STRUCTURAL ANALYSIS OF KINETIC SYSTEMS WITH APPLICATION TO CELL-FREE EXPRESSION SYSTEMS

Zoltán András Tuza

Theses of the Ph.D. Dissertation

Pázmány Péter Catholic University
Faculty of Information Technology

Thesis Advisor:
Dr. Gábor Szederkényi, D.Sc.

Budapest, 2015

1. Introduction

Dynamical models play an important role in many fields of science and engineering. The aim of applying these models is to solve real-life problems, using that they are able to reproduce the important observed phenomena as accurately as required. Nonnegative systems form a special class of dynamical systems where all the state variables remain in the nonnegative orthant if the coordinates of the initial condition are nonnegative [11]. Kinetic dynamical models originate from chemistry as descriptors of chemical processes, but their range of applicability reaches far beyond (bio)chemical models as they are suitable to describe all important dynamical phenomena [7].

It is possible to associate a directed graph structure to a kinetic system which enables us to investigate not only the graph theoretic properties, but the dynamical properties of the kinetic system as well. It has been known that the associated sparse directed graph structure is not necessarily unique that is in contrast with the unique dense structure of a kinetic system. This non-uniqueness may hamper the successful identification of a kinetic system—especially a (bio)chemical system—because a unique sparse structure is often implicitly assumed. On the other hand dynamical equivalence enables us to compute directed graph structures with prescribed properties such as maximum or minimum number of edges in the graph, deficiency, weak reversibility or complex balance.

The theory to investigate these properties has been developed for decades [10] but a set of optimization based approaches—exploiting that the graph structure corresponding to a given kinetic dynamics is non-unique—have been developed relatively recently [16]. Building upon these, this thesis presents two new algorithms utilizing mathematical optimization.

Dynamical modeling in the field of systems biology and synthetic biology are good examples where nonnegative, especially, kinetic systems are typically applied [17]. In systems biology the aim is to understand and eventually control biomolecular processes such as signal transduction or metabolism. Thus, dynamical models can support this process by accurately describing the observed phenomena and the inherent properties of the biological system [1]. On the other hand, in synthetic biology rational designing and creating novel interaction networks, e.g. gene regulatory networks is the main aim [14]. If these interaction networks or so called biocircuits are successfully designed and tested they may be capable to sense external or internal signals, compute the necessary res-

ponse and actuate the molecular system accordingly. Meanwhile, all of these steps are based on molecular computation [13].

Therefore, in both fields, but especially in synthetic biology dynamical models are becoming essential tools to carefully investigate and understand the biological processes, to predict the possible dynamical properties and to support the rational design process with appropriate feedback. Moreover, recent advances in measurement technology provides us with a rich source of data for revealing the structure and behavior of biochemical processes. Using real time measurements of both transcriptional and translational stages of gene expression provide us the necessary insight to build a kinetic model for a cell-free system.

2. Applied Tools and Methods

2.1. Mathematical Optimization

Linear Programming Linear programing (LP) is a constrained convex optimization technique, where a linear objective function of the real-valued optimization variables is minimized (or maximized) with respect to linear equality and inequality constraints.

The linear programming framework is very versatile, thus the practical applications are ranging from engineering to social sciences, e.g. production optimization, transportation and assignment problems, etc [2]. Moreover, many efficient solvers are available to solve linear programming problems even with millions of decision variables and hundreds of thousands constraints. These solvers are based on the simplex method or lately the interior point method, reviewed in e.g. [5].

Mixed Integer Linear Programming Some problems require decision variables with integer values. This constraint makes the optimization problem NP-hard, although thanks to efficient solvers many practical problems can be solved [6]. Linear programs with integrality constraints arise in many fields, e.g. in transportation, scheduling, etc.

A connection between linear integer programming and propositional calculus can be made. A propositional logic problem can be solved by means of a linear program with integrality constraints, by translating the original compound statements into linear inequalities involving logical variables [19].

2.2. Kinetic Systems

A widely used class within nonnegative systems is the class of kinetic systems, where the exponent of the monomials are nonnegative integers and there are additional relations between the monomial coefficients and exponents.

Kinetic dynamical models are suitable to describe all important dynamical phenomena such as stability/instability and multiplicity of equilibria, bifurcation, oscillatory and even chaotic behavior. Many of these phenomena have actually been observed in real chemical experiments where the practical constraints are much more severe than in the case of pure mathematical models [7].

Furthermore, kinetic models can effectively be used in the description of numerous natural processes such as disease dynamics, population dynamics, compartmental models, or certain transportation phenomena. On the top of that, kinetic systems can be used to describe pure chemical reactions or the complex dynamics of intracellular processes, metabolic or cell signaling pathways. Kinetic models have also been useful in performing complex non-conventional computation tasks. Moreover, their simple algebraic structure make these models attractive both for rigorous mathematical analysis and for efficient computational techniques [8, 12], as well as certain strong statements of the structural and dynamical properties of the model can be made about kinetic systems using Chemical Reaction Network Theory, even without knowing the parameters of the kinetic model [9].

Directed Graph Structure We can associate a graph representation to kinetic models. A kinetic system equipped with this graph structure is called a Chemical Reaction Network (CRN) as it is described in e.g. [10]. The vertices in the graph represent the complexes of the reaction network. Whereas the directed edges representing the reactions between the complexes and the corresponding reaction rate coefficient is assigned as weight to each edges.

Linear programming based optimization techniques exist to calculate certain graph structures. Some of these structural properties are directly connected to the dynamical behavior of the kinetic system. Therefore, based on the structure of the graph some dynamical properties, e.g. stability can be determined.

Dynamical Equivalence It has been known since at least the 1970's that multiple different structures (parametrizations) of a CRN can generate exactly the same dynamics of the concentrations [7, 12]. This phenomenon is called *macro-equivalence* or *dynamical equivalence*. However, the exact geometric conditions of macro-equivalence were not studied until relatively recently in [4].

From an optimization point of view the dynamical equivalence defines a polyhedron in the space of reaction rate coefficients where all dynamically equivalent realizations exist. Therefore, we can define such a linear programming problem where the constraint set contains the definition of dynamical equivalence, then we can search for realizations with required properties that can be translated into linear constraints.

2.3. Parameter Estimation

The parameters of a kinetic system is often needed to be determined from measurement data, but the process of parameter estimation is often challenging. Generally, these challenges can be classified into two main categories. First, the selected process model may have structural identifiability issues, namely the model structure is capable to produce exactly the same output for different sets of parameters [18]. The second challenge stems from the poor excitation of the dynamics or from the poor quality of the available measurements, which is often labeled as practical identifiability problem.

Before the parameter estimation a type of model has been selected, typically a nonlinear state space model where we can measure a projection (usually a subset) of the state variables. In this model the unknown parameter vector may not only include the dynamical parameters but also the unknown initial conditions.

Due to this projection, often not all state variables can be measured directly and observation function greatly influences which parameters are possible to compute. Generally, the output of the systems can be measured at certain frequency, thus we have the output of the system at discrete time instance.

The goal of the parameter estimation is to find a suitable vector of parameters that generates the minimal distance between the model output and the measurements. Multiple metrics exist to measure the distance between the model output and the measurements, for example quadratic or absolute distance can be used [15]. Once we have a way to measure the distance between the model output and the measurements at

each time instance, we can define a function that assigns a nonnegative real number to any possible parameter vector, this is called the cost function.

Structural Identifiability Once we have selected a model structure, the question of parameter identifiability has to be considered, i.e. whether it is theoretically possible to determine the model parameters based on the model structure and the observables [3].

Structural identifiability depends only on the structure of the model including the output functions. Because of that structural identifiability analysis can be carried out before collecting the data—if the model structure is known. Unfortunately, this analysis is often neglected and still not a standard practice of modelers. On the other hand, several approaches and software tools exist for structural identifiability analysis and they are suitable for many different model structures [18].

Parameter Estimation of Kinetic Systems Due to the structure of the kinetic models, if we assume that all the state variables are directly measured or the parameter of the output function are known, then the model is linear in the monomial coefficients. Hence many standard parameter estimation techniques can be used [18].

In case of the Cell-free system, most of the experiments are initial condition experiments, i.e. the conditions can be changed only at the on-set of an experiment. Therefore, a typical set of data contains several different initial scenarios with the same system, e.g. change of the initial concentration of one or more species. This set of data is considered all at once in the cost function in order to have enough information about the parameters of the system. These parameters include the rate coefficients and the initial conditions as well.

3. New scientific results

Thesis I. *I have developed a mathematical optimization based efficient algorithm to compute all dynamically equivalent sparse realizations of a kinetic system.*

Using combinatorial and mathematical optimization techniques, I have developed the first algorithm in the literature to compute all sparse realizations of dynamically equivalent kinetic systems. This algorithm uses mixed integer linear programming (MILP) and linear programming (LP) steps to compute all the sparse realizations.

Corresponding publications: [J2], [C2].

Thesis I.a *I have proposed an effective reduction of the combinatorially possible search space by using appropriate constraint-pairs and the properties of constrained sparse realizations.*

The special properties of dense and sparse realizations made it possible to reduce the original search space which consists of all directed graphs with fixed set of nodes to a computationally tractable number of candidate structures.

Thesis I.b *By applying state-dependent time-rescaling and X-factorable transformation, I have computed all sparse CRN structures for two different kinetic models of the well-known Lorenz system showing chaotic behavior. I have compared the obtained realizations from a structural point of view and determined the minimum and maximum number of linkage classes and deficiencies corresponding to the sparse realizations. I have shown that the complete search space was reduced to 0.01 and 0.0037 percents in the case of state-dependent time-rescaling and X-factorable transformation, respectively.*

The Lorenz system was a good candidate to show the application of the developed algorithm. Two distinct approaches yielded different number of sparse structures in each case, but with similar structural properties. It was computationally checked that, the chaotic behavior of the system was preserved in all representations.

Thesis II. *I have developed structural analysis tools for kinetic systems with parametric uncertainty.*

I have developed optimization based tools for the structural analysis of uncertain kinetic systems. The uncertainty in these systems are represented as a multi-dimensional interval in the space of monomial coefficients.

Corresponding publications: [C5], [C1].

Thesis II.a *I have proposed a new algorithm for the computation of dense and sparse reaction network structures for kinetic polynomial models, where the uncertainties are represented as parameter intervals. The problem is traced back to mixed integer linear programming where the parameter uncertainties are given by linear inequalities.*

The current computational framework has been extended to accommodate parametric uncertainties while certain structural properties of the kinetic system—as well as the associated directed graph structure—can be effectively calculated. This approach opens up the possibility to extend several previous optimization based results to uncertain kinetic systems such as weak reversibility or complex balance.

Thesis II.b *I have developed an algorithm with polynomial time complexity to calculate the structurally invariant elements, called core reactions that are present in any reaction network belonging to the model set defined by the uncertainties. The proposed algorithm is based on linear programming and incorporates the parametric uncertainty of the system as element-wise boundary constraints.*

The core reactions are one of the most important elements of the reaction graph, since if the set of core reactions is non-empty, its elements are present in each dynamically equivalent realization. Thus, the computation of distinct dynamically equivalent reaction network structures satisfying a given property can effectively be supported by utilizing the core reaction set.

Thesis III. *I have built a kinetic model for an in vitro cell-free gene expression system.*

I have built a first principle kinetic model for an in vitro cell-free gene expression system using a specific experimental setup. I have shown that with the estimated and validated parameters the kinetic model effectively captures the dynamical features of the cell-free system.

Corresponding publications: [C4], [J1], [C3].

Thesis III.a *I have built combinations of molecular probes containing RNA aptamer and fluorescent proteins for the investigation of the cell-free system. I have designed and carried out a comprehensive study of the cell-free system utilizing concurrent measurement of transcription and translation.*

Concurrent measurement of transcription and translation was important for the development of a reliable kinetic model. The library of molecular probes made possible to study the dynamical features of the cell-free system in detail. As a result of that, this measurement set up serves as a benchmark for testing different versions of the cell-free system.

Thesis III.b *Based on the observations from the experiments, I have built a kinetic model for the studied cell-free system, which is capable of capturing the transient behavior of the system taking the finite resources into consideration. I have shown that the model is structurally identifiable using the applied measurement setup. Finally, I have determined and validated the parameters of the dynamical models using constrained least squares based parameter estimation.*

Utilizing the library of molecular probes and domain knowledge about the process of gene expression, the resulting dynamical model is structurally identifiable. Also, the fact that the developed model structure is linear in parameters and the quality of the data enabled us to use the least squares based parameter estimation.

4. Application of the results

Given the broad topic of this thesis, several fields of application are possible. We investigated the non-uniqueness of the dynamically equivalent sparse structure of a kinetic systems. The result illustrates that the sparse structure of kinetic systems may be non-unique. Moreover, the Lorenz example has a large number of sparse realizations. A possible direction of future work might be the graph theoretical investigation of these sparse structures. That could include the average incoming and outgoing connection of a complex in each realizations or the reaction distribution considering a complex in all the realizations.

As another direction, we could extend the proposed algorithm for calculation of all sparse realizations to accommodate uncertain kinetic systems. This would offer an algorithmic way to calculate all dynamically equivalent sparse realizations at a given level of uncertainty. This algorithm would be a powerful tool for network reconstruction where usually only noisy time series data and the set of vertices of the directed graph is known.

We briefly investigated the application of the core-reactions in network reconstruction. As a result of that, we have concluded that the core reaction set can be at least partially restored from the time series data. On the other hand, we did not extend the concept of core complexes to uncertain kinetic systems and did not include it to the computation. Thus, the application of the core reactions and the core complexes in an algorithm where they are calculated in an alternating fashion could be the backbone of an improved method for network reconstruction within the class of kinetic systems.

The thesis elaborated on the challenges of kinetic system identification. Among many important open questions, the relevance of optimization based experiment design for a molecular breadboard was highlighted. The modeling of the *in vitro* system laid the foundation of that work by introducing a dynamical model for the cell-free system on which the molecular breadboard is built on. Utilizing this kinetic model, an optimization based framework could be developed where the optimization task would consider all the physical limitations of the molecular breadboard.

Publications Related to The Thesis

[J1] D. Siegal-Gaskins, Z. A. **Tuza**, J. Kim, V. Noireaux, and R. M. Murray. "Gene Circuit Performance Characterization and Resource Usage in a Cell-Free "Breadboard"". In: *ACS Synth Biol Impact Factor: 3.951* 3 (2014), pp. 416–25. DOI: 10.1021/sb400203p.

[C1] G. Szederkényi, Z. A. **Tuza**, and K. M. Hangos. "Determining biochemical reaction network structures for kinetic polynomial models with uncertain coefficients". In: *AIP Conference Proceedings-American Institute of Physics* 1479 (2012), pp. 2427–2430. DOI: 10.1063/1.4756685.

[C2] G. Szederkényi, Z. A. **Tuza**, and K. M. Hangos. "Dynamical Equivalence and Linear Conjugacy of Biochemical Reaction Network Models". In: *Biological and Medical Systems 2012*. Vol. 8. 2012, pp. 125–130. DOI: 10.3182/20120829-3-HU-2029.00042.

[C3] Z. A. **Tuza**, D. Siegal-Gaskins, J. Kim, and G. Szederkényi. "Analysis-Based Parameter Estimation of an in Vitro Transcription-Translation System". In: *European Control Conference 2015 (ECC2015)*. 2015.

[C4] Z. A. **Tuza**, V. Singhal, J. Kim, and R. M. Murray. "In Silico Modeling Toolbox for Rapid Prototyping of Circuits in a Biomolecular "Breadboard" System". In: *Decision and Control (CDC), 2013 51th IEEE Conference on*. 2013, pp. 1404–1410. DOI: 10.1109/CDC.2013.6760079.

[C5] Z. A. **Tuza** and G. Szederkényi. "Computing Core-Reactions of Uncertain Polynomial Kinetic Systems". In: *23rd Mediterranean Conference on Control and Automation*. IEEE, 2015, pp. 1140–1147.

[J2] Z. A. **Tuza**, G. Szederkényi, K. M. Hangos, A. A. Alonso, and J. R. Banga. "Computing All Sparse kinetic Structures for a Lorenz System using Optimization." In: *I. J. Bifurcation and Chaos Impact Factor: 1.017* 23.8 (2013). DOI: 10.1142/S0218127413501411.

Other Publications of The Author

[J3] B. Ács, G. Szederkényi, Z. A. **Tuza**, and Z. Tuza. "Computing Linearly Conjugate Weakly Reversible Kinetic Structures Using Optimization and Graph Theory". In: *MATCH Communications in Mathematical and in Computer Chemistry Impact Factor: 1.829* (2015).

[C6] J. Rudan, Z. **Tuza**, and G. Szederkényi. "Using LMS-100 laser rangefinder for indoor metric map building". In: *IEEE International Symposium on Industrial Electronics - ISIE 2010, Bari, Italy, 4-7 July*. 978-1-4244-6391-6/10. IEEE, 2010, pp. 525–530. DOI: 10.1109/ISIE.2010.5637842.

[C7] Z. **Tuza**, J. Rudan, and G. Szederkényi. "Developing an integrated software environment for mobile robot navigation and control". In: *2010 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2010; Zurich; 15 September 2010 through 17 September 2010*. ISBN: 978-142445864-6. 2010, no. 5647506 (on CD). DOI: 10.1109/IPIN.2010.5647506.

References

[1] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, CRC, 2007, p. 320.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] O. Chis, J. R. Banga, and E. Balsa-Canto. "Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods". In: *PLoS ONE* 27 (2011), pp. 2610–2611.

[4] G. Craciun and C. Pantea. "Identifiability of chemical reaction networks". In: *Journal of Mathematical Chemistry* 44 (2008), pp. 244–259.

[5] G. B. Dantzig and M. N. Thapa. *Linear programming 1: Introduction*. Springer-Verlag, 1997.

[6] G. B. Dantzig and M. N. Thapa. *Linear Programming 2: Theory and Extensions*. Springer-Verlag, 2003.

[7] P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models.* Manchester, Princeton: Manchester University Press, Princeton University Press, 1989.

[8] M. Feinberg. "Chemical reaction network structure and the stability of complex isothermal reactors - I. The deficiency zero and deficiency one theorems". In: *Chemical Engineering Science* 42 (10) (1987), pp. 2229–2268.

[9] M. Feinberg. "Chemical reaction network structure and the stability of complex isothermal reactors - II. Multiple steady states for networks of deficiency one". In: *Chemical Engineering Science* 43 (1988), pp. 1–25.

[10] M. Feinberg. *Lectures on chemical reaction networks.* Notes of lectures given at the Mathematics Research Center, University of Wisconsin, 1979.

[11] W. M. Haddad and V. Chellaboina. "Stability and dissipativity theory for nonnegative dynamical systems: A unified analysis framework for biological and physiological systems". In: *Nonlinear Analysis: Real World Applications* 6 (2005), pp. 35–65.

[12] F. Horn and R. Jackson. "General mass action kinetics". In: *Archive for Rational Mechanics and Analysis* 47 (1972), pp. 81–116.

[13] A. S. Khalil and J. J. Collins. "Synthetic biology: applications come of age." In: *Nat Rev Genet* 11.5 (May 2010), pp. 367–379.

[14] M. E. W. A. Lim. "Build life to understand it". In: *Nature* 468 (2010), pp. 889–890.

[15] L. Ljung. *System Identification (2Nd Ed.): Theory for the User.* Ed. by L. Ljung. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999. ISBN: 0-13-656695-2.

[16] G. Szederkényi, J. R. Banga, and A. A. Alonso. "Inference of complex biological networks: distinguishability issues and optimization-based solutions". In: *BMC Systems Biology* 5 (2011), p. 177. DOI: 10.1186/1752-0509-5-177.

[17] D. del Vecchio and R. M. Murray. *Biomolecular Feedback Systems.* 2012.

[18] É. Walter, L. Pronzato, and J. Norton. *Identification of parametric models from experimental data*. Communications and control engineering. Translated from an updated French version : Identification de modèles paramétriques à partir de données expérimentales. Berlin: Springer Paris, 1997. ISBN: 978-1-84996-996-3.

[19] H. P. Williams. *Model Building in Mathematical Programming*. Wiley, 2013, p. 432.