Projection-Based Model Order Reduction for Biochemical Systems

By

Anum Javed

FALL-2015-MS-CSE00000118041

Supervisor

Dr. Mian Ilyas Ahmad

Department of Computational Engineering

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computational Science and Engineering (MS CS&E)

In

Research Center for Modeling and Simulation (RCMS),

National University of Sciences and Technology (NUST),
Islamabad, Pakistan.
(November 2018)

Approval

It is certified that the contents and form of the thesis entitled "**PROJECTION-BASED MODEL ORDER REDUCTION FOR BIOCHEMICAL SYSTEMS** " submitted by **Anum Javed** have been found satisfactory for the requirement of the degree.

Advisor: **Dr. Mian Ilyas Ahmad**

Signature: _____

Date: _____

Committee Member 1: **Dr. Abdul Ghafoor**

Signature: _____

Date: _____

Committee Member 2: **Dr. Salma Sherbaz**

Signature: _____

Date: _____

Dedication

I dedicate this effort to my family who have assisted me in any possible way to become what I am today. Their sacrifices seeded my success especially my ***mom dad*** *who showed their devoted attention and my brother and twin sisters who inspired me all the way.*

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at RCMS NUST or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at RCMS NUST or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Anum Javed**

Signature: _____

Acknowledgement

Glory be to **Allah (S.W.A)**, owner of the sovereignty, creator of the Universe. Who has the power to exalt whom He wills and to abase whom He wills. Verily, Man is nothing without His blessings, support and help. From the day, I came to NUST till my departure, He was the one Who opened ways for me, elevated my morale and guided me towards the best. Nothing can pay back the gratitude of His blessings.

Firstly, I want to show my deep indebtedness to my **beloved parents**, who struggled hard to give me an opportunity to be benefited from the renowned institute. Whose blind trust on me, boundless love for me and continuous motivation helped me to achieve the status that I am today. May Allah have mercy upon them (Ameen).

Then I would like to express my gratitude to my supervisor **Dr. Mian Ilyas Ahmad** whose vision has always been the most important pillar in establishing this research work. This work would not have been completed without his unwavering support, motivation, encouragement and guidance..

I wish to express the deepest appreciation to my GEC members, **Dr. Salma Sherbaz** for always bringing up new ideas and providing valuable insights regarding this research and **Dr. Abdul Ghafoor** for his support and encouragement.

Furthermore, I am extremely grateful to **Madiha Muzaffar**, who is not just a friend but family to me and whom I can always count on. Apart from the research team, I would like to thanks my close friends exclusively **Madiha Amjad, Mehmona Arsalan, Seema Mir Akbar and Saima Amjad** whose kind support has always ensured my true potential. Their spiritual and moral support throughout the thesis phase will truly be remembered in my life.

Anum Javed

Contents

List of Abbreviations

ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
RNA	RiboNucleic Acid
mRNA	Messenger RiboNucleic Acid
BRN	Biological Regulatory Networks
MOR	Model Order Reduction
IRKA	Iteractive Rational Krylov Algorithm
QSSA	Quasi Steady State Approximation
CRN	Chemical Reaction Network
DAEs	Differential Algebraic Equations
SVD	Singular Value Decomposition

List of Tables

List of Figures

Abstract

Biochemical systems represent a process that involves different biological species linked by a network of chemical reactions. To analyze the behavior of the system, we perform experiments either on the actual system or on the mathematical model of the system. In this thesis, our focus is on modeling and analysis (computer simulation) of biochemical systems. The problem with mathematical models is their complexity. The desire for more details and accurate results often generate large scale complex models. Numerical simulation of such complex models is computationally expensive. Model order reduction can be utilized to tackle this issue of complexity by trying to take out those parts of a reaction network that are mathematically contributing very little in our parameters of interest. In this thesis we are using an important projection based model reduction technique that is called IRKA for model reduction of biochemical systems. To clarify the application of IRKA in reduction of biochemical systems, we consider an example of biochemical system from the literature and presents the key steps of modeling, conservation, linearization and reduction. The results of IRKA are compared with lumping, which is a common reduction technique for chemical reactions. It is observed that the approximation error through IRKA is much better as compared to the lumping technique.
Keywords: model order reduction, complexity, mathematical modeling, chemical reaction.

Chapter 1
Introduction

Biochemical systems represent a process that involves different biological species linked by a network of chemical reactions. To analyze the behavior of the system, we perform experiments either on the actual system or on the mathematical model of the system. In this thesis, our focus is on modeling and analysis (computer simulation) of biochemical systems. One of the problem with mathematical models is their complexity. The desire for more details and accurate results often generate large scale complex models. Numerical simulation of such complex models is computationally expensive. Model order reduction can be utilized to tackle this issue of complexity by trying to take out those parts of a reaction network that are mathematically contributing very little in our parameters of interest. In the reduced model, variables and parameters are less as compared to the original model but the behavior of both models are almost same. In this thesis we are using an important projection based model reduction technique that is called iterative rational Krylov algorithm (IRKA) for model reduction of biochemical systems. The advantage of IRKA is that, it can be extended to very large-scale settings because it involves only matrix vector multiplications.

1.1 Biochemical Reactions

A biochemical reaction is a process that leads to the chemical transformation of one set of chemical substances to another. Biochemical reactions are controlled by enzymes [1], which are biological catalysts that can modify the rate of chemical reactions. Computational analysis of biochemical reactions [2] is commonly used in computational biology to observe important biological process such as cell signaling, metabolism and the rules of gene expression. A network of biochemical reactions that define a specific process is often used to predict the behavior of the network. The example of such a biochemical networks is discussed by **Gheorghe Craciun and Martin Feinberg** in 2006 [3] and is shown below.

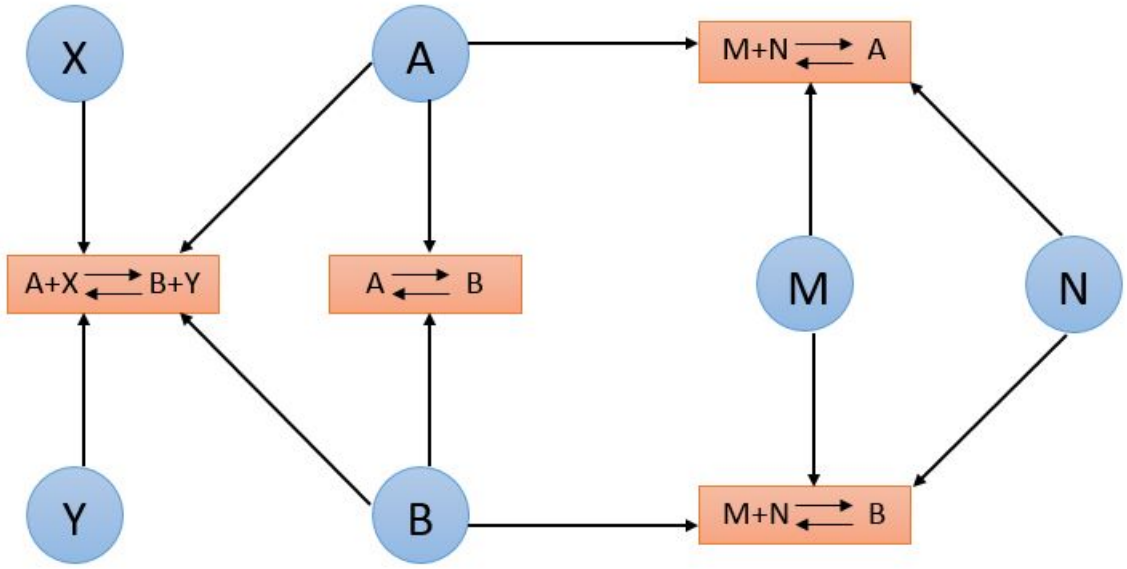


Figure (1.1) Example of biochemical reaction networks

Biochemical reactions networks are usually very complex networks [4]. These networks consist of hundreds of thousands of components with complicated interactions. The quantitative description of such networks involve nonlinear ordinary

differential equations (ODEs). However, the utilization of such large set of ODEs is often prohibited, since we have to face the challenge of numerically solving a very large nonlinear set of differential equations.

1.2 Model Order Reduction

Model order reduction (MOR) is a process where a large-scale mathematical model [5] is reduced to a low order model such that the response of these two models are almost equivalent or comparable. The reduced model is then used as a surrogate model to obtain useful information about the actual system. There are different types of techniques that are used for model order reduction. For example linear reduction techniques [6] are used to reduce linear models and nonlinear reduction techniques [7], [8] are used to reduce nonlinear models. In this thesis, our focus is on linear reduction techniques with application in biochemical systems. Since the ODE representation of biochemical systems involve nonlinear terms, we have to linearize the model before proceeding to the model order reduction technique. We have used two techniques for the purpose of MOR, one is lumping and second is iterative rational Krylov algorithm (IRKA) [9], which are discussed in Chapter 2. The main purpose of model order reduction is to simplify the model representation so that we can get fast simulation response and in case of control, easily tune the controller parameters.

1.2.1 MOR Formulation for Linear Systems

MOR for linear time-invariant systems can be formulated in both time and frequency domain. We begin with the time domain [10] representation. Consider a linear state

space model of the form :

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t),$$

where $x \in \mathbb{R}^n, y \in \mathbb{R}^p, u \in \mathbb{R}^m$ are n internal states, p outputs, m inputs of the system respectively, and A, B, C are constant matrices of appropriate dimensions. These matrices are either directly obtained from modeling or computed from a linearization procedure of the corresponding nonlinear system [11]. The pictorial representation of state space form of original model is shown in Figure 1.2 .

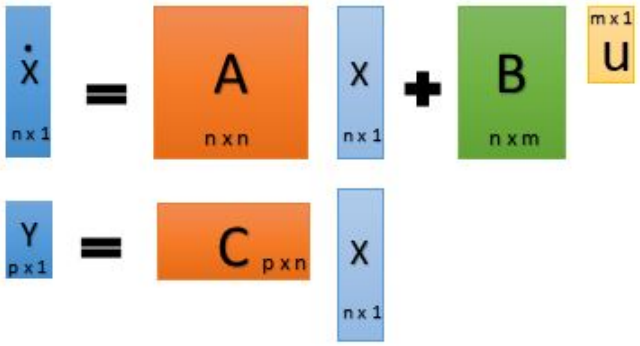


Figure (1.2) State space form of original model

.

The problem of MOR is to compute an order 'r' model (r $\ll$ n) from the original model the form:

$$\dot{x}_r(t) = A_r x_r(t) + B_r u_r(t)$$
$$y_r(t) = C_r x_r(t),$$

where $x_r \in \mathbb{R}^r, y_r \in \mathbb{R}^p$, r internal states, p outputs, m inputs of the reduced system

respectively and A_r, B_r, C_r are constant matrices of appropriate dimensions. The pictorial representation of state space form of reduced model is shown in Figure 1.3.

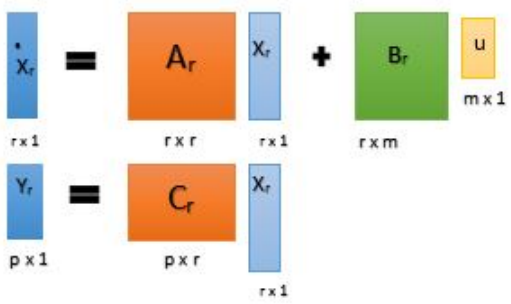


Figure (1.3) State space form of reduced model

1.2.2 Importance of MOR

The main reasons for computing low-order models can be grouped as follows:

1. To have low-order models so as to simplify the understanding of a system.
2. To reduce computational efforts in simulation problems.
3. To decrease computational efforts required for design of a numerically efficient controller, for the system to be reduced.

1.3 Problem Statement

Since biochemical systems are often very complex with hundreds of thousands components with complex interactions, their models are not easy to simulate. MOR can provide solution to this problem. Since we have different model reduction techniques, we need to observe which reduction techniques perform better for biochemical systems. The performance of the reduced model is measured by observing the following qualities of the reduction technique.

- The response of the original and reduced systems should be similar.

- The reduction technique should be computationally efficient. In fact, the computation time in constructing the reduced system and the time in simulating the reduced system should be much less then the simulation time of the original system.
- The reduction technique should be extendable to very large scale systems.
- The reduction technique should ensure the properties and structure of the original system. For example if the original system is stable and have linear structure, the reduced system should also be stable and linear.

1.4 Research Motivation

MOR is an important computational tool for efficient simulation and the control of different large scale dynamical systems. Efficient simulation of large scale systems decrease the computational cost and identify to the real-time response of the system. Computer simulations are now used in almost every physical, chemical, biological and other processes. It is good idea to simplify the model, either in size or in complexity to speed-up the computation time.

1.5 My Research Contribution

The main contributions of this thesis are as follows:

- Convert biochemical systems into ODE for reduction.
- Use reduction techniques on ODE to reduce the nonlinear ODE.
- We propose the use of IRKA to reduce the complexity of large scale systems.
- Reduce computational time of ODE through reduction techniques.

1.6 Thesis Outline

This thesis is organized as a collection of articles, hence each chapter can be read individually. An outline of the thesis follows next.

Chapter 1 is an introduction that gives an overview about the thesis, the problem statement, some background about biochemical systems and stat space model, research motivation and my contribution on this work.

Chapter 2 presents detail overview of previous literature o MOR.

Chapter 3 describe proposed methodology .

Chapter 4 discuss the results.

Chapter 5 draws the conclusion and present the direction for future work.

Chapter 2

Literature Review

2.1 Modeling of Chemical reaction

A chemical reaction [12] is a process in which chemical substances are converted into other substances with some rate of reaction [13]. A reaction rate is a concentration of the produced chemical per unit time [14] or the concentration of consumed reactants. There are different types of chemical reactions, a chemical reaction that complete in a single step is known as elementary reactions and a reaction that complete in more than one step is called composite reaction or complex reactions. The chemical reaction models change the physical knowledge into a mathematical form so that the knowledge can be used according to the academic problems in computational simulation. There are several types of mathematical models of chemical reactions [15]. Some of them are as follows:

2.1.1 The Law of Mass Action

The law of mass action [16] states that the rate of chemical reaction is directly proportional to the product of the activities or concentrations of the reactants [17]. It defines and predicts the solutions in dynamic equilibrium. It means that for a chemical reaction, the ratio between the concentration of reactants and the product is constant [18].

For example, there are three chemical A, B and C. Chemical A and chemical B

reacts to produce chemical C:

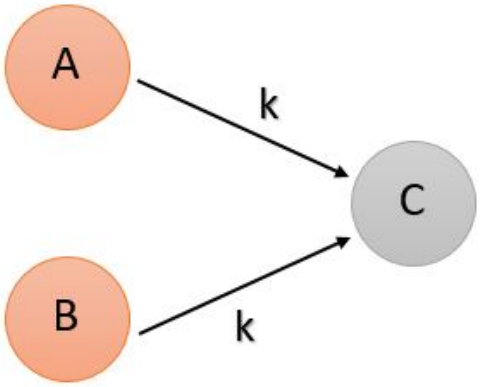


Figure (2.1) Model of 3 chemicals

.

$$A + B \xrightarrow{k} C$$

The k is the constant rate that determines the rate of the reaction. The probability of collision between that reactants produces the results is described by reaction rate. The law of mass action given as :

$$\frac{d[C]}{dt} = k[A][B]$$

$$\frac{d[A]}{dt} = -k[A][B]$$

$$\frac{d[B]}{dt} = -k[A][B]$$

Where the product of [A] [B] represents the probability of a collision.

2.1.2 Reversible reactions

A reversible reaction is a reaction where the reactants form the products, which react with the reactants and it reacts together to give the reactants back [19]. Suppose

we have A, B, C and D are four chemicals. A and B can be react to form C and D or, in the reverse reaction, C and D can reacts to form A and B.
For example, There are 3 chemicals A, B and C. A and B react with C to produce the chemical and C and vice versa.

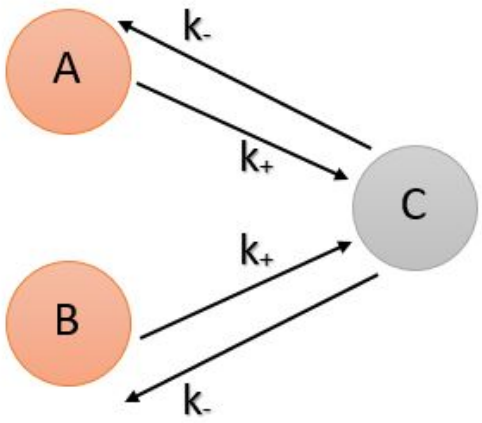


Figure (2.2) Model of 3 chemicals
.

The $k+$ and $k-$ is the constant rate that determines the rate of the reaction. Then the production rate is as following:

$$\frac{d[C]}{dt} = k_+[A][B] - k_-[C]$$
$$\frac{d[A]}{dt} = -k_+[A][B] + k_-[C]$$
$$\frac{d[B]}{dt} = -k_+[A][B] + k_-[C]$$

2.2 Examples of Chemical Reactions

Suppose that you have a container which contains the 4 chemicals. The name of chemicals are A, B, X, and Y. All of these chemicals are involved in chemical

reactions as shown below [14].

$$A + X \xrightarrow{k_1} 2X \tag{2.1}$$

$$X + Y \xrightarrow{k_2} 2Y \tag{2.2}$$

$$Y \xrightarrow{k_3} B \tag{2.3}$$

The task is to construct a set of differential equations shows the variations of each chemical with respect to time. Start with the differential equation for the first chemical, that is A. To do this, Firstly identify all chemical reactions that either used as a consumer the chemical A. And then construct the differential equation for chemical A is given below in Figure 2.3

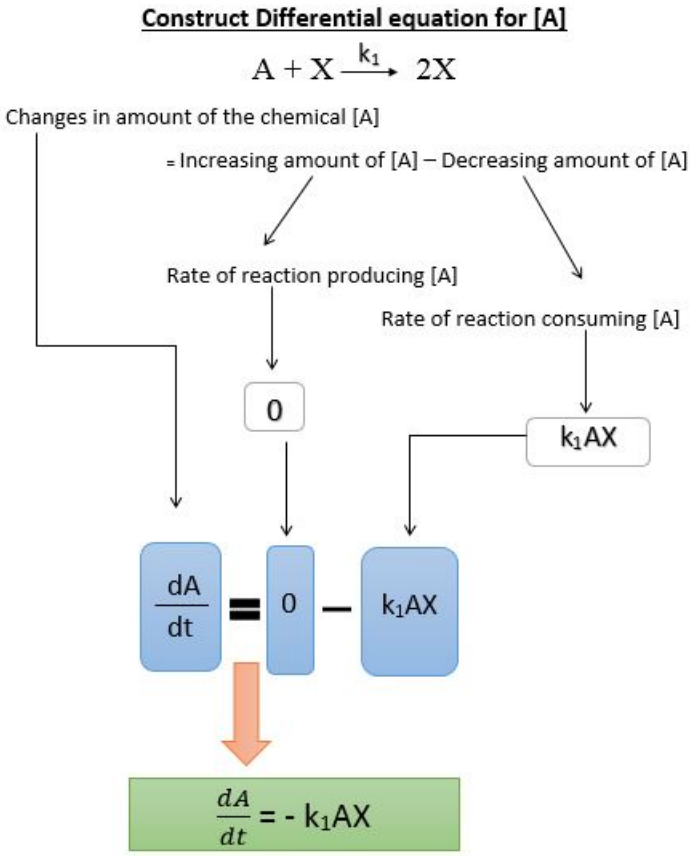


Figure (2.3) Construct the differential equation of first chemical "A"

.

Next, suppose build the differential equations for the second chemical that is B. To do this, first of all, all chemical reactions identify that either they used as a product or consumes the chemical (i.e., recognized all the chemical reactions in which chemical B is used). And then the construct the differential equation according to the governing equation is given below Figure 2.4

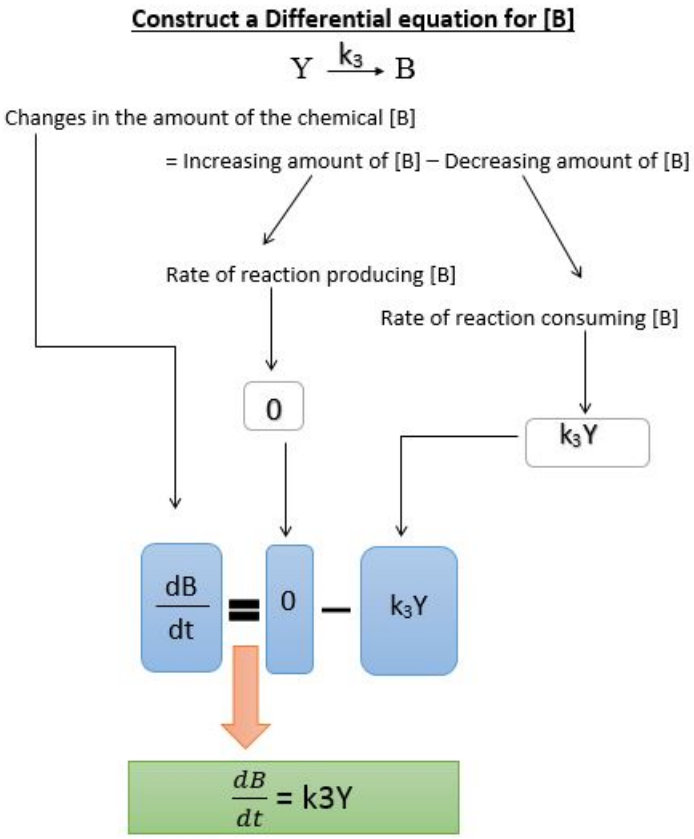


Figure (2.4) Construct the differential equation of second chemical "B"

.

Next, suppose build the differential equations for the third chemical that is X. To do this, first of all, all chemical reactions identify that either they used as a product or consumes the chemical (i.e., recognized all the chemical reactions in which chemical X is used). And then the construct the differential equation according to

the governing equation is given below in Figure 2.5

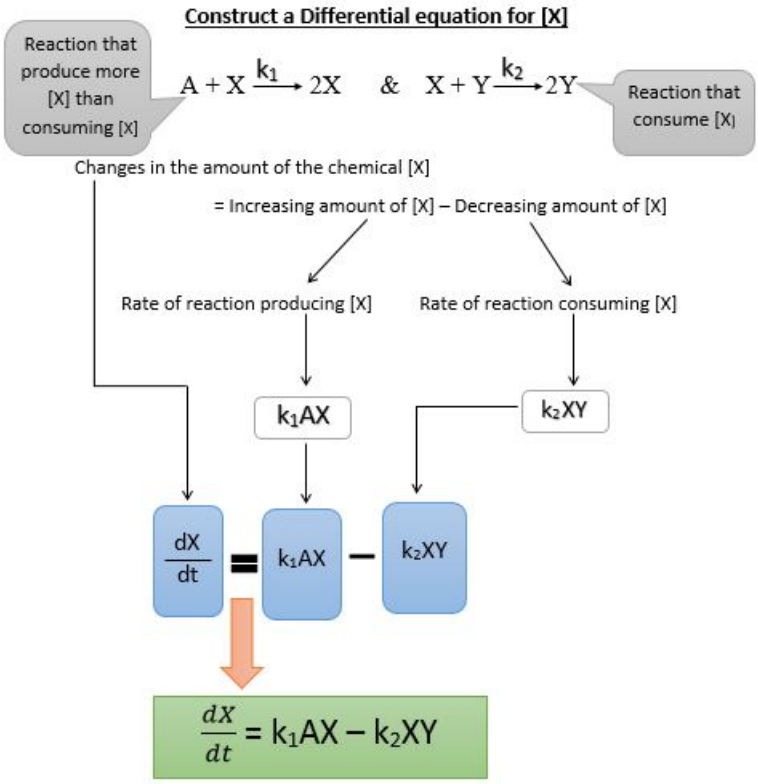


Figure (2.5) Construct the differential equation of third chemical "X"

.

Next, suppose build the differential equations for the fourth and last chemical that is Y. To do this, first of all, all chemical reactions identify that either they used as a product or consumes the chemical (i.e., recognized all the chemical reactions in which chemical Y is used). And then the construct the differential equation according to the governing equation is given below in Figure 2.6

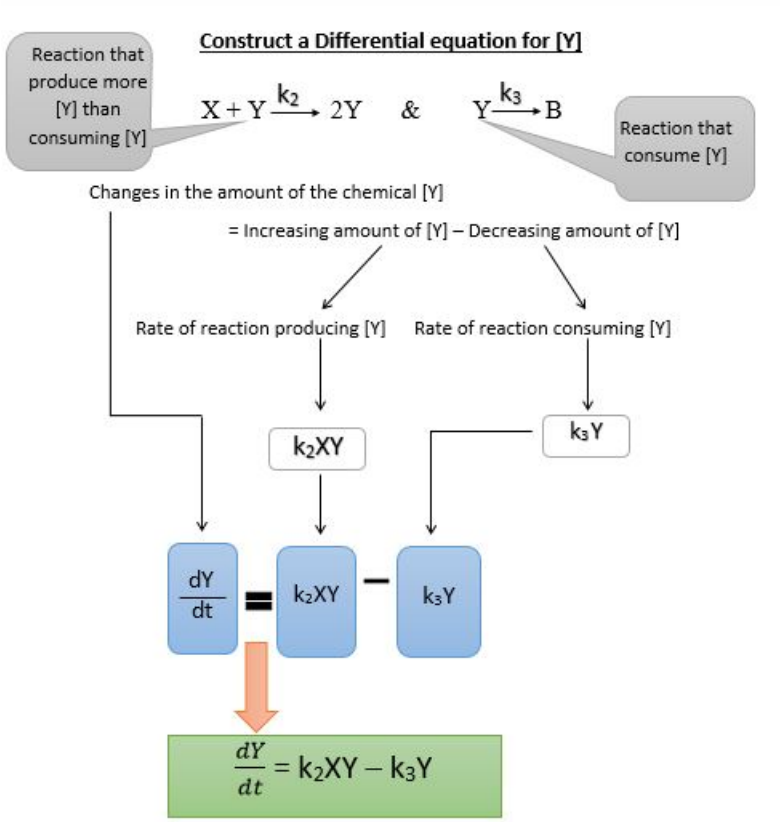


Figure (2.6) Construct the differential equation of fourth chemical "Y"

.

Now all the differential equations that representing the concentration of all the chemicals individually is received. The last step is to mix all the equations together and put them as a simultaneous equations as shown below.

$$\begin{aligned}
\frac{dA}{dt} &= -k_1 AX \\
\frac{dB}{dt} &= -k_3 Y \\
\frac{dX}{dt} &= -k_1 AX - k_2 XY \\
\frac{dY}{dt} &= -k_2 XY - k_3 Y
\end{aligned} \tag{2.4}$$

2.3 State Space Modeling

State space modeling convert higher order differential equations into a set of first order differential equations. It is a mathematical model of a physical system in which there is a set of inputs, outputs and state variables associated by first order differential equation [22]. General nonlinear state space representation of a system is

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t))$$

Now take equation (2.4) as an example of nonlinear state space, so we get

$$\begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \dot{\mathbf{x}}_3 \\ \dot{\mathbf{x}}_4 \end{bmatrix} = \begin{bmatrix} f_1(x(t), u(t)) \\ f_2(x(t), u(t)) \\ f_3(x(t), u(t)) \\ f_4(x(t), u(t)) \end{bmatrix} = \begin{bmatrix} -k_1 x_1 x_3 \\ -k_3 x_4 \\ -k_1 x_1 x_3 - -k_2 x_3 x_4 \\ -k_2 x_3 x_4 - -k_3 x_4 \end{bmatrix} \tag{2.5}$$

where $A = x_1$, $B = x_2$, $X = x_3$ and $Y = x_4$. In case of linear systems, the generalized state space representation is of the form shows in Figure 2.7

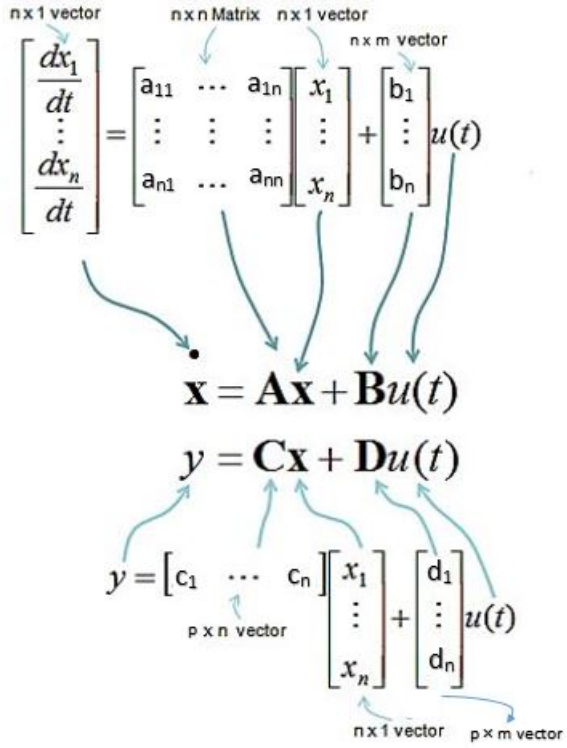


Figure (2.7) Differential equation into state space model

.

In the above figure, the first equation is known as the state equation and it has a first order derivative of the state variables on the left side, and the state variables and inputs, multiplied by matrices on the right side. And the second equation is known as the output equation and it has the output on the left hand side and the state variables and inputs, multiplied by matrices on the right hand side.In these equations:

- **x** is $\mathbf{n} \times \mathbf{1}$ (n rows and 1 column); **x** is called the state vector and it is a function of time.
- **A** is $\mathbf{n} \times \mathbf{n}$ (n rows and n columns); **A** is the state matrix and it is constant.
- **B** is $\mathbf{n} \times \mathbf{m}$ (n rows and m columns); **B** is the input matrix and it is constant.

- **u** is $\mathbf{m} \times \mathbf{1}$ (m rows and 1 column); **u** is the input and in general it is a function of time.
- **C** is $\mathbf{p} \times \mathbf{n}$ (p rows and n columns); **C** is the output matrix.
- **D** is $\mathbf{p} \times \mathbf{m}$ (p rows and m columns); **D** is the direct transition matrix.
- **y** is $\mathbf{p} \times \mathbf{1}$ (p rows and 1 column); **y** is the output of the system and it is a function of time.

Now we represent the example discussed in the previous section in the state space form. Equation (2.4) has a set of nonlinear differential equations. These nonlinear equations can be converted in linear state space form as

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dX}{dt} \\ \frac{dY}{dt} \end{bmatrix} = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -k_3 \\ -k_1 & 0 & -k_2 & 0 \\ 0 & 0 & -k_2 & -k_3 \end{bmatrix} \begin{bmatrix} A \\ B \\ X \\ Y \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t) \tag{2.6}$$

2.4 Techniques of Model Order Reduction

MOR is a computational technique that reduce large scale systems that are represented by a set of ODEs or DAEs, to make its simulation easy and smooth. There are many methods or techniques of MOR that are used to reduce large scale systems. A common approach for model order reduction is projection-based reduction. Some techniques that fall in this category are as following:

- Proper orthogonal decomposition
- Balanced truncation
- Transfer function interpolation
- Krylov subspace method in particular IRKA
- Lumping

In the following we discuss lumping and IKRA techniques for model order reduction. For details on other techniques, we refer [23].

2.5 Lumping

Wei and Kuo in 1960s proposed a methodology for reduction of dynamical systems, that is known as lumping [24]. Lumping remove atleast one set of a state-variable and replace it with a single 'lumped' variable which shows some direct mapping from the original variables in the system.

There are different types of lumping, each type define specific rules of combining the state variables during reduction and their details are discussed in the following:

Proper Lumping And Improper Lumping [25]

In proper lumping every original state of system is presented in only one lumped variable of the reduced model, while in improper lumping every original state of the system can be presented in to one or more then one lumped variables of the reduced model. Figure 2.8 represent the symbolic picture of proper lumping and improper lumping. Proper lumping is a subdivision of original species, in which every subdivision can be reduced in only one independant dynamicaal variable in the reduced model. In literature most of the paper discussed only proper lumping

methodologies. In majority literature, proper lumping methodologies are discussed, which can be required to manage a certain degree of biological interpretability in the reduced structure of the network.

Linear Lumping And Nonlinear Lumping [24]

In lumping when the lumped variable consist of only linear combinations of the original species is known as linear lumping. When the lumped variable in lumping consist of both linear or nonlinear combinations of the combinations of the original species is known as nonlinear lumping [26]. In most of the literature,authors are discussed that linear lumping is similar to the proper lumping and this approach produces a reduced networks that can easy to translate biologically.

Exact Lumping And Approximate Lumping

In lumping when the lumped variable consist of only time-invariant parameters in reduced system that can be exactly mapped to the original species is said to be exact lumping schemes [24] [26]. Exact lumping schemes are divided into three categories that are; proper lumping, improper lumping [27] and semi-proper lumping. In lumping when the lumped variable consist of small group of positive parameters corresponding to different time scale in reduced system that can be exactly mapped to the original species is said to be approximate lumping schemes [28]. There are two types of approximate lumping [29]; linear and nonlinear approximate lumping.

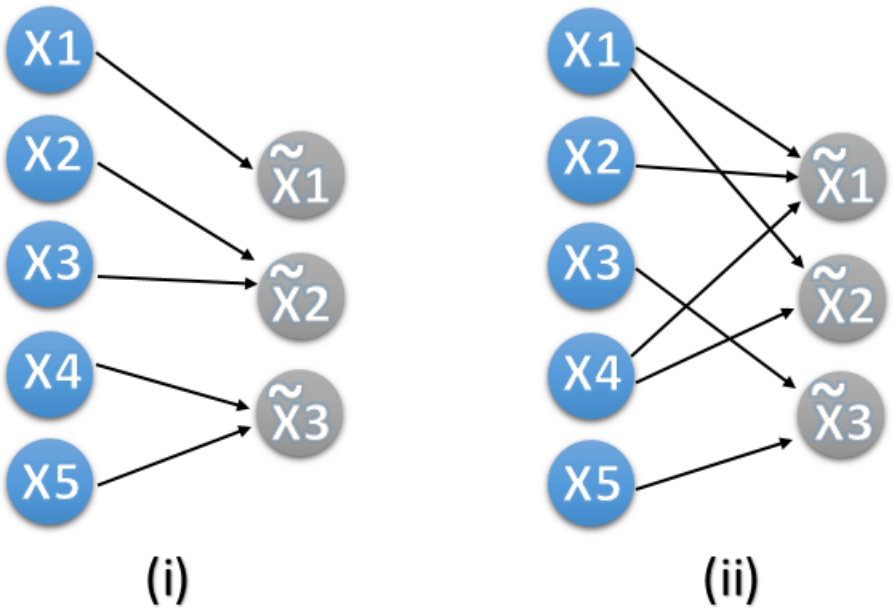


Figure (2.8) Symbolic picture of proper lumping and improper lumping. **I** Proper lumping **II** Improper lumping .

A general state space model are

$$\dot{x}(t) = f(x(t), \mathbf{p}, \mathbf{u}(\mathbf{t})) \tag{2.7}$$

$$y(t) = g(x(t), \mathbf{p}) \tag{2.8}$$

where $\mathbf{u}(t) \in \mathbb{R}^m$ represents a input vector, $\mathbf{y} \in \mathbb{R}^p$ represents the output, $p_i \in \mathbf{p}$ are real proportionality constant that is equal to the corresponding kinetic parameter and the output function $\mathbf{g}(\mathbf{x}(t))$ involve the original state-variables and parameter p. Reduction through any linear projection $\mathbf{L} \in \{0,1\}^{r\times n}$, where each row of $\mathbf{L}$ is pairwise orthogonal. The reduced state-variables $x_r(t)$ can be computed as

$$x_r(t) = \mathbf{L}\mathbf{x}(t) \tag{2.9}$$

The dynamics of the system can be represented through the reduced variables $x_r(t)$.

So the reduced form of the system is

$$x_r(t) = \mathbf{L}f(\mathbf{L_r x_r}(t), \mathbf{p}, \mathbf{u}(t)),\ \mathbf{x_r}(0) = Lx(0) = \mathbf{x_{r0}} \tag{2.10}$$

$$y_r(t) = g(\mathbf{L_r x_r}(t), \mathbf{p}) \tag{2.11}$$

where L_r is the generalized inverse of $\mathbf{L}$ and there are different ways to construct this inverse. In the paper of **Wei and Kuo** [24], they are suggesting to select the L_r that rebuild the fixed state of the system, such that $\hat{\mathbf{x}} = \mathbf{L_r}\hat{\mathbf{x}}_\mathbf{r}$ with $\hat{\mathbf{x}} = \lim_{t \longrightarrow +\infty} \mathbf{x}(t)$. In contradistinction, **Dokoumetzidis and Aarons** [25] follows the work of **Li and Rabitz** [30], and suggested to use the Moore-Penrose inverse $\mathbf{L}^+$ for the reasons of clarity and simplicity of calculation. However, this selection of lumping inverse, has a major impact on the efficiency of the model reduction.

In the literature of recent years lumping are used to reduce a number of biochemical systems. A lumping approach and successive optimization to a 20 dimensional model of yeast glycolysis was applied by **Dano** in 2006 [31]. It was established that this system may be reduced to 8 dimensions by maintaining good accuracy. An algorithmic approach for linear, proper lumping was presented by **Dokoumetzidis and Aarons** in 2009 [25]. This is an optimization-based reduction method, which add two state-variables at every step and testing of each possible pair by the help of simulating the results of reduced model and matching its output with the original. This method was applied to a 26-dimensional model of the $NF -_k B$ signaling pathway. The methodology of **Dano** [31] was applied on a 62-dimensional model that studying the effect of snake vitriol management by **Gulati** in 2009 [32]. It was presented that the 5-dimensional model can be formed which reproduced the original system dynamics to within a maximal relative error of 20%. A lumping style approach termed layer-based reduced modeling was applied by **Koschorreck** in 2007 [33]. Under this approach, finding a lumping is comparatively good for

understanding of the model in order to decompose it into lumpable modules. The ability to switch between particular dimensionality of reduced models depending upon the application and accuracy preferred was presented in 2010 by **Sunnker** [34]. This method is demonstrated via application to a 26-dimensional model of fluorescence production in photosynthesis, which is reduced to 6 dimensions yielding only an insignificant difference in the output profile of the reduced model. They extend their method to the nonlinear model in their second paper [35]. To reduce a model of glycolysis from 9 down to 5 state-variables which still offers an excellent description of the state dynamics then this methodology is used.

2.6 Iterative Rational Krylov Algorithm

Iterative Rational Krylov Algorithm (IRKA) is an interpolatory model reduction technique that link the problem of optimal $\mathcal{H}_2$ approximation to projection [36]. In that paper, the authors address the optimal $\mathcal{H}_2$ approximation of a stable, single-input single-output large-scale dynamical system. They observe that for an n^{th} order linear dynamical system with transfer function

$$G(s) = C(sI - A)^{-1}B$$

a stable r^{th} order reduced system

$$G_r(s) = C_r(sI_r - A_r)^{-1}B_r$$

can be computed with $r << n$ and with the $\mathcal{H}_2$ error satisfying

$$G_r(s) = arg \quad \min_{deg(\hat{G}=r)} \parallel G(s) - \hat{G}(s) \parallel_{\mathcal{H}_2} \tag{2.12}$$

where $\parallel G \parallel_{\mathcal{H}_2} := (\int_{-\infty}^{+\infty} |G(jw)|^2 dw)^{\frac{1}{2}}$.

The problem of computing $G_r(s)$ is lined to Krylov projection. methods. Two projection matrices V $\in \mathbb{R}^{n\times r}$ and W $\in \mathbb{R}^{n\times r}$ are constructed certain with columns

spanning the Krylov subspaces and satisfying $W^TV = I_r$. Then the state matrices of reduced order model are

$$A_r = W^TAV, \quad B_r = W^TB \quad and \quad C_r = CV. \tag{2.13}$$

The problem is how to construct V and W that link this projection framework with the $\mathcal{H}_2$ optimal model reduction problem defined in equation (2.12). To solve this issue, the iterative rational Krylov algorithm has been proposed which efficiently construct the projection matrices V and W through an iterative framework. IRKA is acceptable for large-scale system as it involve matrix vector multiplications only. The rational interpolation concept [Grimme] construct V and W through

Im(V) = Span $\{(\sigma_1 I - A)^{-1}B, ..., (\sigma_r I - A)^{-1}B\}$

Im(W) = Span $\{(\sigma_1 I - A)^{-T}C^T, ..., (\sigma_r I - A)^{-T}C^T\}$

with $W^TV = I_r$.

IRKA use similar rational interpolation framework but these points $\sigma_i's$ are iteratively updated until $\sigma_i's$ are the negative eigenvalues of A_r.

IRKA basic steps

1. Make an initial shift selection σ_i for i = 1,...,r.
2. W = [$(\sigma_1 I - A^T)^{-1}C^T, ..., (\sigma_r I - A^T)^{-1}C^T$]
3. V = [$(\sigma_1 I - A)^{-1}B, ..., (\sigma_r I - A)^{-1}B$]
4. while (not converged)
   a) $A_r = W^TAV$
   b) $\sigma_i \leftarrow -\lambda_i(A_r)$ for i = 1,...,r

c) $\mathrm{W} = [\ (\sigma_1 I - A^T)^{-1} C^T, ..., (\sigma_r I - A^T)^{-1} C^T]$

d) $\mathrm{V} = [\ (\sigma_1 I - A)^{-1} B, ..., (\sigma_r I - A)^{-1} B]$

5. $A_r = W^T A V, \quad B_r = W^T B$ and $C_r = CV$

Chapter 3

Research Methodology

In this chapter we discuss the methodology used to find our results of reduction of biochemical reaction networks. Some basic steps that are followed in the methodology are represented through a flowchart, shown in Figure 3.1.

3.1 Chemical Reaction Network

A CRN consists of a set of reactants, a set of products and a set of reactions. It can be modeled by means of nonlinear, parameter dependent systems of ordinary differential equations. To explain the modeling of chemical reaction, we consider a nonlinear example from the literature [37], where the network consists of 9 species, 6 reactions, 10 kinetic rate constants and 1 input as shown in Figure 3.2. In a simple way we describe this chemical reaction network by a set of chemical equations as

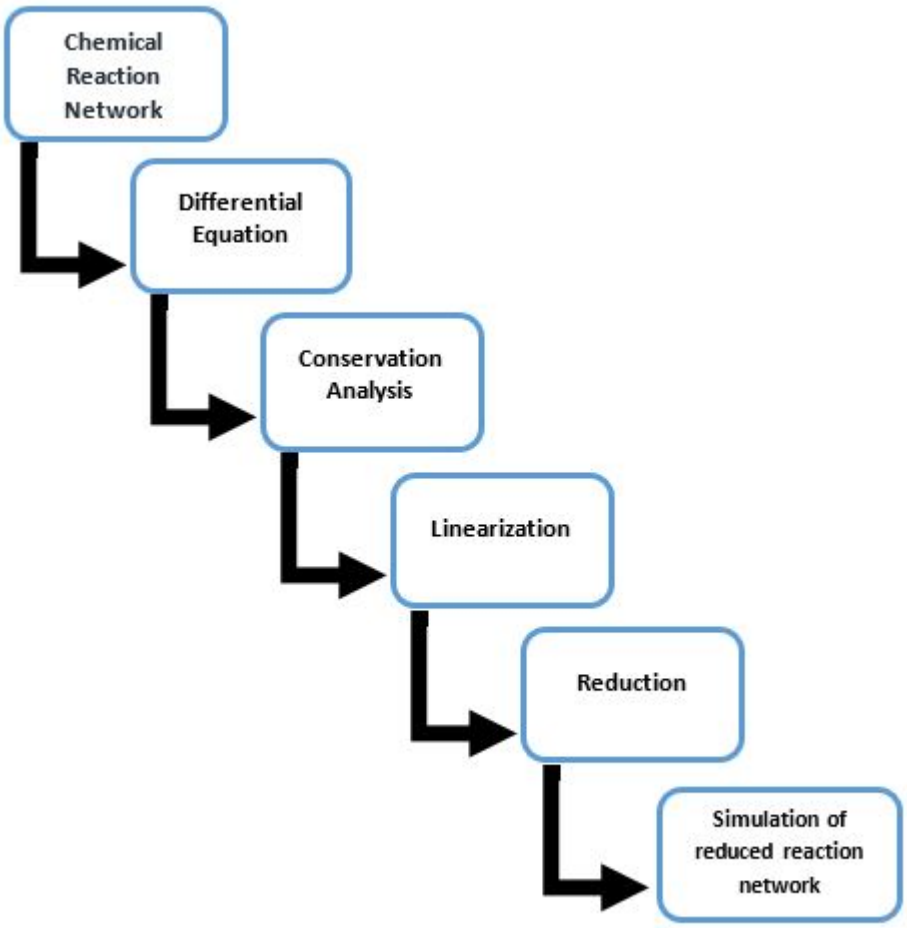


Figure (3.1) Flowchart of my methodology

.

follows :

$$A + B \underset{k_2}{\overset{k_1}{\rightleftarrows}} AB \overset{k_3}{\rightarrow} C + B,$$

$$C + D \underset{k_5}{\overset{k_4}{\rightleftarrows}} CD,$$

$$C + E \underset{k_7}{\overset{k_6}{\rightleftarrows}} CE \overset{k_8}{\rightarrow} A + E,$$

$$A + U \overset{k_9}{\rightarrow} U + F,$$

$$F \overset{k_{10}}{\rightarrow} A,$$

While this is only a test example, it could be assumed that the example represents a process where an enzyme **B** is catalyzing the transformation of a substrate **A** to the form **C**. The enzyme **C** in turn can bind with **E** to revert to **A** or can bind with **D** to undergo degradation. The specie **U** represents a molecule that catalyzes the transformation of specie **A** to specie **F** hence sequestering the substrate from performing the autonomous process described in Figure 3.2.

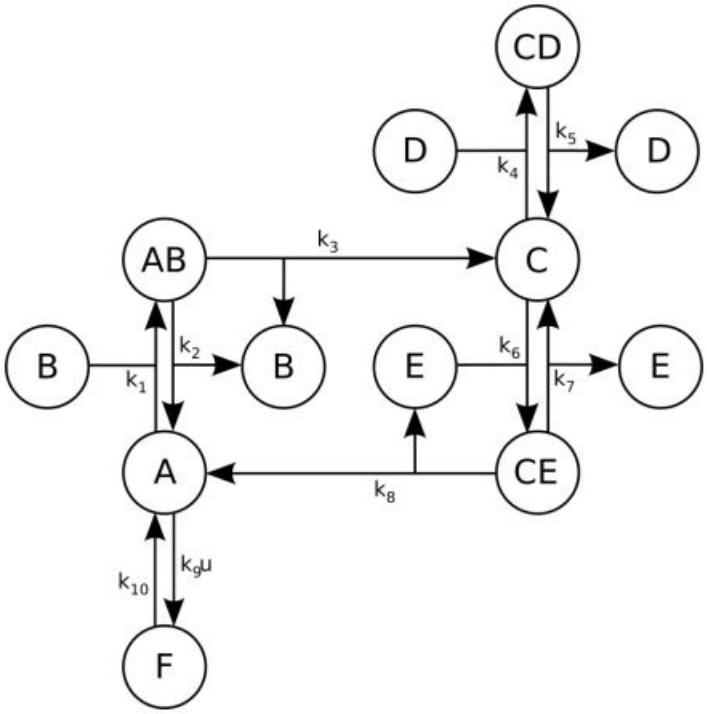


Figure (3.2) A nonlinear, system for the demonstration of model reduction methodologies.

.

3.2 Differential Equation Model

An ordinary differential equation is an equation containing a function of one independent variable and its derivatives. The term "ordinary" is used in contrast to the term partial differential equation which has more than one independent variable. It is a fundamental tool for studying the dynamics of linear as well as nonlinear

systems. To understand the behavior of any chemical system, we model that system in the mathematical form through ODEs via application of the Law of mass action. The set of chemical equations given in section 3.1 are modeled as a set of ODEs, which are as follows

$$\begin{aligned}
\frac{dx_1}{dt} &= k_2x_3 - k_1x_1x_2 + k_8x_8 - Uk_9x_1 + k_{10}x_9 \\
\frac{dx_2}{dt} &= k_2x_3 - k_1x_1x_2 + k_3x_3 \\
\frac{dx_3}{dt} &= -k_2x_3 + k_1x_1x_2 - k_3x_3 \\
\frac{dx_4}{dt} &= k_3x_3 + k_5x_5 - k_4x_4x_6 + k_7x_8 - k_6x_4x_7 \\
\frac{dx_5}{dt} &= -k_5x_5 + k_4x_4x_6 \\
\frac{dx_6}{dt} &= k_5x_5 - k_4x_4x_6 \\
\frac{dx_7}{dt} &= k_7x_8 - k_6x_4x_7 + k_8x_8 \\
\frac{dx_8}{dt} &= -k_7x_8 + k_6x_4x_7 - k_8x_8 \\
\frac{dx_9}{dt} &= Uk_9x_1 - k_{10}x_9
\end{aligned}$$

Where the state-variables are described as [A] = $x_1(t)$, [B] = $x_2(t)$, [AB] = $x_3(t)$, [C] = $x_4(t)$, [CD] = $x_5(t)$, [D] = $x_6(t)$, [E] = $x_7(t)$, [CE] = $x_8(t)$, and [F] = $x_9(t)$. Moreover, let u represents the concentration of the input molecule U , that is [U]=u. In matrix-vector form, we have

$$\dot{x}(t) = \mathcal{S}\mathbf{v}(\mathbf{x}(t), \mathbf{k}), \tag{3.1}$$

where

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \; v = \begin{pmatrix} k_2 x_3 - k_1 x_1 x_2 \\ k_3 x_3 \\ k_5 x_5 - k_4 x_4 x_6 \\ k_7 x_8 - k_6 x_4 x_7 \\ k_8 x_8 \\ U k_9 x_1 \\ k_{10} x_9 \end{pmatrix}$$

Also let us define a single output, y = $x_6(t)$= [CD]. So this system can be expressed clearly in a state-space form

$$\dot{x}(t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{g}(\mathbf{x}, t)u,$$

$$y = x_6(t).$$

where

$$\mathbf{f}(\mathbf{x}, t) = \begin{bmatrix} k_2 x_3 - k_1 x_1 x_2 + k_8 x_8 - k_{10} x_9 \\ k_2 x_3 - k_1 x_1 x_2 + k_3 x_3 \\ -k_2 x_3 + k_1 x_1 x_2 - k_3 x_3 \\ k_3 x_3 + k_5 x_5 - k_4 x_4 x_6 + k_7 x_8 - k_6 x_4 x_7 \\ -k_5 x_5 + k_4 x_4 x_6 \\ k_5 x_5 - k_4 x_4 x_6 \\ k_7 x_8 - k_6 x_4 x_7 + k_8 x_8 \\ -k_7 x_8 + k_6 x_4 x_7 - k_8 x_8 \\ k_{10} x_9 \end{bmatrix}, \mathbf{g}(\mathbf{x}, t) = \begin{bmatrix} -k_9 x_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ k_9 x_1 \end{bmatrix} u.$$

The initial values of variables are as follows:

$x_1(0) = x_4(0) = 1.54, x_2(0) = x_7(0) = 0.566, x_3(0) = x_8(0) = 0.435, x_5(0) = 6.06, x_6(0) = 3.94$, and $x_9(0) = 0$. These represents the steady-state of the system under the condition, U = 0. The value of the parameters are fixed and are given in Table 3.1

Table (3.1) The set of parameter values associated with the nonlinear example model as defined by equation (3.1)

Parameters	**Values**
k_1	1
k_2	1
k_3	1
k_4	1
k_5	1
k_6	1
k_7	1
k_8	1
k_9	100
k_{10}	1

3.3 Conservation Analysis

It is easy to see that, in chemical reactions the rate of rise in concentration of one specie is exactly equal to the rate of decay in concentration of another specie. This means that some linear combination of the rate of change of specific species will be zero. That is

$$\Gamma \dot{\mathbf{x}}(t) = 0 \tag{3.2}$$

where Γ represents the conservation matrix of size $h \times n$. By integration, we have

$$\Gamma \mathbf{x}(\mathbf{t}) = \mathbf{c}, \tag{3.3}$$

These conservation relations can be solved for some targeted species dependent on other species such that the dependent species are completely replaced in the original model. To obtain this partition $\mathbf{x}$ into two subsets: $\mathbf{x}_d$ and $\mathbf{x}_i$. $\mathbf{x}_d$ is h dimensional subset of the species with every element included in a given conservation relation and describe the dependent specie. And $\mathbf{x}_i$ is $(n-h)$ dimensional subset that represents the independent species. Thus

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_i(t) \end{bmatrix} \tag{3.4}$$

Then from the equation (3.3)

$$\Gamma \begin{bmatrix} \mathbf{x}_d(t) \\ \mathbf{x}_i(t) \end{bmatrix} = \mathbf{c} \tag{3.5}$$

This is a system of linear equations and so if Γ is expressed in reduced row echelon form, such that

$$\Gamma = \begin{bmatrix} I_h & N_0 \end{bmatrix}, \tag{3.6}$$

Where I_h is an h dimensional identity matrix and N_0 is an $h \times (n-h)$ matrix. Simplification implies

$$\mathbf{x}_d(t) = \mathbf{c} - N_0 \mathbf{x}_i(t). \tag{3.7}$$

This indicates that the subset of dependent species $\mathbf{x}_d$ can be excluded from the governing system of ODEs by substituting it in the appropriate element of equation (3.7). A system exhibiting conservation relations can be expressed in the form of a semi-explicit system of DAEs, such that

$$\dot{\mathbf{x}}_i = S_i \mathbf{v}(\mathbf{x}_i(t)), \tag{3.8}$$

$$\mathbf{x}_d(t) = N_0\mathbf{x}_i(t) - \mathbf{c}, \tag{3.9}$$

where equation (3.9) has been utilized in equation (3.8) to get a system of ODEs such that state-variables $\mathbf{x}_d$ are no longer given. And S_i represents the row of the stoichiometric matrix proportional to the independent state-variables $\mathbf{x}_i$.

To identify the conservation matrix Γ, especially for large systems, a more algorithmic approach is possible through the stoichiometric form the model. Decompose the stoichiometric matrix through the same partition as the set of species leads to

$$\begin{pmatrix} \dot{\mathbf{x}}_d(t) \\ \dot{\mathbf{x}}_i(t) \end{pmatrix} = \begin{pmatrix} \mathbf{S}_d \\ \mathbf{S}_i \end{pmatrix} \mathbf{v}(\mathbf{x}_d(t), \mathbf{x}_i(t)). \tag{3.10}$$

However, through differentiation of equation (3.7), we have

$$\dot{\mathbf{x}}_d(t) = -N_0\dot{\mathbf{x}}_i(t) = -N_0\mathbf{S}_i\mathbf{v}(\mathbf{x}_d(t), \mathbf{x}_i(t)). \tag{3.11}$$

So, $\mathbf{S}_d = -N_0\mathbf{S}_i$. As, conservation relations can be found by finding the left null space Z_n of S (i.e by finding the null space of S^T) such that

$$Z_n = \{\mathbf{z} \in \mathcal{R}^n | \mathbf{S^T z} = 0\}, \tag{3.12}$$

and so $Z_n^T S = \mathbf{0}$. This implies that

$$Z_n^T Sv(x(t)) = 0 = Z_n^T(\dot{x}(t)) \tag{3.13}$$

and then by comparison to equation (3.2), it is clear that

$$Z_n^T = \Gamma, \tag{3.14}$$

such that the transpose of the left null space of the stoichiometry matrix is equal to the conservation matrix. In case of the biochemical system represented by equation

(3.1), calculation of the left null-space is

$$Z_n = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.15}$$

which implies that we have conservation relationships

$$B_T = x_2(t) + x_3(t), \tag{3.16a}$$

$$C_T = x_5(t) + x_6(t), \tag{3.16b}$$

$$E_T = x_7(t) + x_8(t), \tag{3.16c}$$

$$S_T = x_1(t) + x_3(t) + x_4(t) + x_5(t) + x_8(t) + x_9(t) \tag{3.16d}$$

Finally, to get the simplified realization, just substituting these conservation relations into the system implies

$$\begin{aligned}\frac{dx_1(t)}{dt} &= k_2x_3(t) + k_{10}(x_1(t) - S_T + x_3(t) + x_4(t) + x_5(t) + x_8(t)) \\ &\quad + k_8x_8(t) - Uk_9x_1(t) - k_1x_1(t)(B_T - x_3(t)),\end{aligned} \tag{3.17a}$$

$$\frac{dx_3(t)}{dt} = k_1x_1(t)(B_T - x_3(t)) - k_3x_3(t) - k_2x_3(t), \tag{3.17b}$$

$$\begin{aligned}\frac{dx_4(t)}{dt} &= k_3x_3(t) + k_5x_5(t) + k_7x_8(t) - k_4x_4(t)(C_T - x_5(t)) - \\ &\quad k_6x_4(t)(E_T - x_8(t)),\end{aligned} \tag{3.17c}$$

$$\frac{dx_5(t)}{dt} = k_4x_4(t)(C_T - x_5(t)) - k_5x_5(t), \tag{3.17d}$$

$$\frac{dx_8(t)}{dt} = k_6x_4(t)(E_T - x_8(t)) - k_8x_8(t) - k_7x_8(t). \tag{3.17e}$$

The initial values of B_T, C_T, E_T and S_T are given in Table 3.2

Table (3.2) The initial values of conservation relationships as defined in equation (3.17)

Parameters	**Values**
B_T	1
C_T	10
E_T	1
S_T	10

3.4 Linearization

The process of taking the slope of a nonlinear function with respect to all variables and creating a linear representation at that specific point is called linearization. We use linearization to allow the use of linear theory for analysis and design in a specific range. Consider a nonlinear differential equation that is obtained from

balance equations with the input u and output y .

$$\frac{dy}{dt} = f(y, u)$$

The right hand side of the equation is linearized by a Taylor series extension, with utilizing just the initial two terms.

$$\frac{dy}{dt} = f(y, u) \approx f(\bar{y}, \bar{u}) + \frac{\partial f}{\partial y} \mid_{\bar{y},\bar{u}} (y - \bar{y}) + \frac{\partial f}{\partial u} \mid_{\bar{y},\bar{u}} (u - \bar{u})$$

If the values of $\bar{u}$ and $\bar{y}$ are selected at steady state conditions, then $f(\bar{y}, \bar{u}) = 0$ because at steady state $\frac{dy}{dt} = 0$. Deviation variables are defined as $y^{'} = y - \bar{y}$ and $u^{'} = u - \bar{u}$ to simplify the final linearized expression. A deviation variable is a change from the nominal steady state conditions. The derivatives of the deviation variable is describe as $\frac{dy^{'}}{dt} = \frac{dy}{dt}$ because $\frac{d\bar{y}}{dt} = 0$ in $\frac{dy^{'}}{dt} = \frac{d(y - \bar{y})}{dt} = \frac{dy}{dt} - \frac{d\bar{y}}{dt}$. If there are additional variables such as a disturbance variable d then it is added as another phase in deviation variable form $d^{'} = d - \bar{d}$

$$\frac{d\bar{y}}{dt} = \alpha\bar{y} + \beta\bar{u} + \gamma\bar{d}$$

The values of α, β and γ are constants and the partial derivatives of $f(y, u, d)$ analyzed at steady state conditions.

$$\alpha = \frac{\partial f}{\partial y} \mid_{\bar{y},\bar{u},\bar{d}} \quad \beta = \frac{\partial f}{\partial u} \mid_{\bar{y},\bar{u},\bar{d}} \quad \gamma = \frac{\partial f}{\partial d} \mid_{\bar{y},\bar{u},\bar{d}}$$

Now we apply this method to linearize the nonlinear system obtained after conservation. The systems consists of ODEs that can be linearized around a given state $\mathbf{x}_c$ by Jacobian matrix

$$\mathbf{J}_{\mathbf{x}_c} = \mathbf{SE} \mid_{\mathbf{x}(t)=\mathbf{x}_c} \tag{3.18}$$

In which, the matrix E is represented as the elasticity matrix, with entries

$$\mathbf{E} = \{e_{ij} = \frac{\partial v_i(\mathbf{x}, \mathbf{p})}{\partial x_j}\} \tag{3.19}$$

So by the first Taylor series extension, the system can be approximated in the neighborhood of $\mathbf{x}_c$ by

$$\dot{\mathbf{x}}(t) \approx \mathbf{Sv}(\mathbf{x}_c, \mathbf{p}) + \mathbf{J}_{\mathbf{x}_c}(\mathbf{x}(t) - \mathbf{x}_c) \tag{3.20}$$

The Jacobian matrix with the initial condition can be written as

$$\mathbf{J}_{\mathbf{x}_c} = \begin{pmatrix} -1.57 - 100u & 1.54 & -1 & -1 & 0 \\ 0.566 & -3.54 & 0 & 0 & 0 \\ 0 & 1 & -4.51 & 2.54 & 2.45 \\ 0 & 0 & 3.94 & -2.54 & 0 \\ 0 & 0 & 0.566 & 0 & -3.54 \end{pmatrix} \tag{3.21}$$

The linear state space form of equation (3.21) is written as

$$\dot{x} = \underbrace{\begin{bmatrix} -1.57 & 1.54 & -1 & -1 & 0 \\ 0.566 & -3.54 & 0 & 0 & 0 \\ 0 & 1 & -4.51 & 2.54 & 2.45 \\ 0 & 0 & 3.94 & -2.54 & 0 \\ 0 & 0 & 0.566 & 0 & -3.54 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_3 \\ x_4 \\ x_5 \\ x_8 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} -100 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{B} u \tag{3.22}$$

$$y = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} x_1 \\ x_3 \\ x_4 \\ x_5 \\ x_8 \end{bmatrix}}_{x}$$

3.5 Model Order Reduction

As discussed before, model order reduction is a technique for reducing the computational complexity of mathematical models in numerical simulations. As such it is almost related to the idea of surrogate model with applications in all areas of mathematical modeling [23]. So the model reduction is to compute a simpler model that reduce the set of state-variables $\mathbf{x} \in \mathbb{R}^n$ such that $r < n$. There are number of techniques that are used to reduced biochemical reaction networks but we used IRKA to reduced the system. IRKA is a robust model reduction technique which is used to reduce stable linear dynamical systems [40]. A brief outline on the working of IRKA is given below

Algorithm IRKA. *Iterative Rational Krylov Algorithm*

Given a full-order system with transfer function `H(s)`, a reduced order r, and convergence tolerance **tol**, the following steps are followed.

1. Make an initial selection of r distinct interpolation points, $\{\mathrm{s_i}\}_1^r$, that are closed under complex conjugation.

2. Construct $\mathbf{V_r}$ and $\mathbf{W_r}$.
   $V_r = [\ (\sigma_1 I - A)^{-1}B, ..., (\sigma_r I - A)^{-1}B]$
   $W_r = (W_r^T V)^{-T}$ (to make $W_r^T V = I_r$)

3. while (relative change in $\{s_i\} > tol$)

   a.) $\mathbf{A_r} = (\mathbf{W}_r^T \mathbf{V}_r)^{-1} \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$.

   b.) Solve $r \times r$ eigenvalue problem $\mathbf{A_r u} = \lambda \mathbf{u}$ and assign $s_i \leftarrow \lambda_i(\mathbf{A_r})$ for $i = 1, \ldots, r$.

   c.) Update $\mathbf{V_r}$ and $\mathbf{W_r}$ with new $s_i's$.

4. On convergence compute reduced system state matrices $\mathbf{A_r} = (\mathbf{W}_r^T\mathbf{V}_r)^{-1}\mathbf{W}_r^T\mathbf{A}\mathbf{V}_r$, $\mathbf{b_r} = (\mathbf{W}_r^T\mathbf{V}_r)^{-1}\mathbf{W}_r^T\mathbf{b}$ and $\mathbf{c_r} = \mathbf{V}_r^T\mathbf{c}$

Chapter 4

Results and Discussions

In this chapter, we discuss the results of MOR for system of Chemical Reactions. In particular we have applied the model order reduction technique, IRKA, after linearizing the model, representing the biochemical reaction network. The results are compared with lumping, a well used model reduction technique for system of chemical reactions.

4.1 Linearized Model

The linearized model discussed in chapter 3 has a special form that can be written in the standard form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

,

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$$

, with

$$\mathbf{A} = \begin{bmatrix} -101.9573 & 0.0892 & -1.0000 & -1.0000 & 0 \\ 0.9573 & -2.0892 & 0 & 0 & 0 \\ 0 & 1.0000 & -10.1356 & 1.0895 & 1.0895 \\ 0 & 0 & 9.1784 & -1.0895 & 0 \\ 0 & 0 & 0.9572 & 0 & -2.0895 \end{bmatrix}$$

,

$$\mathbf{B} = \begin{bmatrix} -8.9181 & 9.9962 \\ 0 & 0.0038 \\ 0 & -0.0774 \\ 0 & 0.0735 \\ 0 & 0.0038 \end{bmatrix}$$

,

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } D = 0.$$

It is clear from the size of B that u(t) involve two inputs with second fixed to 1. The first input is taken as a unit step function but it can be changed to any other signal. With the above C, we have $y(t) = x_5(t)$. If we want to observe the concentration of $x_1(t)$, we will choose

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

.

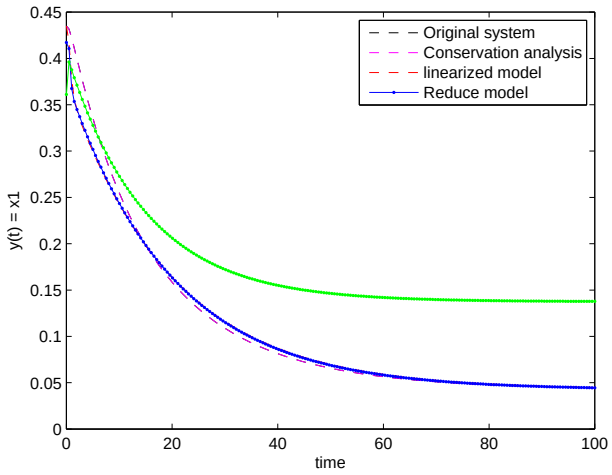
4.2 Lumping and IRKA Techniques

We are implementing the two reduction techniques, the lumping method and IRKA in MATLAB version 2015. We are using the reduced size of 3 and 2 with two different outputs, $\mathbf{y}(t) = x_1$ and $\mathbf{y}(t) = x_5$. In each case, the response of the original and the reduced systems are plotted along with the approximation error. Also the computational time are shown for both lumping and IRKA. We first show the results of reduction of order of 3 and then of order 2.

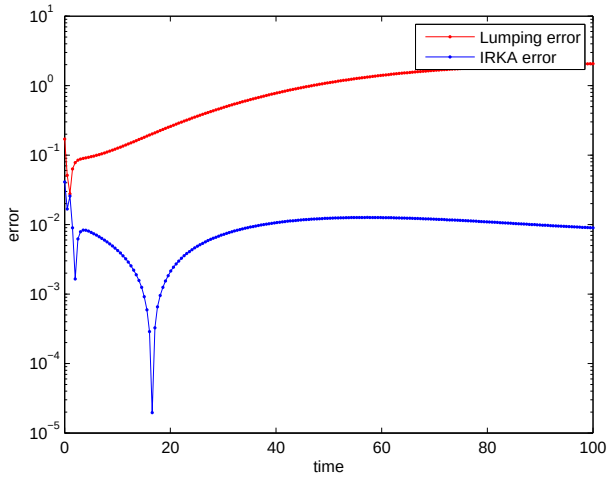
4.3 Reduction to order 3

We consider the chemical system of size 9 and observe tow outputs $\mathbf{y}(t) = x_1$ and $\mathbf{y}(t) = x_5$. This system is reduced to order 3 with both IRKA and lumping. It

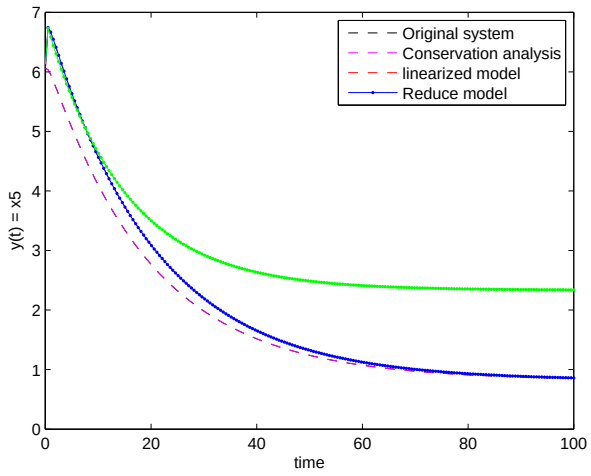
is observed that when we reduce the model with $\mathbf{y}(\mathrm{t})=\mathbf{x}_1(\mathrm{t})$ IRKA shows better approximation error and less computational time as compared to lumping. Similar behavior is obtained for $\mathbf{y}(\mathrm{t})=\mathbf{x}_5(\mathrm{t})$. These results are shown in Figure 4.1.



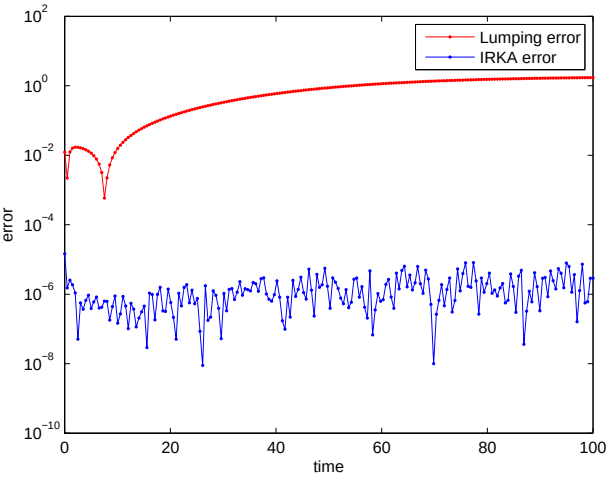
(a) Actual response when $\mathrm{y}(\mathrm{t}) = x_1$



(b) Relative error when $\mathrm{y}(\mathrm{t}) = x_1$



(c) Actual response when $\mathrm{y}(\mathrm{t}) = x_5$



(d) Relative error when $\mathrm{y}(\mathrm{t}) = x_5$

Figure (4.1) Change in concentration of [A] ($y(t) = x_1$) and [CD] ($y(t) = x_5$) using actual nonlinear systems, linear system and reduced systems(order 3) with both IRKA and lumping

Notice that for the same size of the reduced system, IRKA outperforms the lumping technique. The computational time and percentage decrease in the simulation time of original system are shown in Table 4.1.

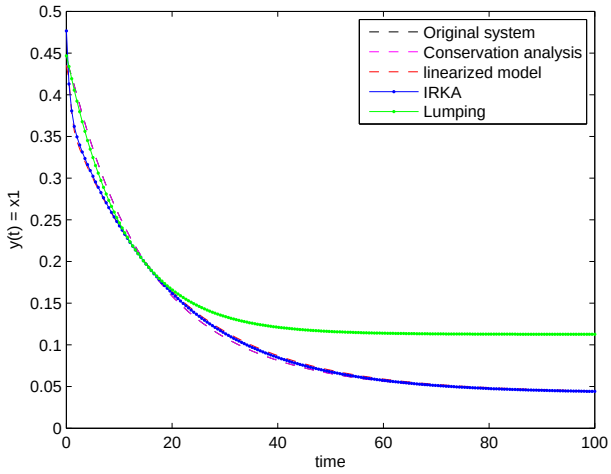
Table (4.1) Comparison of computational time at 3 dimensional model

	Computational time	**Percentage Decreases**
Original System	0.860	—
IRKA when $\mathbf{y(t) = x_1}$	0.084	90%
IRKA when $\mathbf{y(t) = x_5}$	0.063	93%
Lumping when $\mathbf{y(t) = x_1}$	0.249	71%
Lumping when $\mathbf{y(t) = x_5}$	0.267	69%

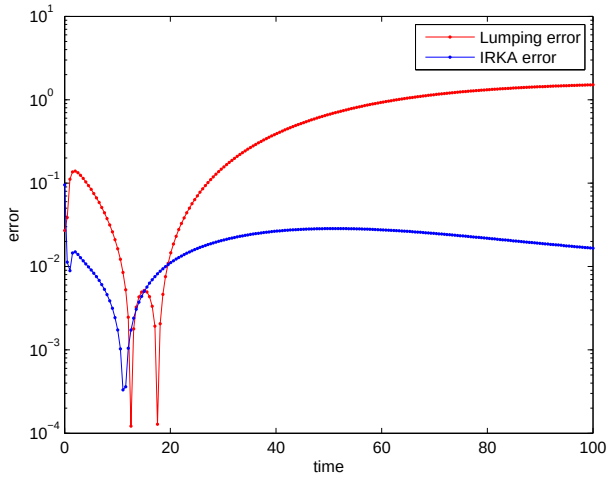
It is clear from the Table 4.1 that the computational time of IRKA is much better then the computational time of lumping in both cases. So IKRA is performing much better then lumping method for the reduction of biochemical systems.

4.4 Reduction to order 2

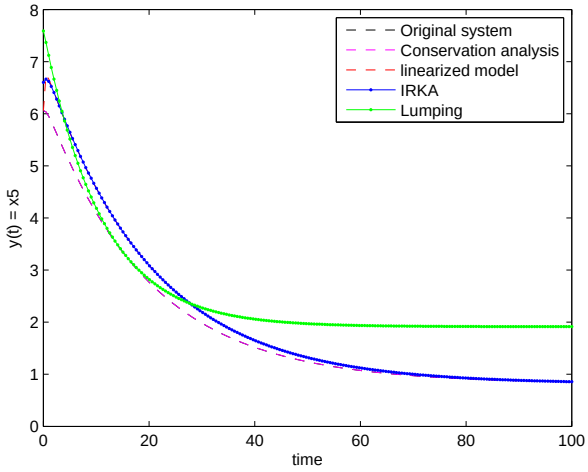
We consider the chemical system of size 9 and observe tow outputs $\mathbf{y}(t) = x_1$ and $\mathbf{y}(t) = x_5$. This system is reduced to order 2 with both IRKA and lumping. It is observed that when we reduce the model with $\mathbf{y}$(t)=$\mathbf{x}_1$(t) IRKA shows better approximation error and less computational time as compared to lumping. Similar behavior is obtained for $\mathbf{y}$(t)=$\mathbf{x}_5$(t). These results are shown in Figure 4.2.



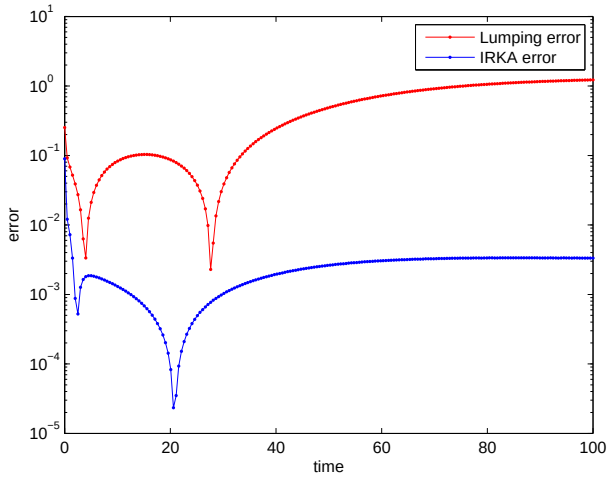
(a) Actual response when y(t) = x_1



(b) Relative error when y(t) = x_1



(c) Actual response when y(t) = x_5



(d) Relative error when y(t) = x_5

Figure (4.2) Change in concentration of [A] ($y(t) = x_1$) and [CD] ($y(t) = x_5$) using actual nonlinear systems, linear system and reduced systems(order 2) with both IRKA and lumping

Notice that for the same size of the reduced system, IRKA outperforms the lumping technique. The computational time and percentage decrease in the simulation time of original system are shown in Table 4.2.

Table (4.2) Comparison of computational time at 2 dimensional model

	Computational time	**Percentage Decreases**
Original System	0.851	—
IRKA at $\mathbf{y(t) = x_1}$	0.093	89%
IRKA at $\mathbf{y(t) = x_5}$	0.024	97%
Lumping at $\mathbf{y(t) = x_1}$	0.247	71%
Lumping at $\mathbf{y(t) = x_5}$	0.258	70%

It is clear from the Table 4.2 that the computational time of IRKA is much better then the computational time of lumping in both cases. So IKRA is performing much better then lumping method for the reduction of biochemical systems.

Chapter 5
Conclusion and Future work

In this chapter, we present the conclusions based on the observations of our results and also show some interesting future research directions.

5.1 Conclusion

For an example of biochemical system taken from literature, the applicability of IRKA has been tested and compared with lumping technique to obtain reduced order representation of the system. The methodology also involve conservation analysis and linearization. It is observed that linearization at steady state give us better result as compared to linearization at initial state. Once we have linearized model, we can perform the reduction techniques on the linearized model. The advantage of IRKA is that, it can be extended to very large-scale settings because it involves only matrix vector multiplications. It is observed that for the biochemical system of size 9, the computational time for construction as well as simulation time of reduced order model via IRKA is **89%** less than the simulation time of full order model. In case of lumping, the simulation time is **71%** less than the simulation time of full order models. Also the approximation error for the IRKA technique is much better as compared to the lumping technique. So IRKA is performing much better than the lumping technique.

5.2 Future Work

An important future work is the implementation of the reduction technique IRKA on an actual large scale biochemical system. That is, to utilize a framework which can be used to analyze the reactions that involve different type of diseases like chromosomal abnormalities, muscular dystrophy and polycystic kidney disease. Since there are some nonlinear versions of IRKA, it will also be important to see the utilization of nonlinear reduction techniques on biochemical system, avoiding approximation in linearization of the actual system.

Bibliography

[1] G. Lente, *Deterministic kinetics in chemistry and systems biology: the dynamics of complex reaction networks*. Springer, 2015.

[2] S. K. Hahl and A. Kremling, "A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: On fixed points, means, and modes," *Frontiers in genetics*, vol. 7, p. 157, 2016.

[3] G. Craciun and M. Feinberg, "Multiple equilibria in complex chemical reaction networks: Ii. the species-reaction graph," *SIAM Journal on Applied Mathematics*, vol. 66, no. 4, pp. 1321–1338, 2006.

[4] J. Ross and A. P. Arkin, "Complex systems: from chemistry to systems biology," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6433–6434, 2009.

[5] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Applied numerical mathematics*, vol. 43, no. 1-2, pp. 9–44, 2002.

[6] C. Himpe, "emgrthe empirical gramian framework," *Algorithms*, vol. 11, no. 7, p. 91, 2018.

[7] C. Gu, "Qlmor: A new projection-based approach for nonlinear model order reduction," in *Computer-Aided Design-Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*, pp. 389–396, IEEE, 2009.

[8] A. Munjal, J.-M. Philippe, E. Munro, and T. Lecuit, "A self-organized biomechanical network drives shape changes during tissue morphogenesis," *Nature*, vol. 524, no. 7565, p. 351, 2015.

[9] R. Choudhary and K. Ahuja, "Stability analysis of bilinear iterative rational krylov algorithm," *Linear Algebra and its Applications*, vol. 538, pp. 56–88, 2018.

[10] L. Fortuna, G. Nunnari, and A. Gallo, *Model order reduction techniques with applications in electrical engineering*. Springer Science & Business Media, 2012.

[11] C. Nowakowski, J. Fehr, M. Fischer, and P. Eberhard, "Model order reduction in elastic multibody systems using the floating frame of reference formulation," *IFAC Proceedings Volumes*, vol. 45, no. 2, pp. 40–48, 2012.

[12] J. Proppe, T. Husch, G. N. Simm, and M. Reiher, "Uncertainty quantification for quantum chemical models of complex reaction networks," *Faraday discussions*, vol. 195, pp. 497–520, 2017.

[13] M. W. Baig, "Effect of strength of gravitational field on the rate of chemical reactions," *arXiv preprint arXiv:1708.05285*, 2017.

[14] S. Zarra, D. M. Wood, D. A. Roberts, and J. R. Nitschke, "Molecular containers in complex chemical systems," *Chemical Society Reviews*, vol. 44, no. 2, pp. 419–432, 2015.

[15] P. M. Gschwend *et al.*, *Environmental organic chemistry*. John Wiley & Sons, 2016.

[16] E. O. Voit, H. A. Martens, and S. W. Omholt, "150 years of the mass action law," *PLoS computational biology*, vol. 11, no. 1, p. e1004012, 2015.

[17] S. A. J. Marsden, L. S. S. Wiggins, L. Glass, R. Kohn, and S. Sastry, *Interdisciplinary Applied Mathematics*, vol. 3. Springer, 1993.

[18] G.-J. Cheng, X. Zhang, L. W. Chung, L. Xu, and Y.-D. Wu, "Computational organic chemistry: bridging theory and experiment in establishing the mechanisms of chemical reactions," *Journal of the American Chemical Society*, vol. 137, no. 5, pp. 1706–1725, 2015.

[19] S. Mondal and B. Mandal, "Procedures for obtaining characteristic polynomials of the kinetic graphs of reversible reaction networks," *Bulletin of the Chemical Society of Japan*, vol. 91, no. 4, pp. 700–709, 2018.

[20] Y. Ding, D. Ceglarek, J. Shi, *et al.*, "Modeling and diagnosis of multistage manufacturing processes: part i: state space model," in *Proceedings of the 2000 Japan/USA symposium on flexible automation*, pp. 23–26, 2000.

[21] D. Rowell, "State-space representation of lti systems," *URL: http://web. mit. edu/2.14/www/Handouts/StateSpace. pdf*, 2002.

[22] V. G. Kulkarni, *Modeling and analysis of stochastic systems*. Chapman and Hall/CRC, 2016.

[23] A. C. Antoulas, *Approximation of large-scale dynamical systems*, vol. 6. Siam, 2005.

[24] J. Wei and J. C. Kuo, "Lumping analysis in monomolecular reaction systems. analysis of the exactly lumpable system," *Industrial & Engineering chemistry fundamentals*, vol. 8, no. 1, pp. 114–123, 1969.

[25] A. Dokoumetzidis and L. Aarons, "Proper lumping in systems biology models," *IET systems biology*, vol. 3, no. 1, pp. 40–51, 2009.

[26] A. S. Tomlin, G. Li, H. Rabitz, and J. Toth, "A general analysis of approximate nonlinear lumping in chemical kinetics. ii. constrained lumping," *The Journal of chemical physics*, vol. 101, no. 2, pp. 1188–1201, 1994.

[27] H. Li *et al.*, *Applications of lumping kinetics methodology to plastic waste recovery via pyrolysis*. PhD thesis, Heriot-Watt University, 2017.

[28] G. Li, A. S. Tomlin, H. Rabitz, and J. Tóth, "Determination of approximate lumping schemes by a singular perturbation method," *The Journal of chemical physics*, vol. 99, no. 5, pp. 3562–3574, 1993.

[29] G. Großmann, C. Kyriakopoulos, L. Bortolussi, and V. Wolf, "Lumping the approximate master equation for multistate processes on complex networks," *arXiv preprint arXiv:1804.02981*, 2018.

[30] G. Li and H. Rabitz, "A general analysis of approximate lumping in chemical kinetics," *Chemical engineering science*, vol. 45, no. 4, pp. 977–1002, 1990.

[31] S. Danø, M. F. Madsen, H. Schmidt, and G. Cedersund, "Reduction of a biochemical model with preservation of its basic dynamic properties," *The FEBS journal*, vol. 273, no. 21, pp. 4862–4877, 2006.

[32] A. Gulati, G. Isbister, and S. Duffull, "Scale reduction of a systems coagulation model with an application to modeling pharmacokinetic–pharmacodynamic

data," *CPT: pharmacometrics & systems pharmacology*, vol. 3, no. 1, pp. 1–8, 2014.

[33] M. Koschorreck, H. Conzelmann, S. Ebert, M. Ederer, and E. D. Gilles, "Reduced modeling of signal transduction–a modular approach," *BMC bioinformatics*, vol. 8, no. 1, p. 336, 2007.

[34] M. Sunnåker, H. Schmidt, M. Jirstrand, and G. Cedersund, "Zooming of states and parameters using a lumping approach including back-translation," *BMC systems biology*, vol. 4, no. 1, p. 28, 2010.

[35] M. Sunnåker, G. Cedersund, and M. Jirstrand, "A method for zooming of nonlinear models of biochemical systems," *BMC systems biology*, vol. 5, no. 1, p. 140, 2011.

[36] S. Gugercin, C. Beattie, and A. Antoulas, "Rational krylov methods for optimal h2 model reduction," *submitted for publication*, 2006.

[37] T. J. Snowden, P. H. van der Graaf, and M. J. Tindall, "Methods of model reduction for large-scale biological systems: a survey of current methods and trends," *Bulletin of mathematical biology*, vol. 79, no. 7, pp. 1449–1486, 2017.

[38] C. Reder, "Metabolic control theory: a structural approach," *Journal of theoretical biology*, vol. 135, no. 2, pp. 175–201, 1988.

[39] R. Heinrich and S. Schuster, *The regulation of cellular systems*. Springer Science & Business Media, 2012.

[40] K. Sinani and S. Gugercin, "Iterative rational krylov algorithms for unstable dynamical systems and optimality conditions for a finite-time horizon (2017)."