estimation problems.

3.1.4 Contents of Chapter

In this study, we make use of the special structure of the LFM as in Eq. (3.1): the numerator and the denominator are linear with respect to parameters. Briefly, the remainder of this chapter is organized as follows. In Section 3.2, the separation approach proposed for the S-system is applied here to the LFM, and a modified version with a numerical example is also included. Section 3.3 introduces a weighted least squares algorithm for the estimation of parameters in the LFM, denoted by LFM-WLS [1]. By designing a special weight matrix for the weighted least squares, parameters in the numerator and the denominator can be estimated by solving two linear least squares problems. Illustrative examples are provided to show the effectiveness of the proposed LFM-WLS algorithm. Finally we give summaries of the chapter in Section 3.4.

3.2 Separation Estimation Method

3.2.1 Ordinary Approach

Suppose that at a series of time points we obtain a sequence of measurements (observations) of the variable y_t $(t = 1, 2, \ldots, n)$, which can be represented by an LFM of

[1] This part of work has been published as in the journal paper [51].

variables X_t and parameter vector β. In practice, any measurements can be contaminated by some random noise. For simplicity, we assume that measurement errors are additive. Thus we have the relationships

$$y_t = \eta(X_t, \beta) + \varepsilon_t = \eta_t(\beta) + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{3.2}$$

where ε_t $(t = 1, 2, \ldots, n)$ stand for the measurement errors at time point t, and X_t $(t = 1, 2, \ldots, n)$ stand for the measured or known values of variables at time point t.

Define the two parameter vectors β_N and β_D for parameters in the numerator and in the denominator, respectively,

$$\begin{aligned} \beta_N &= [\beta_{N_1}, \beta_{N_2}, \ldots, \beta_{N_{p_N}}]^{\mathrm{T}} \in \mathbb{R}^{p_N}, \\ \beta_D &= [\beta_{D_1}, \beta_{D_2}, \ldots, \beta_{D_{p_D}}]^{\mathrm{T}} \in \mathbb{R}^{p_D}. \end{aligned}$$

Define the information vectors $\varphi_N(X_t)$ and $\varphi_D(X_t)$ as follows:

$$\begin{aligned} \varphi_N(X_t) &= [N_1(X_t), N_2(X_t), \ldots, N_{p_N}(X_t)] \in \mathbb{R}^{p_N}, \\ \varphi_D(X_t) &= [D_1(X_t), D_2(X_t), \ldots, D_{p_D}(X_t)] \in \mathbb{R}^{p_D}, \end{aligned}$$

and

$$\begin{aligned} Y &= [y(1), y(2), \ldots, y(n)]^{\mathrm{T}} \in \mathbb{R}^n, \\ \eta(\beta) &= [\eta_1(\beta), \eta_2(\beta), \ldots, \eta_n(\beta)]^{\mathrm{T}} \in \mathbb{R}^n, \end{aligned}$$

$$\Phi_N = \begin{bmatrix} \varphi_N(X_1) \\ \varphi_N(X_2) \\ \vdots \\ \varphi_N(X_n) \end{bmatrix} \in \mathbb{R}^{n \times p_N}, \qquad \Phi_{N_0} = \begin{bmatrix} N_0(X_1) \\ N_0(X_2) \\ \vdots \\ N_0(X_n) \end{bmatrix} \in \mathbb{R}^n,$$

$$\Phi_D = \begin{bmatrix} \varphi_D(X_1) \\ \varphi_D(X_2) \\ \vdots \\ \varphi_D(X_n) \end{bmatrix} \in \mathbb{R}^{n \times p_D}, \qquad \Phi_{D_0} = \begin{bmatrix} D_0(X_1) \\ D_0(X_2) \\ \vdots \\ D_0(X_n) \end{bmatrix} \in \mathbb{R}^n,$$

$$\Psi(\beta_D) \;=\; \mathrm{diag}\begin{bmatrix} D_0(X_1)+\varphi_D(X_1)\beta_D \\ D_0(X_2)+\varphi_D(X_2)\beta_D \\ \vdots \\ D_0(X_n)+\varphi_D(X_n)\beta_D \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

From the above definitions, we get

$$y_t = \frac{N_0(X_t)+\varphi_N(X_t)\beta_N}{D_0(X_t)+\varphi_D(X_t)\beta_D} + \varepsilon_t. \tag{3.3}$$

Construct the cost function as the sum of squared errors

$$J(\beta) = J(\beta_N, \beta_D) = [Y-\eta(\beta)]^{\mathrm{T}}[Y-\eta(\beta)]. \tag{3.4}$$

Minimizing $J(\beta)$ with respect to β can give the least squares estimation of parameters β_N and β_D.

As the LFM $\eta(\beta)$ shows nonlinearity with respect to the parameter vector β, the Gauss-Newton iteration method and its variants [6] can typically be applied to estimation of these parameters by minimizing the cost function (3.4). However, it is well known that the Gauss-Newton method may fall into a local minimum and thus cannot find the correct estimates of the parameters. Also, when the number of parameters is large, it will take quite a long time to seek the estimates using the Gauss-Newton method. We have observed from the model structure that the parameters β_N are linear in the model $\eta(X,\beta)$. Let $F(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_N$ and $G(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_{N_0}$. The cost function (3.4) becomes

$$J(\beta) = J(\beta_N, \beta_D) = [Y-G(\beta_D)-F(\beta_D)\beta_N]^{\mathrm{T}}[Y-G(\beta_D)-F(\beta_D)\beta_N]. \tag{3.5}$$

Now the cost function has the structure similar to that of Eq. (2.17), and therefore we can apply the separation estimation method, which is stated in Section 2.4, to the LFM here. The estimation procedure includes three steps. Firstly, for given values of $\beta_D = \bar{\beta}_D$, by the ordinary least squares method, the estimates of β_N can be computed as

$$\hat{\beta}_N = [F(\bar{\beta}_D)^{\mathrm{T}}F(\bar{\beta}_D)]^{-1}F(\bar{\beta}_D)^{\mathrm{T}}[Y-G(\bar{\beta}_D)]. \tag{3.6}$$

Substitute Eq. (3.6) into Eq. (3.5) and we can get

$$J(\bar{\beta}_D) = [Y - G(\bar{\beta}_D)]^{\mathrm{T}}\{I - F(\bar{\beta}_D)[F(\bar{\beta}_D)^{\mathrm{T}}F(\bar{\beta}_D)]^{-1}F(\bar{\beta}_D)^{\mathrm{T}}\}[Y - G(\bar{\beta}_D)]. \quad (3.7)$$

Secondly, find the optimal values $\hat{\beta}_D$ of the parameter set β_D, using the Gauss-Newton, the Levenberg-Marquardt or other methods. Finally, with the optimal $\hat{\beta}_D$ obtained, estimates for β_N can be computed from Eq. (3.6).

3.2.2 Numerical Example

The expression of a gene is regulated by regulatory proteins and/or RNA polymerase (RNAP) which are binding to the gene's regulatory binding sites [46]. The regulatory binding site is a short sequence of DNA close to associated genes. One gene can have a number of binding sites. The binding sites for regulatory proteins are called operators while those for RNAP are called promoters. A gene regulatory network is composed of such a collection of genes, whose expression rates are regulated by each other. A protein encoded by each gene acts in the role of a regulator, affecting the rates during this biochemical process. To illustrate the proposed estimation algorithm, this section will consider the parameter identification of a simple gene regulatory network, with one gene, two operators and one promoter as shown in Fig. 3.1. Its corresponding mathematical model is described by an ODE as shown in Eq. (3.8).

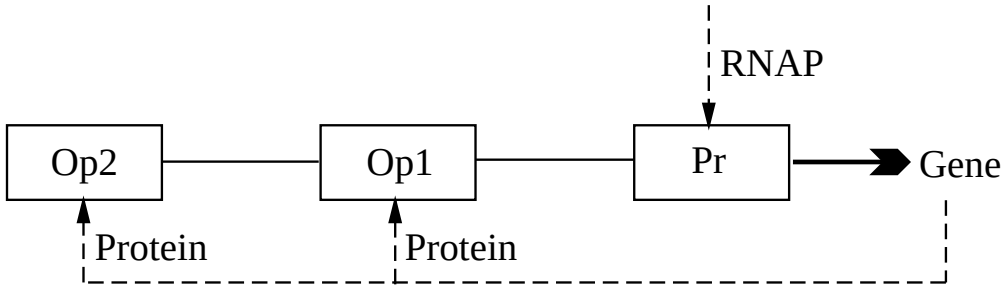

Figure 3.1: A gene regulatory network with one gene, two operators (Op1 and Op2) and one promoter (Pr) (redrawn from [46])

Based on the statistical thermodynamic theory and the biochemical kinetics, the

model of this network can be expressed as follows

$$\dot{x} = \frac{a_0 + a_1 x + a_2 x^2}{1 + b_1 x + b_2 x^2} - \lambda x, \tag{3.8}$$

where x is the concentration of protein encoded by the gene, a_i $(i = 0, 1, 2)$ and b_i $(i = 1, 2)$ are positive constants related to the biochemical kinetics and λ is a positive constant representing the protein degradation rate. One can see that model (3.8) is a linear fractional model with positive parameters λ, a_i and b_i. Note that the expression of equation is slightly different from the one in reference [46]. In order to uniquely identify parameters in the model, according to [6], we rescale the parameters such that the constant term in the denominator is 1.

The nominal values of parameters are set as: $a_0 = 0.4, a_1 = 2.8, a_2 = 0.24, b_1 = 0.5, b_2 = 1.4, \lambda = 0.4$. In this example, we use the nominal values to generate the time-course of $x(t)$ shown as in Fig. 3.2, with the sampling interval 0.1 min. The time starts at $t = 0$. From Fig. 3.2, system (3.8) is stable at its steady state $x^\star = 2.18$ after $t = 10$ min. Therefore, we only use the synthetic data on the interval $0 \sim 10$ min.

In Eq. (3.8), the RHS is a linear fractional function minus a term linear in one parameter, which is not in the same format as in Eq. (3.1). We can transform it into the following form

$$\dot{x} = \frac{a_0 + (a_1 - \lambda)x + (a_2 - b_1\lambda)x^2 - b_2\lambda x^3}{1 + b_1 x + b_2 x^2}, \tag{3.9}$$

although the numerator is no longer linear in the 6 original parameters in model (3.9). Nonetheless, if we regard a single coefficient as a new parameter, the numerator in model (3.9) is linear in the new parameters. In addition, there are 6 unknown parameters in 6 new coefficients in model (3.9). Therefore, using our proposed method, the 6 original parameters can be uniquely identified.

In this study, a group of artificial/synthetic data is generated from the model of gene regulatory networks, with nominal parameter values and initial states provided. There is no noise added on the artificial data in the simulation, so they can be considered as noise-free measurements. The 5-point central differences formula is adopted

to approximate derivatives of state variables, as in Eq. (2.3). After obtaining derivatives at different sampling points, we can apply the proposed method to estimating parameters in the model. The relative estimation error (REE) is employed to gauge the performance of estimation accuracy, defined the same as in Eq. (2.9).

Simulations are conducted using both the conventional LMA method and the proposed separation method to make a comparison. Initial values are randomly selected from the range with true values plus a relative Gaussian distribution, i.e. true_values $\cdot (1 + \sigma \cdot \varepsilon)$, where ε follows the standard normal distribution and σ is the standard deviation. Comparison results are shown as listed in Table 3.1. When it reaches convergence, each run of 100 trials achieves the same optimal results: the minimum of the cost function is 5.9027×10^{-6}. The proposed separation method shows a faster speed and a higher convergence percentage, especially when the initial guess is selected from a wider range.

Table 3.1: Comparisons of the conventional method and the separation method applied on the LFM (results show the average of 100 runs with various initial guesses).

	LMA_C			LMA_S		
σ	iterations	CPU time (s)	not converge	iterations	CPU time (s)	not converge
0.1	∼6	0.0172	0%	∼2	0.0106	0%
0.3	∼12	0.0267	2%	∼2	0.0112	0%
0.5	∼17	0.0348	3%	∼3	0.0122	0%
1.0	∼26	0.0476	22%	∼4	0.0261	1%

The relative estimation errors are listed in Table 3.2 (the REE for the whole parameter vector is 0.0219). From this table, the estimation results are generally very good, except the estimate for a_2 somewhat deviates from the true value. One possible explanation is that the estimate of a_2 is calculated indirectly from the new coefficient $(a_2 - b_1\lambda)$, and the effect of estimation errors from b_1 and λ magnifies in the process.

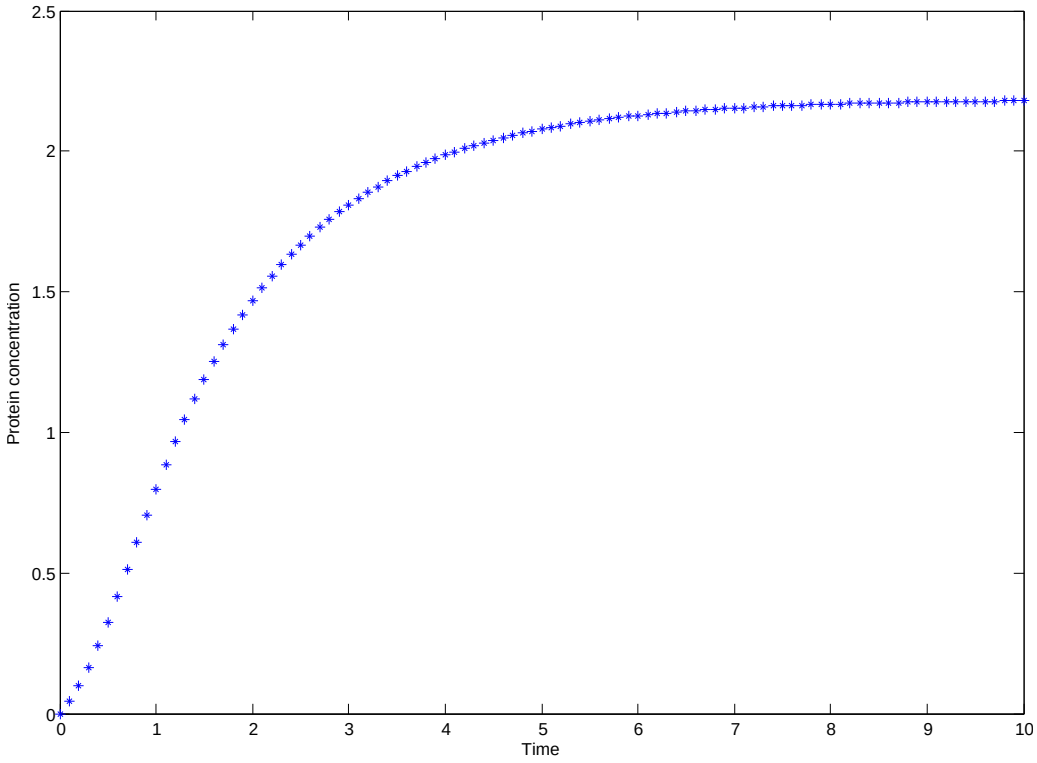

Figure 3.2: State profile of model (3.8)

Table 3.2: Estimation results using the separation method on the LFM

True Value	Estimate	REE	True Value	Estimate	REE
0.40	0.4002	0.0467%	2.80	2.7989	0.0381%
0.24	0.1801	24.94%	0.50	0.5032	0.64%
1.40	1.3630	2.64%	0.40	0.3926	1.86%

3.2.3 Modified Iterative Approach

We have obtained the expression of cost function as shown in Eq. (3.5). From Theorem 2.1 in reference [28], if there exists an $(n-k) \times n$ matrix, say $X_1(\beta_D)$, having the rank of $(n-k)$ and satisfying

$$X_1(\beta_D)F(\beta_D) = 0, \tag{3.10}$$

then

$$\min_{\beta_N, \beta_D} J(\beta_N, \beta_D) = \min_{\beta_D} K(\beta_D), \tag{3.11}$$

where the equivalent cost function $K(\beta_D)$, with

$$\begin{aligned} K(\beta_D) &= [X_1(\beta_D)(Y - G(\beta_D))]^{\mathrm{T}} \cdot \\ &\quad [X_1(\beta_D)X_1(\beta_D)^{\mathrm{T}}]^{-1} X_1(\beta_D)(Y - G(\beta_D)), \end{aligned} \tag{3.12}$$

is independent of parameters in the numerator, i.e., β_N. Let $\hat{\beta}_D$ be the optimizer of Eq. (3.12), and then $\hat{\beta}_D$ and $\hat{\beta}_N$ will minimize the cost function (3.5), where $\hat{\beta}_N$ is calculated as follows:

$$\hat{\beta}_N = [F(\hat{\beta}_D)^{\mathrm{T}} F(\hat{\beta}_D)]^{-1} F(\hat{\beta}_D)^{\mathrm{T}} [Y - G(\hat{\beta}_D)]. \tag{3.13}$$

Let $\Phi_N^{\perp}$ be an $(n-k) \times n$ matrix, with the rank of $(n-k)$, and orthogonal to the matrix Φ_N. Then we can construct

$$X_1(\beta_D) = \Phi_N^{\perp} \Psi(\beta_D). \tag{3.14}$$

From the definition of $\Psi(\beta_D)$ we have

$$\begin{aligned} \Psi(\beta_D)(Y - G(\beta_D)) &= \mathrm{diag}[Y]\Phi_{D_0} - \Phi_{N_0} + \mathrm{diag}[Y]\Phi_D \beta_D \\ &= b + A\beta_D, \end{aligned} \tag{3.15}$$

where the constant vector $b = \mathrm{diag}[Y]\Phi_{D_0} - \Phi_{N_0} \in \mathbb{R}^n$, and the constant matrix $A = \mathrm{diag}[Y]\Phi_D \in \mathbb{R}^{n \times p_D}$, both of which are independent of the unknown parameters.

Substituting Eqs. (3.14) and (3.15) into Eq. (3.12) yields

$$K(\beta_D) = [b + A\beta_D]^{\mathrm{T}} {\Phi_N^{\perp}}^{\mathrm{T}} M(\beta_D)^{-1} \Phi_N^{\perp} [b + A\beta_D], \tag{3.16}$$

where the matrix $M(\beta_D) = \Phi_N^{\perp}\Psi(\beta_D)\Psi(\beta_D)^{\mathrm{T}}{\Phi_N^{\perp}}^{\mathrm{T}}$. The necessary condition for minimizing $K(\beta_D)$ with respect to β_D is that $\partial K(\beta_D)/\partial\beta_D = 0$, which gives

$$\begin{aligned} A^{\mathrm{T}}{\Phi_N^{\perp}}^{\mathrm{T}}M(\beta_D)^{-1}\Phi_N^{\perp}b + A^{\mathrm{T}}{\Phi_N^{\perp}}^{\mathrm{T}}M(\beta_D)^{-1}\Phi_N^{\perp}A\beta_D \\ -U(\beta_D)^{\mathrm{T}}u_0(\beta_D) - U(\beta_D)^{\mathrm{T}}U(\beta_D)\beta_D = 0, \end{aligned} \tag{3.17}$$

where the matrix $U = [u_1, \ldots, u_{p_D}]$ and its i-th column vector is defined as

$$u_i = \mathrm{diag}[\Phi_{D_i}]{\Phi_N^{\perp}}^{\mathrm{T}}M(\beta_D)^{-1}\Phi_N^{\perp}[b + A\beta_D] \in \mathbb{R}^n, \quad \text{for } i = 0, 1, \ldots, p_D. \tag{3.18}$$

From Eq. (3.17), we propose an iteration formula to solve the optimization problem (3.11) as follows:

$$\begin{aligned} \beta_D^{k+1} \;=\; & [A^{\mathrm{T}}{\Phi_N^{\perp}}^{\mathrm{T}}M(\beta_D^k)^{-1}\Phi_N^{\perp}A]^{-1} \cdot \\ & [U(\beta_D^k)^{\mathrm{T}}u_0(\beta_D^k) + U(\beta_D^k)^{\mathrm{T}}U(\beta_D^k)\beta_D^k - A^{\mathrm{T}}{\Phi_N^{\perp}}^{\mathrm{T}}M(\beta_D^k)^{-1}\Phi_N^{\perp}b]. \end{aligned} \tag{3.19}$$

Select an initial starting point for the parameter vector β_D, and the optimal estimate could be obtained by using the iteration formula (3.19). After that, the optimal estimate for β_N will be easily accomplished via Eq. (3.13).

3.2.4 Numerical Example

Here we still adopt the simple one gene regulatory network as an example (the same one of Section 3.2.2). The synthetic data are generated with nominal parameter values and the sampling frequency is set as 100 min^{-1}. In the following, we mainly compare the proposed iterative method with a method directly minimizing the objective function (3.12), i.e., the Matlab embedded function `fminsearch()` which is directly called in the optimization. These two methods only need the initial values of parameters in the denominator of the LFM. The initial values of two parameters in the denominator are initialized by true_values$\cdot(1+\sigma\cdot\varepsilon)$ with $\sigma = 1$. The results [2] are listed in Table 3.3, showing the average over 20 runs. 'CPU time' is the average running time, 'REE' stands for the minimum relative estimation error for the whole parameter vector, and 'Robustness' is the percentage of runs converging with the minimum REE.

[2] Some of the material was presented in the conference paper [48].

From Table 3.3, the proposed method shows the same estimation accuracy as the direct method when both methods converge to the minimum of cost function. However, the proposed method uses much less CPU time to converge and is more robust (insensitive) to the initial values than the direct method.

3.3 Weighted Least Squares Method

3.3.1 Case A: with a Single Dependent Variable

Take the simple one gene regulatory network as the analysis target, which has only one dependent variable in the model. We can form the cost function as the sum of the weighted squared errors

$$J(\beta) = J(\beta_N, \beta_D) = [Y - \eta(\beta)]^{\mathrm{T}} W(\beta_D)[Y - \eta(\beta)], \tag{3.20}$$

where $W(\beta_D)$ is a weight matrix. Minimizing $J(\beta)$ can give the least squares estimation of parameters β_N and β_D.

We observe that the parameters β_N are linear in the model $\eta(X, \beta)$. Let $F(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_N$ and $G(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_{N_0}$, and the cost function (3.20) becomes

$$\begin{aligned} J(\beta) &= J(\beta_N, \beta_D) \\ &= [Y - G(\beta_D) - F(\beta_D)\beta_N]^{\mathrm{T}} W(\beta_D)[Y - G(\beta_D) - F(\beta_D)\beta_N]. \end{aligned} \tag{3.21}$$

From optimization principles [7], we have

$$\min_{\beta} J(\beta) = \min_{\beta_N, \beta_D} J(\beta_N, \beta_D) = \min_{\beta_D} \min_{\beta_N} J(\beta_N, \beta_D). \tag{3.22}$$

Table 3.3: Comparisons of the proposed iterative method and the direct method applied on the LFM (results show the average of 20 runs with various initial guesses).

	CPU time (s)	REE	Robustness
Proposed iterative method	3.3657	2.36%	90%
Direct method `fminsearch()`	26.3656	2.36%	80%

Therefore, minimizing the cost function (3.21) with respect to β can be achieved by solving two optimization problems: one with respect to β_N and the other with respect to β_D. Furthermore, given that $\beta_D = \bar{\beta}_D$, the estimation of parameter β_N becomes minimizing the following cost function

$$\begin{aligned} J(\beta) &= J(\beta_N, \bar{\beta}_D) \\ &= [Y - G(\bar{\beta}_D) - F(\bar{\beta}_D)\beta_N]^{\mathrm{T}} W(\bar{\beta}_D)[Y - G(\bar{\beta}_D) - F(\bar{\beta}_D)\beta_N], \end{aligned} \quad (3.23)$$

which is quadratic in β_N. Therefore, from the standard linear least squares method, we obtain the estimates of β_N as follows:

$$\hat{\beta}_N(\bar{\beta}_D) = [F(\bar{\beta}_D)^{\mathrm{T}} W(\bar{\beta}_D) F(\bar{\beta}_D)]^{-1} F(\bar{\beta}_D)^{\mathrm{T}} W(\bar{\beta}_D)[Y - G(\bar{\beta}_D)]. \quad (3.24)$$

Substituting Eq. (3.24) into Eq. (3.23) yields

$$\begin{aligned} J(\bar{\beta}_D) &= J(\hat{\beta}_N, \bar{\beta}_D) \\ &= [Y - G(\bar{\beta}_D) - F(\bar{\beta}_D)\hat{\beta}_N]^{\mathrm{T}} W(\bar{\beta}_D)[Y - G(\bar{\beta}_D) - F(\bar{\beta}_D)\hat{\beta}_N] \\ &= [Y - G(\bar{\beta}_D)]^{\mathrm{T}} \{W(\bar{\beta}_D) - W(\bar{\beta}_D)F(\bar{\beta}_D)[F(\bar{\beta}_D)^{\mathrm{T}} W(\bar{\beta}_D) F(\bar{\beta}_D)]^{-1} \cdot \\ &\quad F(\bar{\beta}_D)^{\mathrm{T}} W(\bar{\beta}_D)\}[Y - G(\bar{\beta}_D)]. \end{aligned} \quad (3.25)$$

However, the cost function $J(\bar{\beta}_D)$ in Eq. (3.25) is not quadratic in $\bar{\beta}_D$ for an arbitrary weight matrix $W(\beta_D)$, and thus minimizing it is a nonlinear optimization problem.

In this study, we design the weight matrix $W(\beta_D) = \Psi(\beta_D)^{\mathrm{T}}\Psi(\beta_D)$. Using $F(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_N$ and $G(\beta_D) = [\Psi(\beta_D)]^{-1}\Phi_{N_0}$, the cost function (3.25) becomes

$$J(\bar{\beta}_D) = [\Psi(\bar{\beta}_D)Y - \Phi_{N_0}]^{\mathrm{T}} \{I - \Phi_N[\Phi_N^{\mathrm{T}}\Phi_N]^{-1}\Phi_N^{\mathrm{T}}\}[\Psi(\bar{\beta}_D)Y - \Phi_{N_0}], \quad (3.26)$$

and the estimation of β_N in Eq. (3.24) becomes

$$\hat{\beta}_N(\bar{\beta}_D) = [\Phi_N^{\mathrm{T}}\Phi_N]^{-1}\Phi_N^{\mathrm{T}}[\Psi(\bar{\beta}_D)Y - \Phi_{N_0}]. \quad (3.27)$$

As $\Psi(\beta_D)$ is linear in β_D from its definition, the cost function (3.26) is quadratic in β_D. Furthermore, from the definition of $\Psi(\beta_D)$ we have

$$\Psi(\bar{\beta}_D)Y = \mathrm{diag}[Y]\Phi_{D_0} + \mathrm{diag}[Y]\Phi_D\bar{\beta}_D. \quad (3.28)$$

Substituting Eq. (3.28) into Eq. (3.26) yields

$$\begin{aligned} J(\bar{\beta}_D) &= [\Psi(\bar{\beta}_D)Y - \Phi_{N_0}]^{\mathrm{T}}\{I - \Phi_N[\Phi_N^{\mathrm{T}}\Phi_N]^{-1}\Phi_N^{\mathrm{T}}\}[\Psi(\bar{\beta}_D)Y - \Phi_{N_0}] \\ &= [b + A\bar{\beta}_D]^{\mathrm{T}}U[b + A\bar{\beta}_D], \end{aligned} \tag{3.29}$$

where $b = \mathrm{diag}[Y]\Phi_{D_0} - \Phi_{N_0} \in \mathbb{R}^n$, $A = \mathrm{diag}[Y]\Phi_D \in \mathbb{R}^{n\times p_D}$, and $U = I - \Phi_N[\Phi_N^{\mathrm{T}}\Phi_N]^{-1}\Phi_N^{\mathrm{T}} \in \mathbb{R}^{n\times n}$, which are constant vectors or matrices.

Minimizing the cost function (3.29) gets the estimation of $\bar{\beta}_D$ as follows

$$\hat{\beta}_D = -[A^{\mathrm{T}}UA]^{-1}A^{\mathrm{T}}Ub. \tag{3.30}$$

Substituting Eq. (3.30) into Eq. (3.27) yields

$$\hat{\beta}_N = [\Phi_N^{\mathrm{T}}\Phi_N]^{-1}\Phi_N^{\mathrm{T}}[b + A\hat{\beta}_D]. \tag{3.31}$$

From the above derivation, by designing the weight matrix $W(\beta_D) = \Psi(\beta_D)^{\mathrm{T}}\Psi(\beta_D)$, minimizing the cost function (3.21) is reduced to solving two linear least squares problems (3.23) and (3.29). Furthermore, the minimizers (i.e., the estimates of the parameters) of the cost function (3.21) can be analytically expressed in Eqs. (3.30) and (3.31).

3.3.2 Example A

Take the simple one gene regulatory network as a numerical test example, as shown in Section 3.2.2. Simulations are conducted using both the proposed method LFM-WLS and the conventional Gauss-Newton algorithm (GNA) to make a comparison. The most obvious advantage of the proposed method is that there is no need to provide initial guess values for parameters. For GNA, the initial values are selected from the range with true values plus a relative Gaussian distribuation, i.e., true_values$\cdot(1+\sigma\cdot\varepsilon)$, where ε follows the standard normal distribution and σ is the standard deviation. We have tried to take the initial guess with a large value of the standard deviation. However, even with $\sigma = 0.5$, the GNA sometimes could not converge. In this study, the standard deviation is chosen as 0.3. The performance of the GNA shown in the comparison table is the average of 100 runs with randomly selected initial values.

Table 3.4 presents the comparison of performance between the proposed method LFM-WLS and the GNA. The proposed method shows good estimation accuracy compared to the GNA in terms of the relative estimation error. Considering that multiple runs are needed when using the GNA to determine the optimal parameter values, the proposed method LFM-WLS will cost less time in total than the GNA method.

3.3.3 Case B: with Multiple Dependent Variables

Sometimes we may obtain a sequence of measurements (observations) of multiple dependent variables: $y_k \in \mathbb{R}^n$ $(k = 1, 2, \ldots, m)$, which can be represented by the LFM of variables X_t and parameters β, and all equations in the LFM have the same denominator. In this case,

$$y_{kt} = \eta_k(X_t, \beta) + \varepsilon_{kt} = \frac{N_{k0}(X_t) + \varphi_{kN}(X_t)\beta_{kN}}{D_0(X_t) + \varphi_D(X_t)\beta_D} + \varepsilon_{kt}, \\ k = 1, 2, \ldots, m, \quad t = 1, 2, \ldots, n. \tag{3.32}$$

In these equations we assume that parameter vectors β_{kN} $(k = 1, \ldots, m)$ are independent. The definitions of $N_{k0}(X_t)$ and $\varphi_{kN}(X_t)$ for the k-th equation are similar to those in Case A and Section 3.2.1. The sum of weighted squared errors (the cost function) for this case becomes

$$\begin{aligned} J(\beta) &= J(\beta_{1N}, \ldots, \beta_{mN}, \beta_D) \\ &= \sum_{k=1}^{m} [Y_k - \eta_k(\beta)]^{\mathrm{T}} W(\beta_D) [Y_k - \eta_k(\beta)]. \end{aligned} \tag{3.33}$$

Table 3.4: Comparisons of the proposed method and the GNA in Example A

Parameter	REE_LFM-WLS	REE_GNA	Parameter	REE_LFM-WLS	REE_GNA
a_0	0.76%	0.0469%	a_1	1.38%	0.0384%
a_2	3.78%	24.93%	b_1	4.91%	0.64%
b_2	0.0732%	2.64%	λ	0.0532%	1.86%

Similar to the derivation in Case A, we will have

$$\hat{\beta}_{kN}(\bar{\beta}_D) = [\Phi_{kN}^{\mathrm{T}}\Phi_{kN}]^{-1}\Phi_{kN}^{\mathrm{T}}[\Psi(\bar{\beta}_D)Y_k - \Phi_{kN_0}], \tag{3.34}$$

and

$$J(\bar{\beta}_D) = \sum_{k=1}^{m}[b_k + A_k\bar{\beta}_D]^{\mathrm{T}}U_k[b_k + A_k\bar{\beta}_D], \tag{3.35}$$

where $b_k = \mathrm{diag}[Y_k]\Phi_{D_0} - \Phi_{kN_0} \in \mathbb{R}^n$, $A_k = \mathrm{diag}[Y_k]\Phi_D \in \mathbb{R}^{n\times p_D}$, and $U_k = I - \Phi_{kN}[\Phi_{kN}^{\mathrm{T}}\Phi_{kN}]^{-1}\Phi_{kN}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ are constant vectors or matrices. Minimizing the cost function (3.35) gets the estimation of $\bar{\beta}_D$ as follows

$$\hat{\beta}_D = -[\sum_{k=1}^{m} A_k^{\mathrm{T}}U_kA_k]^{-1}\sum_{k=1}^{m} A_k^{\mathrm{T}}U_kb_k. \tag{3.36}$$

Substituting Eq. (3.36) into Eq. (3.34) yields

$$\hat{\beta}_{kN} = [\Phi_{kN}^{\mathrm{T}}\Phi_{kN}]^{-1}\Phi_{kN}^{\mathrm{T}}[b_k + A_k\hat{\beta}_D]. \tag{3.37}$$

3.3.4 Example B

Here we take a gene regulatory network with two genes, one operator and two promoters as a test example. This gene regulatory network is shown in Fig. 3.3, which is a simplified version of the Lambda phage switch topology [33, 46]. In this network, it is assumed that promoters (Pr1 and Pr2) can be occupied by RNAP and that the operator (Op) can be in any of three states: either empty, occupied by protein x, or occupied by protein y. There are also some additional assumptions: 1) proteins x and y cannot bind to Op at the same time; 2) if protein x is bound to Op, then RNAP cannot bind to Pr2; and similarly 3) in case that protein y is bound to Op, RNAP cannot bind to Pr1.

Based on the statistical thermodynamic theory and the biochemical kinetics, the model of this network can be expressed as a group of differential equations

$$\begin{aligned}\dot{x} &= \frac{a_0 + a_1x}{1 + b_1x + b_2y} - \lambda_1x,\\ \dot{y} &= \frac{c_0 + c_1y}{1 + b_1x + b_2y} - \lambda_2y,\end{aligned} \tag{3.38}$$

where x and y are the concentrations of proteins encoded by gene X and gene Y, respectively, b_i $(i = 1, 2)$, a_i and c_i $(i = 0, 1)$ are constants related to the biochemical kinetics and λ_i $(i = 1, 2)$ are constants representing the degradation rates of proteins. For the same reason as in Example A, we rescale the parameters such that the constant term in the denominator is 1. Again, two equations in model (3.38) are also not in the same format as model (3.1). The model (3.38) can be rewritten in the form of the linear factional model (LFM), to utilize our proposed method to estimate parameters in the LFM.

The nominal values of parameters are selected as provided in reference [46]: $a_0 = 0.4, a_1 = 2.8, b_1 = 0.5, b_2 = 1, c_0 = 0.4, c_1 = 3.6, \lambda_1 = 0.4, \lambda_2 = 0.4$. In this example, nominal values are adopted to generate time profiles of $x(t)$ and $y(t)$ with the sampling interval 0.1 min. The time-courses of concentrations of proteins x and y are shown in Fig. 3.4. For parameter estimation we just use the data during $0 \sim 50$ min, as after that the system stays with its stable state: $x^\star = 0.47$, $y^\star = 7.89$.

Table 3.5 presents the comparison of performance between the proposed method LFM-WLS and the GNA. The proposed method again gives a satisfactory estimation accuracy compared to the GNA in terms of the relative estimation error.

3.4 Summary

In this chapter two estimation methods are proposed for the linear fractional model (LFM). The first method is based on the partially linear structure of the LFM and

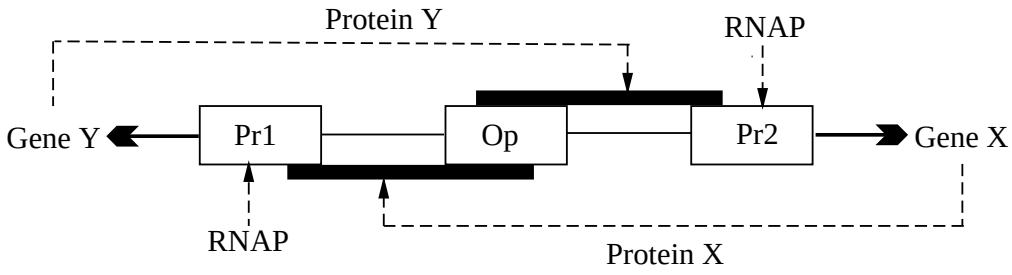

Figure 3.3: A gene regulatory network with two genes (X and Y), one operator (Op) and two promoters (Pr1 and Pr2) (redrawn from [46])

the separation principle. The idea and the procedure are almost the same as that stated in Section 2.4, designed for the S-system. Furthermore, a novel iterative approach is proposed and it gives good performance.

In the following part, we present a weighted least squares method for estimating the parameters in gene regulatory networks with linear fractional reaction rates. The presented method has made use of the special structure of the linear fractional models: both the denominator and the numerator are linear in parameters respectively. By designing a weight matrix, estimating parameters in the linear fractional functions, which are essentially nonlinear in parameters, is transformed into solving two linear least squares problems. The highlight is the estimates of parameters can be analytically expressed. Compared to the traditional Gauss-Newton method and its variants, the presented method does not need any initial guess of parameters. Therefore, with this method there is no worry about problems with the Gauss-Newton method such as the sensitivity to initial guess values. In fact, two illustrated examples have also shown that the presented method gives a good performance compared with the traditional Gauss-Newton method in terms of the relative estimation errors.

In models (3.8) and (3.38), the RHS of the equation is one linear fractional function minus one linear term. If there is more than one linear term or more than one linear fractional function with different denominators, problems become more complicated. As a consequence, the parameters in such a model may not be directly identified using the presented method. One direction of future work is to develop a method to identify the parameters in the models with the above-mentioned structures. Although illustrated on simple gene regulatory networks in this study, the presented methods can be applied to many other biological systems which are on the basis of biochemical kinetics such as metabolic networks and complex gene regulatory networks. Another direction of future work is to apply the presented method to some more complex biological networks.

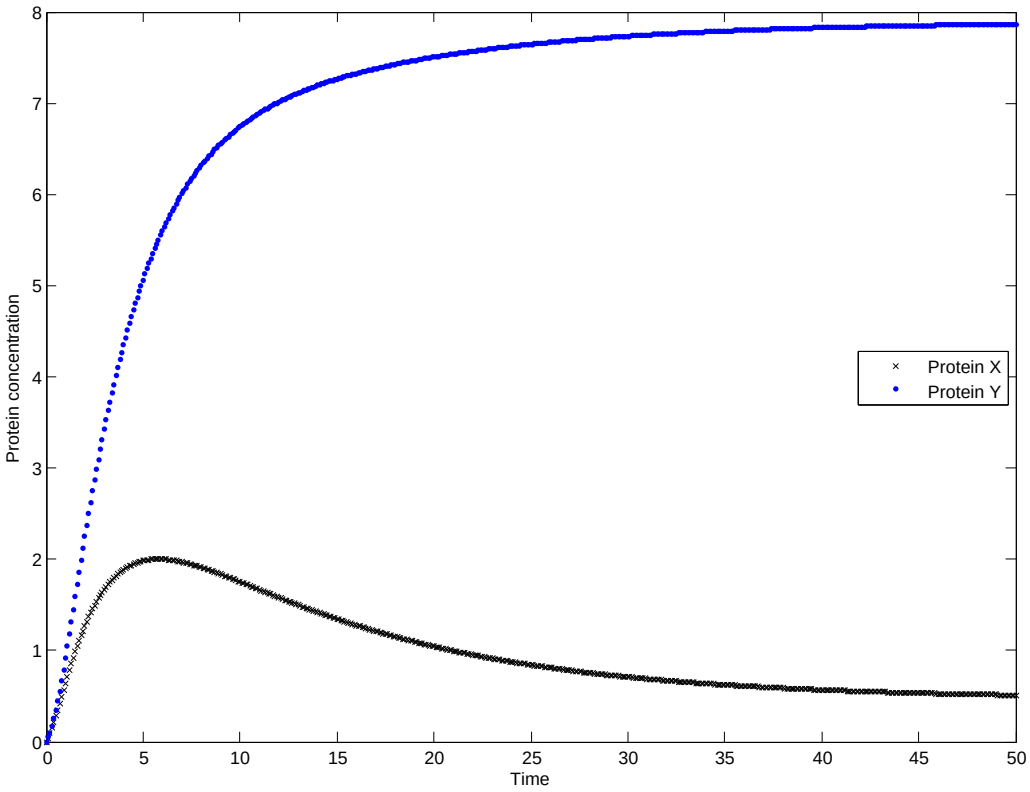

Figure 3.4: State profiles of model (3.38)

Table 3.5: Comparisons of the proposed method and the GNA in Example B

Parameter	REE_LFM-WLS	REE_GNA	Parameter	REE_LFM-WLS	REE_GNA
a_0	0.21%	0.11%	a_1	0.78%	0.22%
b_1	2.47%	0.50%	b_2	0.66%	0.22%
c_0	0.60%	0.21%	c_1	0.75%	0.21%
λ_1	0.0793%	0.0538%	λ_2	0.18%	0.0949%

Chapter 4

More Complex Case: with State and Reaction Rate Estimation

4.1 Introduction

To estimate the parameters in the reaction rate functions, it requires the measurement of the reaction rates and all components (states) at a series of time points. Generally, it is difficult (or costly) to measure all states in a biological network at all time points in an experiment. Furthermore the reaction rates in a biological system are usually unmeasurable. Therefore, it is very important to design experiments such that unmeasured states and all reaction rates can be uniquely estimated and thus as many as possible parameters in the reaction rates can be estimated.

Recently, Gadkar *et al.* proposed a method to select the optimal measurement set of states using the Fisher Information Matrix (FIM) along with the Cramer-Rao theorem, and to estimate all unmeasured states and reaction rates using a state regulator problem (SRP) formulation [12]. In their method, to calculate the FIM, one must know the nominal values of all parameters, which may not be available in practice. In addition, the reaction rates in the model are not linearly independent and thus cannot be estimated uniquely. Therefore, the estimates may not be as accurate as they appeared, and the number of identifiable parameters may be greater than that in their work.

In this chapter, we propose a new methodology for estimating states and reaction rates in the caspase-activated apoptosis system. The core of this chapter is to present the methodology of estimating unmeasured states and reaction rates. Only after this

procedure we can conduct the estimation of parameters within the reaction model. The remainder of Chapter 4 is organized as follows: In Section 4.2 we present the framework of our methodology which consists of five modules. In Sections 4.3, each module is explained in detail. In Section 4.4, we use the synthetic data to illustrate the proposed methodology. Some conclusions and summaries are given in Section 4.5.

4.2 Framework of Methodology

The framework of the proposed methodology is shown in Fig. 4.1: 1) describing the caspase-activated apoptosis system and its model which is a group of nonlinear differential equations; 2) presenting a method to analyze the complexity of the system using linear algebra; 3) proposing the optimal experiment design aimed at measuring as few as possible system states and identifying as many as possible parameters; 4) estimating the unmeasured states and reaction rates; 5) presenting a separation method to estimate the parameters in the system model.

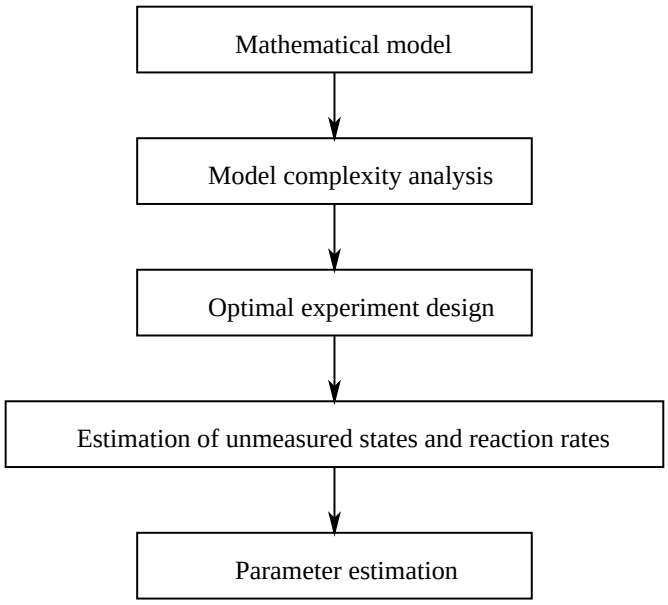

Figure 4.1: Proposed framework for model development using less experimental data

The proposed methodology does not need the nominal values of the parameters.

Furthermore, by combining some reaction rates together through the complexity analysis, we obtain a group of combined reaction rates, which are linearly independent and can be uniquely estimated. In addition, we introduce a separation method to estimate parameters in the fractional reaction rates, which makes use of the special structure of fractional functions: parameters show linearity in the numerator and the denominator respectively, although they show nonlinearity in the whole fractional functions.

4.3 Example and Implementation of Methodology

4.3.1 Caspase Model

Apoptosis, which means programmed cell death, is a biological process carried out in living organisms. The process is in the form of cell suicide, to prevent malfunctions caused by cell stress, damage or conflicting division signals [10]. A balance is sought and maintained between cell perishing and growing, which is quite important for the organisms to function properly. On one hand, failure to respond to apoptotic signals is partly the reason why some cancers are so hard to eliminate. While on the other hand, excessive apoptotic activities would cause some neuron degenerative disorder diseases [37, 17]. A family of 19 proteases, named caspases (cysteine containing aspartate-specific proteases), is most important to the mechanism of apoptotic cell death, as shown in Fig. 4.2 [11].

In the apoptosis system, caspases are thought to be an important player in the execution process of apoptosis, and are present as inactive proteins under normal circumstances. Once activated, these caspases will cleave key protein targets, resulting in their activation, and consequently a caspase cascade. Therefore, understanding the caspase activation cascade process is helpful not only in learning the mechanism of cancers and other autoimmune diseases, but also in promoting the development of anti-cancer drug research. Based on the caspase-dependent apoptosis mechanism as shown in Fig. 4.2, a state-space model was proposed for this process [10, 11, 12],

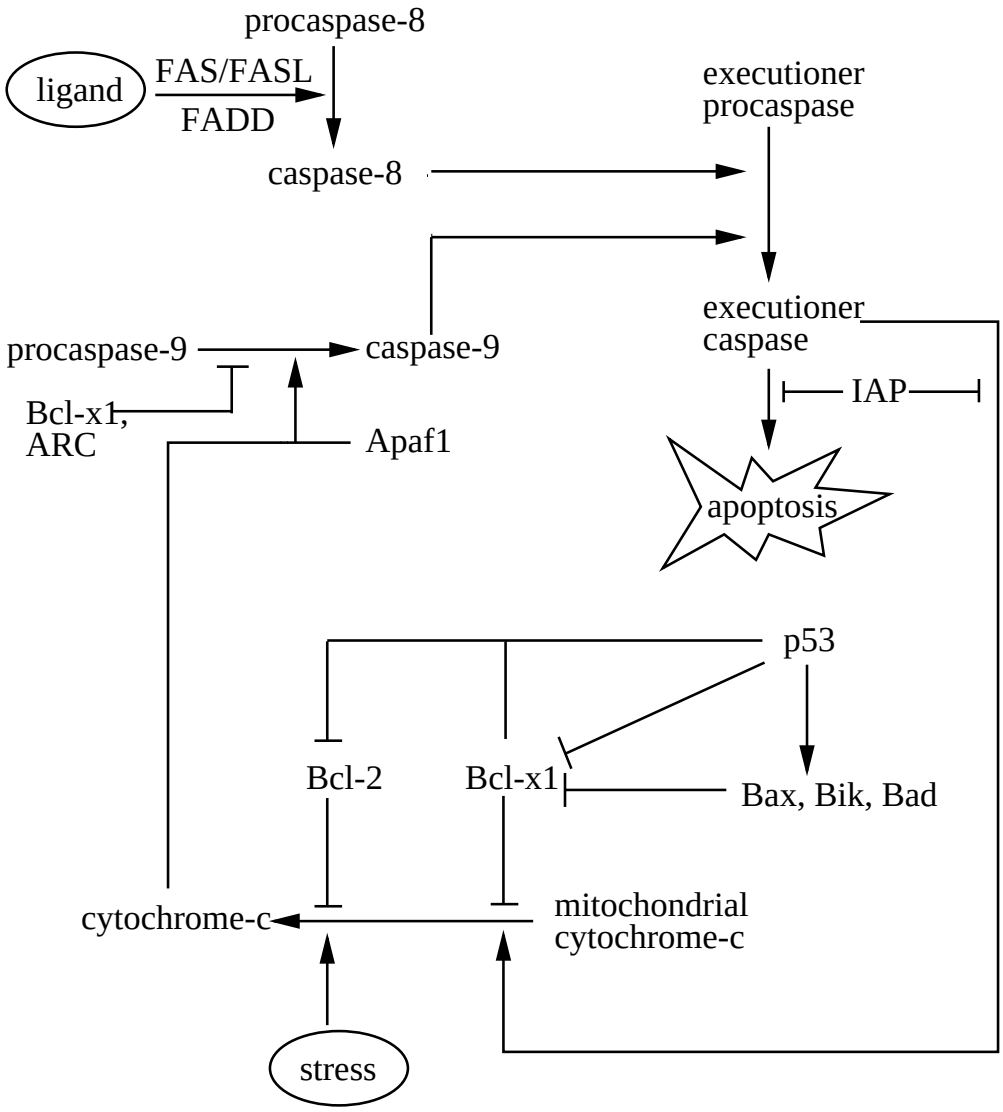

Figure 4.2: Caspase-dependent apoptosis mechanism described in [10]. The model includes two triggers for the activation of cell suicide mechanism, extracellular death ligand and stress-related factor. The cell will die when executioner caspase is activated by caspase-8 (ligand effector) or caspase-9 (stress-related effector). (redrawn from [11])

in the matrix-vector form as follows:

$$\begin{aligned} \dot{x} &= Ax + Br + C, \\ r &= f(x, p), \end{aligned} \tag{4.1}$$

where the vector x represents 19 states (the concentrations of proteins), and their representing proteins are listed in Table 4.1. The vector r represents 11 reaction rates and has the expression as follows:

$$\begin{aligned} r_1 &= \frac{k_l(x_1 - x_2)L}{K_S^{-1} + L}, \\ r_2 &= \frac{k_a x_3 x_2}{1 + K_A x_3 + K_A K_B x_3^2} - \frac{x_4}{K_A K_B x_3}, \\ r_3 &= \frac{k_h x_5 x_6}{1 + K_H x_3 + K_I \frac{x_{19}}{1 + K_J x_{17}}} - \frac{x_7}{K_H}, \\ r_4 &= \frac{k_{8za1} x_8^2 x_4}{K_C^{-1} K_D^{-1} + K_D^{-1} x_8 + x_8^2 + K_F K_C^{-1} k_D^{-1} x_{15} + K_G K_D^{-1} x_8 x_{15}}, \\ r_5 &= \frac{k_{9za1} x_9^2 x_7}{K_K^{-1} K_L^{-1} + K_L^{-1} x_9 + x_9^2 + K_N K_K^{-1} K_L^{-1} x_{16} + K_O K_L^{-1} x_9 x_{16}}, \\ r_6 &= k_{8za2} x_8^2, \\ r_7 &= k_{9za2} x_9^2, \\ r_8 &= \frac{k_{83a} x_{10} x_{11}}{K_P^{-1} + K_R K_P^{-1} x_{14} + x_{10}}, \\ r_9 &= \frac{k_{93a} x_{10} x_{12}}{K_P^{-1} + K_R K_P^{-1} x_{14} + x_{10}}, \\ r_{10} &= \alpha_{CE}[v(x_{13}, x_{18}) + v(X, x_{18})], \\ r_{11} &= k_u x_{13} \frac{[\text{IAPs}]}{1 + K_U[\text{IAPs}]}, \end{aligned} \tag{4.2}$$

where

L is the free ligand concentration,

$$v(x_{13}, x_{18}) = \begin{cases} 1, & \forall x_{13}/x_{18} > 0.25 \\ 0, & \forall x_{13}/x_{18} \leq 0.25 \end{cases},$$

$$v(X, x_{18}) = \begin{cases} 1, & \forall X/x_{18} > 0.025 \\ 0, & \forall X/x_{18} \leq 0.025 \end{cases},$$

$$X = \text{chemical/nutrition factor (stress)},$$

$$\frac{[\text{IAPs}]}{1 + K_U[\text{IAPs}]} = 0.1765.$$

The vector p represents 27 parameters (11 reaction rate constants and 16 saturation constants) in this model, and the corresponding notations are listed in Table 4.2.

Table 4.1: List of states in model (4.1)

No.	Name of Protein	No.	Name of Protein
1	total receptor ligands	11	caspase-8
2	FAS/FASL	12	caspase-9
3	FADD	13	executioner caspase
4	FAS/FASL-FADD	14	decoy protein
5	cytochrome c	15	decoy protein
6	Apaf1	16	decoy protein
7	Apaf1- cytochrome c	17	activator protein
8	procaspase-8	18	Bcl-2
9	procaspase-9	19	Bcl-x1
10	executioner procaspase		

The diagonal matrix A and the vector C describe degradation and auto-generation respectively, whereas the matrix B represents the stoichiometric matrix of the bio-

logical network and has the value as follows:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The key challenge is to identify the parameters in the nonlinear reaction rate functions. In this research, the model (4.1) for the caspase system with the 'true' parameter values is considered as the 'actual' system. As measurements are taken at a series of discrete time points, we will consider a discrete version of the model, which can be derived by using the zero-order hold technique. The resulting discrete model is represented as:

$$\begin{aligned} x(k+1) &= \hat{A}x(k) + \hat{B}r(k) + \hat{C}, \\ r(k) &= f(x(k), p), \end{aligned} \tag{4.3}$$

where $\hat{A} = e^{A\Delta t}$, $\hat{B} = (e^{A\Delta t} - I)A^{-1}B$, $\hat{C} = (e^{A\Delta t} - I)A^{-1}C$. The scalar Δt is the sampling interval and therefore $1/\Delta t$ is the sampling frequency.

4.3.2 Model Complexity Analysis

In model (4.1) or (4.3), the parameters of interest are in the nonlinear reaction rate functions. Unfortunately, none of these rates can be directly measured. If we have the measurements of all states and know the values of matrix A and vector C, we could estimate the values of reaction rates from the top equation in model (4.3). However, practically it is almost impossible (or too costly) to measure all states in a biological network at all time points in an experiment. In addition, one reaction rate (and thus parameters involved) may be contained in several equations. It is too complicated to estimate the parameters directly from the state-space model (4.1) or (4.3). The purpose of model complexity analysis is to provide some insights for designing optimal experiments such that unmeasured states and all reaction rates can be estimated uniquely and as many as possible parameters in rate functions can be estimated accurately.

By using the sensitivity analysis, Gadkar *et al.* selected the optimal measurement set of states, $\{x_2, x_3, x_5, x_7, x_{10}, x_{11}, x_{12}\}$ [11, 12]. With measurements of the states in the optimal set at a series of time points and measurements of all other states at the initial time point, they estimate all 11 reaction rates and all 12 states other than those in the optimal set, at all time points by using the state regulator method. However, our study shows that the rank of matrix B is 8, which is less than 11 (the number of reaction rates). This indicates that all 11 reaction rates can not be uniquely estimated. In addition, we have observed that the reaction rate r_{11} is only in the equation of state x_{13} and that state x_{13} is not in the optimal set selected by Gadkar *et al.* in [11, 12]. This implies that the reaction rate r_{11} cannot be correctly estimated.

In our previous study [50], we analyzed the stoichiometric matrix B or $\hat{B}$ using linear algebra and concluded that reaction rates in each of the three pairs (r_4, r_6), (r_5, r_7), and (r_8, r_9) were dependent, while other five rates were independent. Combining each dependent pair as a new reaction rate, we can rewrite the model (4.3)

as follows:

$$\begin{aligned} x(k+1) &= \bar{A}x(k) + \bar{B}\bar{r}(k) + \bar{C}, \\ \bar{r}(k) &= f(x(k), p), \end{aligned} \tag{4.4}$$

where $\bar{A} = \hat{A}$, $\bar{C} = \hat{C}$, and the matrix $\bar{B}$ is a 19×8 matrix which has the following expression:

$$\bar{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \beta, \quad \text{when } A = \alpha \times I^{1}.$$

The new reaction rates $\bar{r} = [r_1, r_2, r_3, r_4 + r_6, r_5 + r_7, r_8 + r_9, r_{10}, r_{11}]^{\mathrm{T}}$ are defined

[1] α and β are constant numbers.

as follows:

$$
\begin{aligned}
\bar{r}_1 &= \frac{k_l(x_1 - x_2)L}{K_S^{-1} + L}, \\
\bar{r}_2 &= \frac{k_a x_3 x_2}{1 + K_A x_3 + K_A K_B x_3^2} - \frac{x_4}{K_A K_B x_3}, \\
\bar{r}_3 &= \frac{k_h x_5 x_6}{1 + K_H x_3 + K_I \frac{x_{19}}{1 + K_J x_{17}}} - \frac{x_7}{K_H}, \\
\bar{r}_4 &= \frac{k_{8za1} x_8^2 x_4}{K_C^{-1} K_D^{-1} + K_D^{-1} x_8 + x_8^2 + K_F K_C^{-1} k_D^{-1} x_{15} + K_G K_D^{-1} x_8 x_{15}} + k_{8za2} x_8^2, \\
\bar{r}_5 &= \frac{k_{9za1} x_9^2 x_7}{K_K^{-1} K_L^{-1} + K_L^{-1} x_9 + x_9^2 + K_N K_K^{-1} K_L^{-1} x_{16} + K_O K_L^{-1} x_9 x_{16}} + k_{9za2} x_9^2, \\
\bar{r}_6 &= \frac{k_{83a} x_{10} x_{11} + k_{93a} x_{10} x_{12}}{K_P^{-1} + K_R K_P^{-1} x_{14} + x_{10}}, \\
\bar{r}_7 &= \alpha_{CE}[v(x_{13}, x_{18}) + v(X, x_{18})], \\
\bar{r}_8 &= k_u x_{13} \frac{[\text{IAPs}]}{1 + K_U[\text{IAPs}]}.
\end{aligned}
\tag{4.5}
$$

It can be verified that $\text{rank}(\bar{B}) = 8$, which indicates that the newly defined 8 reaction rates are linearly independent, and thus can be estimated uniquely.

4.3.3 Optimal Measurement Set

To estimate the parameters in the reaction rate functions, we may first estimate the reaction rates from the measurements of state variables. In model (4.4), the rank of matrix $\bar{B}$ is 8, which means eight reaction rates can be uniquely identified from the measurements of at least 8 states. In the caspase system, 11 of the 19 states are correlated with 8 reaction rates. From a further analysis of model (4.4), we get Table 4.3 to show the relationship between the measured states and the estimated reaction rates.

From Table 4.3, one can select a number of equivalent optimal measurement sets. One of the optimal measurement sets could be $\{x_2, x_3, x_5, x_7, x_{10}, x_{11}, x_{12}, x_{13}\}$, which has one more state than that obtained by Gadkar *et al.* in [11, 12]. Collecting the dynamic equations of all states in the optimal measurement set, we have

$$
\tilde{x}(k+1) = \widetilde{A}\tilde{x}(k) + \widetilde{B}\bar{r}(k) + \widetilde{C}, \tag{4.6}
$$

Table 4.2: List of parameters in model (4.1)

No.	Parameter	No.	Parameter	No.	Parameter
1	k_l	10	α_{CE}	19	K_D
2	k_a	11	k_u	20	K_F
3	k_h	12	K_S	21	K_G
4	k_{8za1}	13	K_A	22	K_K
5	k_{9za1}	14	K_B	23	K_L
6	k_{8za2}	15	K_H	24	K_N
7	k_{9za2}	16	K_I	25	K_O
8	k_{83a}	17	K_J	26	K_P
9	k_{93a}	18	K_C	27	K_R

Table 4.3: Relationship between the measured states and the estimated reaction rates

Measured states	Estimated reaction rates
x_2	$\bar{r}_1$
x_3 or x_4	$\bar{r}_2$
x_7	$\bar{r}_3$
x_8 or x_{11}	$\bar{r}_4$
x_9 or x_{12}	$\bar{r}_5$
x_{10}	$\bar{r}_6$
x_5	$\bar{r}_7$
x_{13}	$\bar{r}_8$

where the vector $\tilde{x} = [x_2, x_3, x_5, x_7, x_{10}, x_{11}, x_{12}, x_{13}]^{\mathrm{T}}$ consists of the optimal set of states.

Accordingly, the matrix $\widetilde{A}$ is a sub-matrix of $\bar{A}$, the vector $\widetilde{C}$ is a sub-vector of $\bar{C}$, and the matrix $\widetilde{B}$ is a sub-matrix of $\bar{B}$, which has the following expression:

$$\widetilde{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix} \times \beta.$$

4.3.4 Estimation of Unmeasured States and Reaction Rates

As the matrix $\widetilde{B}$ in Eq. (4.6) is nonsingular, one can solve eight reaction rates in $\bar{r}$ from Eq. (4.6) to get

$$\bar{r}(k) = \widetilde{B}^{-1}[\tilde{x}(k+1) - \widetilde{A}\tilde{x}(k) - \widetilde{C}]. \tag{4.7}$$

This means that as long as we get the measurements of states in the optimal set at time point $(k+1)$, we can estimate all reaction rates at time point k.

In addition, to estimate the parameters in the reaction rates, we also need values of all states (not only the states in the optimal measurement set). To estimate the unmeasured states (not in the optimal measurement set), we need the initial values of these states. With these initial values, we can submit Eq. (4.7) into model (4.4) to get the estimates of the unmeasured states step after step.

4.3.5 Parameter Estimation

From the reaction rate expression (4.5), although there are two parameters K_S and k_l in $\bar{r}_1$ given the free death ligand concentration value L, these two parameters

cannot be independently estimated. In this study, we consider $k_l L/(K_S^{-1}+L)$ as one parameter denoted by K_1. Therefore, each of the three reaction rates ($\bar{r}_1$, $\bar{r}_7$, and $\bar{r}_8$) contains only one linear parameter, and can be easily estimated by the least squares method. The other five reaction rates $\bar{r}_2 \sim \bar{r}_6$ are or can be transferred into the linear fractional models (LFM). If we consider the coefficients of each state function as one combined parameter (e.g., $K_N K_K^{-1} K_L^{-1}$ and $K_O K_L^{-1}$), these five reaction rates are also in the form of the LFM, i.e., its numerator and denominator are linear with respect to the combined parameters respectively.

These parameters can be estimated using the proposed separation method for the LFM, which is detailed in Section 3.2.1.

4.4 Numerical Example and Simulation Results

To illustrate the proposed methodology, we use the same parameter values and initial conditions as those in [11, 12] to produce the measurements from the 'real' system, with the sampling frequency of 100 min^{-1}. That is, the degradation matrix is set as $A = -0.5 \times I$, the auto-generation vector is set as $C = 0.1 \times [1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0.3, 1, 1, 0.1, 1, 0.35]^{\text{T}}$, and the parameter vector composed in the order as in Table 4.2 is set as $k = [1.05, 2, 0.3, 1.25, 1.25, 0.00001, 0.00001, 0.5, 0.5, 0.1, 0.5, 10, 0.1, 100, 10, 100, 5, 100, 100, 2000, 2000, 100, 100, 2000, 2000, 1.5, 5]^{\text{T}}$. Then we estimate the reaction rates and those unmeasured states by the proposed framework, using only the values of states in the optimal measurement set. The average error (calculated via formula (20) in [11]) for all states is 8.1×10^{-13}, while the average error for all reaction rates is 3.8×10^{-5}.

Figure 4.3 shows a comparison of the actual and the estimated states while Fig. 4.4 shows a comparison of the actual and the estimated reaction rates. From these two figures we can hardly tell the difference between the actual and the estimated states (or reaction rates) when the sampling frequency is 100 min^{-1}. However, when the sampling frequency is 1 min^{-1}, the difference between the actual and the estimated states (or reaction rates) is somewhat large [50], yet is much smaller than

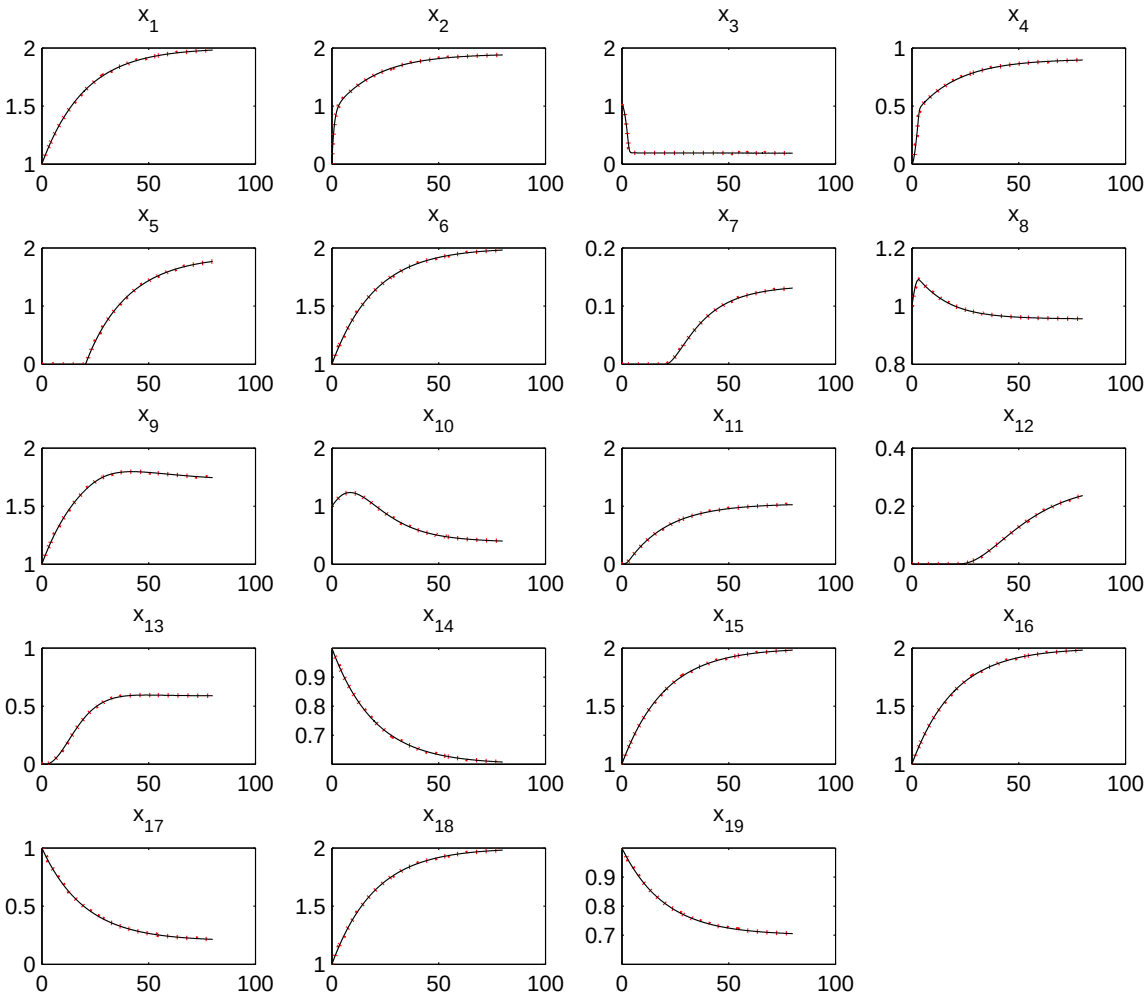

Figure 4.3: Comparisons of the actual and the estimated states: solid lines(black) from the actual system, dotted lines(red) from the estimates.

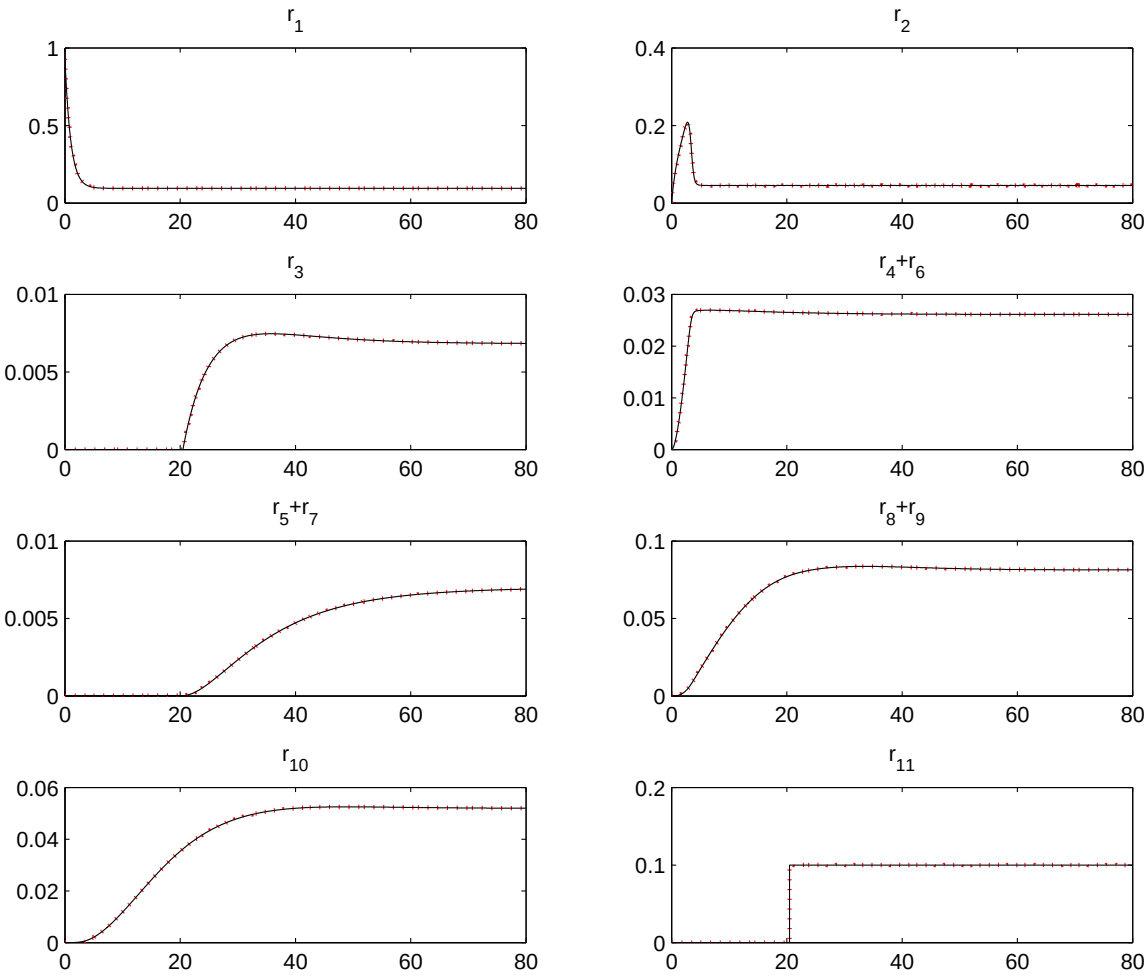

Figure 4.4: Comparisons of the actual and the estimated reaction rates: solid lines(black) from the actual system, dotted lines(red) from the estimates.

the result from [11, 12]. This indicates that our proposed methodology for estimating the unmeasured states and reaction rates is better than that proposed by Gadkar *et al.* in [11, 12]. The more accurate the estimated states and reaction rates are, the more accurate the parameter estimation will be.

With the estimated reaction rates and states, we can estimate the parameters in the reaction rate functions by the method mentioned in Section 4.3.5. Using the methods proposed in [11, 12], most parameters in reaction rates $\bar{r}_4$ (i.e., $r_4 + r_6$) and $\bar{r}_5$ (i.e., $r_5 + r_7$) cannot be identified. To illustrate our proposed method, this study only estimates the parameters in $\bar{r}_4$, $\bar{r}_5$, and $\bar{r}_6$. The results are listed in Table 4.4. The average of relative estimation errors (AREE) are calculated over 20 different initial guess values of parameters which are drawn from the normal distributions with the mean of the true parameter values and the standard deviation of 0.5.

Table 4.4: Estimation results showing the average of relative estimation errors (AREE)

No.	Parameter	True Value	AREE	No.	Parameter	True Value	AREE
4	k_{8za1}	1.25	0.13%	20	K_F	2000	34.26%
5	k_{9za1}	1.25	0.01%	21	K_G	2000	0.00%
6	k_{8za2}	10^{-5}	360.7%	22	K_K	100	32.53%
7	k_{9za2}	10^{-5}	2.72%	23	K_L	100	0.00%
8	k_{83a}	0.5	0.93%	24	K_N	2000	32.53%
9	k_{93a}	0.5	0.83%	25	K_O	2000	0.00%
18	K_C	100	34.26%	26	K_P	1.5	2.92%
19	K_D	100	0.00%	27	K_R	5	5.22%

From Table 4.4, almost all of these parameters can be identified. In particular, four parameters K_D, K_G, K_L, and K_O can be exactly estimated, other four parameters k_{8za1}, k_{9za1}, k_{83a}, and k_{93a} are estimated with the AREE less than 1%, three parameters k_{9za2}, K_P, and K_R are estimated with the AREE of 2$\sim$5%, four parameters K_C, K_F, K_K, and K_N are estimated with the AREE of about 30%. Due to

its unreasonably large AREE, the parameter k_{8za2} can be considered unidentified. In summary, most of these parameters can be identified with the AREE values less than those in [11, 12].

4.5 Summary

This chapter describes the development of a new methodology for estimating states and reaction rates in the model of the caspase-activated apoptosis system. As the reaction rates are unmeasurable, it is crucial to estimate these rates for the subsequent identification of the parameters within them. Because of limitations of the experimental condition and cost, not all states are measured in practice. Thus, it is essential to estimate also the unmeasured states. This chapter focuses on giving estimations for unmeasurable variables of the biological model using less experimental measurement. The proposed methodology could be generalized for a wide range of complex biological processes, especially in cases where only limited experimental data are available. A numerical example, i.e., estimating parameters in the caspase model, is given to show the effectiveness of the whole framework.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this thesis, several parameter estimation methods have been developed for two classes of nonlinear biological systems, specifically the S-system and the linear fractional model (LFM). These methods are primarily based upon the separation approach, proposed to utilize the special structures of target models. The simulation results show that these methods, applied on small-scale or middle-scale model examples, give better performances in estimation accuracy and computation time than conventional nonlinear estimation methods.

In Chapter 2, we develop two methods for estimating parameters within the S-system. These methods fully make use of the structure specialty of the S-system. The alternating least squares (ALS) method treats the positive term or the negative term separately in the model. By taking the logarithm, the estimation problem is transformed into a linear least squares problem, which has greatly reduced the computational complexity. However, as the distribution property of measurement errors could be destroyed when taking the logarithm, this method can only be applied on noise-free or small noise cases. The second method proposed for the S-system is called the separation estimation method. This method concentrates on the model's partially linear structure. Using this method, a subset of the parameters will be eliminated in the optimization process and thus the computation cost is reduced. This separation method can be ported to other models which possess the partial linearity property.

In Chapter 3, two estimation methods are designed for the LFM. Firstly, as the

structure of the LFM is also linear with respect to some of the parameters, specifically parameters in the numerator, the separation estimation method can be applied in this case. Furthermore, a modified version of the separation method is developed to make the estimation in an iterative approach. Secondly, using the weighted least squares method and the separation principle, the estimates can be achieved in an analytical closed-form expression, by introducing a specific weight matrix.

In Chapter 4, the separation estimation method is applied to a more complex case – caspase model with incomplete measurements. The model complexity is analyzed and an optimal measurement set is designed. With these partial measurements of states, the reaction rates and the unmeasured states could be estimated under the proposed framework of methodology. After this, parameters within the caspase model are estimated using the separation method for the LFM.

In a nutshell, these proposed parameter estimation methods give better performances in terms of estimation accuracy and computation cost, compared to the conventional nonlinear optimization approaches. Simulations have been performed and show the effectiveness of the proposed methods. After determining the values of parameters within the model, further understanding of the system characteristics can be achieved upon verifying hypotheses from the derived model.

5.2 Future Work

There are still some problems awaiting to be tackled for the parameter estimation problem. First, we need to consider the low sampling problem. As limitations exist for the experimental measurement frequency, the measured data may be sampled at relatively large time spans. In this case, we can use a curve-fit or interpolation technique to connect the measurement points by a smooth curve. Therefore the intermediate values can be restored and used.

Second, the measurement noise is a problem to deal with as it will undermine the efficiency of parameter estimation methods. For such a problem, smoothing or filtering techniques could be adopted to remove the noise signal. The conducted

simulations shown in the thesis are mainly based on noise-free data sets. So the performances of the proposed methods can be evaluated, without the effect of noise-induced errors. In this research, we do make trials of the proposed methods on noise-contaminated data, and make use of different smoothing or filtering strategies. As there is no panacea for denoising, the performances of the proposed methods fluctuate a lot when facing different noise data and need to be analyzed case by case. In general, the noise data will not affect the speed benefit of the proposed methods, although the estimation accuracy cannot be guaranteed.

Lastly, if the initial range or boundary of parameters can be obtained, the non-convergence problem due to unreasonable initial guess values could be greatly reduced.

There are also some suggestions for future work. 1) Some stochastic methods, e.g., Bayesian approaches, Monte Carlo simulations and simulated annealing, can be developed to deal with some large-scale models. As these methods usually require a huge amount of computation resources, some estimation algorithms need to be developed and ported to a high performance computer or a parallel computing cluster. 2) As there are different sources of experimental data for the same biological model or biochemical reaction, such a question arises and needs to be considered: how to incorporate multiple sources of data in the estimation process? 3) Currently this estimation work is performed under the assumption that the exact structure of model is already known. One direction for extended study is to combine the structure identification and the parameter estimation together. All these questions still need to be considered in further research.

References

[1] Bioinformatics station. http://bioinformaticstation.org. cited on Oct 19, 2009.

[2] Genome project. http://en.wikipedia.org/wiki/Genome_project. cited on Jan 25, 2010.

[3] Systems biology. http://en.wikipedia.org/wiki/Systems_biology. cited on Jan 25, 2010.

[4] Systems biology - the 21st century science. http://www.systemsbiology.org/Intro_to_ISB_and_Systems_Biology/Systems_Biology_--_the_21st_Century_Science. cited on Oct 19, 2009.

[5] H. Abdi. Least squares. http://www.utdallas.edu/~herve/Abdi-LeastSquares-pretty.pdf. cited on Oct 19, 2009.

[6] J.V. Beck and K.J. Arnold. *Parameter estimation in engineering and science*. Wiley New York, 1977.

[7] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[8] C. Corradi and L. Stefanini. Computational experience with algorithms for separable nonlinear least squares problems. *Calcolo*, 15(3):317–330, 1978.

[9] P.H. Eilers. A perfect smoother. *Analytical Chemistry*, 75(14):3631–3636, 2003.

[10] M. Fussenegger, J.E. Bailey, and J. Varner. A mathematical model of caspase function in apoptosis. *Nature Biotechnology*, 18(7):768–774, 2000.

[11] K.G. Gadkar, R. Gunawan, and F.J. Doyle. Iterative approach to model identification of biological networks. *BMC Bioinformatics*, 6:155, 2005.

[12] K.G. Gadkar, J. Varner, and F.J. Doyle. Model identification of signal transduction networks from data using a state regulator problem. *IEE Proceedings Systems Biology*, 2(1):17–30, 2005.

[13] M. Gautier and P. Poignet. Extended Kalman filtering and weighted least squares dynamic identification of robot. *Control Engineering Practice*, 9(12):1361–1372, 2001.

[14] G.H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973.

[15] R.N. Gutenkunst, F.P. Casey, J.J. Waterfall, C.R. Myers, and J.P. Sethna. Extracting falsifiable predictions from sloppy models. *Annals of the New York Academy of Sciences*, 1115:203–211, 2007.

[16] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, and J.P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):1871–1878, 2007.

[17] C. Haass. Apoptosis: Dead end for neurodegeneration? *Nature*, 399(6733):204–207, 1999.

[18] S.Y. Ho, C.H. Hsieh, F.C. Yu, and H.L. Huang. An intelligent two-stage evolutionary algorithm for dynamic pathway identification from gene expression profiles. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 648–660, 2007.

[19] L. Kaufman. A variable projection method for solving separable nonlinear least squares problems. *BIT Numerical Mathematics*, 15(1):49–57, 1975.

[20] B.N. Kholodenko, F.J. Bruggeman, and H.M. Sauro. Mechanistic and modular approaches to modeling and inference of cellular regulatory networks. *Topics in Current Genetics*, 13:143–159, 2005.

[21] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, 19(5):643–650, 2003.

[22] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems biology in practice: concepts, implementation and application*. Wiley-VCH, 2005.

[23] G. Lillacci and M. Khammash. Parameter estimation and model selection in computational biology. `http://cnls.lanl.gov/q-bio/wiki/images/3/35/119_Lillacci2.pdf`. cited on Oct 19, 2009.

[24] Y. Matsubara, S. Kikuchi, M. Sugimoto, and M. Tomita. Parameter estimation for stiff equations of biosystems using radial basis function networks. *BMC Bioinformatics*, 7:230, 2006.

[25] C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13(11):2467–2474, 2003.

[26] A.R. Neves, A. Ramos, M.C. Nunes, M. Kleerebezem, J. Hugenholtz, W.M. De Vos, J. Almeida, and H. Santos. *In vivo* nuclear magnetic resonance studies of glycolytic kinetics in *Lactococcus lactis*. *Biotechnology and Bioengineering*, 64(2):200–212, 1999.

[27] J. Nielsen, J. Villadsen, and G. Lidén. *Bioreaction engineering principles*. Kluwer Academic/Plenum Publishers New York, 2003.

[28] M.R. Osborne. Some special nonlinear least squares problems. *SIAM Journal on Numerical Analysis*, 12(4):571–592, 1975.

[29] M. Piazza, X.J. Feng, J.D. Rabinowitz, and H. Rabitz. Diverse metabolic model parameters generate similar methionine cycle dynamics. *Journal of Theoretical Biology*, 251(4):628–639, 2008.

[30] M.A. Savageau. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology*, 25(3):365–369, 1969.

[31] M.A. Savageau. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology*, 25(3):370–379, 1969.

[32] M.A. Savageau. *Biochemical systems analysis: a study of function and design in molecular biology*. Addison Wesley Publishing Company, 1976.

[33] M.A. Shea and G.K. Ackers. The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181(2):211–230, 1985.

[34] J.L. Snoep and H.V. Westerhoff. From isolation to integration, a systems biology approach for building the silicon cell. *Topics in Current Genetics*, 13:13–30, 2005.

[35] G.N. Stephanopoulos, A.A. Aristidou, and J. Nielsen. *Metabolic Engineering: Principles and Methodologies*. Academic Press, 1998.

[36] J. Swartz and H. Bremermann. Discussion of parameter estimation in biological modelling: Algorithms for estimation and evaluation of the estimates. *Journal of Mathematical Biology*, 1(3):241–257, 1975.

[37] C.B. Thompson. Apoptosis in the pathogenesis and treatment of disease. *Science*, 267(5203):1456–1462, 1995.

[38] N.V. Torres and E.O. Voit. *Pathway analysis and optimization in metabolic engineering*. Cambridge University Press, 2002.

[39] W. Tucker, Z. Kutalik, and V. Moulton. Estimating parameters for generalized mass action models using constraint propagation. *Mathematical Biosciences*, 208(2):607–620, 2007.

[40] W. Tucker and V. Moulton. Parameter reconstruction for biochemical networks using interval analysis. *Reliable Computing*, 12(5):389–402, 2006.

[41] M. Verhaegen and V. Verdult. *Filtering and system identification: a least squares approach*. Cambridge University Press, 2007.

[42] M. Vilela, C.C.H. Borges, S. Vinga, A.T.R. Vasconcelos, H. Santos, E.O. Voit, and J.S. Almeida. Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics*, 8:305, 2007.

[43] E.O. Voit. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press, 2000.

[44] E.O. Voit and J. Almeida. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*, 20(11):1670–1681, 2004.

[45] H. Wang, L. Qian, and E. Dougherty. Inference of gene regulatory networks using S-system: A unified approach. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 82–89, 2007.

[46] D.M. Wolf and F.H. Eeckman. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *Journal of Theoretical Biology*, 195(2):167–186, 1998.

[47] F.X. Wu. Estimation of parameters in the linear-fractional models. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1086–1089, 2007.

[48] F.X. Wu and L. Mu. Parameter estimation in rational models of molecular biological systems. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3263 –3266, 2009.

[49] F.X. Wu and L. Mu. Separable parameter estimation method for nonlinear biological system. In *The 3rd International Conference on Bioinformatics and Biomedical Engineering*, 2009. 4 pages on disc.

[50] F.X. Wu, L. Mu, and R. Luo. Complexity analysis and optimal experimental design for parameter estimation of biological systems. In *Canadian Conference on Electrical and Computer Engineering*, pages 393–398, 2008.

[51] 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*, 41(1):73–80, 2010.

[52] D. Zwillinger. *CRC standard mathematical tables and formulae*. CRC Press, 2002.