SET BASED UNCERTAINTY ANALYSIS AND PARAMETER ESTIMATION OF BIOLOGICAL NETWORKS WITH THE BIOSDP TOOLBOX

Steffen Waldherr, Jan Hasenauer, and Frank Allgöwer

Institute for Systems Theory and Automatic Control, University of Stuttgart
Pfaffenwaldring 9, 70550 Stuttgart, Germany

{waldherr, hasenauer, allgower}@ist.uni-stuttgart.de

ABSTRACT

We present the *bioSDP* toolbox for Matlab, which provides methods for the analysis of biological networks with parametric uncertainty. Its set based approach allows to compute guaranteed bounds on the network behaviour under uncertainty, or on the parameter values consistent with uncertain measurement data. These two applications of the toolbox are illustrated with two case studies for specific biological networks.

1. INTRODUCTION

The main goal of system biology is a quantitative description of cellular processes. Unfortunately, achieving this is rather difficult, as models of biological systems are subject to a significant degree of uncertainty. This uncertainty arises from limited knowledge about the system, mostly due to limitations in the experimental technology, and/or large variations in environmental and internal boundary conditions. These are manifested by the large parametric uncertainty shown by most models of biochemical networks. Drawing dedicated conclusions about various system properties is a major computational challenge under such uncertainty.

We present the Matlab toolbox *bioSDP* which provides methods to analyse biological networks modeled by ordinary differential equations with polynomial and rational terms. It allows to study the variation in steady states of a network under parametric uncertainty, and to upper-bound the size of the remaining parametric uncertainty from noise corrupted measurement data. This allows to derive guaranteed predictions about important system properties despite the uncertainty in the input data.

The *bioSDP* toolbox is available under an open source license from [1].

2. THEORETICAL BACKGROUND

In this section, we present the steady state uncertainty analysis problem, the parameter estimation problem and the algorithms used to solve them.

2.1. The steady state uncertainty analysis problem

The steady state uncertainty analysis problem [2, 3] is defined by a system of polynomial or rational equalities

$$F(x,p) = 0, \tag{1}$$

where $x \in \mathbb{R}^n$ is the network's state vector (e.g. protein or metabolite concentrations), $p \in \mathbb{R}^m$ is the parameter vector, and F is a vector of polynomial functions in x and p, typically n-dimensional, describing the network's steady state conditions.

The values of the parameters p in (1) are uncertain in that only a bounding box for the parameters is given, but not the exact values. This box is defined by element-wise inequalities, yielding the set

$$\mathcal{P} = \left\{p \in \mathbb{R}^m \mid \check{p} \leq p \leq \hat{p}\right\}, \tag{2}$$

where $\check{p}$ and $\hat{p}$ are element-wise lower and upper bounds on the parameter vector p.

The task in the steady state uncertainty analysis problem is to compute a tight outer approximation $\widehat{\mathcal{X}}$ for the set $\mathcal{X}$ of all feasible solutions x of (1), i.e.,

$$\widehat{\mathcal{X}} \supset \mathcal{X} = \left\{x \in \mathbb{R}^n \mid \exists p \in \mathcal{P} : F(x,p) = 0\right\}. \tag{3}$$

In the algorithm implemented in the *bioSDP* toolbox, the approximation $\widehat{\mathcal{X}}$ is either computed as one big bounding box for all feasible steady states, or as the union of many small boxes, depending on the user's choice. The second option generally yields a tighter approximation at the expense of a higher computational effort.

2.2. The parameter estimation problem

In the parameter estimation problem [4, 5], we consider a dynamic network defined by the difference equation

$$\begin{aligned} x_k &= F(x_{k-1}, p) \\ y_k &= H(x_k, p), \end{aligned} \tag{4}$$

where $x_k \in \mathbb{R}^n$, $p \in \mathbb{R}^m$, and F are as in Section 2.1, and $y_k \in \mathbb{R}^q$ is a vector of measurements, which depends on the network's state via the measurement function H. The index $k = 1, \ldots, N$ denotes the discrete time steps. We assume that both F and H are polynomial or rational functions in both the state x and the parameter p.

Uncertain measurement data are given as a bounding box $\mathcal{Y}_k$ on the output vector y_k for each time point k,

$$\mathcal{Y}_k = \left\{y \in \mathbb{R}^q \mid \check{y}_k \leq y \leq \hat{y}_k\right\}, \quad k = 1, \ldots, N. \tag{5}$$

Also, a bounding box on the states x_k is given as

$$\mathcal{X} = \left\{x_k \in \mathbb{R}^n \mid \check{x} \leq x_k \leq \hat{x}\right\}. \tag{6}$$

A parameter p is called consistent, if there exist a sequence of states $x_k \in \mathcal{X}$ and a sequence of outputs $y_k \in \mathcal{Y}_k$ with $k = 1, \ldots, N$ satisfying (4). The goal in the parameter estimation problem is to compute a tight outer approximation $\widehat{\mathcal{P}}$ of the set of consistent parameters $\mathcal{P}$:

$$\widehat{\mathcal{P}} \supset \mathcal{P} = \left\{ p \in \mathbb{R}^m \mid p \text{ is consistent} \right\}. \tag{7}$$

As in the steady state uncertainty analysis problem, the algorithm implemented in *bioSDP* can either compute an outer approximation $\widehat{\mathcal{P}}$ as one big bounding box, or as the union of many smaller boxes.

2.3. Iterative set exclusion with an infeasibility test

The algorithms implemented in *bioSDP* compute the outer approximations $\widehat{\mathcal{X}}$ and $\widehat{\mathcal{P}}$ with an iterative set exclusion approach. As *a priori* information, an initial estimate $\widehat{\mathcal{X}}_0$ or $\widehat{\mathcal{P}}_0$ is required. The basis of the set exclusion approach is an infeasibility test, which applies to a system of polynomial equalities and a box constraint of the form

$$\begin{aligned} G(\chi) &= 0 \\ \check{\chi} \leq \chi &\leq \hat{\chi}, \end{aligned} \tag{8}$$

where χ is a vector of uncertain variables, e.g., the state x and the parameters p for the steady state uncertainty analysis problem, or the state and output sequences x_k, y_k and parameters p for the parameter estimation problem. G is a polynomial vector-valued function representing the equality constraints.

In each subsequent iteration step, the approximation is refined by excluding subsets which pass an infeasibility test. In iteration i, the algorithm generates an appropriate list of test sets $\widetilde{\mathcal{X}}_{i,1}, \ldots \widetilde{\mathcal{X}}_{i,r}$ (or $\widetilde{\mathcal{P}}_{i,1}, \ldots, \widetilde{\mathcal{P}}_{i,r}$ for the parameter estimation problem), and applies the infeasibility test to each test set. Those test sets which pass the infeasibility test are then united to form the exclusion set $\widetilde{\mathcal{X}}_i$, and the refined approximation for the next iteration is obtained as

$$\widehat{\mathcal{X}}_{i+1} = \widehat{\mathcal{X}}_i \setminus \widetilde{\mathcal{X}}_i. \tag{9}$$

In each iteration, the test sets are reduced in size, and the algorithm terminates when the size drops beneath a predefined threshold.

The infeasibility tests are solved by a quadratic reformulation of the problem (8) and semidefinite programming [6]. Details on the construction of (8) and the infeasibility test are given in [2, 3] for the steady state uncertainty analysis, and in [4] for the parameter estimation.

3. STRUCTURE OF THE *bioSDP* TOOLBOX

3.1. Analysis tasks

The main functionality of *bioSDP* is to offer methods for solving the steady state uncertainty analysis problem from Section 2.1, and the parameter estimation problem from Section 2.2.

The main *bioSDP* routine for steady state uncertainty analysis is `stationary_uncertainty`. This routine takes the problem setup specified in Matlab structure variables called `system` and `uncertainty`, and computes an outer approximation of the set of feasible steady states. The `system` variable contains the definition of the model variables and equations, while the `uncertainty` variable describes the parameter range to consider as well as the *a priori* state bounds. The behaviour of the uncertainty analysis function is further controlled by several options, which are passed as an additional structure variable, here called `options`, to the function.

A required option is `set_exclusion.method`, which specifies how regions in state space are removed from the feasible set. The default choice is `'box shrinkage'`, which just tries to reduce the size of the initial box as far as possible. A method which allows to obtain a more refined uncertainty set is `'bisection'`, in which the uncertainty set is computed by multi-dimensional bisection. However, the bisection method is only recommended for state spaces up to dimension three due to significantly increased computational effort in higher dimensions.

The steady state uncertainty analysis is then simply performed by a call to `stationary_uncertainty`, passing the problem setup in the variables `system` and `uncertainty` as well as the algorithm options as arguments to the function.

The parameter estimation problem is handled by the function `parameter_estimation`. It also takes two problem definition variables as arguments. The first one, `system`, contains the model definition, and the second one, `uncertainty`, contains the measurement bounds as well as the *a priori* state and parameter bounds. As in the steady state uncertainty analysis, either `'box shrinkage'` or `'bisection'` can be chosen as set exclusion method.

In addition to computing outer bounds by set exclusion, *bioSDP* offers methods to compute samples of feasible steady states for the uncertainty analysis or consistent parameters for the parameter estimation. These can for example be used to evaluate the tightness of the computed outer approximations.

3.2. Generic SDP problem solver

At the core of the *bioSDP* toolbox is an optimization algorithm which solves set exclusion problems as discussed in Section 2.3 with semidefinite programming methods. This algorithm does the iteration required for the set exclusion, constructs the appropriate semidefinite programs for the infeasibility tests, and refines the bounding sets based on the solutions to the semidefinite programs. The semidefinite programs are not solved by *bioSDP* itself, but are handed to specialised optimisation toolboxes, for example SeDuMi [7].

In short, the algorithm takes a system of constraints as defined in (8), together with an initial bounding box on the uncertain variables χ. The vector χ is thereby structured in variables where the uncertainty remains fixed to the initial bounds (e.g., the parameter bounds in the steady state uncertainty analysis problem), and variables where the uncertainty is to be reduced as much as possible, while ensuring that all feasible solutions to (8) are retained in the set (e.g., the steady state bounds for the steady state uncertainty analysis problem). The algorithm then iteratively prunes subsets of this initial uncertainty set, based on results from the infeasibility tests.

Table 1. Nominal parameter values for the insulin pathway model (10).

k_1	k_0	k_2	k_R	k_3	k_{m3}	ins
0.05	10^{-6}	1.0	0.5	1.0	30	1.0

In typical use cases, this generic algorithm need not to be called directly by the user. Instead, the wrapper methods for individual analysis tasks as described in Section 3.1 should be used. These methods take care of constructing the constraints (8) and the vector of uncertain variables χ as appropriate for the specific task from the problem-specific user input data.

3.3. Visualisation routines

The *bioSDP* toolbox offers two main possibilities for visualising the set based estimates resulting from the implemented analysis algorithms: box plots for two- and three-dimensional state or parameter spaces, and parallel coordinates plots for problems of any dimension. The output of the analysis functions described in Section 3.1 can directly be passed to the *bioSDP* routine `visualize_uncertainty` for this purpose.

4. EXAMPLE APPLICATIONS

We present the two main applications of the *bioSDP* toolbox—steady state uncertainty analysis and parameter estimation—with two exemplary studies. In the interest of brevity, we discuss for each example only the problem setup, some key points of their implementation in the *bioSDP* toolbox, and significant conclusions drawn from the analysis. For the full implementation of both examples, we refer the reader to the toolbox' software package available from [1], which contains both examples implemented as Matlab scripts.

4.1. Uncertainty analysis of an insulin pathway model

The steady state uncertainty analysis is exemplified with a simple insulin pathway model taken from [8]. The model equations are given as

$$\begin{aligned} \dot{\mathrm{IR}} &= -k_0\mathrm{IR} - k_1\, ins\, \mathrm{IR} + k_R\,(\mathrm{IR}_{tot} - \mathrm{IR} - \mathrm{IRP}) \\ \dot{\mathrm{IRP}} &= k_0\mathrm{IR} + k_1\, ins\, \mathrm{IR} - k_2\, \mathrm{IRP} \\ \dot{\mathrm{IRSP}} &= k_3\, \mathrm{IRP}\,(\mathrm{IRS}_{tot} - \mathrm{IRSP}) - k_{m3}\, \mathrm{IRSP}, \end{aligned} \tag{10}$$

with IR, IRP, and IRSP the concentrations of insulin receptor, phosphorylated insulin receptor, and phosphorylated insulin receptor substrate, respectively. The total protein amounts are conserved and given by $\mathrm{IR}_{tot} = 10$ and $\mathrm{IRS}_{tot} = 10$. Note that the conservation relations can directly be used to compute an *a priori* outer bound $\widehat{\mathcal{X}}_0$ independent of any parameter values.

For the steady state uncertainty analysis problem, we assume that all of the model parameters may vary by a factor of 2 around their nominal values, which are given in Table 1. Using both the multi-dimensional bisection method and the simpler box shrinkage, we let *bioSDP* compute an outer approximation to the set of feasible steady states under this uncertainty. The resulting outer estimate

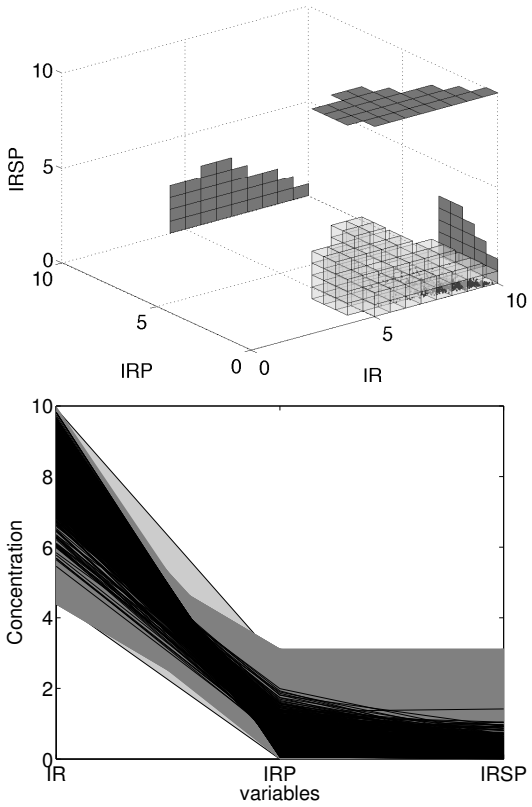

Figure 1. Results for the steady state uncertainty analysis of the insulin pathway model (10). Top: Outer bounds to the set of feasible steady states in a box plot. Bottom: Outer bounds to the set of feasible steady states in a parallel coordinates plot together with sampled steady states. Feasible intervals from the box shrinkage algorithm are shown as light gray, and from the bisection algorithm as dark gray.

$\widehat{\mathcal{X}}$ on the feasible steady states is shown in Figure 1, together with some sampled steady states. These plots are generated directly by *bioSDP*'s builtin visualisation routines discussed in Section 3.3. The comparison between the outer bounds obtained with the set exclusion methods and the sampled steady states shows that the outer bounds are reasonably tight. The computation time on a standard desktop computer was about 50 seconds for the box shrinkage method and 140 seconds for the bisection method. Both methods achieved similar bounds in this example. In the parallel coordinates plot, the thinning waist between IR and IRP, from the bounds obtained by bisection, indicates a slight negative correlation between these two variables. This correlation is also seen more explicitly in the three-dimensional box plot.

4.2. Parameter estimation for a reversible modification reaction

The set based parameter estimation with *bioSDP* is illustrated with a very simplistic biological model of a reversible modification reaction. The time-continuous model for this reaction is given by the scalar differential equation

$$\dot{x} = -k_m\, x + k_d(1 - x), \tag{11}$$

where x is the concentration of the unmodified variant of the considered molecular species measured relative to its total concentration, and k_m and k_d are unknown parameters to be estimated.

The parameter estimation is done for artificial measurement data for x from five time points which are 0.1 time units apart. The measurement is uncertain in that only upper and lower bounds are available. For the purpose of this example, we set $\check{x} = (0.0, 0.44, 0.57, 0.6, 0.56)$ as lower bound and $\hat{x} = (0.1, 0.54, 0.67, 0.70, 0.66)$ as upper bound.

First, we transform the differential equation model (11) to a difference equation using the Euler-forward discretisation scheme. In *bioSDP*, this is simply done by calling the auxiliary function `discretize_ode`, passing the continuous model (11), its state variable x, the length of the time step (here 0.1) and the number of steps to take (here 4) as arguments. *bioSDP* then automatically generates the discrete equations (4) and the internally required sequence variables x_k.

In the next step, the set based parameter estimation is carried out by a call to the `parameter_estimation` function. The results from the set based analysis are complemented by samples from the set of consistent parameters obtained through Monte Carlo sampling. Result plots from *bioSDP*'s visualisation routines are shown in Figure 2. From the parallel coordinates plot, we observe that there seems to be a correlation in the consistent parameters: from the sampled parameter vectors, either both elements are low, or both are high within the respective intervals. Using the box plot with the `'bisection'` set exclusion method, this observation is confirmed by the resulting outer bound on the set of consistent parameters.

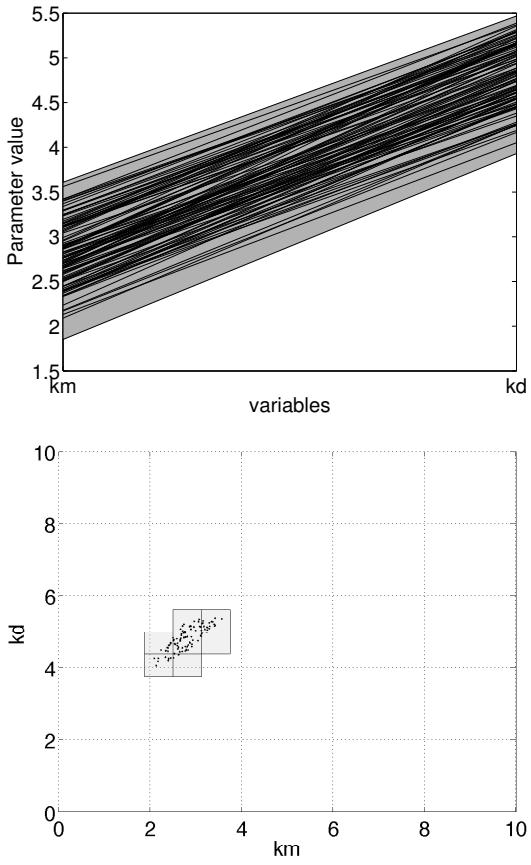

Figure 2. Results for the parameter estimation of the reversible modification reaction (11). Top: Outer bounds to the set of feasible steady states in a box plot. Bottom: Outer bounds to the set of feasible steady states in a parallel coordinates plot together with sampled steady states.

5. CONCLUSIONS

We have presented the *bioSDP* toolbox for Matlab. *bioSDP* provides methods for steady state uncertainty analysis and parameter estimation from uncertain measurement data in biological networks. While the examples presented here only involve small networks, the methods are also applicable to medium-sized networks. For example, a case study for a rather detailed model of tumor necrosis factor signalling is presented in [9].

6. ACKNOWLEDGEMENTS

We thank Ganzhou Wang and Shen Zeng for their help in the implementation of the *bioSDP* toolbox, Pelle Lundberg and Gunnar Cedersund for testing, and Julian Heinrich for helpful comments on the parallel coordinate plots.

7. REFERENCES

[1] "The biosdp toolbox," `http://biosdp.sourceforge.net`, 2012, Accessed 26th March 2012.

[2] S. Waldherr, R. Findeisen, and F. Allgöwer, "Global sensitivity analysis of biochemical reaction networks via semidefinite programming," in *Proc. of the 17th IFAC World Congress, Seoul, Korea*, 2008, pp. 9701–06.

[3] J. Hasenauer, P. Rumschinski, S. Waldherr, S. Borchers, F. Allgöwer, and R. Findeisen, "Guaranteed steady state bounds for uncertain (bio-)chemical processes using infeasibility certificates," *J. Process Control*, vol. 20, no. 9, pp. 1076–1083, 2010.

[4] J. Hasenauer, S. Waldherr, K. Wagner, and F. Allgöwer, "Parameter identification, experimental design and model falsification for biological network models using semidefinite programming," *IET Syst. Biol.*, vol. 4, no. 2, pp. 119–130, 2010.

[5] Ph. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, and R. Findeisen, "Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks," *BMC Syst. Biol.*, vol. 4, 69, 2010.

[6] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.

[7] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11, no. 1, pp. 625–653, 1999.

[8] C. Brännmark, R. Palmer, S.T. Glad, G. Cedersund, and P. Strålfors, "Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework," *J. Biol. Chem.*, vol. 285, no. 26, pp. 20171–20179, 2010.

[9] S. Waldherr, J. Hasenauer, M. Doszczak, P. Scheurich, and F. Allgöwer, "Global uncertainty analysis for a model of TNF-induced NF-κB signalling," in *Advances in the Theory of Control, Signals and Systems with Physical Modeling*, Jean Lévine and Philippe Müllhaupt, Eds., vol. 407 of *LNCIS*, pp. 365–377. Springer Verlag, Berlin Heidelberg, 2010.