Model Invalidation and System Identification of Biochemical Reaction Networks

Steffen Borchers[1,2], Philipp Rumschinski[1], Sandro Bosio[3], Robert Weismantel[3], Rolf Findeisen[1]
[1] Department of Systems Theory and Automatic Control University of Magdeburg, Germany
[2] Max Planck Institute for Dynamic of Complex Technichal Systems, Magdeburg, Germany
[3] Department of Mathematical Optimization, University of Magdeburg, Germany

Corresponding author: R. Findeisen, Department of System Theory and Control, University of Magdeburg, Germany; rolf.findeisen@ovgu.de

Abstract. In systems biology, uncertainties in reaction mechanisms and system structure often result in competing hypotheses. Current approaches to discriminate between model alternatives are often inappropriate for biochemical reaction networks. This is because often only noisy measurements and sparse experimental data are available.

In this work a new method for model invalidation, based on the certification of *non-existence* of a feasible parametrization, is presented. This is achieved by reformulating the model invalidation task into a feasibility problem. As shown, due to the polynomial structure of many biochemical reaction systems, it is possible to relax the non-convex feasibility problem into a semidefinite program. This allows to obtain conclusive results on model invalidity as well as to estimate model parameters and states.

Our framework allows us to consider the arising difficulties posed by biochemical reaction networks by taking the polynomial structure of the dynamics and model outputs into account. An application example of the approach to a well-known biological reaction scheme is also presented, for which we show that it is possible to discriminate between the Michaelis-Menten and the Henri kinetics, and to estimate parameters.

1 Introduction

An important modeling issue in systems biology is system identification, comprising both estimation of model parameters and model structure analysis. System identification can be very difficult, due to uncertainties of the involved reaction mechanisms leading to concurring structural hypotheses. Moreover, kinetic parameters can not be determined experimentally and may therefore be completely unknown. To obtain a model that captures the essential behavior of the process under study, it is important to discriminate between the structural possibilities and to estimate model parameters.

Model validation consists in checking whether a model can represent a given observed experimental behavior. As it is impossible to *prove* model validity (see e.g. [8]), the purpose of model validation techniques is rather to invalidate a model by proving inconsistency of the model hypotheses with a given set of experimental data, or to give only inconclusive results, e.g. via simulation.

A model invalidation framework using barrier certificates which are functions of state-parameter-time has been proposed in [8]. These barriers separate possible model trajectories and measurement data, allowing to conclusively invalidate a model. Finding such barrier certificates however is a nontrivial task. Moreover, the existence of a suitable barrier function cannot be guaranteed for any arbitrary invalid models.

Our work is based on the work of [4], which introduced a semidefinite programming approach for parameter estimation based on stationary measurements. By using a different class of infeasibility certificates for polynomial reaction mechanisms, similar to the one proposed in [12] for sensitivity analysis, our approach allows us to consider a large class of nonlinear systems subjected to bounded, but possibly large, parameter uncertainties and measurement errors. Competing model alternatives can thus be discriminated by *proving* inconsistency with the available data for (some of) the wrong alternatives.

The approach presented here also allows for model parameter and state estimation. By invalidating respectively parameter and state space subsets, it is possible to classify parameter and state space regions into "feasible" regions, consistent with the measurements, and "infeasible" regions, for which the given measurements cannot be produced.

The paper is structured as follows. In Section 2 the considered system class is introduced. In Section 3 the invalidation setting and the basic results using feasible parameter sets are presented. In Section 4 we show how to estimate feasible parameters using semidefinite programming, and in Section 5 the approach is extended to consider feasible states estimation. Examples of model invalidation and parameter estimation are presented in Section 6, and Section 7 summarizes and discusses the approach.

2 Considered System Class

Biochemical reaction networks are commonly described by the occurring reactions in the form

$$\alpha_1 S_1 + \cdots + \alpha_{n_s} S_{n_s} \rightarrow \beta_1 T_1 + \cdots + \beta_{n_p} T_{n_t},$$

where a set of n_s substrates S_i is transformed into a set of n_t products T_j, with factors α_i, β_j defining the stoichiometric relations of the participating compounds.

If spatial and stochastic effects are neglected, such biochemical reaction networks can be expressed as ordinary differential equations. The dynamic models can then be defined considering the reaction fluxes and the corresponding balance equations as

$$\dot{x} = N\nu(x, p), \tag{1}$$

where $x \in \mathcal{X} \subset \mathbb{R}^{n_x}$ denotes the vector of concentrations, $p \in \mathcal{P} \subset \mathbb{R}^{n_p}$ the vector of kinetic parameters, and ν the vector of fluxes. The stoichiometric matrix N is built up from the factors α_i, β_j. We assume the given sets $\mathcal{X}$ and $\mathcal{P}$ to be bounded.

Describing this model with a suitable discretization method we obtain the following discrete-time model

$$\Sigma: \quad x[k+1] = g(x[k], p), \tag{2}$$

where $x[k] \in \mathcal{X}$ denotes the vector of state variables at the time index $k \in \mathbb{N}$. We assume that a measurement of the state obtained at time k, if taken without measurement errors, is given by the model output equation

$$y[k] = h(x[k], p), \tag{3}$$

with $y[k] \in \mathbb{R}^{n_y}$.

There are many possibilities to describe the reaction fluxes (see e.g. [2]), including the law of mass action, Hill kinetic, and Monod kinetic, the first approach being one of the most frequently used. Hereby, the reaction fluxes are proportional to the substrates, resulting in a system (1) with polynomial right-hand side. As our approach takes advantage on this special structure, we focus on mass action models.

In the remainder we consider $g(\cdot,\cdot)$ and $h(\cdot,\cdot)$ to be polynomial functions.

Remark 1 *The setting can be extended to rational functions, as well as to consider model inputs.*

Let an experiment be performed with the real process, and let a sequence of measurements

$$\mathcal{Y} = \{\mathcal{Y}_{k_1} \subset \mathbb{R}^{n_y},\ \ldots,\ \mathcal{Y}_{k_N} \subset \mathbb{R}^{n_y}\},$$

be taken at times $k_1, \ldots, k_N$. In general, the measurements are considered to be sets. This allows to address measurement errors by this concept, whereas in this case the measurement sets are intervals. We assume that each set $\mathcal{Y}_{k_i}$ contains the ideal measurement point $h(x[k_i], p)$, for $i \in \{1, \ldots, N\}$.

Based on this setup, the model invalidation problem can be formulated as follows:

Given a model Σ as in (2), a parameter set $\mathcal{P}$, and a measurement sequence $\mathcal{Y} = \{\mathcal{Y}_{k_1}, \ldots, Y_{k_N}\}$, find a *feasible parametrization* for Σ, i.e., a parameter $p \in \mathcal{P}$ for which Σ can produce a trajectory for which $x[k] \in \mathcal{X}$ for every $k \in \{1, \ldots, k_N\}$ and $y[k_i] \in \mathcal{Y}_{k_i}$ for every $i \in \{1, \ldots, N\}$, or a proof that none exists.

Our approach to prove model invalidity is based on a certificate of *non-existence* of feasible parametrization. Remarkably, this is achieved without direct simulating nor computing the solution of the model.

3 Model Invalidation and Parameter Estimation

The key idea is to reformulate the invalidation problem as a feasibility problem. While taking into account the measurements and parameter bounds as constraints, infeasibility of the feasibility problem can be certified for the considered system class as shown later on. Thus, model invalidity can be proved.

Two Measurements. Let us start considering the simple case of two measurements $\mathcal{Y}_{k_1}, \mathcal{Y}_{k_2}$ taken at distinct time indexes $k_1 < k_2$, and let $m = k_2 - k_1$. The *feasible parameter set* corresponding to $\mathcal{Y}_{k_1}, \mathcal{Y}_{k_2}$ is defined as

$$\mathcal{P}_{k_1,k_2} = \{p \in \mathcal{P} : \exists x_1, x_2 \in \mathcal{X} \mid x_2 = g^m(x_1, p), h(x_1, p) \in \mathcal{Y}_{k_1}, h(x_2, p) \in \mathcal{Y}_{k_2}\},$$

where $g^m(\cdot,\cdot)$ denotes the composition of g m-times. We can then state the following necessary condition for model validity:

Condition 1 *(Model Validity): Given a model* Σ *as in* (2)*, a parameter set* $\mathcal{P}$*, and two measurements* $\mathcal{Y}_{k_1}, \mathcal{Y}_{k_2}$*,* Σ *can be valid for the system under study only if* $\mathcal{P}_{k_1,k_2} \neq \emptyset$*.*

As a consequence, if $\mathcal{P}_{k_1,k_2} = \emptyset$ we can say that the model (2), with admissible parameter set $\mathcal{P}$, is invalidated by the measurements $\mathcal{Y}_{k_1}$,$\mathcal{Y}_{k_2}$. Thus, the condition $\mathcal{P}_{k_1,k_2} = \emptyset$ is sufficient for model invalidity.

The set $\mathcal{P}_{k_1,k_2}$ is the projection on $\mathbb{R}^{n_p}$ of the higher-dimensional set $\mathcal{F}(\mathcal{P}, \mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{(m+1)n_x+2n_y+n_p}$ defined by the feasibility problem

$$F(\mathcal{P}, \mathcal{X}, \mathcal{Y}) := \begin{cases} y[k_1] = h(x[k_1], p) & \\ y[k_2] = h(x[k_2], p) & \\ x[i] = g(x[i-1], p) & i \in M \setminus \{k_1\} \\ x[i] \in \mathcal{X} & i \in M \\ y[k_1] \in \mathcal{Y}_{k_1} & \\ y[k_2] \in \mathcal{Y}_{k_2} & \\ p \in \mathcal{P}, & \end{cases} \tag{4}$$

where $M = \{k_1, \ldots, k_1 + m = k_2\}$. An efficient approach to the solution of this feasibility problem for the case of polynomial systems is discussed in Section 4.

Multiple Measurements. The above approach can be directly extended to a multiple-measurement sequence $\mathcal{Y} = \{\mathcal{Y}_{k_1}, \ldots, Y_{k_N}\}$ by defining $m = k_N - k_1$ and including in (4) the constraints $y[k_i] = h(x[k_i], p)$ and $y[k_i] \in \mathcal{Y}_{k_i}$ for every measurement $i \in \{1, \ldots, N\}$. However, this approach quickly becomes computationally prohibitive as the number of measurements increases, since the number of variables of $F(\mathcal{P}, \mathcal{X}, \mathcal{Y})$ increases.

We therefore consider a less demanding though more conservative approach, for which the key idea is to consider pairwise consecutive measurements. If any of the parameter sets $\mathcal{P}_{k_i,k_{i+1}}$ is empty, then the model (2) is invalid. We additionally demand a consistent parametrization of the model (2), defined as

$$\mathcal{P}_{\mathcal{Y}} = \bigcap_{i=1}^{N-1} \mathcal{P}_{k_i,k_{i+1}}.$$

Note that $P_{\mathcal{Y}}$ provides an estimate of the model parameters. If $P_{\mathcal{Y}} = \emptyset$, then the model (2) and the admissible parameter set $\mathcal{P}$ are invalidated by the measurements $\mathcal{Y}$.

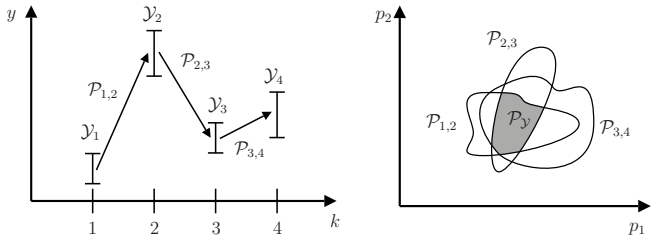

Figure 1: Description of the set-based invalidity approach. (a) Representation of the system flow, with measurements indicated by error bars. (b) Feasible parameter sets for consecutive measurements. The intersection of the feasible parameter sets is not empty, and hence the model is not invalidated by the measurements.

In the remainder, we show how to solve the feasibility problem given by (4) for the class of polynomial systems, and provide an algorithm to outer-approximate the feasible parameter set.

4 Bounding Feasible Parameter Sets by Semidefinite Programming

In this section we focus on how to estimate the feasible parameter set given by the feasibility problem $F(\mathcal{P}, \mathcal{X}, \mathcal{Y})$ (4). In general, this is a hard non-convex optimization problem, where non-convexity results from the nonlinearities of the model equations.

A method that allows to obtain some results on the infeasibility of polynomial optimization problems has been proposed in [4]. By taking advantage of the polynomiality, it is possible to *relax* $F(\mathcal{P}, \mathcal{X}, \mathcal{Y})$ to a convex semidefinite program (SDP). Such SDPs can then be solved efficiently, e.g. via interior point methods.

To do so, first the original feasibility problem is to be reformulated as a quadratic feasibility problem (QFP). For simplicity of notation, let us consider two measurements taken at consecutive time steps. We define the variable vector

$$\begin{array}{rll} \xi^T = & (1, & \\ & p_i, & i \in N_p, \\ & x[k]_j, x[k+1]_j, & j \in N_x, \\ & y[k]_\ell, y[k+1]_\ell, & \ell \in N_y, \\ & x[k]_j \cdot p_i, x[k+1]_j \cdot p_i, & i \in N_p, j \in N_x, \\ & y[k]_l \cdot p_i, y[k+1]_l \cdot p_i) & i \in N_p, l \in N_y, \end{array}$$

containing all the monomials appearing in (4).

Let S^n be the set of real symmetric $n \times n$ matrices, and let $\succeq$ denote the order operator with respect to the cone of positive semidefinite matrices in S^n. The equality constraints

$$\begin{aligned} x_i[k+1] - g_i(x[k], p) = 0 \quad i \in N_x \\ y_i[k] - h_i(x[k], p) = 0 \quad i \in N_y \\ y_i[k+1] - h_i(x[k+1], p) = 0 \quad i \in N_y \end{aligned}$$

can be respectively written as

$$\begin{aligned} \xi^T Q_i \xi = 0 \quad i \in N_x \\ \xi^T R_i^1 \xi = 0 \quad i \in N_y \\ \xi^T R_i^2 \xi = 0 \quad i \in N_y \end{aligned}$$

where $Q_i, R_i^1, R_i^2 \in S^{n_\xi}$ are symmetric matrices. This quadratic decomposition is possible for the considered time-discrete system (2) with polynomial structure. Thus, the approach depends on a discretization method that leads to polynomial $g(\cdot,\cdot)$ and $h(\cdot,\cdot)$.

Note also that some elements of the vector ξ may be dependent from one another. Such dependencies can also be expressed in the quadratic form

$$\xi^T D_i \xi = 0, \; i \in N_d,$$

where $N_d = \{1, \ldots, n_d\}$, n_d being the number of dependencies. The feasibility problem (4) can then be written as

$$QFP(\mathcal{P}, \mathcal{X}, \mathcal{Y}) := \begin{cases} \text{find } \xi \in \mathbb{R}^{n_\xi} \text{ s.t.} & \\ \xi^T Q_i \xi = 0 & i \in N_x \\ \xi^T R_i^1 \xi = 0 & i \in N_y \\ \xi^T R_i^2 \xi = 0 & i \in N_y \\ \xi^T D_i \xi = 0 & i \in N_d \\ B\xi \geq 0 & \\ \xi_1 = 1, & \end{cases}$$

where $B \in \mathbb{R}^{n_\xi \times 2(n_\xi - 1)}$ is constructed to define the bounding constraints for each component of ξ (but the first). Note that B defines the constraints $y[k] \in \mathcal{Y}_k$, $y[k+1] \in \mathcal{Y}_{k+1}$, $p \in \mathcal{P}$, and $x[k], x[k+1] \in \mathcal{X}$.

Problem $QFP(\mathcal{P}, \mathcal{X}, \mathcal{Y})$ can be subsequently relaxed into an SDP (see e.g. [7]) by setting $X = \xi \cdot \xi^T$ and replacing the conditions $rank(X) = 1$ and $tr(X) \geq 1$ with the weaker constraint $X \succeq 0$, resulting in the relaxed formulation

$$SDP(\mathcal{P}, \mathcal{X}, \mathcal{Y}) := \begin{cases} \text{find } \; X \in S^{n_\xi} \text{ s.t.} & \\ tr(Q_i X) = 0 & i \in N_x \\ tr(R_i^1 X) = 0 & i \in N_y \\ tr(R_i^2 X) = 0 & i \in N_y \\ tr(D_i X) = 0 & i \in N_d \\ BXe_1 \geq 0 & \\ tr(e_1 e_1^T) = 1 & \\ BXB^T \geq 0 & \\ X \succeq 0, & \end{cases}$$

where $e_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^{n_\xi}$. This relaxation is based on an image convexification method ([9]). As the relaxation process is conservative, all the solutions that were previously feasible remain feasible. However, "false" solutions may have been introduced. Although this does not lead to wrong invalidation

results, it may lead to considering an invalid model as valid. Constraints $BXB^T \geq 0$ are added to strengthen the relaxation and reduce this problem (see [4]).

It is a standard procedure in convex optimization to use the dual problem to certify infeasibility of the primal problem ([1, 6]). Weak duality of SDP ensures that if the dual problem is unbounded, then the primal problem is infeasible.The corresponding Lagrangian dual to $SDP(\mathcal{P}, \mathcal{X}, \mathcal{Y})$ is given by

$$L_D(\mathcal{P}, \mathcal{X}, \mathcal{Y}) := \begin{cases} \max\ \nu_{n_\nu} \\ \text{s.t. } B^T\lambda_2 B + e_1\lambda_1^T B + B^T\lambda_1 e_1^T + \\ +\lambda_3 + \sum_{i\in N_x}\nu_i Q_i + \sum_{i\in N_y}\nu_{n_x+i}R_i + \\ +\sum_{i\in N_d}\nu_{n_x+n_y+i}D_i + \nu_{n_\nu}e_1e_1^T = 0 \\ \lambda_1 \geq 0,\ \lambda_2 \geq 0,\ \lambda_3 \succeq 0, \end{cases}$$

where $\lambda_1 \in \mathbb{R}^{2(n_\xi-1)}$, $\lambda_2 \in \mathbb{S}^{2(n_\xi-1)}$, $\lambda_3 \in \mathbb{S}^{n_\xi}$, $\nu \in \mathbb{R}^{n_\nu}$ are the free variables, with $n_\nu = n_x + n_y + n_d + 1$.

4.1 Outer-Approximation

As infeasibility of (4) can be certified solving L_D, the feasible parameter set can be outer-approximated by removing infeasible subregions in the surrounding. There are many possibilities to do so. A general approach is to partition the initial parameter space and to check the partitions for infeasibility. This approach derives from the one proposed by [4], where parameter regions containing no steady states are similarly estimated.

In order to reduce the overall computational cost, the following simple bisection algorithm is implemented, so as to check groups of partitions simultaneously. Hereby, $\mathcal{Q}$ is an hyper-rectangle in the space of the bisection variables, which here is the space of the parameter variables p_i, for $i \in N_p$.

Algorithm 1 *Outer-Approximate*$(\mathcal{Q}, \mathcal{Y}_k, \mathcal{Y}_{k+1})$

0. *If volume*$(\mathcal{Q}) \leq \epsilon$, *return* $\mathcal{Q}$
1. *Compute the Lagrangian dual* $L_D(\mathcal{Q}, \mathcal{Y}_k, \mathcal{Y}_{k+1})$
2. *If* L_D *is feasible, then return* $\emptyset$
3. *If* L_D *is not feasible, then*
   a. *Partition* $\mathcal{Q}$ *into* $\mathcal{Q}_1$ *and* $\mathcal{Q}_2$
   b. *Set* $\mathcal{Q}_1' :=$ *Outer-Approximate*$(\mathcal{Q}_1, \mathcal{Y}_k, \mathcal{Y}_{k+1})$
   c. *Set* $\mathcal{Q}_2' :=$ *Outer-Approximate*$(\mathcal{Q}_2, \mathcal{Y}_k, \mathcal{Y}_{k+1})$
   d. *Return* $\mathcal{Q}_1' \cup \mathcal{Q}_2'$

Note that $\mathcal{P}_{k,k+1}$ is contained in the set obtained calling Outer-Approximate$(\mathcal{P}, \mathcal{Y}_k, \mathcal{Y}_{k+1})$, up to the given precision threshold ϵ. This results in a robust and convergent way to explore the parameter space. An illustration of the resulting partition given by the algorithm is given in Figure 2.

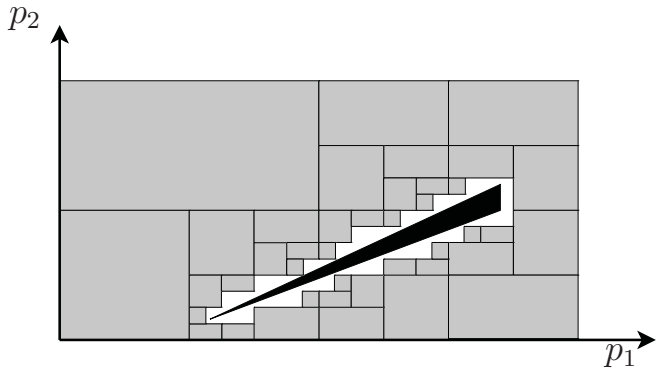

Figure 2: Bipartition algorithm. Grey boxes indicate invalidated parameter regions. The black set corresponds to the set to be approximated.

The computational time T_A of Algorithm 1 is $O(N \cdot T_D)$, where T_D is the time taken by a single Lagrangian dual computation and N is the number of dual evaluations, which depends on the precision threshold and the number of bisection variables considered. Increasing the dimensionality of the bi-partitioning or

reducing the threshold increases exponentially the time required, but it is important to note that the algorithm can be easily and efficiently paralleled.

The proposed algorithm can be modified to obtain upper and lower bounds on each variable considered corresponding to a one dimensional bi-sectioning. This can be advantageous if for example several time steps between the measurements are considered. Then, the number of bisection variables includes not only the unknown parameters n_p, but also the intermediate $(m-1)n_x$ system states. Considering only the physical bounds of the intermediate states may lead to too conservative results. It is therefore advantageous to combine less demanding upper and lower bounds approximation of the states, which are considered in the following section, with the computationally more demanding bi-sectioning of the parameter space.

5 State Estimation

Formulating the invalidation problem as a feasibility problem allows also to outer–approximate the initial conditions and the system states of the discrete system from measurements, even if the parameters and the experimental data are uncertain. This is basically achieved considering the feasibility problem (4) while keeping the parameters fixed and choosing the corresponding system states as bisection variables.

Initial State Estimation. Given each measurement $\mathcal{Y}_k$, the set X_k of the feasible system state at time index k can be estimated by solving the feasibility problem

$$F(\mathcal{X}_k, \mathcal{Y}_k) := \begin{cases} y[k] = h(x[k], p) \\ x[k] \in \mathcal{X}_k \\ y[k] \in \mathcal{Y}_k \\ p \in \mathcal{P}, \end{cases} \tag{5}$$

Note that this includes also the initial state.

Using the same approach used to bound $\mathcal{P}$ solving the feasibility problem $F(\mathcal{P}, \mathcal{X}, \mathcal{Y})$, we can similarly bound the set $\mathcal{X}_k$, using a semidefinite relaxations of $F(\mathcal{X}_k, \mathcal{Y}_k)$ and modifying the bisection algorithm so as to consider as bisection variables the state variables $x[k]_j$, for $i \in N_x$.

Intermediate States. Given an initial state $\mathcal{X}_k$, the system state at time $k+1$ can be estimated by solving the feasibility problem

$$F(\mathcal{X}_{k+1}, \mathcal{X}_k) := \begin{cases} x[k+1] = g(x[k], p) \\ x[k+1] \in \mathcal{X}_{k+1} \\ x[k] \in X_k \\ p \in \mathcal{P}, \end{cases} \tag{6}$$

As for the previous problem, we can bound the set $\mathcal{X}_{k+1}$, given $\mathcal{X}_k$, using a semidefinite relaxations of $F(\mathcal{X}_{k+1}, \mathcal{X}_k)$, and modifying the bisection algorithm so as to consider as bisection variables the state variables $x[k]_j$, for $i \in N_x$. By estimating the state space for consecutive time indexes allows to evaluate the flow of the system.

Note that the system state at time $k-1$ can be similarly estimated. As the quality of the bounds obtained decreases as the distance from the last measurement increases, this allows to obtain better estimates for states that are far from the closest previous measurement, but near to the following one.

6 Example

In this section we show the application of our method to a well-known example system. We consider two possible reaction mechanisms proposed by [3] between an enzyme (E) and a substrate (S) forming an enzyme-substrate complex (C):

$$E + S \underset{p_{-1}}{\overset{p_1}{\rightleftharpoons}} C \xrightarrow{p_2} E + P, \tag{7}$$

and

$$C \underset{\tilde{p}_{-1}}{\overset{\tilde{p}_1}{\rightleftharpoons}} E + S \xrightarrow{\tilde{p}_2} E + P. \tag{8}$$

These reaction schemes are known nowadays respectively as the Michaelis-Menten (MM) mechanism of enzyme activation (7) and as the Henri (H) mechanism (8). Both reaction schemes and their relevance are discussed in detail e.g. in [10], in which a main result states that both reaction mechanisms are analytically distinguishable, if the transient initial dynamic of two independent states is considered. We therefore consider measurements in the transient phase.

The reaction mechanisms are modeled according to the law of mass action. Since both mechanisms obey the two conservation laws

$$\begin{aligned} e_0 &= e(t) + c(t), \\ s_0 &= s(t) + c(t) + p(t), \end{aligned}$$

both models can be expressed as second order systems. Let us then consider a simple first order Euler discretization scheme (see e.g. [5]), given by

$$\dot{x}[k] \approx \frac{x[k+1] - x[k]}{h},$$

where $h > 0$ denotes the size of the discretization step and $x[k]$ the value of $x(t)$ for $t = kh$. Fixing $e_0 = 1$, the corresponding difference equations for the MM mechanism are given by

$$\begin{aligned} s[\mathrm{k}+1] &= s[k] + p_1 h((c[k]-1)s[k] + K_S c[k]) \\ c[\mathrm{k}+1] &= c[k] + p_1 h((1-c[k])s[k] - K_M c[k]), \end{aligned} \tag{9}$$

where $K_S = p_{-1}/p_1$, $K_M = (p_{-1} + p_2)/p_1$. For the Henri mechanism we obtain

$$\begin{aligned} s[\mathrm{k}+1] &= s[k] + h(\tilde{p}_1 + \tilde{p}_2)((c[k]-1)s[k] - K_H c[k]) \\ c[\mathrm{k}+1] &= c[k] + \tilde{p}_1 h((1-c[k])s[k] - \tilde{K}_S c[k]), \end{aligned} \tag{10}$$

where $\tilde{K}_S = \tilde{p}_{-1}/\tilde{p}_1$ and $K_H = \tilde{p}_{-1}/(\tilde{p}_1 + \tilde{p}_2)$.

6.1 Model Discrimination

In order to discriminate between the models, we consider the ideal (noise-free) measurements

$$\mathcal{Y}^H = \{\mathcal{Y}_k = y[k] = (s[k], c[k])^T,\ 3 \le k \le 7\},$$

taken from the Henri mechanism (H) with $(s[0], c[0]) = (1, 0)$, $h = 0.1$sec and the parameter values $\tilde{p}_1 = 1$, $\tilde{p}_{-1} = 1$, $\tilde{p}_2 = 1$. Next, we show that the MM is invalid with respect to these measurements.

For the invalidation setup, we consider as initial parameter space $\mathcal{P}^{MM}$ a range of one order of magnitude for each parameter:

$$\frac{1}{3} \le p_1,\ p_{-1},\ p_2 \le 3.$$

In order to show model invalidity, the feasible parameter sets $\mathcal{P}_{3,4}$, $\mathcal{P}_{4,5}$, $\mathcal{P}_{5,6}$, and $\mathcal{P}_{6,7}$ for the corresponding time points are calculated via Algorithm 1. As partition of the parameter space $\mathcal{P}^{\mathrm{MM}}$ we chose unitary hypercubes, whereas each cube is of size 0.05. The feasible parameter sets are shown in Figure 3.

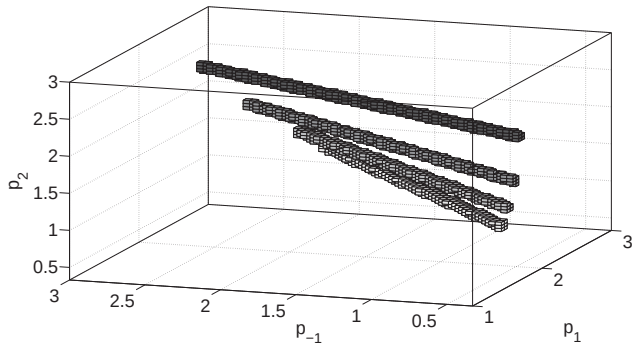

Figure 3: Feasible parameter sets from $\mathcal{P}_{3,4}$ to $\mathcal{P}_{6,7}$ from to dark to light gray matching the measurements.

As can clearly be seen from the Figure 3, the intersection

$$\mathcal{P}_{\mathcal{Y}^H} = \bigcap_{i=3}^{6} \mathcal{P}_{i,i+1}$$

of the feasible parameter sets is empty. Thus, the Michaelis-Menten mechanism with the admitted $\mathcal{P}^{MM}$ is invalidated with respect to the measurements $\mathcal{Y}^H$.

6.2 Parameter Estimation

In order to estimate the model parameters, we consider now the ideal measurements $\mathcal{Y}^{\mathrm{MM}} = \{\mathcal{Y}_k = y[k] = (s[k], c[k])^T,\ 1 \leq k \leq 6\}$ taken from the Michaelis-Menten mechanism (9) with $(s[0], c[0]) = (1, 0)$, $h = 0.1$sec and the parameter values $p_1 = 1$, $p_{-1} = 1$, $p_2 = 1$.

The feasible parameter sets $\mathcal{P}_{i,j}$ are estimated considering Algorithm 1 and depicted in Figure 4. In this case, the intersection $\mathcal{P}_{\mathcal{Y}}$ according to Corollary 3 is highlighted in black corresponding to the consistent feasible parameter set.

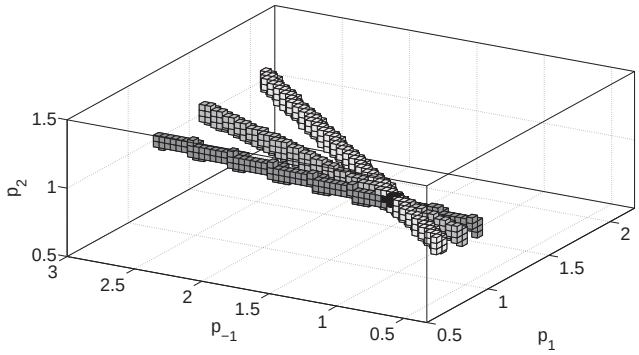

Figure 4: Parameter sets. The intersection of the parameter sets is indicated black.

The complement $\mathcal{P} \setminus \mathcal{P}_{\mathcal{Y}}$ of $\mathcal{P}_{\mathcal{Y}}$ is therefore classified as infeasible.

As the number of measurements considered increases, the parameter estimates improve, see Figure 5. Notably, with this approach large regions of the parameter space can be discarded considering only a few measurements.

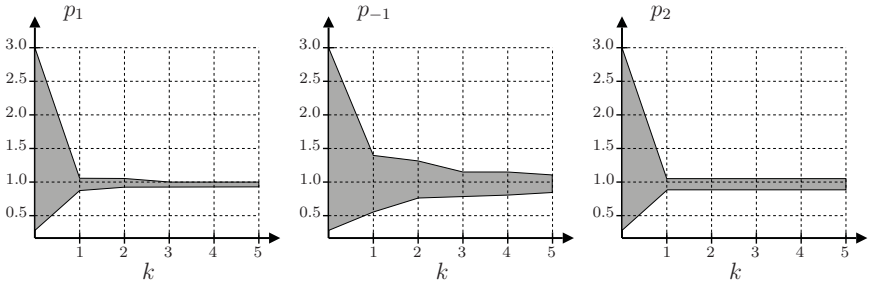

Figure 5: Feasible parameter regions (shaded in the figure) with respect to the number of measurements taken into account.

Remark 2 *For the examples considered, the Lagrangian dual is computed within less than one second on a standard desktop computer using MATLAB 2008 and the SEDUMI ([11]) solver.*

7 Discussion and Outlook

In this contribution we have studied the problems of model invalidation, parameter and state estimation of discrete-time polynomial systems, and we proposed a method that allows to discriminate between different model hypotheses by invalidation. Conclusive results can be achieved on the model invalidation problem even if only imprecise or sparse measurements are available. The method also allows to discard large parameter regions, and therefore complements existing identification methods. Furthermore, we have also shown how to use this method to estimate the model states from the measurements.

The proposed method is based on the outer-approximation of the feasible parameter space, or of the feasible state space, for state estimation, into a set of regions that are consistent with the experimental data. These regions correspond to the solution space of a nonlinear feasibility problem, which we relax into a semidefinite program. By solving the corresponding Lagrangian dual, and applying a bisection algorithm, the feasible regions can then be efficiently bounded.

We have shown with an example, considering two alternative enzyme-substrate reaction mechanisms, that with this method it is possible to prove invalidity of one of the two mechanisms. Extending this example, we also demonstrated the applicability of the approach to estimate the model parameters.

In summary, our approach is a reliable and computationally manageable method for dynamical model invalidation, and parameter and state estimation. Future work considered including model inputs and addressing the issue of experimental design of biochemical reaction networks.

8 References

[1] S. Boyd and L. Vandenberghe. *Convex Optimisation*. Cambridge University Press, UK, 2004.

[2] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, 3. edition edition, 2004.

[3] V. Henri. General theory of action of certain hydrolases. *C. R. H. Acad. Sci. Paris*, 135:916–919, 1902.

[4] L. Kuepfer, U. Sauer, and P.A. Parrilo. Efficient classification of complete parameter regions based on semidefinite programming. *BMC Bioinformatics*, 8:12, 2007.

[5] C. Letellier, S. Elaydi, L.A. Aguirre, and A. Alaoui. Difference equations versus differential equations, a possible equivalence for the Rössler system? *Physica D*, 195:29–49, 2004.

[6] D.G. Luenberger. *Linear and Nonlinear Programming*. Springer, 2003.

[7] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program., Ser. B*, 96:293–320, 2003.

[8] S. Prajna. Barrier certificates for nonlinear model validation. *Automatica*, 42:117–126, 2006.

[9] M.V. Ramana. *An Algorithmic analysis of multiquadratic and semidefinite programming problems*. PhD thesis, John Hopkins University, 1994.

[10] S. Schnell, M.J. Chappell, N.D. Evans, and M.R. Roussel. The mechanism distinguishability problem in biochemical kinetics: The single-enzyme, single-substrate reaction as a case study. *Comptes rendus-Biologies*, 329:51–61, 2006.

[11] J.F. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11:625–653, 1999.

[12] S. Waldherr, R. Findeisen, and F. Allgöwer. Global sensitivity analysis of biochemical reaction networks via semidefinite programming. *Proceedings of the 10th IFAC World Congress*, 2008.