

PARAMETER IDENTIFIABILITY OF BIOCHEMICAL  
REACTION NETWORKS IN SYSTEMS BIOLOGY

by

DARA GEFFEN

A thesis submitted to the  
Department of Chemical Engineering  
in conformity with the requirements for  
the degree of Master of Science

Queen's University  
Kingston, Ontario, Canada

August 2008

Copyright © Dara Geffen, 2008

# Abstract

In systems biology, models often contain a large number of unknown or only roughly known parameters that must be estimated through the fitting of data. This work examines the question of whether or not these parameters can in fact be estimated from available measurements. Structural or *a priori* identifiability of unknown parameters in biochemical reaction networks is considered. Such systems consist of continuous time, nonlinear differential equations. Several methods for analyzing identifiability of such systems exist, most of which restate the question as one of observability by expanding the state space to include parameters. However, these existing methods were not developed with biological systems in mind, so do not necessarily address the specific challenges posed by this type of problem.

In this work, such methods are considered for the analysis of a representative biological system, the NF- $\kappa$ B signal transduction pathway. It is shown that existing observability-based strategies, which rely on finding an analytical solution, require significant simplifications to be applicable to systems biology problems that are seldom feasible. The analytical nature of the solution imposes restrictions on the size and complexity of systems that these methods can handle. This conflicts with the fact that most currently studied systems biology models are rather large networks containing many states and parameters. In this thesis, a new simulation based method using an empirical observability Gramian for determining identifiability is proposed. Computational and numerical sensitivity issues for this method are considered. An algorithm, based on this method, is developed and demonstrated on a simple biological example of microbial growth with Michaelis-Menten kinetics. The new method is applied to the motivating NF- $\kappa$ B example to show its suitability for use in systems biology.

# Acknowledgments

The successful completion of this thesis would not have been possible without the contributions of the following people. I would like to thank my advisor Martin Guay for his direction, encouragement and for giving me a lot of flexibility in terms of research direction. I also want to thank him for allowing me the opportunity to complete part of my studies at the University of Stuttgart, Germany, an experience which I found both academically and personally rewarding. I also want to thank my supervisor while in Germany, Rolf Findeisen for all of his guidance and support and for helping me get off to a good start on my research project. I greatly appreciate all of the time he spent reviewing and editing my papers and presentations as well.

I would like to thank all of my colleagues both here at Queen's University and at the University of Stuttgart: Monica Schliemann for all of the collaboration and assistance on biological topics (and speaking German), Frank Allgöwer for hosting my stay in Stuttgart and including me in all of the group's activities, my officemates, especially Paolo, Timm, Nick and Veronica for thought-provoking discussions, help with ideas, implementation issues, and programming and for making my time at the office more enjoyable.

I would like to acknowledge funding from the Natural Sciences and Engineering Research Council which contributed to the successful completion of my degree.

To my parents, Penny and Lorrie, I am very grateful for all of your love and support and for encouraging me to continue with my studies. To my sister Alyssa and Jenna, I want to thank you for your friendship, fun times and talking me through everything. To Sean, I want to thank you for all of your support, for always encouraging me to follow my own path, for letting me run everything by you and for patiently waiting for me to finish.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>ii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review</b>	<b>6</b>
2.1 Introduction . . . . .	6
2.2 Parameter Identifiability . . . . .	6
2.2.1 Practical Applications . . . . .	7
2.3 Definition of System and Problem Class . . . . .	8
2.4 Identifiability Challenges for Biological Systems . . . . .	10
2.5 Review of Parameter Identification Methods . . . . .	11
2.5.1 Observability Based Methods . . . . .	11
2.5.2 Functional Relationship Based Methods . . . . .	16
2.6 Motivating Example: NF- $\kappa$ B Signal Transduction Pathway . . . . .	17
<b>3 Application of Existing Parameter Identifiability Methods</b>	<b>22</b>
3.1 Introduction . . . . .	22
3.2 Reduced NF- $\kappa$ B Model . . . . .	23
3.3 Identifiability of Reduced NF- $\kappa$ B Model . . . . .	25
3.3.1 Algebraic Identifiability . . . . .	27
3.3.2 Observable Representation . . . . .	28
3.3.3 Algebraic Identifiability of the NF- $\kappa$ B Example . . . . .	28
3.4 Summary . . . . .	32

<b>4</b>	<b>Empirical Gramian Based Identifiability Method</b>	<b>34</b>
4.1	Introduction . . . . .	34
4.2	System Description . . . . .	36
4.3	Empirical Gramian for Identifiability . . . . .	37
4.3.1	Linear Observability Gramian . . . . .	37
4.3.2	Empirical Observability Gramian . . . . .	38
4.3.3	Empirical Gramian for Identifiability Analysis . . . . .	39
4.3.4	Illustrative Examples . . . . .	41
4.4	Description of Method . . . . .	44
4.4.1	Calculation of Gramian . . . . .	45
4.4.2	Identifiability Analysis . . . . .	49
4.5	Biological Example . . . . .	51
4.5.1	Application of the Identifiability Gramian . . . . .	52
4.5.2	Determination of Unidentifiable Parameters . . . . .	54
4.6	Summary . . . . .	58
<b>5</b>	<b>Application of Empirical Identifiability Gramian to NF-<math>\kappa</math>B Example</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.2	Application of Empirical Identifiability Gramian to Reduced NF- $\kappa$ B Example . . . . .	60
5.3	Application of Empirical Identifiability Gramian to Full NF- $\kappa$ B Example	62
5.4	Summary . . . . .	65
<b>6</b>	<b>Summary and Conclusions</b>	<b>67</b>
<b>A</b>	<b>Description of NF-<math>\kappa</math>B Model Parameters</b>	<b>73</b>
<b>B</b>	<b>Nominal NF-<math>\kappa</math>B Parameters Values for Gramian Calculation</b>	<b>75</b>

# List of Tables

4.1	Results for use of algorithm on microbial growth example . . . . .	55
5.1	Results for use of empirical identifiability Gramian method on reduced NF- $\kappa$ B model with same unknown parameters as in Section 3.2. . . . .	60
5.2	Comparison of identifiable parameter of the reduced NF- $\kappa$ B model using the empirical identifiability Gramian and the differential algebra based methods . . . . .	61
5.3	Comparison of identifiable parameters for the reduced NF- $\kappa$ B model using differential algebra based methods with different observable representations . . . . .	62
5.4	Results for use of empirical identifiability Gramian method on full NF- $\kappa$ B model with all 29 parameters as unknown . . . . .	63
A.1	Summary of NF- $\kappa$ B signal transduction pathway parameters . . . . .	74
B.1	Nominal NF- $\kappa$ B Parameter Values for Gramian . . . . .	76

# List of Figures

2.1	Schematic diagram of input-output map of system . . . . .	10
2.2	Schematic pathway representation of the full NF- $\kappa$ B model . . . . .	18
3.1	Schematic pathway representation of the reduced NF- $\kappa$ B model . . .	26
4.1	Effect on the observed output of perturbing nominal parameters in the direction of the eigenvectors of the identifiability Gramian . . . . .	56
4.2	Effect of perturbing identifiable and unidentifiable parameters on the observed output . . . . .	57
5.1	Full NF- $\kappa$ B model showing outputs and identifiable parameters . . .	64

# Chapter 1

## Introduction

Systems biology is an interdisciplinary approach in which tools from traditional systems theory and control are applied to biological systems in an effort to obtain a systems level understanding of biology. Ideally, this will lead to a greater understanding of biological processes such as diseases and could offer new tools for drug development (Klipp et al., 2005). It is an emerging research area so many challenges remain. To understand biology at a systems level, one must look at the structure and dynamics of cellular and organism function as a whole rather than the characteristics of the isolated parts themselves (Kitano, 2002). Due to the emergence of new experimental techniques, there is currently an abundance of high throughput data. It is important to integrate such data into models of a complete system to provide a global view of the underlying mechanisms (Klipp et al., 2005). For this reason, much of systems biology is based on the mathematical modeling and the simulation of biochemical networks.

In systems biology, models generally contain a large number of unknown or only

roughly known parameters. Accurate knowledge of these parameter values is important for describing and analyzing the dynamics and behaviour of biological systems. This can be done using one of several existing parameter identification strategies (see eg. Gadkar et al., 2005; Moles et al., 2003; Vayttaden et al., 2004) which all involve the fitting of measurement data. However, these methods are often difficult to apply in practice and offer no guarantee that available measurements will yield meaningful values for the desired parameters. For this reason, it is important to first consider structural or *a priori* identifiability; whether or not parameters can in fact be determined (uniquely), in the ideal noise free case, from a given model and available outputs or measurements. This is also a valuable tool for the design of experiments. Most relevant phenomenological biological systems models result in first principle models which generally take the form of nonlinear continuous-time dynamical systems. The analysis of parameter identifiability in such models is the primary focus of this thesis. Several methods exist for checking structural identifiability of nonlinear dynamical systems, many of which restate the question of identifiability as one of observability by expanding the state space to include parameters (see eg. Chappell et al., 1990; Denis-Vidal et al., 2001; Godfrey and Fitch, 1984; Ljung and Glad, 1994; Pohjanpalo, 1978; Xia and Moog, 2003). However, these methods, developed for the analysis of general nonlinear dynamical systems, do not necessarily address the specific challenges posed by biological systems.

In this thesis, existing methods are considered in the context of a representative systems biology problem: the nuclear factor  $\kappa\text{B}$  (NF- $\kappa\text{B}$ ) signal transduction pathway. This is part of a larger project focusing on mathematical modeling to clarify the link between the tumor necrosis factor (TNF- $\alpha$ ) induced apoptotic and anti-apoptotic

signalling pathways in mammalian cells. Understanding the underlying mechanism plays an important role in gaining insight about cancer, diabetes, osteoporosis and other autoimmune diseases. Models for the involved processes are often very complex and involve many influencing (external) factors. Advancements in measurement techniques have allowed for the *in vivo* investigation of the appearing phenomena. However, many limitations remain. Only a restricted number of concentrations and reaction rates can be measured and many of the parameters describing the dynamics of these pathways are unknown. The *a priori* identifiability of the NF- $\kappa$ B pathway is performed to identify the measurements that are required for estimation of unknown parameters. This will assist in the design of future experiments to be carried out.

The NF- $\kappa$ B example highlights some of the limitations of the use of previous methods for biological systems. The observability of a nonlinear system in continuous time constitutes a challenging problem. As a result, most existing observability based identifiability techniques, which compound the complexity by adding the effect of the parameters, can often lead to very difficult computational issues, even for small problems. The limitation of these methods, such as those based on differential algebra (Ljung and Glad, 1994; Denis-Vidal et al., 2001; Xia and Moog, 2003), power series expansion (Godfrey and Fitch, 1984; Pohjanpalo, 1978), or the local state isomorphism theory (Chappell et al., 1990; Godfrey and Fitch, 1984), are related to both the size of the problem and the degree of nonlinearity of the models considered. Although the size of the problem will always pose a problem in practice, the degree of nonlinearity of the models can severely restrict the application of such techniques since most require the computation of some analytical solution. While this does not prove to be a problem for smaller systems (see Xia and Moog, 2003; Ljung and Glad,

1994; Pohjanpalo, 1978), it is not ideal for models currently encountered in systems biology which often have a large number of states and parameters. For example, the NF- $\kappa$ B signal transduction pathway considered in Lipniacki et al. (2004) contains 15 states and 29 parameters and as is shown here cannot be analyzed using the previous methods. For this reason, it is necessary to develop simulation based methods that are less dependent on the size and complexity of the system.

The concept of empirical observability and controllability Gramians is introduced in Lall et al. (2002) for the purpose of model reduction of nonlinear systems. It is a ‘data based’ approach that makes use of data either from simulation or experiments. It is used to overcome some of the computational complexity associated with other methods. It has been successfully applied to systems with dozens of states (Hahn and Edgar, 2002). The use of such Gramians for observability (and controllability) analysis itself is proposed in Singh and Hahn (2005) and compared to the use of linear Gramians and Lie algebra based methods. There are currently no results of using these Gramians for the purpose of identifiability analysis.

In this thesis, a new approach for checking identifiability based on the use of the ‘empirical observability Gramian’ is presented. This method is demonstrated using the relatively simple biological example of microbial growth with Michaelis-Menten kinetics as described in Denis-Vidal and Joly-Blanchard (2000); Holmberg (1982); Chappell and Godfrey (1992). This method allows for the identifiability analysis of relatively large and complex systems such as those encountered in systems biology. It is applied to the larger NF- $\kappa$ B example to demonstrate its applicability.

This thesis is structured as follows. A literature review and the fundamental theory related to this thesis are presented in Chapter 2. Specifically, topics such as

identification and parameter identifiability are discussed in the context of the specific challenges encountered in the analysis of biological systems. The NF- $\kappa$ B signal transduction network system is introduced as a motivating example. In Chapter 3, existing observability based parameter identifiability methods are applied to the NF- $\kappa$ B model. This provides the motivation for the following chapters. In Chapter 4, a new method for checking identifiability based on the use of an 'empirical observability Gramian' is proposed to overcome the challenges highlighted in Chapter 3. An algorithm for the use of this method is outlined and demonstrated using a simple biological example of microbial growth with Michaelis-Menten kinetics. Chapter 5 deals with the application of the new method to the NF- $\kappa$ B model. Conclusions and directions of future work are discussed in Chapter 6.

# Chapter 2

## Literature Review

### 2.1 Introduction

This chapter serves as a brief review of parameter identification and identifiability in the context of systems biology. The biological example considered in this work, the NF- $\kappa$ B signal transduction pathway, is introduced here as well.

### 2.2 Parameter Identifiability

In systems biology, modeling of biochemical networks is the basis for analysis. However, many parameters are unknown or only roughly known so must be estimated. Accurate knowledge of these parameter values is important for describing and analyzing the dynamics and behaviour of biological systems. There are several methods for estimating parameters as described, for example, in Gadkar et al. (2005); Moles et al. (2003); Vayttaden et al. (2004), which depend on the fitting of measurement

data. However, as stated in Ljung and Glad (1994), a fundamental problem of identification is to be able, before the data have been analyzed, to decide if all the unknown parameters of a model structure can be (uniquely) recovered from data. This is the question of identifiability that is focused on here. This thesis is specifically concerned with structural or *a priori* identifiability: whether or not parameters can be identified from a given model structure and outputs without considering possible measurement noise (as opposed to ‘practical identifiability’ in Muller et al. (2002); Quaiser et al. (2007); Hengl et al. (2007) which considers the quality of the data as well).

### 2.2.1 Practical Applications

Determining the structural identifiability of a model serves several practical purposes. Since model parameters have physical significance, it is important to know if it is possible to determine their values from observed data. This analysis can be used for experimental design to solve the inverse problem of what needs to be measured to estimate the model parameters (as is the case in Xia and Moog, 2003). If there are limitations on what can be measured (as is the case in Denis-Vidal et al., 2001; Ljung and Glad, 1994; Hengl et al., 2007; Pohjanpalo, 1978; Audoly et al., 2001) the analysis can offer information about parameters that can be identified, thus, giving information about the validity of the solution to the parameter identification problem. This is normally carried out as a step prior to parameter identification using experimental data (as in Muller et al., 2002). This is an important step as numerical search procedures for parameter estimation (such as those found in Moles et al., 2003) may not be effective for problems in which a (unique) solution does not exist.

The system and problem class considered in this thesis are defined in the next

section prior to a more detailed discussion of identifiability methods.

## 2.3 Definition of System and Problem Class

Consider a nonlinear dynamic system of the following form:

$$\Sigma : \begin{cases} \dot{x}(t) = f(x(t), u(t), \theta) , & x(0) = x_0, \\ y(t) = h(x(t), u(t), \theta), \end{cases} \quad (2.1)$$

where  $x \in \mathbb{R}^n$ ,  $u \in \mathbb{R}^r$ ,  $y \in \mathbb{R}^p$ , and  $\theta \in \mathbb{R}^q$  are the states, inputs, outputs, and unknown parameters respectively.

The question considered is whether the parameters can be identified from the measurements available. A common definition of structural identifiability is as follows:

**Definition 1** (*Structural identifiability I*)

*A given parameter  $\theta_i$  is a priori or structurally globally identifiable if there exists a unique solution to (2.1) for  $\theta_i$ . A parameter with a countable or uncountable number of solutions is considered locally identifiable or unidentifiable (Hengl et al., 2007).*

For global identifiability especially, this is a very restrictive definition as identifiability has to hold for all possible measurement trajectories or realizations of the output trajectory. A more relaxed definition of identifiability, suitable for application, can be found in Jacquez and Greif (1985). Note that for the system in 2.1,  $f$  specifies the model,  $u$  specifies the input and  $y$  the observations for a particular experiment. An experiment is defined by the initial conditions  $x_0$  and (in the case of a controlled system) inputs  $u$  from a set of (persistently excitable) inputs as well as the observations  $y$ .

**Definition 2** (*Structural identifiability II*)

Let  $\hat{\theta}$  be the value of  $\theta$  that gives the response of a model to a given experiment (i.e. the solution of 2.1). A parameter vector is (structurally) locally identifiable if for almost any  $\hat{\theta}$  the solution is unique in some neighbourhood of  $\hat{\theta}$ .

**Remark 1** Note that the term “almost any” only excludes subsets of zero measure in  $\theta$ . This enables the definition to be used practically for analyzing identifiability by omitting irrelevant special cases that are unlikely to occur in practice.

The definition of (structural) global identifiability is the same except that a solution must be unique for the entire domain of  $\theta$  not just a neighbourhood of  $\theta$ .

For local identifiability, one could consider identifiability along specific trajectories subject to specific choices of inputs (persistence of excitation) or of specific regions of initial conditions in the state space. The initial state  $x_0 \in X \in \mathbb{R}^n$  can be considered to exclude values of the state variables that are not physically possible (such as  $x_0 < 0$ ) for concentrations in biochemical pathways (Audoly et al., 2001).

Several methods exist for checking identifiability of nonlinear continuous time systems, such as those in Denis-Vidal et al. (2001); Ljung and Glad (1994); Pohjanpalo (1978). However, such methods were not developed with biological systems in mind and, therefore, do not necessarily address the specific challenges posed by this type of problem. These challenges are discussed in the following section.

## 2.4 Identifiability Challenges for Biological Systems

There are many specific challenges to consider when dealing with parameter identification and identifiability in systems biology. Continuous time nonlinear dynamical systems are considered as such systems arise naturally as a result of first-principles-modeling of biochemical reaction networks (i.e. mass action, Michaelis-Menten kinetics). Unfortunately, there is no well established identifiability method for this class of systems. In systems theory, systems are often linearized or discretized in time for analysis purposes (Ljung, 1999). This is not a desirable approach for biological systems as parameters have physical significance, such as describing reaction rates, which would be lost. Systems theoretic analysis methods are often based on the input-output map of the system shown in Fig. 2.1

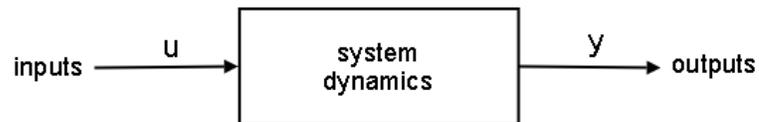


Figure 2.1: Schematic diagram of input-output map of system.

This idealized representation is often possible since it is generally assumed that these systems offer some degree of actuation used to excite the system and that accurate output measurements are available.

In biological systems, however, there are generally limited or no inputs that can persistently excite the system. There are limitations on the quantities that can be physically measured. In addition, these measurements are often subject to a large degree of noise. Another challenge for carrying out the identifiability analysis is that many currently studied systems biology models are large networks that contain many

states and parameters (for example, the NF- $\kappa$ B network in Lipniacki et al. (2004) has 15 states and 29 parameters and the JAK-STAT pathway from Yamada et al. (2003) has 31 states and 52 parameters). Several methods exist for checking identifiability of nonlinear continuous time systems, such as those found in Denis-Vidal et al. (2001); Ljung and Glad (1994); Pohjanpalo (1978).

## 2.5 Review of Parameter Identification Methods

Identifiability methods can be roughly divided into two categories: 1) methods that restate the question of identifiability as one of observability by extending the state space to include parameters (see eg. Chappell et al., 1990; Denis-Vidal et al., 2001; Godfrey and Fitch, 1984; Ljung and Glad, 1994; Pohjanpalo, 1978; Xia and Moog, 2003) and 2) methods that check for a functional relationship between parameters through simulation, optimization and parameter estimation such as in Hengl et al. (2007); Muller et al. (2002); Quaiser et al. (2007); Jacquez and Greif (1985). This thesis is primarily concerned with structural identifiability and focuses on observability based methods. Most of these methods were developed for nonlinear systems and do not consider the specific challenges posed by biological systems outlined in Section 2.4. The existing parameter identifiability methods are discussed here in the context of their relevance to systems biology problems.

### 2.5.1 Observability Based Methods

There are several approaches based on restating the question of identifiability as one of observability. For time invariant parameters, this is done by including the parameters

as states with  $\dot{\theta} = 0$ . Methods based on differential algebra, a power (Taylor) series expansion, and the local state isomorphism theory are discussed here.

### Power Series Expansion

One method that uses this observability based approach is the power series expansion in Godfrey and Fitch (1984); Pohjanpalo (1978). This involves expressing the system outputs and its derivatives in terms of the unknown parameters at time zero ( $t = 0^+$ ) using the following Taylor series expansion:

$$y(t, \theta) = y(0^+, \theta) + y^1(0^+, \theta)t + y^2(0^+, \theta)\frac{t^2}{2} + \dots + y^i(0^+, \theta)\frac{t^i}{i!}$$

Since the Taylor series coefficients are unique, the identifiability problem is reduced to determining the number of solutions of a set of nonlinear algebraic equations of  $\theta$  (Muller et al., 2002). If one, some or  $\infty$  solutions exist then the parameters are globally, locally or unidentifiable respectively. This method has the advantage of simplifying the parameter identifiability problem if the initial values of the states are known. The calculations can be performed using Maple or Mathematica. However this method requires knowledge of initial conditions which is not always easy to obtain for biological systems. The main disadvantage of this method is that the structure of the resulting equations is often complicated so it is difficult to determine identifiability properties even for biological systems of moderate complexity (see application to nonlinear compartmental models in Godfrey and Fitch (1984)). The complexity of this method has been found to increase greatly with system size (as demonstrated in Chappell et al., 1990; Ljung and Glad, 1994; Muller et al., 2002) and it generally only provides local identifiability results (Pohjanpalo, 1978).

### Local State Isomorphism Theory

In Chappell et al. (1990); Godfrey and Fitch (1984) a method using the local state isomorphism theory is employed. This method consists of solving a set of partial differential equations for a parameter set to determine if the solution is unique. In order to use this method, the system must be controllable and observable. Controllability is required to ensure that the system inputs can be chosen to reach any points in the state-space starting from any initial condition. Observability guarantees that every initial condition, and hence every state trajectory, can be estimated uniquely from the output measurements. In this context, the local state isomorphism theory can be used to verify if there exists an input trajectory that makes the parameters identifiable from the available measurements. However, the ability to compute such input trajectories remains a very difficult problem. This is especially true in the study of biological systems where only limited actuation is available. The use of this technique can often lead to inputs that are not implementable in practice for systems biology problems (Tunali and Tarn, 1987).

### Differential Algebra Methods

Methods based on tools from differential algebra including those in Audoly et al. (2001); Denis-Vidal et al. (2001); Ljung and Glad (1994); Xia and Moog (2003) are promising for use in systems biology as they neither require inputs to excite the system nor previous information about the system, such as initial conditions, that would need to be obtained experimentally. This condition is important for *a priori* identifiability. For these reasons, this is the method employed on the NF- $\kappa$ B example in Section 2.6. It is based on using differential algebraic operations and ranking variables to find

an observable representation of the system in a structured manner. One can then determine if it is possible to express parameters in terms of measurable quantities (i.e.  $y$  and its derivatives).

A full description of differential algebra concepts can be found in Ritt (2005). The aspects relating to the use of differential algebra for analyzing identifiability are presented below.

The system in (2.1) can be rewritten in differential polynomial form as follows (Denis-Vidal et al., 2001):

$$\Sigma : \begin{cases} p(\dot{x}, x, u, \theta) = 0 \\ q(x, y, u, \theta) = 0 \\ \dot{\theta} = 0. \end{cases} \quad (2.2)$$

The principle behind the differential algebra based methods is that if outputs satisfy (2.2), they will also satisfy equations obtained by differentiation, adding, scaling, and multiplying of the left hand side of (2.2).

A differential ideal is a set closed by adding, multiplying, scaling, and taking the derivatives of a system. This ideal can be represented by a finite basis which can be determined by following a set of ranking rules. Ranking in this case refers to a 'well-ordering' of the variables and their derivatives (Denis-Vidal et al., 2001). For identifiability purposes the following ranking is considered (where  $<$  means ranked less than):

$$[y_1] < [y_2], [y_1] < [y_1], [y_1] < [y_2] < [y_1] < [y_2]$$

$$[y] < [u] < [\theta] < [x].$$

The variables that can be measured ( $y$  and  $u$ ) are ranked the highest. The states are ranked the lowest and can be excluded from the generated finite basis. This ranking is used to eliminate the states to obtain an observable representation of the system. The observable representation of the system can be determined heuristically (see for example Xia and Moog, 2003; Geffen et al., 2007) or algorithmically as part of the characteristic set (as in Ljung and Glad, 1994; Denis-Vidal et al., 2001; Audoly et al., 2001). The characteristic set is a finite set that is a canonical representation of the differential ideal. A number of algorithms exist for determining the characteristic set such as Ritt's algorithm (used for identifiability analysis in Ljung and Glad, 1994) and the Rosenfeld-Groebner algorithm which has a Maple implementation and applied in Denis-Vidal et al. (2001); Geffen et al. (2007)) for identifiability purposes. In both cases, one obtains an equation of the following form:

$$\phi_{l+1}(y) = \sum_{i=1}^l \phi_i(y)g_i(\theta) \quad (2.3)$$

where  $\phi_i(y)_{i=1,\dots,l+1}$  are differential polynomials in  $y$  (consider case of only one output for simplicity of presentation) and  $g_i(\theta), g_i \neq g_j (i \neq j)$  are rational functions of  $\theta$ . It is shown in Ljung and Glad (1994); Denis-Vidal et al. (2001) that if the functions  $\phi_i(y)_{i=1,\dots,l+1}$  are linearly independent (can be checked by taking the functional determinant of  $\phi(y)$ ) then the lumped parameters  $g(\theta)$  have a unique solution and are therefore identifiable. One is then left with a system of algebraic equations in  $\theta$  so rank considerations can be used to determine if the individual parameters are identifiable as well.

One advantage of this method is that the algebraic nature of the solution identifies

clearly the identifiable and non-identifiable parameters. However, the main disadvantage of this technique is the potential explosion in computational requirement when more states and parameters are considered.

### 2.5.2 Functional Relationship Based Methods

As an alternative to the observability based methods, one can check identifiability by determining if there is a functional relationship between parameters (functionally related parameters are not identifiable). This can be done in several ways. An approach widely used in experimental systems biology literature is based on calculating the correlation of columns of the sensitivity matrix as described in Jacquez and Greif (1985). The advantage of this method, and the reason it is widely used, is that it is relatively easy to compute. The calculation of the sensitivity matrix, however, involves a linearization in the parameter space about the initial estimates. The disadvantage of doing this lies in the fact that it is possible for the nonlinear model to be identifiable and the linear one not to be or vice versa (Jacquez and Greif, 1985).

One can also employ simulation based methods such as the ones discussed in Muller et al. (2002); Quaiser et al. (2007). In these methods, least squares identifiability is checked via parameter estimation in simulation or using experimental data and determining if there is a functional relationship between any of the parameters by considering the covariance matrix described in Muller et al. (2002) or the eigenvalue approach in Quaiser et al. (2007). Since these methods are generally simulation based they have the benefit of avoiding the computational complexity associated with the other methods. However, these methods generally can only check local identifiability. The most troublesome aspect of these techniques is that they require an estimate of

the parameters to perform the analysis. This leads to a problematic circular argument where the identifiability of a given parameter depends on the value of some other parameter. Since the models considered are mostly nonlinear in the parameters, this can easily lead to erroneous identifiability results with very limited practical applicability. This is especially true since parameters are not always known in the context of an *a priori* identifiability analysis. Such techniques are therefore of little use in the analysis of biological networks where parameters values are largely unknown.

Parameter estimation is typically carried out by minimizing a cost function that measures the goodness of fit of a model to experimental data. The Prediction Error Minimization method (PEM) is commonly used in literature (Ljung, 1999). An approach based on optimality of biological systems that employs the Fisher Information Matrix (FIM) is described in Gadkar et al. (2005). Several different global optimization methods for parameter estimation in biochemical pathways are compared in Moles et al. (2003). Overall these approaches have a high computational demand and do not guarantee convergence. The optimization problem is non-convex so there is no guarantee an optimal solution can be found, especially for large system sizes.

## 2.6 Motivating Example: NF- $\kappa$ B Signal Transduction Pathway

Much of systems biology is focused on the modeling and subsequent analysis of signal transduction and metabolic networks. The NF- $\kappa$ B signal transduction network is a relatively complex system that is representative of what is typically encountered in systems biology.

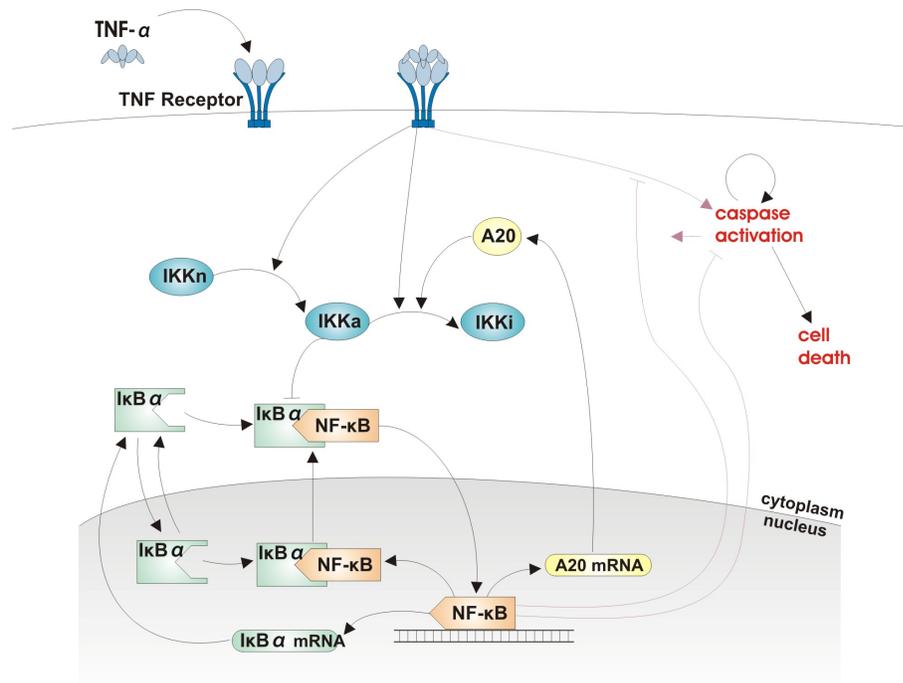


Figure 2.2: Schematic pathway representation of the full NF- $\kappa$ B signal transduction network model (adapted from Schliemann et al., 2007).

Nuclear factor  $\kappa$ B is responsible for regulating genes involved in such areas as apoptosis (programmed cell death), inflammation and other immune responses in mammalian cells (Lipniacki et al., 2004). Gaining a greater understanding of this pathway can allow one to gain insight into cancer, arthritis, asthma and other inflammatory and autoimmune diseases. This pathway has been the focus of many systems biology investigations in areas such as modeling, bistability analysis, identification (see eg. Lipniacki et al., 2004; Quaiser et al., 2007; Eissing et al., 2004; Hoffmann et al., 2002). A mathematical model of the NF- $\kappa$ B signal transduction pathway is proposed in Lipniacki et al. (2004) but the unknown parameters are only roughly estimated through simulation and physical considerations. A schematic diagram of this model is given in Figure 2.2.

It is a two-compartment model consisting of two feedback loops describing the kinetics of the activators IKK and NF- $\kappa$ B and inhibitors A20 and I $\kappa$ B $\alpha$ . In resting cells NF $\kappa$ B is in its inactive form bound to I $\kappa$ B $\alpha$  in the cytoplasm. In response to extracellular signals such as tumor necrosis factor (TNF- $\alpha$ ) IKK is activated (IKKn  $\rightarrow$  IKKa) facilitating the degradation of the protein I $\kappa$ B $\alpha$  freeing NF- $\kappa$ B to enter the nucleus. There it triggers the transcription of several genes, including those of the inhibitors I $\kappa$ B $\alpha$ , which binds to NF- $\kappa$ B and leads it out of the nucleus, and A20, which inactivates the IKKa (IKKa  $\rightarrow$  IKKi). This results in the following model from Lipniacki et al. (2004) which consists of 15 states and 29 parameters:

$$\dot{z}_1 = k_{prod} - k_{deg}z_1 - k_1z_1 \quad (2.4)$$

$$\dot{z}_2 = k_1z_1 - k_3z_2 - k_2z_2z_8 - k_{deg}z_2 - a_2z_2z_{10} + t_1z_4 - a_3z_2z_{13} + t_2z_5 \quad (2.5)$$

$$\dot{z}_3 = k_3z_2 + k_2z_2z_8 - k_{deg}z_3 \quad (2.6)$$

$$\dot{z}_4 = a_2z_2z_{10} - t_1z_4 \quad (2.7)$$

$$\dot{z}_5 = a_3z_2z_{13} - t_2z_5 \quad (2.8)$$

$$\dot{z}_6 = c_{6a}z_{13} - a_1z_6z_{10} + t_2z_5 - i_1z_6 \quad (2.9)$$

$$\dot{z}_7 = i_1k_vz_6 - a_1z_{11}z_7 \quad (2.10)$$

$$\dot{z}_8 = c_4z_9 - c_5z_8 \quad (2.11)$$

$$\dot{z}_9 = c_2 + c_1z_7 - c_3z_9 \quad (2.12)$$

$$\dot{z}_{10} = -a_2z_2z_{10} - a_1z_{10}z_6 + c_{4a}z_{12} - c_{5a}z_{10} - i_{1a}z_{10} + e_{1a}z_{11} \quad (2.13)$$

$$\dot{z}_{11} = -a_1z_{11}z_7 + i_{1a}k_vz_{10} - e_{1a}k_vz_{11} \quad (2.14)$$

$$\dot{z}_{12} = c_{2a} + c_{1a}z_7 - c_{3a}z_{12} \quad (2.15)$$

$$\dot{z}_{13} = a_1z_{10}z_6 - c_{6a}z_{13} - a_3z_2z_{13} + e_{2a}z_{14} \quad (2.16)$$

$$\dot{z}_{14} = a_1z_{11}z_7 - e_{2a}k_vz_{14} \quad (2.17)$$

$$\dot{z}_{15} = c_{2c} + c_{1c}z_7 - c_{3c}z_{15} \quad (2.18)$$

$$(2.19)$$

where  $z_1$  is  $IKKn$ ,  $z_2$  is  $IKKa$ ,  $z_3$  is  $IKKi$ ,  $z_4$  is  $(IKKa|I\kappa B\alpha)$ ,  $z_5$  is  $(IKKa|I\kappa B\alpha|NF-\kappa B)$ ,  $z_6$  is  $NF-\kappa B$ ,  $z_7$  is  $NF-\kappa B_n$ ,  $z_8$  is  $A20$ ,  $z_9$  is  $A20_t$ ,  $z_{10}$  is  $I\kappa B\alpha$ ,  $z_{11}$  is  $I\kappa B\alpha_n$ ,  $z_{12}$  is  $I\kappa B\alpha_t$ ,  $z_{13}$  is  $(I\kappa B\alpha|NF-\kappa B)$ ,  $z_{14}$  is  $(I\kappa B\alpha_n|NF-\kappa B_n)$

and  $z_{15}$  is  $c_{gen}$ . It should be noted that the states for the full NF- $\kappa$ B model are denoted as  $z$  instead of  $x$  to distinguish them from those of the reduced NF- $\kappa$ B model in Section 3.2. The description of the 29 unknown parameters is found in Appendix B. The question this thesis is trying to answer is whether or not the unknown parameters are identifiable with the given model and choice of outputs. As mentioned previously, this work is part of a larger project (see Schliemann et al., 2007) which combines experiments and modeling to clarify the link between the apoptotic and anti-apoptotic signal transduction pathway. The choice of outputs is based on the feasibility of current measurement techniques such as Western blotting. It is initially assumed that one can measure the concentration of free NF- $\kappa$ B in both the nucleus and the cytoplasm as well as the concentration of (I $\kappa$ B $\alpha$ |NF- $\kappa$ B) and IKK $\alpha$  in the cytoplasm so the output  $y(t, \theta)$  for purposes of parameter identifiability is given by:

$$y^T = [z_2 \quad z_6 \quad z_7 \quad z_{13}].$$

The identifiability analysis of this network can give information on how to design future experiments. It can also be used as a framework to assess the performance of existing methods for determining nonlinear, continuous time identifiability on biological systems.

# Chapter 3

## Application of Existing Parameter Identifiability Methods to the NF- $\kappa$ B Signal Transduction Pathway<sup>1</sup>

### 3.1 Introduction

In this chapter, the application of existing parameter identification methods to a representative systems biology example is considered. The *a priori* identifiability

---

<sup>1</sup>The contents of this chapter were originally found in:

D. Geffen, R. Findeisen, M. Schliemann, F. Allgöwer, and M. Guay, “The question of parameter identifiability for biochemical reaction networks considering the NF- $\kappa$ B signal transduction pathway”, in *Proceedings 2nd Foundations of Systems Biology in Engineering*, Stuttgart, Germany, 2007

D.Geffen, R. Findeisen, M. Schliemann, F. Allgöwer, and M. Guay, “Observability based parameter identifiability for biochemical reaction networks”, in *Proceedings 2008 American Control Conference*, Seattle, Washington, USA, 2008

of the NF- $\kappa$ B signal transduction pathway, introduced in Section 2.6, is considered to determine the model parameters that can be identified with a given choice of outputs. This also serves the broader purpose of determining the suitability of a current identifiability method, outlined in Section 2.5 for systems of the complexity and nonlinearity typically encountered in systems biology.

This chapter is structured as follows. Section 3.2 describes the reduction and resulting properties of the NF- $\kappa$ B model. The application of a current parameter identifiability method to the representative biological system is presented in Section 3.3. Section 3.4 contains results and conclusions.

## 3.2 Reduced NF- $\kappa$ B Model

The mathematical model of the NF- $\kappa$ B signal transduction pathway from Lipniacki et al. (2004) introduced in Section 2.6 consists of 15 states and 29 parameters. In order to facilitate analysis using existing methods, a model reduction is required.

To simplify the model, it is assumed that the concentrations of active IKK (IKKa) can be measured over time since this is a protein that can be physically measured in practice. This quantity is then treated as a measurable disturbance,  $d(t)$ , instead of a state as in Lipniacki et al. (2004). This eliminates the need to include neutral and inactive IKK (IKKn and IKKi respectively) as well as the concentration of A20 protein and mRNA transcript (A20 and A20t respectively) as states for the purpose of parameter estimation. The function of these states in the model is essentially to describe the concentration of IKKa over time. A20 facilitates the inactivation of IKK while IKKn and IKKi have a more direct effect (IKKn  $\rightarrow$  IKKa  $\rightarrow$  IKKi). If IKKa can be measured directly, these other states are not required. (IKKa|I $\kappa$ B $\alpha$ )

is also removed as a state because it does not appear elsewhere in the system. The identifiability of the parameters related to IKKa can now be considered a separate problem that is not covered here. This results in the following simplified model which contains 8 states and 15 parameters:

$$\dot{x}_1 = c_{6a}x_7 - a_1x_1x_3 + t_2x_6 - i_1x_1 \quad (3.1)$$

$$\dot{x}_2 = p_4x_1 - a_1x_2x_4 \quad (3.2)$$

$$\dot{x}_3 = -a_2d(t)x_3 - a_1x_1x_3 + c_{4a}x_5 - c_{5a}x_3 - i_{1a}x_3 + e_{1a}x_4 \quad (3.3)$$

$$\dot{x}_4 = -a_1x_2x_4 + p_2x_3 - p_3x_4 \quad (3.4)$$

$$\dot{x}_5 = c_{2a} + c_{1a}x_2 - c_{3a}x_5 \quad (3.5)$$

$$\dot{x}_6 = a_3d(t)x_7 - t_2x_6 \quad (3.6)$$

$$\dot{x}_7 = a_1x_1x_3 - c_{6a}x_7 - a_3d(t)x_7 + e_{2a}x_8 \quad (3.7)$$

$$\dot{x}_8 = a_1x_2x_4 - p_1x_8 \quad (3.8)$$

where  $x_1(= z_6)$  is  $NF-\kappa B$ ,  $x_2(= z_7)$  is  $NF-\kappa B_n$ ,  $x_3(= z_{10})$  is  $I\kappa B\alpha$ ,  $x_4(= z_{11})$  is  $I\kappa B\alpha_n$ ,  $x_5(= z_{12})$  is  $I\kappa B\alpha_t$ ,  $x_6(= z_5)$  is  $(IKKa|I\kappa B\alpha|NF-\kappa B)$ ,  $x_7(= z_{13})$  is  $(I\kappa B\alpha|NF-\kappa B)$  and  $x_8(= z_{14})$  is  $(I\kappa B\alpha_n|NF-\kappa B_n)$  are the states of the model and represent the concentration of that species, in either the cytoplasm or nucleus (subscript  $n$ ), of an individual mammalian cell. The model parameters are kept the same as in Lipniacki et al. (2004) with the exception of the introduction of lumped parameters  $p_1 = k_v e_{2a}$ ,  $p_2 = k_v i_{1a}$ ,  $p_3 = k_v e_{1a}$  and  $p_4 = k_v i_1$  where  $k_v$  is the ratio of the cytoplasmic to nuclear volume of the cell. The nonlinearities in the system are caused by two bilinear terms ( $x_1x_3$  and  $x_2x_4$ ).

For the purpose of parameter identification, and based on the feasibility of current measurement techniques, it is assumed that one can measure the concentration of free NF- $\kappa$ B in both the nucleus and the cytoplasm as well as the concentration of (I $\kappa$ B $\alpha$ |NF- $\kappa$ B) in the cytoplasm. This means the output  $y(t, \theta)$  is given by:

$$y^T = \begin{bmatrix} x_2 & x_7 & x_1 \end{bmatrix}$$

To simplify the task of parameter identification, values that are already estimated from other sources in Lipniacki et al. (2004) ( $a_1, c_{2a}, c_{3a}, c_{5a}, c_{6a}, e_{2a}, i_{1a}, p_1, p_2$ ) are assumed to be known. This leaves the following nine parameters as unknown:  $a_2, a_3, c_{1a}, c_{4a}, t_2, e_{1a}, i_1, p_3, p_4$ . The meaning of these parameters can be found in Lipniacki et al. (2004) and in the diagram of the structure of the reduced model in Fig. 3.1.

This results in an identifiability problem of a nonlinear (bilinear) system with 8 states and 9 unknown parameters. Despite the simplifications made, the size and complexity of the problem is still significantly greater than those previously treated by the current identifiability analysis methods for example in Audoly et al. (2001); Denis-Vidal et al. (2001); Ljung and Glad (1994). Note that there is no manipulated input variable ( $u$ ) for the system.

### 3.3 Identifiability of Reduced NF- $\kappa$ B Model

The following example shows the parameter identifiability analysis for the reduced NF- $\kappa$ B model. The differential algebra methods described in Section 2.5.1 are employed. The definition and conditions for algebraic identifiability used here are adapted from Xia and Moog (2003) for the case of no input variables.

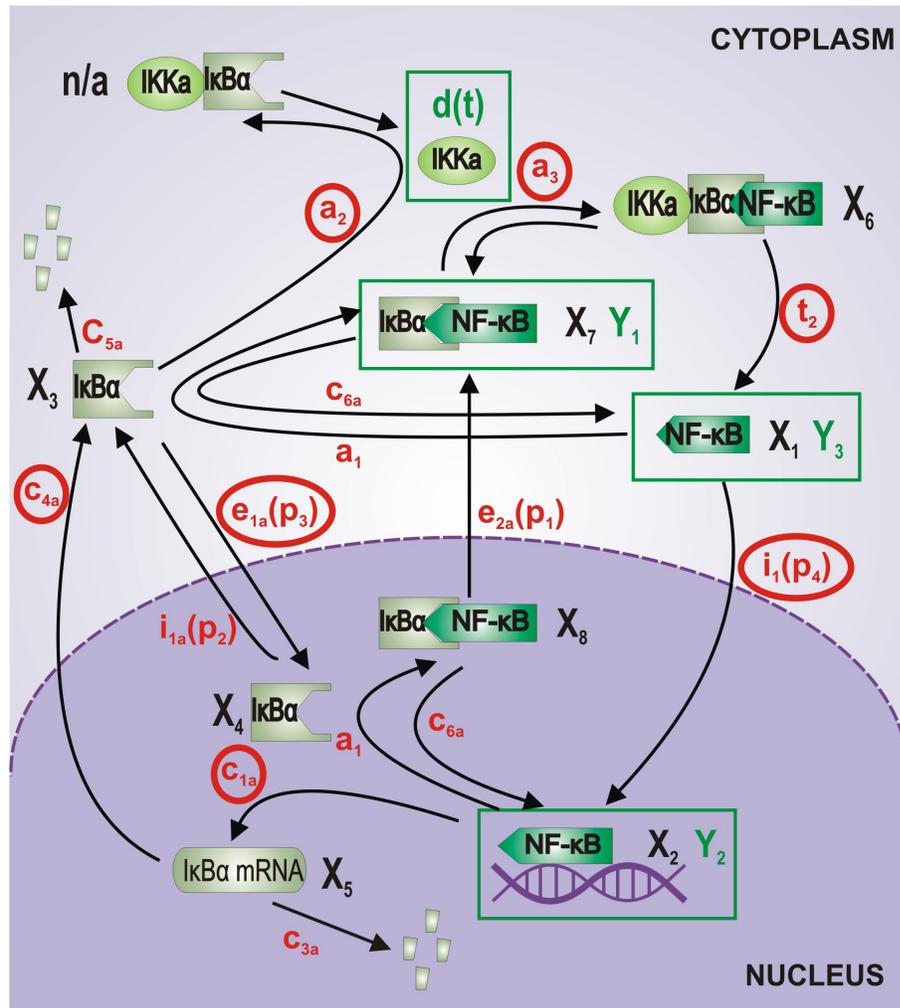


Figure 3.1: Schematic pathway representation of the reduced NF- $\kappa$ B model with unknown parameters circled. The green squares represent the measurable variables.

### 3.3.1 Algebraic Identifiability

Consider a nonlinear system with no input variables of the form:

$$\Sigma : \begin{cases} \dot{x} = f(x, \theta), & x(0) = x_0, \\ y = h(x, \theta), \end{cases} \quad (3.9)$$

where  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^p$ , and  $\theta \in P \subset \mathbb{R}^q$  are the states, outputs, and unknown parameters respectively and  $P$  is a simply connected open subset of  $\mathbb{R}^q$  containing the allowable parameter values. The functions  $f(x, \theta)$  and  $h(x, \theta)$  are meromorphic functions on a simply connected open subset  $X \times P$  of  $\mathbb{R}^n \times \mathbb{R}^q$  where  $X$  is the subset obtained by considering restrictions on  $x$ . Assume that:

$$\text{rank} \frac{\partial h(x, \theta)}{\partial x} = p \quad (3.10)$$

holds for all  $x$  and that  $x_0$  is independent of  $\theta$ .

The system  $\Sigma$  is said to be algebraically identifiable if there exists a  $T > 0$ , positive integer  $k$ , and representation of that system as a meromorphic function  $\phi : \mathbb{R}^q \times \mathbb{R}^{(k+1)p} \rightarrow \mathbb{R}^q$  such that

$$\det \frac{\partial \phi}{\partial \theta} \neq 0 \quad (3.11)$$

and

$$\phi(\theta, y, \dot{y}, \dots, y^{(k)}) = 0 \quad (3.12)$$

hold on  $[0, T]$ , for all  $(\theta, y, \dot{y}, \dots, y^{(k)})$  where  $(\theta, x_0)$  belong to an open and dense subset of  $P \times X$  and  $y, \dot{y}, \dots, y^{(k)}$  are derivatives of the output  $y(t, \theta, x_0)$  (Xia and Moog, 2003).

Algebraic identifiability in this case means that one can express the parameters in terms of the measurable outputs of the system. Using (3.11) one can locally solve (3.12) for  $\theta$  using the implicit function theorem.

### 3.3.2 Observable Representation

The differential algebra method involves the elimination of the state variables  $x$  using the observability map of the system. The degree of differentiation required to eliminate  $x$  is referred to as the observability index  $k$  (Xia and Moog, 2003).

In practice, this can be accomplished heuristically by differentiating  $y$  until enough independent equations are generated to solve for all  $x$ . These values of  $x$  can then be substituted into  $y^{(k)}$  to obtain the observable representation

$$\phi(\theta, y_1, \dot{y}_1, \dots, y_1^{(k_1)}, y_2, \dot{y}_2, \dots, y_2^{(k_2)}, \dots, y_p, \dot{y}_p, \dots, y_p^{(k_p)}) = 0. \quad (3.13)$$

### 3.3.3 Algebraic Identifiability of the NF- $\kappa$ B Example

This method is applied to the reduced NF- $\kappa$ B model. In Section 3.2, a term  $d(t)$  that is considered a measurable disturbance is introduced. It is included in the same manner as  $y$  for the purpose of identifiability (i.e.  $d(t)$  and its derivatives are known). The calculations are done in Maple.

**Remark 2** *Introduction of the third output variable ( $y_3 = x_1$ ) is due to the fact that it is not possible to obtain the observability map of the system with only the two output variables ( $y_1 = x_2$  and  $y_2 = x_7$ ). Due to the presence of bilinear terms in the system, derivatives  $y^{(k)}$  with  $k > 2$  result in equations containing polynomials of order  $\geq 3$  which cannot be solved algebraically (numerical solvers are required). This means that*

for a system with at least one bilinear term the highest order derivative that can be used to find the observability map is  $y^{(2)}$ . This imposes an upper limit of  $3m$  states (where  $y_j \propto x_i$ ) on the system size for which a solution can be found.

**Remark 3** *The limitation on the number of states is only the case when the term is bilinear in terms of  $x$  despite the choice of  $y$ . For example,  $x_2x_4 = y_1x_4$  is linear in terms of the states so does not pose a problem. For the case of the NF- $\kappa$ B model, one requires an additional output to find the observable representation required for use of the differential algebra identifiability method (preferably one of the states in the remaining bilinear term  $x_1x_3$ ). Since it was already assumed that the concentration of NF- $\kappa$ B in the nucleus can be measured, it would be reasonable to assume that the concentration in the cytoplasm could be measured as well and, hence,  $y_3 = x_1$ .*

Taking into consideration the specific structure of the NF- $\kappa$ B model, a simplified observable representation can be found.

From the equation for  $\dot{y}_1$  below, one can see that the bilinear term  $x_2x_4$  is observable:

$$a_1(x_2x_4) = p_4y_3 - \dot{y}_1. \quad (3.14)$$

The equations for  $\dot{y}_2$  and  $\dot{y}_3$  both contain  $a_1x_1x_3$ . If one equates these expressions, the other bilinear term in the model can be canceled out obtaining

$$\dot{y}_3 = -\dot{y}_2 - i_2y_3 - a_3q_1 + t_2x_6 + e_{2a}x_8 \quad (3.15)$$

where  $q_1 = d(t)y_2$ . Taking the derivative of (3.15):

$$\ddot{y}_3 = -\ddot{y}_2 - i_2\dot{y}_3 - a_3\dot{q}_1 + t_2\dot{x}_6 + e_{2a}\dot{x}_8 \quad (3.16)$$

where

$$\dot{x}_6 = a_3 q_1 - t_2 x_6 \quad (3.17)$$

$$\dot{x}_8 = a_1(x_2 x_4) - p_1 x_8 \quad (3.18)$$

and substituting (3.14), (3.17) and (3.18) into (3.16) one gets

$$\ddot{y}_3 = \ddot{y}_2 - i_2 \dot{y}_3 - a_3 \dot{q}_1 + t_2 a_3 q_1 - t_2^2 x_6 + e_{2a} p_4 y_3 - e_{2a} \dot{y}_1 - e_{2a} p_1 x_8. \quad (3.19)$$

This results in two linear equations and two unknowns ( $x_6$  and  $x_8$ ) that can be solved for and substituted into  $y_3^{(3)}$  to obtain the reduced observable equation

$$\begin{aligned} y_3^{(3)} &= y_2^{(3)} + (e_{2a})y_1^{(2)} + (t_2 + p_1)y_2^{(2)} + (t_2 + i_1 + p_1)y_3^{(2)} + (a_3)q_1^{(2)} + (t_2 e_{2a})y_1^{(1)} \\ &+ (t_2 p_1)y_2^{(1)} + (p_1 i_1 - e_{2a} p_4 + t_2 p_1 + t_2 i_1)y_3^{(1)} + (a_3 p_1)q_1^{(1)} + (t_2 i_1 p_1 - t_2 e_{2a} p_4)y_3 \end{aligned} \quad (3.20)$$

This equation can be rewritten in the form

$$y_3^{(3)} = y_2^{(3)} + \Psi_1 y_1^{(2)} + \Psi_2 y_2^{(2)} + \Psi_3 y_3^{(2)} + \Psi_4 q_1^{(2)} + \Psi_5 y_1^{(1)} + \Psi_6 y_2^{(1)} + \Psi_7 y_3^{(1)} + \Psi_8 q_1^{(1)} + \Psi_9 y_3 \quad (3.21)$$

where

$$\Psi = \begin{bmatrix} e_{2a} \\ t_2 + p_1 \\ t_2 + p_1 + i_1 \\ a_3 \\ t_2 e_{2a} \\ t_2 p_1 \\ i_1 p_1 - e_{2a} p_4 + t_2 p_1 + i_1 t_2 \\ a_3 p_1 \\ t_2 i_1 p_1 - e_{2a} p_4 t_2 \end{bmatrix} \quad (3.22)$$

with higher order derivatives expressed as

$$y_3^{(i)} = y_2^{(i)} + \Psi_1 y_1^{(i-1)} + \Psi_2 y_2^{(i-1)} + \Psi_3 y_3^{(i-1)} + \Psi_4 q_1^{(i-1)} + \Psi_5 y_1^{(i-2)} \\ + \Psi_6 y_2^{(i-2)} + \Psi_7 y_3^{(i-2)} + \Psi_8 q_1^{(i-2)} + \Psi_9 y_3^{(i-3)}. \quad (3.23)$$

One can obtain a system of 9 equations and 9 unknowns by further differentiating (3.22). This means the lumped parameters are identifiable and can be computed from any  $y(t)$  such that:

$$\text{rank} \partial(y_3^{(3)}, \dots, y_3^{(11)}) / \partial(\Psi_1, \dots, \Psi_9) = 9.$$

Once the identifiability of the lumped parameters is known, the unknown parameters  $(a_2, a_3, c_{1a}, c_{4a}, t_2, e_{1a}, i_1, p_3, p_4)$  are identifiable if the following conditions are satisfied:

$$\text{rank}\left(\frac{\partial \Psi}{\partial \theta}\right) = q, \quad \det\left(\frac{\partial \Psi}{\partial \theta}\right) = 0$$

where  $q$  is the number of unknown parameters.

In this case  $\text{rank}\left(\frac{\partial \Psi}{\partial \theta}\right) = 4$ . One can see that this is due to the fact that only  $a_3, p_4, t_2$  and  $i_1$  appear in Equation (3.22). The rank of the Jacobian considering only those parameters is the same meaning they are identifiable as long as:  $t_2(t_2 - p_1)e_{2a} \neq 0$ .

From a biological standpoint, this means the rate of formation and degradation of (IKKa|IκBα|NF-κB), the transport coefficient for movement of free NF-κB into the nucleus, and the ratio of cytoplasmic to nuclear volume ( $k_v = p_4/i_i$ ) can be determined with the given choice of outputs.

This method is largely heuristic in nature as it involves manipulation of the equations to determine the observability map. Similar results are obtained by applying the Rosenfeld-Groebner algorithm, as used in Denis-Vidal et al. (2001), to obtain the observable representation as the characteristic set of the system of equations. The full results are not presented here but this results in a more complicated observability map that also serves to verify the identifiability of the parameters described above. However, the same limitations in finding the observable representation, and hence the identifiability of the parameters, remain.

### 3.4 Summary

In this chapter, the identifiability of unknown parameters in a reduced model of the NF- $\kappa$ B signal transduction pathway is analyzed using existing methods. However, these methods were not developed with systems biology problems in mind and have only been proven to work for simple examples. Using differential algebra one can determine that 4 of the 9 parameters are identifiable considering the choice of outputs. Specifically these are the rate of formation and degradation of (IKK $\alpha$ |I $\kappa$ B $\alpha$ |NF- $\kappa$ B), the transport coefficient for movement of free NF- $\kappa$ B into the nucleus and the ratio of cytoplasmic to nuclear volume.

However, it is important to note that the identifiability analysis itself can only be carried out after making significant simplifications to the model. The number of states have to be reduced through the model reduction and finding a solution is highly dependent on the choice of outputs. In biological systems, one is extremely limited in terms of what can be measured experimentally so the addition of more outputs is not necessarily feasible. One must also consider that this system is of moderate size

and complexity in comparison to the range of systems biology problems available. This suggests existing observability based methods are not well suited for analyzing identifiability of biological systems.

The difficulty with the current observability based methods examined is that finding the observability of nonlinear systems in continuous time is a highly challenging problem in itself without even considering the parameter values. In fact, the limitations of these methods lay in the number of states they can handle as they require some sort of analytical solution to be found. For this reason, it is necessary to develop methods that are not as dependent on the size and nonlinearity for their computation. In the next section, a simulation based method which considers the empirical observability Gramian as a measure of identifiability is proposed.

# Chapter 4

## Empirical Gramian Based Identifiability Method<sup>1</sup>

### 4.1 Introduction

As mentioned in the previous chapter, the difficulty with the current observability based identifiability methods considered is that they require an analytical solution to be found. This restricts the size and complexity of problems that can be handled. This conflicts with the fact that many currently studied biological models (such as the JAT-STAT pathway in Yamada et al. (2003) and the NF- $\kappa$ B network in Lipniacki et al. (2004) discussed earlier) contains many parameters and states (31 states, 52

---

<sup>1</sup>The contents of this chapter were originally found in:

D. Geffen, R. Findeisen, M. Schliemann, F. Allgöwer, and M. Guay, “Observability based parameter identifiability for biochemical reaction networks”, in *Proceedings 2008 American Control Conference*, Seattle, Washington, USA, 2008

D. Geffen, M. Guay, R. Findeisen, and F. Allgöwer, “A method for the use of empirical Gramians to determine parameter identifiability in systems biology”, submitted for acceptance to: *2008 Conference on Decision and Control*, Cancun, Mexico, 2008

parameters and 15 states, 29 parameters respectively). This suggests the existing methods considered are not suitable for systems biology problems.

For this reason, much of the current focus in identification or identifiability research in systems biology research has shifted to simulation based methods. The approaches in Hengl et al. (2007); Muller et al. (2002); Quaiser et al. (2007) use a combination of simulation, optimization, and parameter estimation to check for functional relationships between parameters as a measure of identifiability (if a functional relationship between two or more parameters exists, those parameters are unidentifiable). However, as mentioned in Section 2.5.2, simulation based methods only provide local identifiability and require an accurate estimate of the experimental error. They also require parameters to first be estimated before a decision about their identifiability is made which is in itself a challenge for biological systems. These methods also look primarily at ‘practical identifiability’ which is an *a posteriori* method that considers the quality and possible measurement noise of the data as well. No simulation based methods exist that consider structural or *a priori* identifiability which is the focus here. In order to achieve this goal an observability based approach, such as those in Chappell et al. (1990); Denis-Vidal et al. (2001); Godfrey and Fitch (1984); Ljung and Glad (1994); Pohjanpalo (1978); Xia and Moog (2003), but which uses a simulation based approximation instead of an analytical solution is proposed.

Empirical observability and controllability Gramians are first introduced in Lall et al. (2002) for the purpose of model reduction. The use of such Gramians for observability (and controllability) analysis itself was proposed in Singh and Hahn (2005) and found to perform well compared to linear Gramians and Lie algebra based methods. However, there are currently no results for using such Gramians for identifiability

analysis. Such a technique is described in this chapter. Issues such as the computation of the Gramian, its subsequent analysis to extract the desired information, and numerical sensitivity of the method are discussed. The approach is demonstrated using a simple biological example of microbial growth with Michaelis-Menten kinetics.

This chapter is structured as follows. Section 4.2 defines the system class considered. In Section 4.3 the observability Gramian is defined in the linear and empirical case and its use for identifiability is commented on. Specific computational considerations for the empirical identifiability Gramian, as well as a proposed method to determine the unidentifiable parameters, are discussed in Section 4.4. The derived concepts are demonstrated through the use of a biological example, a model of microbial growth with Michaelis-Menten kinetics, in Section 4.5. This is followed by a summary of the chapter in Section 4.6.

## 4.2 System Description

A nonlinear dynamic system of the following form is considered:

$$\Sigma : \begin{cases} \dot{x} = f(x, \theta) , & x(0) = x_0, \\ y = h(x, \theta), \end{cases} \quad (4.1)$$

where  $x \in X \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^p$ , and  $\theta \in P \in \mathbb{R}^q$  are the states, outputs, and unknown parameters respectively.  $f$  and  $h$  are assumed to be sufficiently smooth vector valued functions of  $x$  and  $\theta$ .  $P$  is a simply connected open subset of  $\mathbb{R}^q$  containing allowable parameter values and  $X$  is a simply connected open subset of  $\mathbb{R}^n$  containing allowable states. This is referred to as the operating region of the system. Note that inputs

are not included here for simplicity of presentation however the results are easily extended to include this case.

For the purpose of the identifiability Gramian approach, the question of identifiability is restated as one of observability by including parameters as states. Since the parameters are time invariant, one obtains the following augmented system:

$$\tilde{\Sigma} = \begin{cases} \dot{\tilde{x}} = \begin{bmatrix} \dot{x} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} f(x, \theta) \\ 0 \end{bmatrix}, \quad \tilde{x}(0) = \begin{bmatrix} x_0 \\ \theta_0 \end{bmatrix}, \\ y = h(x, \theta), \end{cases} \quad (4.2)$$

where  $\tilde{x} \in \mathbb{R}^{\tilde{n}}$ ,  $\tilde{n} = n+q$  is the augmented state vector containing both the parameters and the original states.

## 4.3 Empirical Gramian for Identifiability

This novel identifiability approach is based on analyzing the observability of the expanded system. For this an empirical observability Gramian based approach is used. The results for linear Gramians are first reviewed.

### 4.3.1 Linear Observability Gramian

For linear systems of the form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) \\ y(t) &= Cx(t) \end{aligned}$$

where  $A$  and  $C$  are matrices of appropriate size. The linear observability Gramian (also referred to here as the analytical observability Gramian) is given by:

**Definition 3** (Analytical observability Gramian)

$$W_{obs_{analytical}} = \int_0^{\infty} e^{A^T t} C^T C e^{At} dt. \quad (4.3)$$

For stable systems, the system is observable if the observability Gramian is of full rank (Fairman, 1998). However, it is not possible to apply this Gramian to nonlinear systems as, in general, nonlinear energy functions are computationally intensive to compute. Furthermore, it is not desirable to linearize biological models as parameters have physical significance such as describing reaction kinetics or transport that would be lost. For this reason, the empirical observability Gramian introduced in Lall et al. (2002) is considered here.

### 4.3.2 Empirical Observability Gramian

The empirical observability Gramian is essentially an approximation of the analytical Gramian that can handle nonlinear systems. The empirical observability Gramian, as introduced in Lall et al. (2002), is defined as follows:

**Definition 4** (Empirical observability Gramian)

$$W_{obs_{empirical}} = \sum_{l=1}^r \sum_{m=1}^s \frac{1}{r s c_m^2} \int_0^{\infty} T_l \Psi^{lm}(t) T_l^T dt$$

where  $\Psi^{lm}(t) \in \mathbb{R}^{n \times n}$  is given by:

$$\Psi_{ij}^{lm}(t) = (y^{ilm}(t) - y_{ss}^{ilm})^T (y^{jlm}(t) - y_{ss}^{jlm})$$

The Gramian,  $W_{obs_{empirical}}$ , is constructed by simulation using output values,  $y^{ilm}(t)$ , obtained by perturbing the initial states of the system within a range of interest. Here  $y^{ilm}(t)$  is the time varying output and  $y_{ss}^{ilm}$  is the steady state output of the system for the particular ‘experiment’ given by the initial state:

$$x(0)^{ilm} = c_m T_l e_i + x_{nom}. \quad (4.4)$$

These initial states are obtained through perturbations around the nominal value  $x_{nom}$ .

The perturbations are defined by the following sets:

$$\begin{aligned} T^n &= \{T_1, \dots, T_r; \quad T_l \in \mathbb{R}^{n \times n}, \quad T_l^T T_l = I, \quad l = 1, \dots, r\}, \\ M &= \{c_1, \dots, c_s; \quad c_m \in \mathbb{R}, \quad c_m > 0, \quad m = 1, \dots, s\}, \\ E^n &= \{e_1, \dots, e_n; \quad \text{standard unit vectors in } \mathbb{R}^n\}. \end{aligned}$$

Here  $r$  are the number of matrices that describe the direction of the perturbations and  $s$  are the number of different perturbation magnitudes for each direction. The perturbations are related to the reasonable or desired operating range of the system.

The following section outlines how the empirical Gramians can be used for parameter identifiability analysis.

### 4.3.3 Empirical Gramian for Identifiability Analysis

The empirical observability Gramian can be used to determine parameter identifiability by applying it to the augmented system in (4.2).

**Remark 4** *Note that for the application of the empirical Gramian, it is typically required that the complete system is exponentially stable (at least locally over the region considered) so the Gramian does not approach infinity. However, due to the addition of the time invariant parameter terms as states, the augmented system is only neutrally stable. This does not pose a problem as long as the outputs considered are stable since they are being subtracted by their steady state values for the construction of the Gramian. Alternatively, the parameters can be included as  $\dot{\theta}_i = -10^{-8}(\theta_i - \theta_{i,nom})$  to make the system exponentially stable while still approximating the time invariant case.*

As in the linear/analytical case, the system is observable if the observability Gramian is of full rank. In the case of the empirical Gramian, the system is only locally observable over the operating range considered. As a direct consequence, it follows, that since the augmented system state consists of both states and parameters, local observability of the augmented system implies the identifiability over the operating range considered. Thus one can check identifiability by employing the observability Gramian.

**Remark 5** *As opposed to other methods which require starting points for parameters such as those based on sensitivity analysis, the Gramian based identifiability method considers a region in the parameter (and state) space not just a single point so accurate estimates for the nominal values are not required. This means that local identifiability using the empirical Gramian refers to the entire region considered not just for the nominal values chosen which is the case for many of the existing simulation based methods.*

**Remark 6** *Additional simplifications can be made to the Gramian of the augmented*

system so that only the part dealing with identifiability is isolated. This is beneficial as it eliminates the effect of the cross terms (for example, the effect of changing the parameters on the time course of the states, which can indirectly effect the output). In this case,  $W_{obs\_empirical}$  of the augmented system can be decomposed as follows:

$$W_O^{(\tilde{n} \times \tilde{n})} = \left[ \begin{array}{c|c} W_X^{(n \times n)} & W_{X\theta}^{(n \times q)} \\ \hline W_{\theta X}^{(q \times n)} & W_{\theta}^{(q \times q)} \end{array} \right]. \quad (4.5)$$

The  $(q \times q)$  empirical identifiability Gramian can then be defined as:

$$W_{ident} = W_{\theta}. \quad (4.6)$$

If this matrix has full rank (rank =  $q$  where  $q$  is the number of unknown parameters) then those parameters are identifiable.

#### 4.3.4 Illustrative Examples

The use of empirical Gramians for identifiability analysis can be demonstrated through the use of simple examples.

##### Example 1: Parameters Not Identifiable

The following is a simple example of a system in which the parameters are not identifiable.

$$\begin{cases} \dot{x} &= -(k_1 + k_2)x \\ y &= x \end{cases}$$

It is clear by looking at the equations that  $k_1$  and  $k_2$  cannot be identified individually as they only appear in a functional relationship with each other. Perturbations in the region of  $\pm 1$  about the following nominal values for the states and parameters are considered:

$$\tilde{x}_{nom} = \begin{bmatrix} x_0 \\ k_1 \\ k_2 \end{bmatrix}_{nom} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The empirical identifiability Gramian in this case is:

$$W_{ident} = \begin{bmatrix} 15.5521 & 15.5521 \\ 15.5521 & 15.5521 \end{bmatrix}.$$

The matrix is not of full rank, meaning the system is not identifiable. By looking at the following eigenvalue/eigenvector decomposition, one can gain more insight into the system.

$$\lambda = \begin{bmatrix} 0 & 0 \\ 0 & 31.1041 \end{bmatrix}$$

$$v = \begin{bmatrix} -0.7071 & 0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}.$$

One can see that the eigenvector corresponding with the identifiable eigenvalue contains both parameters in equal parts which shows that while the parameters themselves are not identifiable the combination  $(k_1 + k_2)$  is. The parameters can be identified individually with the introduction of a second state and output as shown in the

next example.

### Example 2: Identifiable Parameters

A simple example of a system in which the parameters are identifiable is as follows:

$$\begin{cases} \dot{x}_1 &= -(k_1 + k_2)x_1 \\ \dot{x}_2 &= -k_2x_2 \\ y &= (x_1, x_2)^T. \end{cases}$$

It is clear that since both  $x_1$  and  $x_2$  are outputs,  $k_2$  is identifiable from the second expression, allowing  $k_1$  to be individually identified from the first. As in the first example, perturbation of up to  $\pm 1$  are considered about nominal values:

$$\tilde{x}_{nom} = \begin{bmatrix} x_{10} \\ x_{20} \\ k_1 \\ k_2 \end{bmatrix}_{nom} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The empirical identifiability Gramian in this case is:

$$W_{ident} = \begin{bmatrix} 15.552 & 15.552 \\ 15.552 & 78.090 \end{bmatrix}.$$

The eigenvalue/eigenvector decomposition gives:

$$\lambda = \begin{bmatrix} 11.898 & 0 \\ 0 & 81.1744 \end{bmatrix}$$

$$v = \begin{bmatrix} -0.9735 & -0.2287 \\ 0.2287 & -0.9735 \end{bmatrix}.$$

The Gramian is of full rank meaning the augmented system is observable and, in turn, that the parameters are identifiable. It should be noted that this form of identifiability analysis could also be considered a type of sensitivity analysis. This is due to the fact that instead of giving a yes or no answer to the question of identifiability, the Gramian also gives an indication of ‘how identifiable’ a parameter is based on the relative sizes of the eigenvalues. While the results and analysis of the systems considered so far are relatively straight forward, this is not the case when considering larger biological systems. For this reason, it is necessary to develop a structured approach for using the Gramian to determine which parameters are identifiable from a given choice of outputs.

## 4.4 Description of Method

Two questions arise when considering the outlined approach. Firstly, how to efficiently calculate  $W_{obs_{empirical}}$  and secondly how to decide based on the Gramian if the parameters are identifiable.

### 4.4.1 Calculation of Gramian

The empirical Gramian can be calculated using experimental or simulated data. The difficulty with this numerical method, however, lies in the number of degrees of freedom one is presented with in terms of computation. In addition to this, there are a number of other considerations that must be made when including parameters as states for identifiability purposes. For these reasons, a step by step approach for the use of this method is presented.

#### Algorithm for calculating the empirical identifiability Gramian

1. *Choose an operating region of interest and select nominal values for states and parameters :*

Consider a region of biological interest for parameters and states such as what would be examined experimentally for the purpose of parameter estimation. For the nominal values of the states one could use the midpoint of the desired operating range or a steady state point in the system (at the chosen parameter values).

2. *Define perturbation directions:*

There are several possibilities for how to define the perturbation directions. The use of a  $2^{\tilde{n}}$  factorial design is proposed here (for larger systems a fractional factorial design such as the Plackett-Burman design (see eg. Montgomery, 1997) must be used to speed up the computation). This results in an  $(\tilde{n} \times 2^{\tilde{n}})$  T matrix made up of  $\pm 1$  which is made orthogonal by dividing it by  $\sqrt{2^{\tilde{n}}}$ .

**Remark 7** *Following Definition 4,  $T$  is an  $(n \times n)$  orthogonal matrix and multiple ( $l = 1, \dots, r$ )  $T$  matrices are used to define the perturbation directions considered. These requirements can be relaxed as long as the following adjustments are made to ensure that the dimensions of the Gramian remain the same. The sets describing the perturbation direction become:*

$$\begin{aligned} T^{\tilde{n}} &= \{T \in R^{\tilde{n} \times 2\tilde{n}}, T^T T = I\} \\ E^{2\tilde{n}} &= \{e_1, \dots, e_{2\tilde{n}}; \text{ standard unit vectors in } R^{2\tilde{n}}\}. \end{aligned}$$

**Remark 8** *There are several reasons for using this approach when looking at parameter identifiability, as opposed to the ones in (see eg. Lall et al., 2002; Singh and Hahn, 2005). The factorial design allows for the consideration of every perturbation combination, without introducing additional magnitudes as is the case when random  $(n \times n)$  matrices are used. In addition, it facilitates the use of an equilibrium point in the system as a nominal state. In Lall et al. (2002) the use of  $T = [I, -I]$  is suggested which essentially translates into experiments in which only one parameter or state is perturbed at a time. This approach proves to be disadvantageous when considering parameter identifiability of biological systems, as the steady state value for at least some states is often zero. It is then impossible to obtain information about the parameters associated with those states. If the steady state is zero, one must also be careful in using it as a nominal state value since this will only allow for biologically*

*meaningful perturbations in one (the positive) direction.*

3. *Define perturbation magnitude and scaling:*

The different perturbation magnitudes are described by the set:

$$M = \{c_1, \dots, c_s; c_m \in \mathbb{R}, c_m > 0, m = 1, \dots, s\}.$$

It is important to note that the system is scaled prior to perturbation. One can include scaling considerations when determining the perturbations by rewriting the first term of (4.4) as:

$$\tilde{x}_{pert}^{im} = c_m S^{-1} T e_i$$

where  $S$  is the scaling matrix. The scaling can also reflect the different feasible operating regions of the states and parameters. One would want to consider a larger range of parameter values as the exact nominal values are not known.

4. *Calculate initial conditions for each ‘experiment’:*

For the scaled case in which a factorial design is used, this can be written as:

$$\tilde{x}(0)^{im} = c_m S^{-1} T e_i + \tilde{x}_{nom} = \tilde{x}_{pert}^{im} + \tilde{x}_{nom}.$$

5. *Scale the states and parameters to use for Gramian calculations.*

The system is scaled by including the augmented states as follows:

$$\hat{\tilde{x}}_i = \frac{\tilde{x}_i - \tilde{x}_{i,low}}{\Delta\tilde{x}_i}$$

where  $\tilde{x}$  is the actual state or parameter,  $\tilde{x}_{low}$  and  $\Delta\tilde{x}$  are the minimum value and range of  $\tilde{x}$  respectively for all ‘experiments’ considered, and  $\hat{\tilde{x}}$  is the scaled state or parameter value. The scaling is such that  $\hat{\tilde{x}} \in [0, 1]$  over time for all perturbations. These scaled values are used to construct the empirical Gramian to ensure that parameters of different orders of magnitude are not disproportionately represented in the Gramian. The use of such scaling is also beneficial in the case where more than one output is considered so that each output is given equal weighting in the Gramian calculation.

6. *Integrate the system for the initial conditions defined by each ‘experiment’ and use the resulting outputs to construct the Gramian according to Definition 4.*

**Remark 9** *While the initial Gramian definition considers the integral from zero to infinity, a finite end time,  $t_F$ , must be used for calculation purposes. As long as  $t_F$  is greater than the time it takes for the output(s) to reach steady state, the results will be the same.*

The calculations lead to the empirical Gramian which must be analyzed further in order to obtain information about which parameters are unidentifiable.

### 4.4.2 Identifiability Analysis

Since the empirical Gramians are originally introduced for the purpose of model reduction, little attention has been paid to criteria for determining observability (or controllability) using such Gramians. For analytical Gramians, observability is determined by checking the rank of the matrix. However, due to the nature of the empirical Gramian calculations, which involve numerical approximation and simulation, singular value or eigenvalue/eigenvector decomposition should be used as a measure of the effective rank of the matrix (Konstantinides, 1988; Stewart, 1984).

The entries of the empirical observability Gramian give an indication of the energy obtained by observing the output of the system with initial conditions of  $x$ . The states that are difficult to observe are those that lie (or have a significant component) in the span of the eigenvectors of the observability Gramian corresponding to small eigenvalues. However, when looking at the eigenvalues or singular values for the empirical Gramian, the question of how small is significantly small to be considered unobservable, or in turn unidentifiable, remains.

In theory, the rank of a matrix,  $A$ , can be determined by counting the number of nonzero singular values. If  $r$  is the actual rank of an  $(n \times n)$  matrix, the singular values can be denoted as:

$$\Sigma = \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_r > \alpha_{r+1}, \dots, \alpha_n = 0.$$

However, in practice, the observed matrix,  $B$ , consists of the actual matrix in addition to the rounding error caused by finite precision numerical operations and other sources of (numerical) noise (see Konstantinides, 1988) which can be represented

by:  $B = A + E$  where  $E$  is the error term. The singular values of this error influenced matrix are:

$$\beta_1 \geq \beta_2 \geq \dots \geq \beta_r \geq \beta_{r+1} \geq \dots \geq \beta_n \geq 0$$

where  $(\beta_{r+1}, \dots, \beta_n)$  are generally small but not necessarily zero. Since  $\beta_r$  may also be small, a threshold for the magnitude of negligible terms must be established. From Konstantinides (1988), one can establish the following criteria for the ‘effective rank’,  $r_{\text{effective}}$ , of the matrix:

$$\beta_1 \geq \beta_2 \geq \dots \geq \beta_{r_{\text{effective}}} > \epsilon \geq \beta_{r_{\text{effective}}+1} \geq \dots \geq \beta_n \quad (4.7)$$

where  $\epsilon$  is the Frobenius norm or 2–norm of  $E$ .

In the case of the empirical observability Gramian, the matrix also contains the error associated with approximating the analytical Gramian for the nonlinear system with the empirical one. One must determine a way to quantify the approximation error to determine reasonable bounds for determining the effective rank using singular value (or eigenvalue/eigenvector) decomposition. Since the analytical Gramian cannot be computed in the nonlinear case, the error of approximation can be estimated by comparing  $Wobs_{\text{empirical}}$  and  $Wobs_{\text{analytical}}$  for the system after it is linearized about its nominal operating point. In this case  $E$  can be estimated as follows:

$$E \approx Wobs_{\text{empirical}} - Wobs_{\text{analytical}}. \quad (4.8)$$

The threshold is given by the Frobenius-norm of  $E$  as

$$\epsilon = \|E\|_F. \quad (4.9)$$

## 4.5 Biological Example

For a biological example, a model of microbial growth derived using Michaelis-Menten kinetics from Chappell and Godfrey (1992); Holmberg (1982) is used. The system is given by:

$$\begin{cases} \dot{x} = \frac{\mu_m b_2 l(t) x(t)}{K_s + b_2 l(t)} - K_d x(t) \\ \dot{l} = -\frac{\mu_m l(t) x(t)}{Y(K_s + b_2 l(t))} \\ x(0) = x_0, \quad l(0) = 1 \end{cases} \quad (4.10)$$

where  $l(t) = \frac{1}{b_2} s(t)$ ,  $s(t)$  is the concentration of substrate and  $x(t)$  the concentration of product. The question being examined is whether or not the unknown parameter set:

$$\theta = (\mu_m, k_d, b_2, K_s, Y)^T$$

is identifiable over the range of system states and parameters considered. The measurable output in this case is given by:

$$y(t) = x(t).$$

The structural identifiability of this system has been previously analyzed in Denis-Vidal and Joly-Blanchard (2000) using a different approach based on several existing methods and it is determined that the parameters  $b_2$ ,  $K_s$ , and  $Y$  are unidentifiable. This can be used to verify the results of the new method.

### 4.5.1 Application of the Identifiability Gramian

For the microbial growth example, the nominal values for the states and parameters are taken from those used in the simulation and sensitivity analysis in Holmberg (1982) and are as follows:

$$\tilde{x}_{nom} = \begin{bmatrix} x_0 \\ l_0 \\ \mu_m \\ K_d \\ b_2 \\ K_s \\ Y \end{bmatrix}_{nom} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.5 \\ 0.05 \\ 10 \\ 3.0 \\ 0.6 \end{bmatrix}.$$

In this case, perturbations about the nominal value of  $\pm 20\%$  for the states and  $\pm 50\%$  for the parameters are used for the calculation. A larger range of parameter values are considered as the exact or real values are not known. A  $2^7$  full factorial experimental design is used for defining the perturbation directions. The empirical identifiability Gramian in this case is as follows:

$$W_{ident} = \begin{bmatrix} 0.3277 & -0.1056 & 0.3164 & -0.2273 & 0.1311 \\ -0.1056 & 0.7844 & -0.6228 & 0.0933 & -0.5298 \\ 0.3164 & -0.6228 & 0.6738 & -0.2331 & 0.4715 \\ -0.2273 & 0.0933 & -0.2331 & 0.1607 & -0.1024 \\ 0.1311 & -0.5298 & 0.4715 & -0.1024 & 0.3798 \end{bmatrix}.$$

The singular value decomposition of the matrix leads to:

$$\Sigma = \begin{bmatrix} 1.000 \\ 0.252 \\ 4.57 \times 10^{-3} \\ 8.17 \times 10^{-4} \\ 7.41 \times 10^{-4} \end{bmatrix} .$$

Since  $W_{ident}$  is a symmetrical matrix that only has positive eigenvalues, the singular values and eigenvalues are identical and can thereby be used interchangeably. It should also be noted that both singular values and eigenvalues are expressed here as relative values. One can see that the smaller singular values,  $\sigma_4$  and  $\sigma_5$  are several orders of magnitude smaller than the larger ones so the parameters associated with them could be unidentifiable. This distinction becomes increasingly unclear in the case of  $\sigma_3$  which is only two orders of magnitude smaller than the larger values. In this case, the threshold,  $\epsilon$ , needed to identify the negligible singular values must be calculated.

The system is linearized about the operating point used for calculation of  $W_{obs_{empirical}}$  in the nonlinear case. Both the analytical and empirical Gramian are then calculated for the linearized system and compared to find the error of the approximation as the threshold described in (4.7). In this case,  $\epsilon = 0.0079$ . Note that in this case the threshold has also been divided by the norm of the eigenvalues to put it on a comparable scale to the relative eigenvalues/singular values considered. This threshold results in  $r_{effective} = 2$  for the Michaelis-Menten example which suggests that 3 out of 5 parameters are unidentifiable. However, the question of which parameters are

actually unidentifiable still remains and is answered in the following section.

### 4.5.2 Determination of Unidentifiable Parameters

The parameters with significant components in the span of the eigenvectors of the identifiability Gramian corresponding to small eigenvalues can be considered candidates for being unidentifiable. An iterative method for determining the unidentifiable parameters using the calculated threshold  $\epsilon$  is proposed.

#### Algorithm for finding which parameters are unidentifiable:

1. Compute  $W_{ident}$  for the nonlinear system using all parameters  $\Theta = (\theta_1, \dots, \theta_q)$  (where  $q$  is the number of unknown parameters).
2. Calculate  $W_{obs_{analytical}}$  and  $W_{obs_{empirical}}$  for the linearized system to determine the threshold:  $\epsilon = \|W_{obs_{empirical}} - W_{obs_{analytical}}\|_F$ .
3. Find the eigenvalue/eigenvector decomposition and order so  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$ .
4. If  $\lambda_q \geq \epsilon$  then all parameters can be considered identifiable and the iterations stop. Else find the parameter with the maximum component in the eigenvector corresponding to the smallest eigenvalue ( $\max(v_q)$ ). If there are several eigenvalues of the same order of magnitude, the maximum component of a weighted average of the corresponding eigenvectors (weighted using the corresponding eigenvalues) can also be used to find the unidentifiable parameter candidate.
5. Remove the parameter (i.e. assume it is known, do not perturb and set  $q = q - 1$ ) and recalculate the Gramian. If the minimum eigenvalue increases, then add that parameter to the set of unidentifiable parameters:  $\Theta_{unidentifiable} \in$

$\mathbb{R}^{q-r_{\text{effective}}}$ . Else, look at the second highest component of that vector as a candidate.

6. Repeat steps 3-5 until  $\min(\lambda_i) \geq \epsilon$ . The unidentifiable parameters are given by  $\Theta_{\text{unidentifiable}}$ .

The results of this approach for the Michaelis-Menten example are summarized in Table 4.1. Recall that  $\epsilon$  in this case is  $7.90 \times 10^{-3}$ .

Table 4.1: Results for use of algorithm on microbial growth example

Iteration	Removed Parameter	Minimum Eigenvalue
0	-	$7.406 \times 10^{-4}$
1	$K_s$	$1.535 \times 10^{-3}$
2	$Y$	$1.378 \times 10^{-3}$
3	$b_2$	$3.316 \times 10^{-1}$

This suggests that the parameters  $b_2$ ,  $K_s$ , and  $Y$  are unidentifiable over the operating region considered. This is consistent with the results of Denis-Vidal and Joly-Blanchard (2000) obtained using a different parameter identifiability approach so can be used to verify the results of the new method.

The results can be demonstrated by plotting the simulation of the output over time. Figure 4.1 shows a comparison between the output for the nominal parameter values and the output when the parameters are perturbed in the direction of the eigenvectors corresponding to the maximum eigenvalue (identifiable) and the minimum eigenvalue (unidentifiable) respectively.

It is clear that the output obtained by perturbing the parameters in the direction of  $\nu$  for  $\max(\lambda)$  deviates greatly from the nominal case while the output obtained through perturbations in the direction of  $\nu$  for  $\min(\lambda)$  stays roughly the same. This

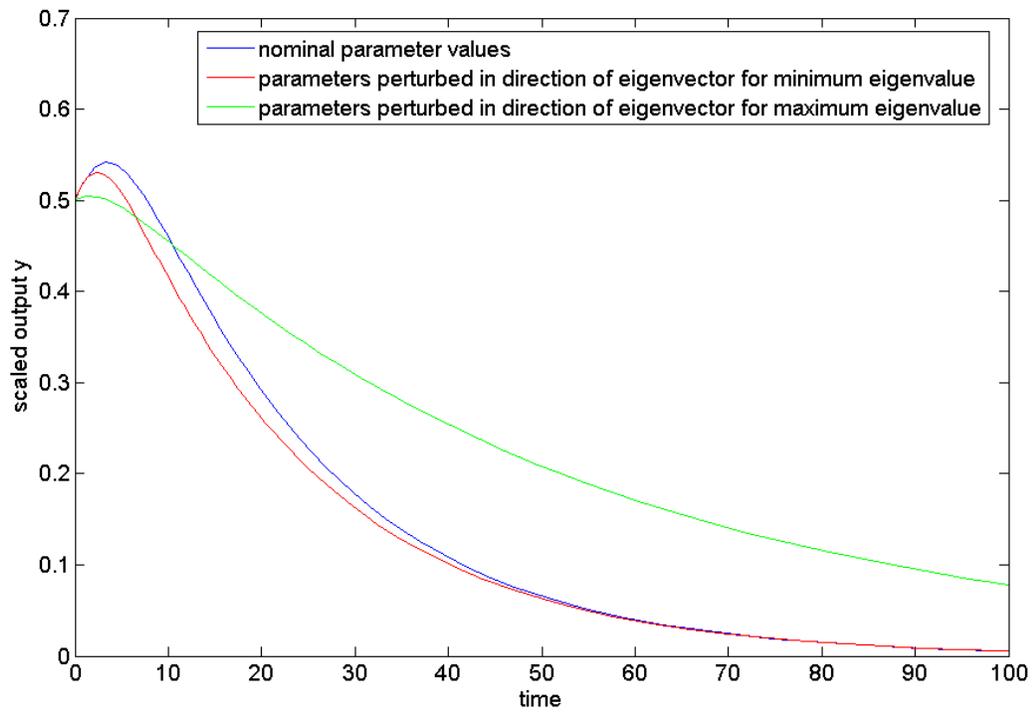


Figure 4.1: Effect on the observed output of perturbing nominal parameters in the direction of the eigenvectors of the identifiability Gramian

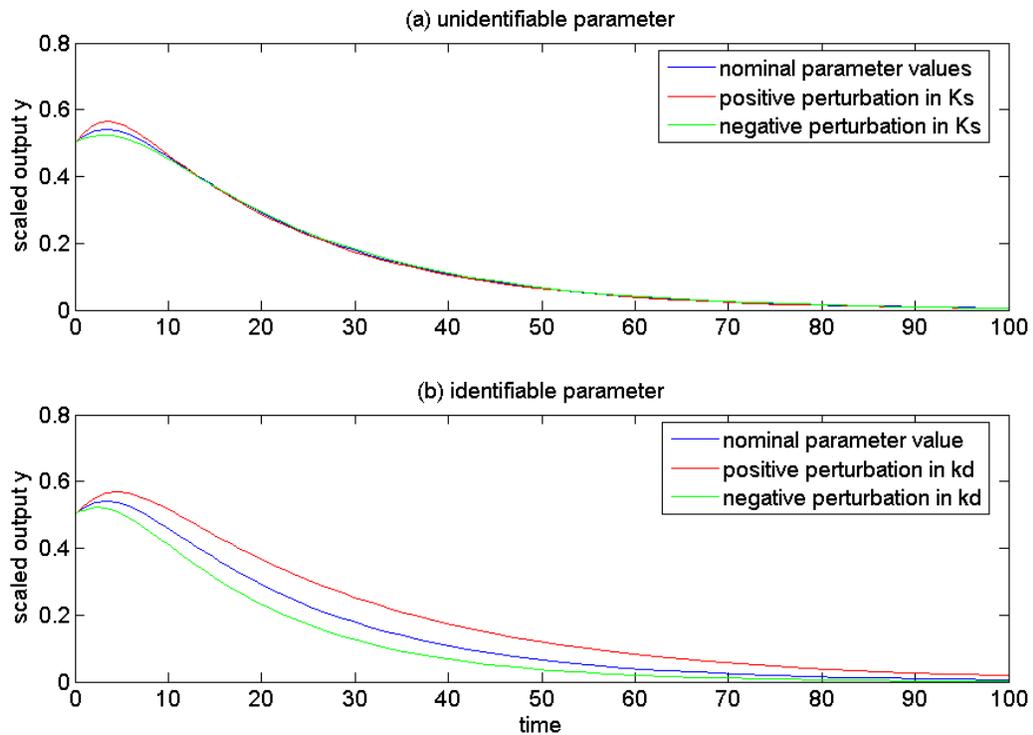


Figure 4.2: Effect of perturbing identifiable and unidentifiable parameters on the observed output

shows that perturbing the parameters has a large effect on the output in the identifiable case and almost no effect in the unidentifiable case as expected.

Perturbations in individual parameters themselves can also be examined to verify the results. Figure 4.2 compares the observed output for perturbations of a parameter determined to be unidentifiable using the empirical identifiability Gramian method ( $K_s$ ) with that of an identifiable parameter ( $k_d$ ). Perturbations in the unidentifiable parameter  $K_s$  have almost no effect on the measured output. This shows that the output cannot observe changes in the parameter so the parameter is therefore unidentifiable with the given output. Changes in the identifiable parameter  $k_d$ , however, are

shown to have an effect on the output confirming this parameter is identifiable.

## 4.6 Summary

In this chapter, a new observability based method to check the structural identifiability of a system using an empirical observability Gramian is introduced with specific focus on biochemical reaction network systems. The method for computing the Gramian is outlined including considerations that must be made to adapt the Gramian to include parameters for identifiability. Since the empirical Gramians are originally introduced for the purpose of model reduction, previous results focus on the analysis for these purposes. They do not consider how to further analyze the Gramian to extract the information desired here relating to observability (and in turn, for identifiability).

A method to determine which parameters are unidentifiable from the calculated empirical Gramian, using the singular value or eigenvalue/eigenvector decomposition is developed. The numerical sensitivity of the matrix calculation is also considered to compute the effective rank. The proposed approach allows one to distinguish eigenvalues with negligible values. Parameters more closely associated with the negligible eigenvalues can be considered to be unidentifiable. An algorithm is presented to compute the effective rank and identify the unidentifiable parameters. It is applied to the biological example of microbial growth using Michaelis-Menten kinetics from Denis-Vidal and Joly-Blanchard (2000); Holmberg (1982). This application of this algorithm to larger systems such as the full model of the NF- $\kappa$ B signal transduction network introduced in Section 3 is discussed in the next chapter.

# Chapter 5

## Application of the Empirical Identifiability Gramian Method to the NF- $\kappa$ B Example

### 5.1 Introduction

In Chapter 3, existing methods are attempted for an identifiability analysis of the NF- $\kappa$ B example in Section 2.6. The goal of this is to determine the parameters that can be estimated from the measurable outputs. However, the full model proves to be too large for the existing observability based methods to handle. This motivates the development of the simulation based empirical identifiability Gramian approach introduced in Chapter 3. The application of this new approach to the full NF- $\kappa$ B example is covered in this chapter to determine the structural identifiability. This serves a dual purpose. It allows one to obtain more information about the NF- $\kappa$ B pathway and demonstrate the ability of the new method for the analysis of larger and

more complex systems of the sort typically encountered in systems biology.

This chapter is structured as follows. Section 5.2 shows the application of the empirical identifiability Gramian to the reduced NF- $\kappa$ B model and a comparison of the results to those of the existing observability based methods. The new method is applied to the full NF- $\kappa$ B model in Section 5.3. A summary of the contents of the chapter is given in Section 5.4.

## 5.2 Application of Empirical Identifiability Gramian to Reduced NF- $\kappa$ B Example

The empirical identifiability Gramian method is applied to the reduced NF- $\kappa$ B example examined in Chapter 3 in order to check the consistency of the results. For this, the algorithm introduced in Section 4.4 is applied and the results are summarized in Table 5.1. The error of approximation of the nonlinear Gramian can be estimated as  $\epsilon = 5 \times 10^{-3}$  from comparing the linear analytical and empirical Gramian for several examples. Eigenvalues that are less than that threshold are considered to be unidentifiable. Recall that the unknown parameters considered in this

Table 5.1: Results for use of empirical identifiability Gramian method on reduced NF- $\kappa$ B model with same unknown parameters as in Section 3.2.

Iteration	Removed Parameter	Minimum Eigenvalue
0	-	$3.208 \times 10^{-4}$
1	$e_{1a}$	$9.530 \times 10^{-4}$
2	$a_2$	$1.407 \times 10^{-3}$
3	$c_{4a}$	$7.722 \times 10^{-3}$

case are  $\theta = (a_2, a_3, t_2, c_{1a}, c_{4a}, i_1, e_{1a}, k_v)$ . This means that the following parameters are determined to be identifiable using the empirical identifiability Gramian method:

$\theta_{identifiable} = (a_3, t_2, c_{1a}, i_1, k_v)$ . This can be compared to the results obtained using the existing differential algebra methods in Section 3.3. In this case, the identifiable parameters are:  $\theta_{identifiable} = (a_3, t_2, i_1, p_4)$ . Since  $p_4 = k_v i_1$  this can be rewritten as:  $\theta_{identifiable} = (a_3, t_2, i_1, k_v)$ . A comparison of the results for each method is shown below.

Table 5.2: Comparison of identifiable parameter of the reduced NF- $\kappa$ B model using the empirical identifiability Gramian and the differential algebra based methods

Method	Identifiable Parameters
Empirical identifiability Gramian	$a_3, t_2, i_1, k_v, c_{1a}$
Differential algebra	$a_3, t_2, i_1, k_v$

The differential algebra method returns four identifiable parameters while the empirical identifiability Gramian method results in five. It should be noted, however, that the four parameters found by the differential algebra method are a subset of the parameter set determined by the empirical identifiability Gramian approach. This shows some consistency between the parameter identifiability approaches. The fact that the differential algebra method results in one less parameter being identifiable could be due to the fact that the identifiability analysis is actually carried out on the observable representation of the system (in other words the state free description of the system defined by the observability map) instead of the system itself. This could result in the existence of a parameter, such as  $c_{1a}$  in this case, that may in fact be identifiable but does not appear in the observable representation so is not included in the resulting analysis and is assumed to be unidentifiable.

For the differential algebra based methods, different results can be obtained for the identifiability analysis depending on how the observable representation is found.

This is one disadvantage of this approach which can be demonstrated with the reduced NF- $\kappa$ B example. In Section 3.3 the existing differential algebra methods are applied but with the observable representation determined heuristically. The analysis is also carried out for the case in which the observable representation is determined algorithmically as part of the characteristic set obtained using the Rosenfeld-Groebner implementation in Maple (as described in Section 2.5.1). The results for each case are summarized in Table 5.3.

Table 5.3: Comparison of identifiable parameters for the reduced NF- $\kappa$ B model using differential algebra based methods with different observable representations

Method	Identifiable Parameters
Heuristic observable representation	$a_3, t_2, i_1, p_4$
Algorithmic observable representation	$a_3, t_2, i_1, p_4, p_3$

This does not prove to be a problem when the empirical identifiability Gramian is used because the system itself is analyzed so all parameters are considered. This proves to be an advantage of this new method over the existing observability based methods considered. Another advantage of the empirical observability Gramian based method over existing ones is its ability to be used for larger and more complex systems such as the full NF- $\kappa$ B model. This is shown in the next section.

### 5.3 Application of Empirical Identifiability Gramian to Full NF- $\kappa$ B Example

The empirical identifiability Gramian method is applied to the full NF- $\kappa$ B model outlined in Section 2.6. Recall that the measured outputs for the full model are:  $y = (z_2, z_6, z_7, z_{13})^T$ . For the analysis of the reduced model, it is assumed that several

Table 5.4: Results for use of empirical identifiability Gramian method on full NF- $\kappa$ B model with all 29 parameters as unknown

Iteration	Removed Parameter	Minimum Eigenvalue
0	-	$2.372 \times 10^{-7}$
1	$a_2$	$3.056 \times 10^{-7}$
2	$c_1$	$1.232 \times 10^{-6}$
3	$c_{6a}$	$1.405 \times 10^{-6}$
4	$c_{3a}$	$1.471 \times 10^{-6}$
5	$c_{3c}$	$2.475 \times 10^{-6}$
6	$k_3$	$2.750 \times 10^{-6}$
7	$c_2$	$6.304 \times 10^{-6}$
8	$t_1$	$6.818 \times 10^{-6}$
9	$e_{2a}$	$1.076 \times 10^{-5}$
10	$c_{1c}$	$1.861 \times 10^{-5}$
11	$c_{2a}$	$3.328 \times 10^{-5}$
12	$e_{1a}$	$3.289 \times 10^{-5}$
13	$c_{5a}$	$4.020 \times 10^{-5}$
14	$c_3$	$4.897 \times 10^{-5}$
15	$c_4$	$1.035 \times 10^{-4}$
16	$c_5$	$1.212 \times 10^{-4}$
17	$k_{deg}$	$1.243 \times 10^{-4}$
18	$k_2$	$1.925 \times 10^{-4}$
19	$c_{2c}$	$7.008 \times 10^{-4}$
20	$i_{1a}$	$2.056 \times 10^{-3}$
21	$c_{4a}$	$2.521 \times 10^{-3}$
22	$a_1$	$3.459 \times 10^{-3}$
23	$t_2$	$7.090 \times 10^{-3}$

of the parameter values are known to facilitate the identifiability analysis using the existing methods. This simplification is not required when using the empirical identifiability Gramian method so all parameter values are assumed to be unknown. The results of the application of the algorithm introduced in Section 4.4 are summarized in Table 5.4. The algorithm is relatively simple to apply in Matlab and the computation takes approximately 20 minutes for an example of this size. Using this method the following parameters are found to be identifiable:  $\theta_{identifiable} = (k_{prod}, k_1, i_1, c_{1a}, a_3, k_v)$ .

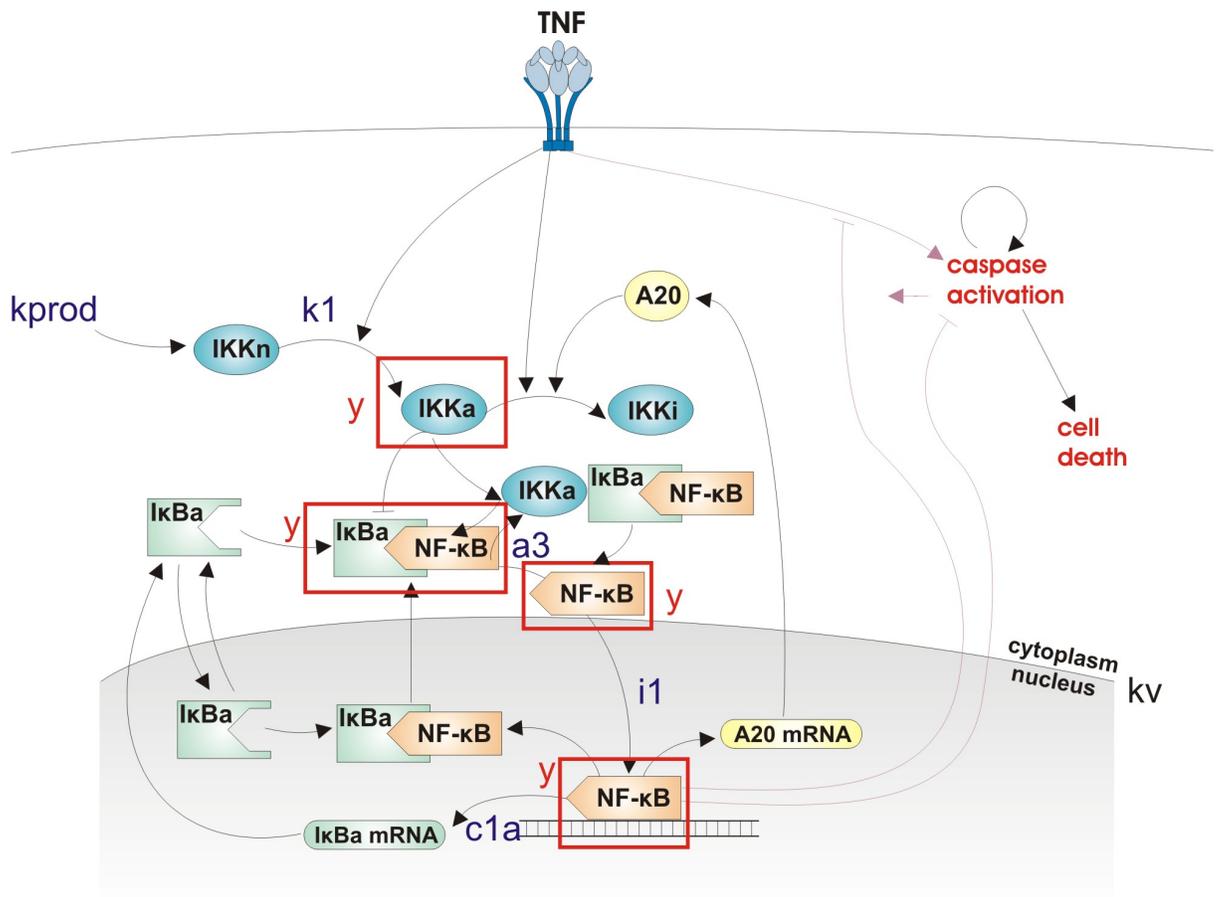


Figure 5.1: Schematic pathway representation of the full  $NF-\kappa B$  signal transduction network model with outputs and identifiable parameters shown.

The physical meaning of these parameters and their relationship to the outputs is shown on a schematic diagram of the full  $NF-\kappa B$  model in Figure 5.1.

In general, the parameters that are identifiable are directly or closely related to the measurable outputs. The parameter  $k_v$  refers to the ratio of the cytoplasmic to nuclear volume so its relationship to the outputs is not as clear. In this case, only 6/29 parameters are found to be identifiable with the choice of outputs considered. This means that a parameter estimation carried out using only these measurements would likely have a large degree of error and possibly would not even converge unless

those unidentifiable parameters were fixed or known from other sources. It may be necessary to reduce the model further to provide an accurate description of the observed dynamics.

It should be noted that the number of identifiable parameters depends greatly on the threshold  $\epsilon$  used to determine which eigenvalues are large enough to correspond to identifiable parameters. This value is estimated to quantify the error of approximation of the empirical Gramian. If one reduces the order of magnitude of  $\epsilon$  to  $10^{-4}$  there could be between four and nine more identifiable parameters. This differs from the analytical observability based methods previously discussed which provide a clear and definite answer to the identifiability question due to the algebraic nature of the solution. This means that the parameters corresponding to eigenvalues near the threshold value should be analyzed further before making a definite decision about whether or not they are identifiable. However, the uncertainty presented by the numerical approximation of the empirical identifiability Gramian method is a necessary tradeoff to analyze systems that are larger and more complex. The Gramian method also has the added advantage of being able to provide information about the degree of identifiability of each parameters which could be useful for parameter estimation.

## 5.4 Summary

In this chapter, the empirical identifiability Gramian method is applied to the example that motivated its development: the NF- $\kappa$ B signal transduction pathway. The method is first applied to the reduced NF- $\kappa$ B model to compare it with the results of the identifiability analysis using existing observability based methods in Section

3.3. Both approaches determine that the parameters  $a_3, k_v, t_2$  and  $i_1$  are identifiable. The results of the Gramian method, however, state that  $c_{1a}$  is identifiable as well. This could be due to the fact that the differential algebra methods use a state free system representation in which certain, potentially identifiable, parameters could be lost. The identifiability Gramian considers the nonlinear system itself so this does not pose a problem. The empirical identifiability Gramian method is also applied to the full NF- $\kappa$ B example to demonstrate its ability to be used on a system of the size and complexity commonly encountered in systems biology. Some uncertainty in the solution exists due to the numerical approximation nature of the method. However, this is a necessary tradeoff to apply the method to larger systems where existing observability based methods cannot be used effectively.

# Chapter 6

## Summary and Conclusions

The problem of parameter identifiability is of great interest in systems biology. It is an important prerequisite to the estimation of unknown parameters from measured data. Accurate knowledge of these parameter values is needed to describe and analyze the behaviour of biological systems. Being able to determine, before analyzing the data, if the unknown model parameters can be determined (uniquely) from given outputs serves several purposes. It is required in numerical search or optimization procedures for parameter estimation to guarantee the existence of a mathematically significant solution. It is also a valuable tool for the design of experiments and can be used to provide information about the suitability of the model structure itself.

In this thesis, a new method for analyzing parameter identifiability of biochemical reaction networks is proposed. The motivation for this method is provided by applying existing observability based methods for structural or *a priori* identifiability to a representative systems biology problem, the NF- $\kappa$ B signal transduction pathway, to determine their suitability for those purposes. Such methods were originally developed for general nonlinear systems and, as a result, do not take into consideration

the specific challenges posed by biological systems. It is found that the analysis itself can only be carried out using existing methods after significant model simplifications are made. The number of states have to be simplified through model reduction and whether or not identifiability can be analyzed depends largely on the choice of outputs. The difficulty with the current observability based methods is that they require analytical solutions that imposes restrictions on the size and complexity of systems that they can handle. The full NF- $\kappa$ B model considered here has 15 states and 29 parameters and is too large to be considered using existing methods. Many currently studied systems biology models are even larger networks (the JAK-STAT pathway in (Yamada et al., 2003) for example has 31 states and 52 parameters) so identifiability methods that can handle larger systems must be developed.

For this reason, an empirical observability Gramian for identifiability analysis is proposed. Since this method is data/simulation based, it can avoid computational complexities associated with the previous methods and be used on biological systems. Empirical Gramians have been computed for nonlinear systems with dozens of states for the purpose of model reduction (see Lall et al., 2002) but no previous results exist for their application in identifiability analysis. In this thesis, the empirical identifiability Gramian method is introduced with special attention being paid to biochemical reaction network systems. Since empirical Gramians were originally introduced for the purpose of model reduction, previous results do not consider the analysis of the Gramian to extract the desired information relating to observability (and in turn, for identifiability). A method for analyzing the Gramian to determine identifiable and non-identifiable parameters is developed here. The method is based on the eigenvalue/eigenvector decomposition and considers the numerical sensitivity

of the calculations. An algorithm is proposed for the computation and analysis of the empirical identifiability Gramian. Its applicability is demonstrated using a simple biological example of microbial growth with Michaelis-Menten kinetics. The new method is successfully applied to the full NF- $\kappa$ B example. This shows the applicability of the empirical identifiability Gramian method for systems of the size and complexity generally encountered in systems biology.

Future research work will focus on the application of this method to a wider range of biological systems. In this work, only time invariant parameters are considered. Since for identifiability purposes the parameters are included as states in the Gramian by considering their time derivatives, this method could easily be extended to the case of time variant parameters. The current empirical observability Gramian definition requires the system in question to be stable. A time varying empirical Gramian could be introduced to consider unstable systems, such as those with oscillations or limit cycles, commonly encountered in systems biology,

# References

- Audoly, S., Bellu, G., D'Angio, L., Saccomani, M. and Cobelli, C. *Global identifiability of nonlinear models of biological systems*. IEEE Transactions on Biomedical Engineering, 2001. **48**(1):55–65
- Chappell, M., Godfrey, K. and Vajda, S. *Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods*. Mathematical Biosciences, 1990. **102**:41–73
- Chappell, M. J. and Godfrey, K. *Structural identifiability of the parameters of a nonlinear batch reactor model*. Mathematical Biosciences, 1992. **108**:241–251
- Denis-Vidal, L. and Joly-Blanchard, G. *An easy to check criterion for (un)identifiability of uncontrolled systems and its applications*. IEEE Transactions on Automatic Control, 2000. **45**(4):768–771
- Denis-Vidal, L., Joly-Blanchard, G. and Noiret, C. *Some effective approaches to check the identifiability of uncontrolled nonlinear systems*. Mathematics in Computers and Simulation, 2001. **57**:35–44
- Eissing, T., Conzelmann, H., Gilles, E., Allgöwer, F., Bullinger, E. and Scheurich, P. *Bistability analysis of a caspase activation model for receptor-induced apoptosis*. Journal of Biological Chemistry, 2004. **279**(35):36892–36897
- Fairman, F. *Linear Control Theory: The State Space Approach*. John Wiley and Sons: New York, 1998
- Gadkar, K., Gunawan, R. and Doyle III, F. *Iterative approaches to model identification of biological systems*. BMC Bioinformatics, 2005. **6**(155)
- Geffen, D., Findeisen, R., Schliemann, M., Allgöwer, F. and Guay, M. *The question of parameter identifiability for biochemical reaction networks considering the NF- $\kappa$ B signal transduction pathway*. In *Proc. 2nd Foundations of Systems Biology in Engineering (FOSBE 2007)*. Stuttgart, Germany, 2007 pp. 509–514

- Godfrey, K. and Fitch, W. *The deterministic identifiability of nonlinear pharmacokinetic models*. Journal of Pharmacokinetics and Biopharmaceutics, 1984. **12**(2):177–191
- Hahn, J. and Edgar, T. *An improved method for nonlinear model reduction using balancing of empirical gramians*. Computers and Chemical Engineering, 2002. **26**:1379–1397
- Hengl, S., Kreutz, C., Timmer, J. and Maiwald, T. *Data-based identifiability analysis of nonlinear dynamical control models*. Bioinformatics, 2007
- Hoffmann, A., Levchenko, A., Scott, M. and Baltimore, D. *The I $\kappa$ B-NF- $\kappa$ B signaling module: Temporal control and selective gene activation*. Science, 2002. **298**:1241–1245
- Holmberg, A. *On the practical identifiability of microbial growth models incorporating michaelis-menton type nonlinearities*. Mathematical Biosciences, 1982. **62**:23–43
- Jacquez, J. and Greif, P. *Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design*. Mathematical Bioscience, 1985. **77**:201–227
- Kitano, H. *Systems biology: A brief overview*. Science, 2002. **295**:1662–1664
- Klipp, E., Herwig, R., Kowald, A., Wierling, C. and Lehrach, H. *Systems Biology in Practice*. Wiley-VCH: Weinheim, 2005
- Konstantinides, K. *Statistical analysis of effective singular values in matrix rank determination*. IEEE Transactions on Acoustics, Speech and Signal Processing, 1988. **36**(5):757–763
- Lall, S., Marsden, J. E. and Glavaski, S. *A subspace approach to balanced truncation for model reduction of nonlinear control systems*. International Journal of Robust and Nonlinear Control, 2002. **12**:519–535
- Lipniacki, T., Paszek, P., Brasier, A., Luxon, B. and Kimmel, M. *Mathematical model of NF- $\kappa$ B regulatory module*. Journal of Theoretical Biology, 2004. **228**:195–215
- Ljung, L. *Systems Identification- Theory for the User*. Prentice Hall: New Jersey, 2nd ed., 1999
- Ljung, L. and Glad, T. *On global identifiability for arbitrary model parameterizations*. Automatica, 1994. **30**(2):265–276

- Moles, C., Mendes, P. and Banga, J. *Parameter estimation in biochemical pathways: A comparison of global optimization methods*. Genome Research, 2003. **13**:2467–2474
- Montgomery, D. *Design and Analysis of Experiments*. Wiley: New York, 1997
- Muller, T., Noykova, N., Gyllenberg, M. and Timmer, J. *Parameter identification in dynamical models of anaerobic waste water treatment*. Mathematical Biosciences, 2002. **177&178**:147–160
- Pohjanpalo, H. *System identifiability based on the power series expansion of the solution*. Mathematical Biosciences, 1978. **41**:21–33
- Quaiser, T., Marquardt, W. and Mönnigmann, M. *Local identifiability analysis of large signalling pathway models*. In *Proc. 2nd Foundations of Systems Biology in Engineering (FOSBE 2007)*. Stuttgart, Germany, 2007 pp. 465–470
- Ritt, J. *Differential Algebra*. American Mathematical Society: Providence, R.I., 2005
- Schliemann, M., Eissing, T., Scheurich, P. and Bullinger, E. *Mathematical modelling of TNF- $\alpha$  induced apoptotic and antiapoptotic signaling pathways in mammalian cells based on dynamic and quantitative experiments*. In *Proc. 2nd Foundations of Systems Biology in Engineering (FOSBE 2007)*. Stuttgart, Germany, 2007 pp. 213–218
- Singh, A. K. and Hahn, J. *On the use of empirical gramians for controllability and observability analysis*. In *Proc. 2005 American Control Conference (ACC)*. Portland, OR, USA, 2005 pp. 140–141
- Stewart, G. W. *Rank degeneracy*. Society for Industrial and Applied Mathematics, 1984. **5**(2):403–413
- Tunali, E. and Tarn, T. *New results for identifiability of nonlinear systems*. IEEE Transactions on Automatic Control, 1987. **32**(2):146–154
- Vayttaden, S., Ajay, S. and Bhalla, U. *A spectrum of models of signalling pathways*. ChemBioChem, 2004. **5**(10):1365–1374
- Xia, X. and Moog, C. *Identifiability of nonlinear systems with application to HIV/AIDS model*. IEEE Transactions on Automatic Control, 2003. **48**(2):330–336
- Yamada, S., Shiono, S., Joo, A. and Yoshimura, A. *Control mechanism of JAK-STAT signal transduction pathway*. FEBS Lett, 2003. **534**:190–196

# Appendix A

## Description of NF- $\kappa$ B Model Parameters

Table A.1: Summary of NF- $\kappa$ B signal transduction pathway parameters

Symbol	Description
$k_{prod}$	rate constant for de novo production of neutral IKK
$k_{deg}$	rate constant degradation of IKK $\alpha$ , IKK $\beta$ , IKK $\gamma$
$k_1$	rate constant activation of IKK caused by TNF
$k_2$	rate constant A20 induced inactivation of IKK
$k_3$	rate constant spontaneous activation of IKK
$a_1$	rate constant formation of (IkBa—NF- $\kappa$ B) complexes in nucleus and cytoplasm
$a_2$	rate constant formation of (IKK $\alpha$ —IkBa) complexes
$a_3$	rate constant formation of (IKK $\alpha$ —IkBa—NF- $\kappa$ B) complexes
$t_1$	rate constant catalytic degradation of (IKK $\alpha$ —IkBa) complexes
$t_2$	rate constant catalytic degradation of (IKK $\alpha$ —IkBa—NF- $\kappa$ B) complexes
$c_1$	rate constant NF- $\kappa$ B inducible A20 synthesis
$c_2$	rate constant constitutive A20 mRNA synthesis (A20 transcript)
$c_3$	rate constant degradation of A20 transcript
$c_4$	rate constant mRNA synthesis/translation rate of A20
$c_5$	rate constant constitutive degradation of A20
$c_{1a}$	rate constant NF- $\kappa$ B inducible IkBa mRNA synthesis
$c_{2a}$	rate constant constitutive IkBa mRNA synthesis
$c_{3a}$	rate constant degradation of IkBa transcript
$c_{4a}$	rate constant IkBa translation rate (formation of IkBa protein)
$c_{5a}$	rate constant constitutive degradation of IkBa
$c_{6a}$	rate constant dissociation of (IkBa—NF- $\kappa$ B) complex
$i_1$	transport coefficient of free NF- $\kappa$ B into nucleus
$i_{1a}$	transport coefficient of IkBa into nucleus
$e_{1a}$	transport coefficient of IkBa out of nucleus
$e_{2a}$	transport coefficient for movement of (IkBa—NF- $\kappa$ B) from nucleus
$c_{1c}$	rate constant for NF- $\kappa$ B inducible cgen synthesis (inducible transcription)
$c_{2c}$	rate constant for constitutive cgen mRNA synthesis (constitutive transcription)
$c_{3c}$	rate constant for degradation of cgen transcript (mRNA degradation)
$k_v$	ratio of cytoplasmic to nuclear volume

## Appendix B

### Nominal NF- $\kappa$ B Parameters Values for Gramian Calculation

Table B.1: Model parameters for NF- $\kappa$ B signal transduction pathway in nominal case

Model Parameters	Nominal Value
$k_{prod}$	$2.5 \times 10^{-5}(\mu M/s)$
$k_{deg}$	$1.25 \times 10^{-4}(1/s)$
$k_1$	$2.5 \times 10^{-3}(1/s)$
$k_2$	$0.1(1/s)$
$k_3$	$1.5 \times 10^{-3}$
$a_1$	$0.5(1/uM/s)$
$a_2$	$0.2(1/uM/s)$
$a_3$	$1(1/uM/s)$
$t_1$	$0.1(1/s)$
$t_2$	$0.1(1/s)$
$c_1$	$5 \times 10^{-7}(1/s)$
$c_2$	$0(uM/s)$
$c_3$	$4 \times 10^{-4}(1/s)$
$c_4$	$0.5(1/s)$
$c_5$	$3 \times 10^{-4}(1/s)$
$c_{1a}$	$5 \times 10^{-7}(1/s)$
$c_{2a}$	$0(1/s)$
$c_{3a}$	$4 \times 10^{-4}(1/s)$
$c_{4a}$	$0.5(1/s)$
$c_{5a}$	$1 \times 10^{-4}(1/s)$
$c_{6a}$	$2 \times 10^{-5}(1/s)$
$i_1$	$2.5 \times 10^{-3}(1/s)$
$i_{1a}$	$1 \times 10^{-3}(1/s)$
$e_{1a}$	$5 \times 10^{-4}(1/s)$
$e_{2a}$	$1 \times 10^{-2}(1/s)$
$c_{1c}$	$5 \times 10^{-7}$
$c_{2c}$	$0$
$c_{3c}$	$4 \times 10^{-4}$
$k_v$	$5$