

# Set-based parameter estimation, model falsification and uncertainty analysis using convex optimization

Jan Hasenauer, Steffen Waldherr, and Frank Allgöwer

Institute for Systems Theory and Automatic Control University of Stuttgart

Workshop "Conclusions despite uncertainties" Linköping, June 2011



### Measurement data





Fluorescence microscopy





Flow cytometry



DNA and Protein microarrays (genomics, proteomics, ...)



X-ray crystallography

### Measurement data





Fluorescence microscopy





Flow cytometry



DNA and Protein microarrays (genomics, proteomics, ...)



X-ray crystallography

measurement devices improved tremendously

BUT always limited and noise corrupted data!





#### Classical approach

- Calculation of *optimal* parameter (e.g. least-square or maximum-likelihood)
- $\Rightarrow$  optimal model of the process
- 2 Model falsification/rejection and prediction using optimal model





#### Classical approach

- Calculation of *optimal* parameter (e.g. least-square or maximum-likelihood)
- $\Rightarrow$  optimal model of the process
- 2 Model falsification/rejection and prediction using optimal model

**Problem:** For sparse, noisy data many parameters are equally plausible.  $\Rightarrow$  How can this be taken into account?





#### Sampling-based approach

- Calculation of a sample of good parameters (e.g. Bayesian methods or bootstrapping)
- $\Rightarrow$  sample of *good* models of the process
- **2** Model falsification/rejection and prediction using a sample of models





#### Sampling-based approach

- Calculation of a sample of good parameters (e.g. Bayesian methods or bootstrapping)
- $\Rightarrow$  sample of *good* models of the process
- **2** Model falsification/rejection and prediction using a sample of models

#### **Advantage:** More robust decision making. **Problem:** The usage of a finite sample makes model falsification difficult.

isto





#### Set-based approach [Jaulin2001]

- Calculation of the set of *consistent* parameters (e.g. interval arithmetics, SOS, or convex optimization)
- $\Rightarrow$  all *consistent* models of the process (with a certain structure)
- 2 Model falsification/rejection and prediction using the set of models





#### Set-based approach [Jaulin2001]

- Calculation of the set of *consistent* parameters (e.g. interval arithmetics, SOS, or convex optimization)
- $\Rightarrow$  all *consistent* models of the process (with a certain structure)
- Odel falsification/rejection and prediction using the set of models

**Advantage:** Model falsification is very easy. **Problem:** Analysis is computationally demanding.

ist





**Goal of this talk:** Illustration of the usage of different methods for the analysis of model and prediction uncertainties!



### Outline



#### 1 Introduction

2 Parameter estimation and model falsification

Steady-state uncertainty





### Outline



#### Introduction

#### 2 Parameter estimation and model falsification

3 Steady-state uncertainty







#### Problem of set-based parameter estimation

Given a set of measurement data  $\bar{y}(t_k) = y(t_k) + \epsilon(t_k)$  and bounds for the measurement noise  $\epsilon(t_k)$ , determine the set of all parameters, which could have generated the measurement data.







#### Problem of set-based parameter estimation

Given a set of measurement data  $\bar{y}(t_k) = y(t_k) + \epsilon(t_k)$  and bounds for the measurement noise  $\epsilon(t_k)$ , determine the set of all parameters, which could have generated the measurement data.

 $\Rightarrow$  set of consistent parameters  $\mathcal{P}^*$ 





#### Problem of set-based parameter estimation

Given a set of measurement data  $\bar{y}(t_k) = y(t_k) + \epsilon(t_k)$  and bounds for the measurement noise  $\epsilon(t_k)$ , determine the set of all parameters, which could have generated the measurement data.

 $\Rightarrow$  set of consistent parameters  $\mathcal{P}^*$ 



**Problem:** It is impossible to compute  $\mathcal{P}^*$  precisely!  $\Rightarrow$  Computation of outer approximation  $\overline{\mathcal{P}}^* \supseteq \mathcal{P}^*$ 

isto



#### Problem of set-based parameter estimation

Given a set of measurement data  $\bar{y}(t_k) = y(t_k) + \epsilon(t_k)$  and bounds for the measurement noise  $\epsilon(t_k)$ , determine the set of all parameters, which could have generated the measurement data.

 $\Rightarrow$  set of consistent parameters  $\mathcal{P}^*$ 



Why is the outer approximation  $\bar{\mathcal{P}}^*$  useful?

# Usage of $\bar{\mathcal{P}}^*$ for model falsification

By definition the set of consistent parameters  $\mathcal{P}^*$  contains all parameter p of a model which can explain the measurement data.

 $\Rightarrow \bar{\mathcal{P}}^*$  contains all parameter which can describe the data!

Let's assume we have two models and the corresponding sets  $\bar{\mathcal{P}}^*$ :



 $\Rightarrow$  Model A is falsified as there are no consistent parameters.

 $\Rightarrow\,$  Model B may be able to reproduce the data.

# Usage of $\bar{\mathcal{P}}^*$ for model falsification

By definition the set of consistent parameters  $\mathcal{P}^*$  contains all parameter p of a model which can explain the measurement data.

 $\Rightarrow \bar{\mathcal{P}}^*$  contains all parameter which can describe the data!

Let's assume we have two models and the corresponding sets  $\bar{\mathcal{P}}^*$ :



 $\Rightarrow$  Model B may be able to reproduce the data.

#### Calculation of $\bar{\mathcal{P}}^*$ ?



We consider nonlinear discrete time systems,

$$x^{(k+1)} = F(x^{(k)}, p), \quad x^{(0)} = x_0$$
  
 $y^{(k)} = H(x^{(k)}, p),$ 

in which  $x^{(k)} \in \mathbb{R}^n$  is the state,  $y^{(k)} \in \mathbb{R}^m$  is the output,  $p \in \mathbb{R}^q$  the vector of unknown parameters, and F and H are rational functions.

F can be obtained by discretizing a continuous process ( $\rightarrow$  ODE solver).

#### Measurement data

$$\bar{y}^{(k)} = y^{(k)} + \epsilon^{(k)}, \quad k \in \{1, \dots, N\},$$

in which  $y^{(k)} \in \mathbb{R}^m$  is the measured noise corrupted output, and  $\epsilon^{(k)} \in \mathbb{R}^m$  is the bounded measurement noise.  $\Rightarrow y^{(k)} \in \mathcal{Y}^{(k)}$ 





# Computation of $\bar{\mathcal{P}}^*$ via a series of feasibility problems



#### Feasibility problem [Hasenauer2010a]

- verification that a set  $\mathcal{P}_i$  cannot contain consistent parameters
- feasibility problem for test set  $\mathcal{P}_i$ :

$$(\mathsf{P}): \begin{cases} \mathsf{find} & p \in \mathcal{P}_i, \ x^{(k)} \in \mathcal{X}, \ y^{(k)} \in \mathcal{Y}^{(k)} \\ \mathsf{subject to} & x^{(k+1)} = F(x^{(k)}, p), \ \forall k \quad \to \mathsf{dynamics} \\ & y^{(k)} = H(x^{(k)}, p), \ \forall k \quad \to \mathsf{measurement} \end{cases}$$

(P) infeasible  $\iff \mathcal{P}_i$  does not contain consistent parameters

isto

# Summary: Reformulation of the feasibility problem





### Quadratic decomposition

#### Reformulation of dynamics and measurement

- Monomial vector:  $\xi^{\mathcal{T}} = (1, p, x^{(k)}, y^{(k)}, p \cdot x^{(k)}, \ldots) \in \mathbb{R}^{\kappa}$
- Dynamics/Measurement:

$$0 = x^{(k+1)} - F(x^{(k)}, p) := G(x^{(k+1)}, x^{(k)}, p)$$

$$\Rightarrow 0 = G_i(x^{(k+1)}, x^{(k)}, p) = \xi^T Q_i \xi, \quad \forall i$$

Quadratic feasibility problem

$$(\mathsf{QP}): \begin{cases} \mathsf{find} & \xi \in \mathbb{R}^{\kappa} \\ \mathsf{subject to} & \xi^{\mathsf{T}} Q_i \xi = 0 & i = 1, \dots, c \\ & B \xi \geq 0 \\ & \xi_1 = 1 \end{cases}$$

in which  $p \in \mathcal{P}_i$ ,  $x^{(k)} \in \mathcal{X}^{(k)}$ ,  $y^{(k)} \in \mathcal{Y}^{(k)} \iff B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)}) \xi \ge 0$ 

(QP) infeasible  $\iff$  (P) infeasible

isto



#### Feasibility problem

- Symmetric monomial matrix:  $X = \xi \xi^T$
- Feasibility problem in X:

$$(\widetilde{\mathsf{QP}}): \begin{cases} \mathsf{find} & X \in \mathcal{S}^{\kappa} \\ \mathsf{subject to} & \mathrm{tr}(Q_{i}X) = 0 & i = 1, \dots, c \\ & BXe_{1} \ge 0 \\ & \mathrm{tr}(e_{1}e_{1}^{T}X) = 1 \\ & \mathrm{rank}(X) = 1 \\ & X \succcurlyeq 0 \end{cases}$$
  
in which  $e_{1} = [1, 0, \dots, 0]^{T}$ 



#### Relaxed feasibility problem

- Symmetric monomial matrix:  $X = \xi \xi^T$
- Feasibility problem in X:





#### Relaxed feasibility problem

- Symmetric monomial matrix:  $X = \xi \xi^T$
- Feasibility problem in X:

$$(\widetilde{\mathsf{QP}}): \begin{cases} \text{find} & X \in \mathcal{S}^{\kappa} \\ \text{subject to} & \operatorname{tr}(Q_{i}X) = 0 \quad i = 1, \dots, c \\ & BXe_{1} \ge 0 \\ & \operatorname{tr}(e_{1}e_{1}^{T}X) = 1 \\ & \underline{\operatorname{rank}(X) = 1} \\ & X \succcurlyeq 0 \\ & X \succcurlyeq 0 \end{cases}$$
  
n which  $e_{1} = [1, 0, \dots, 0]^{T}$ 

(RP) infeasible  $\implies$  (QP) infeasible  $\implies$  (P) infeasible  $\Rightarrow$  (RP) to verify that  $\mathcal{P}_i$  cannot contain consistent parameters

- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- $\bullet$  computation of  $\bar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- $\bullet$  computation of  $\bar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- $\bullet$  computation of  $\bar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



### Algorithm

isto

- ullet computation of  $ar{\mathcal{P}}^*$  based on a bisection algorithm
- in each bisection step the matrix  $B(\mathcal{P}_i, \mathcal{X}^{(k)}, \mathcal{Y}^{(k)})$  is modified
- lower and upper bounds for all parameters known initially



 $\implies$  outer approximation of the set of consistent parameters

### Example: MAP-kinase-cascade (1)



#### Model alternatives

isto

- Model A: without feedback (—)  $\Rightarrow 0 = F^A(x^{(k+1)}, x^{(k)}, p)$
- Model B: with feedback (- / --)  $\Rightarrow 0 = F^{B}(x^{(k+1)}, x^{(k)}, p)$

#### Which model is correct?

### Example: MAP-Kinase-Cascade (2)

#### Artificial measurement data: (generated from Model A)



#### Set of consistent parameters:

Model A (without feedback)



#### Model B (with feedback)

 $\bar{\mathcal{P}}^* = \emptyset \Rightarrow$  Model B cannot describe the above artificial data.

(Only guaranteed for discrete time model!)



### Example: MAP-Kinase-Cascade (2)

#### Artificial measurement data: (generated from Model A)



#### Set of consistent parameters:

Model A (without feedback)



#### Model B (with feedback)

 $\bar{\mathcal{P}}^* = \emptyset \Rightarrow$  Model B cannot describe the above artificial data.

(Only guaranteed for discrete time model!)

How can the knowledge we gained by doing set-based estimation be used?

### Outline



#### Introduction

2 Parameter estimation and model falsification

#### Steady-state uncertainty

#### 4 Summary/Conclusion







key property: asymptotic behavior



# Formulation of uncertainty analysis as feasibility problem



- $\mathcal{X}_s^* =$  set of feasible steady states of an uncertain system (in general not computable analytically!)
- $\mathcal{X}_i = \text{set for which infeasibility}$ certificates can be computed
- $\mathcal{X}_s = \text{obtained outer} \\ \text{approximation of } \mathcal{X}_s^*$

### Feasibility problem

isto

- verification that a set  $\mathcal{X}_i$  cannot contain steady states
- feasibility problem (for  $\dot{x} = f(x, p)$ ):

$$(P): egin{cases} \mathsf{find} & x \in \mathcal{X}_i, \ p \in \mathcal{P} \\ \mathsf{subject to} & f(x,p) = 0 \end{cases}$$

(P) infeasible  $\iff \mathcal{X}_i$  does not contain steady states for  $p \in \mathcal{P}$ 

# Formulation of uncertainty analysis as feasibility problem $\frown$



- $\mathcal{X}_s^* =$  set of feasible steady states of an uncertain system (in general not computable analytically!)
- $\mathcal{X}_i = \text{set for which infeasibility}$ certificates can be computed
- $\mathcal{X}_s = \text{obtained outer} \\ \text{approximation of } \mathcal{X}_s^*$

### Feasibility problem

ist

- verification that a set  $\mathcal{X}_i$  cannot contain steady states
- feasibility problem (for  $\dot{x} = f(x, p)$ ):

$$(P):\begin{cases} \mathsf{find} & x \in \mathcal{X}_i, \ p \in \mathcal{P} \\ \mathsf{subject to} & f(x, p) = 0 \end{cases}$$

Computation of the set of feasible steady states employes the same methods as the computation of the set of consistent parameters!

# Example: TNF-induced anti-apoptotic signaling



- Biological relevance:
  - apoptosis
  - proliferation
  - inflammation
- Components:
  - TNF-receptors
  - NF-κB signaling pathway
- Model:
  - 24 state variables
  - 56 parameter
- Inputs:
  - TNF1
  - TNF2

#### Schematic of antiapoptotic signaling pathway

isto

### Computation of $\mathcal{X}_s$



- Parameter uncertainties of factors:  $\rho^T = (\rho_1, \dots, \rho_q)$
- Parameter set  $\mathcal{P}$  is a hyperrectangle



### Set of feasible steady states for a variation: $\rho^{T} = (2, 2, 2, 2)$



# **Biological findings**





#### Surprising results

The signal transduction process via the *TNF* receptor 1 complex is more sensitive to *TNF*2 than to *TNF*1.  $\Rightarrow$  strong crosstalk!



### Outline



#### Introduction

Parameter estimation and model falsification

3 Steady-state uncertainty





# Summary



#### Coherent framework for set-based ...

- parameter estimation [Küpfer2007, Hasenauer2010a, Rumschinski2010],
- model falsification [Anderson2009, Hasenauer2010a, Rumschinski2010], and
- steady state predictions [Waldherr2008, Hasenauer2010b, Waldherr2011a],

by employing powerful tools from convex optimization.

Alternative to: interval analysis and constraint propagation [Jaulin2001, Tucker2006, Walter2007]

#### Extensions to perform set-based ...

- experimental design [Hasenauer2010a], and
- robustness analysis [Waldherr2011b]

Toolbox: bioSDP for SDP-based analysis of dynamical systems.

### Challenges

ist?

- Discretization error  $\rightarrow$  *Torkel Glad*
- Computational complexity  $\rightarrow$  *Pelle Lundberg*
- Application to real world examples → *Pelle Lundberg*



Uncertainty analysis is a crucial task in systems biology. There are some methods available, but they are limited and not widely used!

 $\Rightarrow$  many challenging open problems

### References



[Jaulin2001]	L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied interval analysis. Springer, 2001.
[Tucker2006]	W. Tucker, Z. Kutalik, and V. Moulton. Estimating parameters for generalized mass action models using constraint propagation. <i>Math. Biosciences</i> , 208:607–620, 2006.
[Walter2007]	E. Walter, and M. Kieffer. Guaranteed nonlinear parameter estimation in knowledge-based models. J. Comp. Appl. Math., 199:277–285, 2007.
[Küpfer2007]	L. Küpfer, U. Sauer and P.A.Parrilo. Efficient classification of complete parameter regions based on semidefinite programming. <i>BMC Bioinf.</i> , 8:12, 2007.
[Waldherr2008]	S. Waldherr, R. Findeisen, and F. Allgöwer. Global sensitivity analysis of biochemical reaction networks via semidefinite programming. In <i>Proc. of the 17th IFAC World Congress, Seoul, Korea</i> , 17(1):9701–9706, 2008.
[Anderson2009]	J. Anderson, and A. Papachristodoulou. On validation and invalidation of biological models. <i>BMC Bioinf.</i> , 10:132, 2009.
[Hasenauer2010a]	J. Hasenauer, S. Waldherr, K. Wagner, and F. Allgöwer. Parameter identification, experimental design and model falsification for biological network models using semidefinite programming. <i>IET Syst. Biol.</i> , 4(2):119–130, 2010.
[Hasenauer2010b]	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. <i>J. Process Control</i> , 20(9):1076–1083, 2010.
[Rumschinski2010]	P. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, and R. Findeisen. Set-based dynamical parameter estimation and model invalidation for biochemical reaction networks. <i>BMC Syst. Biol.</i> , 4:69, 2010.
[Waldherr2011a]	S. Waldherr, J. Hasenauer, M. Doszczak, P. Scheurich, and F. Allgöwer. Global uncertainty analysis for a model of TNF-induced NF- <i>k</i> B signalling. In <i>Advances in the Theory of Control, Signals and Systems with Physical Modeling</i> , volume 407 of <i>Lecture Notes in Control and Information Sciences</i> , pages 365–377. Springer, 2011.
[Waldherr2011b]	S. Waldherr, and F. Allgöwer. Robust stability and instability of biochemical networks with parametric uncertainty. <i>Automatica</i> , 47:1139–1146, 2011.

### References



[Jaulin2001]	L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied interval analysis. Springer, 2001.
[Tucker2006]	W. Tucker, Z. Kutalik, and V. Moulton. Estimating parameters for generalized mass action models using constraint propagation. <i>Math. Biosciences</i> , 208:607–620, 2006.
[Walter2007]	E. Walter, and M. Kieffer. Guaranteed nonlinear parameter estimation in knowledge-based models. <i>J. Comp. Appl. Math.</i> , 199:277–285, 2007.
[Küpfer2007]	L. Küpfer, U. Sauer and P.A.Parrilo. Efficient classification of complete parameter regions based on semidefinite programming. <i>BMC Bioinf.</i> , 8:12, 2007.
[Waldherr2008]	S. Waldherr, R. Findeisen, and F. Allgöwer. Global sensitivity analysis of biochemical reaction networks via semidefinite programming. In <i>Proc. of the 17th IFAC World</i>
[Anc	Thanks for your attention!

Hasen

# Thanks for your attention!

- [Hasenauer2010b] 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, 20(9):1076–1083, 2010.
- [Rumschinski2010] P. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, and R. Findeisen. Set-based dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Syst. Biol.*, 4:69, 2010.
  - [Waldherr2011a] 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, volume 407 of Lecture Notes in Control and Information Sciences, pages 365–377. Springer, 2011.

[Waldherr2011b] S. Waldherr, and F. Allgöwer. Robust stability and instability of biochemical networks with parametric uncertainty. *Automatica*, 47:1139–1146, 2011.