Tutorial on System Identification of Biochemical Reaction Systems

Jan H. van Schuppen
with
Balázs Boros, Dirk De Vos, Jana Němcová

Centrum Wiskunde & Informatica (CWI), Amsterdam

2 June 2009
Lorentz Center, University of Leiden
Leiden, The Netherlands

Outline

Outline

Example of system identification

Example. Red blood cell development

- Example of system identification.
  Erythropoiesis: The development of red blood cells.
- Modeling of the functioning of the GATA-1 network.
  Genetic regulatory network.
- System identification of the formulated model: check the behaviour of the model against measurements.

Blood cells

Hematopoiesis. Blood cells

Multipotent stem cells - intermediate stages -
different types of mature blood cells

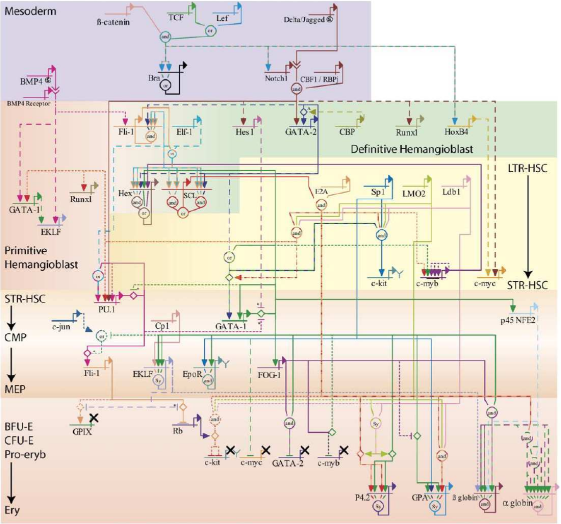
- Lineage specification.
- Leukemia is a form of impaired hemapoiesis.
- Drug-design.

Erythropoiesis. Red blood cells

Development of red blood cells.

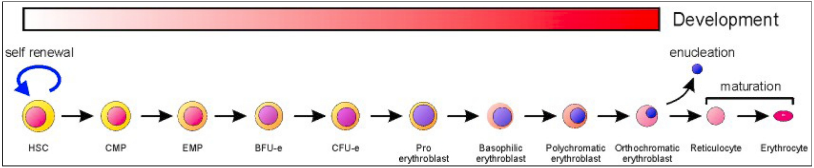
- O2/CO2 transport,
- about 45% of the human blood.

Regulatory network involved in erythropoiesis.



(From G. Swiers et al., Dev. Biol. 294, 2006)

Development process



- A central pool of genes:
  GATA-1, GATA-2, PU1, EKLF, FOG-1, β-globin.
- Erasmus University Rotterdam, Medical Center
  Research Group Frank Grosveld and Sjaak Philipsen.

GATA-1 network. Activation and inhibition

Gene	Gene's activators	Gene's inhibitors
GATA-1	GATA-1, GATA-2, PU1	(PU1+GATA-1)
PU1	PU1, GATA-1	(GATA-1+PU1)
GATA-2	GATA-2	(GATA-1+FOG-1)
FOG-1	GATA-1	
EKLF	GATA-1	(GATA-1+PU1)
β-globin	(GATA-1+FOG-1), EKLF	

GATA-1 - Reaction network

Activations and transcription

$$P_1 + D_1 \overset{k_1}{\leftrightarrow} D_1^{P_1}; \qquad D_1^{P_1} \overset{r_1}{\rightarrow} D_1^{P_1} + M_1$$

$$P_2 + D_1 \overset{k_2}{\leftrightarrow} D_1^{P_2}; \qquad D_1^{P_2} \overset{r_2}{\rightarrow} D_1^{P_2} + M_1$$

$$P_3 + D_1 \overset{k_3}{\leftrightarrow} D_1^{P_3}; \qquad D_1^{P_3} \overset{r_3}{\rightarrow} D_1^{P_2} + M_1$$

Translation

$$M_1 + \eta_1 R \overset{q_1}{\rightarrow} M_1^{\eta_1 R}; \qquad M_1^{\eta_1 R} \overset{l_1}{\rightarrow} M_1^{\eta_1 R} + \eta_1 P_1$$

Inhibitions

$$P_1 + P_2 + D_1 \overset{k_{12}}{\leftrightarrow} D_1^{P_1 P_2};$$

Degradations

$$M_1 \overset{k_{m_1}}{\rightarrow}, \quad P_1 \overset{k_{p_1}}{\rightarrow}$$

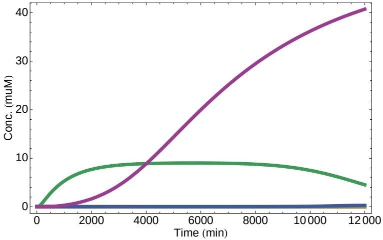
GATA-1 biochemical reaction system

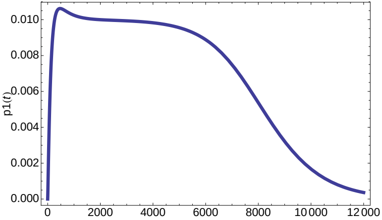
Def. GATA-1 system

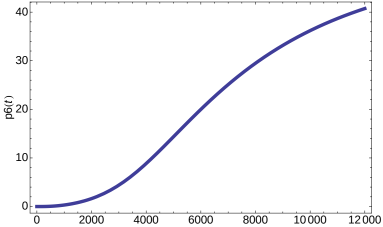
$$
\begin{aligned}
& p_1 = \text{GATA-1}, \quad m_1 = \text{its messenger RNA} \\
dm_1(t)/dt &= -k_{m_1} m_1(t) + f_{11}(x(t)), \\
dp_1(t)/dt &= -k_{p_1} p_1(t) + \frac{Q_1 l_1}{\eta_1} r_1^{\eta_1} m_1(t), \\
f_{11}(x) &= \frac{K_1 r_1 p_1(t) + K_2 r_2 p_2(t) + K_3 r_3 p_3(t)}{1 + K_1 p_1(t) + K_2 p_2(t) + K_3 p_3(t) + K_{12} p_1(t) p_2(t)}, \\
x(t) &= \begin{pmatrix} m_1(t) & p_1(t) & \ldots & m_6(t) & p_6(t) \end{pmatrix}, \\
dx(t)/dt &= f(x(t)), \; x(t_0) = x_0.
\end{aligned}
$$

Simultations of GATA-1 system

Simulation in Matlab.







Development of red blood cells

Further investigation

- Biochemical reaction system, further modeling.
- Mathematical analysis of system.
- System identification: How to obtain numerical values of the parameters from measurements?
- Evaluation of system identification results.

Outline

System identification of biochemical reaction systems

Problem (lightly). System identification

From a time series to a system with parameter values.

Comments

- Need of systems biology for systems with numerical values of parameters.
- Current experience.
- Biology versus control engineering.
- Research areas with system identification problems:
  - Control and system theory: system identification;
  - Econometrics: time series analysis; and
  - Statistics: parameter estimation.

System identification

Procedure. System identification

1. Modeling of phenomenon by a system.
2. Identifiability of a parametrized set of systems.
3. Input design, data collection, and preprocessing of the observations.
4. Approximation
(parameter estimation; Subspace Identification Algorithm).
5. Evaluation and possibly changes to the system class.

Classes of systems discussed:

- Biochemical reaction systems;
- Linear systems;
- Gaussian systems.

Outline

Modeling of biochemical reaction networks

Problem. Modeling phenomenon

Formulate a biochemical reaction system
modeling a biochemical reaction network
which is realistic but not too complex.

Note conflicting modeling objectives!
Realistic vs not too complex.

Example biochemical reaction system - 1

Modeling formalism of Martin Feinberg (1987).

Example

$$A_1 + A_2 \rightleftarrows A_3,$$

- Chemical species: $A_1,\ A_2,\ A_3$.
- Chemical complexes: $C_1 = A_1 + A_2$ and $C_2 = A_3$ and the set of complexes C.
- Relation of chemical species to complexes

$$B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example biochemical reaction system - 2

Example (Continued)

- A reaction is a relation between complexes, associated with a directed edge of a graph

$$\begin{aligned} A_1 + A_2 &\to A_3, & A_3 &\to A_1 + A_2, \\ C_1 &\to C_2, & C_2 &\to C_1. \end{aligned}$$

- Reaction network is a set of reactions $rnet = \{(2,1),(1,2)\} \subseteq C \times C$.
- Kinetics of reactions. Special case of mass action kinetics,

$$r_{2,1}(x) = K(2,1)x_1 x_2, \quad r_{1,2}(x) = K(1,2)x_3.$$

Example biochemical system - 3

Example (Continued)

Biochemical reaction system of example with mass action kinetics

$$dx(t)/dt = \sum_{(i,j) \in rnet} [B(.,i) - B(.,j)] K(i,j) \prod_{s=1}^{n} x_s^{B(s,j)}.$$

$$dx(t)/dt = \begin{pmatrix} K(1,2)x_3(t) - K(2,1)x_1(t)x_2(t) \\ K(1,2)x_3(t) - K(2,1)x_1(t)x_2(t) \\ K(2,1)x_1(t)x_2(t) - K(1,2)x_3(t) \end{pmatrix}, \quad x(t_0) = x_0.$$

Biochemical reaction system - 1

Def. Closed biochemical reaction system (BRS)

$$dx(t)/dt = \sum_{(i,j)\in rnet} [B(.,i) - B(.,j)] r_{i,j}(x(t)) u_{(i,j)}(t),$$

x_s concentration or mass of chemical species, $s \in \mathbb{Z}_n$;

$C(i)$ chemical complex $i \in \mathbb{Z}_m$;

$rnet \subseteq \mathbb{Z}_{n_c} \times \mathbb{Z}_{n_c}$, reaction network,

$(i,j) \in \text{reactionnet}$ if $C(j) \rightarrow C(i)$;

$$B(s,i) = \begin{cases} n_s, & \text{if } x_s^{n_s} \text{ belongs to Complex } C(i); \\ 0, & \text{else;} \end{cases}$$

$B_s(.,q) = B(.,i) - B(.,j),\ q = (i,j);$

stoichiometric matrix,

$u_{i,j} : T \rightarrow \mathbb{R}_+$, enzyme concentration;

Biochemical reaction system 2

Def. Kinetics

- Mass action kinetics.

$$r_{i,j} : \mathbb{R}_+^n \to \mathbb{R}_+ \quad \text{reaction rate of reaction } (i,j) \in rnet;$$

$$r_{i,j}(x) = K(i,j) \prod_{s=1}^{n} x_s^{B(s,j)},$$

$$K(i,j) \in \mathbb{R}_+ \quad \text{value of kinetic constant.}$$

- Michaelis-Menten kinetics (1913).
  Due to presence of fast and slow reactions.
  Example

$$r(x) = \frac{k_1 x}{k_2 + x}.$$

Open biochemical reaction systems

Def. Open biochemical reaction system

$$
\begin{aligned}
dx(t)/dt &= \sum_{(i,j)\in rnet} [B(.,i) - B(.,j)] r_{i,j}(x(t), x_{ex}(t)) u_{(i,j)}(t), \\
& x(t_0) = x_0, \\
y(t) &= h(x(t), x_{ex}(t)), \\
& x, y \ \text{ state and output process.}
\end{aligned}
$$

Notation Σ_{BRS}. System with partial observations.

Outputs can be either measurements of combinations of states or outflows of the system.

Linear systems

Def. Linear system

$$\begin{aligned} dx(t)/dt &= Ax(t) + Bu(t),\ x(t_0) = x_0, \\ y(t) &= Cx(t) + Du(t), \\ & T = [t_0, \infty), X = \mathbb{R}^n, U = \mathbb{R}^m, Y = \mathbb{R}^k,\ n, m, k \in \mathbb{Z}_+, \\ & x : T \to X,\ \text{state trajectory}, \\ & u : T \to U,\ \text{input trajectory}, \\ & y : T \to Y,\ \text{output trajectory}. \\ & \Sigma_{LIN}. \end{aligned}$$

Formally, a time-invariant finite-dimensional linear system.

Gaussian systems

Def. Gaussian system

$$\begin{aligned} x(t+1) &= Ax(t) + Mv(t),\ x(t_0) = x_0, \\ y(t) &= Cx(t) + Nv(t), \\ & T = [t_0, \infty), \quad n, m, k \in \mathbb{Z}_+, \\ & X = \mathbb{R}^n,\ U = \mathbb{R}^m,\ Y = \mathbb{R}^k,\ V = \mathbb{R}^{m_v}, \\ & v : \Omega \times T \to V,\ \text{Gaussian white noise}, \\ & \{v(t), t \in T\} \text{ independent collection and} \\ & v(t) \in G(0, Q_v), \quad \forall t \in T. \\ & \Sigma_G. \end{aligned}$$

Formally, a time-invariant finite-dimensional Gaussian system.

The state process x and the output process y are jointly Gaussian processes.

Outline

Example 2D-MM. Identifiability problem

Consider the biochemical reaction system

$$\begin{aligned} dx(t)/dt &= f(x(t), u(t); p) = \begin{pmatrix} -\frac{p_1 x_1(t)}{p_2 + x_1(t)} + u(t) \\ \frac{p_1 x_1(t)}{p_2 + x_1(t)} - \frac{p_3 x_2(t)}{p_4 + x_2(t)}) \end{pmatrix}, x(0) = (1,1), \\ y(t) &= h(x(t); p) = \begin{pmatrix} \frac{p_1 x_1(t)}{p_2 + x_1(t)} \\ \frac{p_3 x_2(t)}{p_4 + x_2(t)} \end{pmatrix}, \\ X &= \mathbb{R}^2,\ U = \mathbb{R},\ Y = \mathbb{R}^2; \\ P &= \{(p_1, p_2, p_3, p_4) \in \mathbb{R}^4_+\}. \end{aligned}$$

Problem. Identifiability for example

(a) Can one from the observed time series of inputs and outputs in principle determine the parameter values uniquely?

(b) Can one from the input-output map, $u_{[t_0,t)} \mapsto y(t),\ \forall t \in T$, determine the parameter values uniquely?

Structural identifiability of a rational system

Problem. Structural identifiability

Consider a structured rational system.

$$\begin{aligned} dx(t)/dt &= f(x(t), u(t), p), \quad x(t_0) = x_0, \\ y(t) &= h(x(t), p), \quad f,\ h \text{ structured rational maps}, \\ P & \quad \text{irreducible variety in } \mathbb{R}^{n_p}. \end{aligned}$$

Can the parameter vector $p \in P$
be uniquely determined from the input-output map?

Outline

1. Realization theory.
2. Parametrization of a class of systems.
3. Structural identifiability of a parametrized class of systems.

Introduction to realization

Def. Rational system on $\mathbb{R}^n$

$$\begin{aligned} dx(t)/dt &= f(x(t),u(t)),\ x(t_0)=x_0, \\ y(t) &= h(x(t)),\quad f,h \text{ rational maps}, \\ & n,m,k \in \mathbb{Z}_+,\quad X=\mathbb{R}^n,\ U=\mathbb{R}^m,\ Y=\mathbb{R}^k. \\ & \Sigma_{RAT}. \end{aligned}$$

Transform the rational system into an input-output map depending on initial state,

$$\{u_{[t_0,t)} \rightarrow y(t),\ \forall t \in T\}.$$

Realization problem is:
Can one construct a rational system from an input-output map?

Rational system

Def. Rational system on a variety

$$
\begin{aligned}
\sigma &= (X, U, Y, f, h, x_0) \in \Sigma_{RAT}, \\
X & \quad \text{irreducible real affine variety in } \mathbb{R}^n\text{, the state set,} \\
X &= \{x \in \mathbb{R}^n | p_i(x) = 0 \text{ for finitely many } p_i \in \mathbb{R}[X_1, \ldots, X_n]\}, \\
U &= \mathbb{R}^m, \text{ input set, } Y = \mathbb{R}^r, \\
f &= \{f_\alpha, \alpha \in U\}, \text{ family of rational vector fields on } X, \\
& \quad h : X \to Y = \mathbb{R}^r, \text{ rational map, output map,} \\
& \quad x_0 \in X, \text{ initial state.}
\end{aligned}
$$

Real algebraic geometry.

Realization problem

Problem. Realization of a rational system

Consider an input-output map.

- Does there exist a rational system of which the input-output map equals the considered input-output map?
  Call such a system a rational realization.
- Is the rational system a minimal realization?
- What is the relation between minimal realizations?

Literature

- Realization of polynomial systems:
  E.D. Sontag (1979), Z. Bartosiewicz (1987),
  Yuan Wang and E.D. Sontag (1992).
- Realization of rational systems:
  Yuan Wang and E.D. Sontag (1992), affine in input only.

Existence of a realization

Theorem. Existence of a rational realization

An analytic map $p : U_{pc} \to \mathbb{R}^r$ has a rational realization if and only if the field $Q_{obs}(p)$ is finitely generated.

(JN,JHvS,2008a; Theorem 4.7).

Def. Field and field of input-output map

$$\begin{aligned}
\mathcal{A}(U_{pc} \to \mathbb{R}) &= \{\phi : U_{pc} \to \mathbb{R} | \text{analytic}\}, \text{ which is an integral domain,} \\
& \quad \text{analytic at the switching times of inputs,} \\
\mathcal{A}_{obs}(p) & \quad \text{smallest subalgebra of } \mathcal{A}(U_{pc} \to \mathbb{R}), \\
& \quad \text{containing } p_i,\ i \in \mathbb{Z}_r, \\
& \quad \text{closed wrt derivations } \{D_\alpha,\ \alpha \in U\}; \\
\mathcal{Q}_{obs}(p) & \quad \text{field of quotients of } \mathcal{A}_{obs}(p).
\end{aligned}$$

Observability of a rational system

Def. Algebraic observability

$$\Sigma = (X, f = \{f_\alpha, \alpha \in U\}, h, x_0), \text{ rational system},$$

Q field of rational functions on X;

$\mathcal{A}_{obs}(\Sigma)$ observation algebra defined as the smallest subalgebra of Q containing h_i, $i \in \mathbb{Z}_r$, and closed wrt derivations of the rational vector fields $\{f_\alpha,\ \alpha \in U\}$.

$\mathcal{Q}_{obs}(\Sigma)$ observation field of the system field of quotients of $\mathcal{A}_{obs}(\Sigma)$. If,

$Q = \mathcal{Q}_{obs}(\Sigma)$ then system is called algebraically observable.

Controllability of a rational system

Def. Algebraic controllability

Rational system Σ is called
algebraically controllable from the initial state $x_0 \in X$
if X is the smallest variety containing the subset

$$\mathcal{R}(x_0) \quad = \quad \{x(t_u; x_0, u) \in X | u \in \mathcal{U}_{pc}(x_0),\ u : [0, t_u] \to U\},$$

Rational system called algebraically controllable
if it is algebraically controllable from all initial states
(in terms of Zariski topology).

Def. Canonical system

A rational system is called canonical
if it is both algebraically controllable and algebraically observable.

Real algebraic geometry

Def. Transcedence degree

A, Q algebra of polynomials respectively
field of rational functions on X,
$\{\phi_1, \ldots, \phi_s\} \in A(Q)$ algebraically independent over $\mathbb{R}$,
if there does not exist $p \in \mathbb{R}[X]$ such that,

$$0 = p(\phi_1, \ldots, \phi_s).$$

$F \subseteq A$ or $F \subseteq Q$, field.

$\mathrm{trdeg}(F) =$ largest number of elements of F which are algebraically independent over $\mathbb{R}$,
transcedence degree of F.
$\{\phi_1, \ldots, \phi_{\mathrm{trdeg}(F)}\}$ transcedence basis of F, if ...

Algorithm (J. Müller-Quade, R. Steinwandt, J. Symbolic Computation 30 (2000), 469–490.

Minimal realization

Def. Dimension of a rational system

$$\begin{aligned} & X,\ Q; \quad \text{Define dimension of variety as,} \\ \dim(X) &= \mathrm{trdeg}(Q). \\ & \text{Rational system is called minimal realization} \\ & \text{of input-output map } p \text{ if} \\ \dim(X) &= \mathrm{trdeg}(Q_{obs}(p)). \end{aligned}$$

Theorem. Characterization of minimality

Consider a rational realization $\Sigma = (X, f, h, x_0)$
of an analytic input-output map and assume that:
the elements of $Q \backslash \mathcal{Q}_{obs}(\Sigma)$ are not algebraic over $\mathcal{Q}_{obs}(\Sigma)$.

Then the system is a minimal realization of its input-output map
if and only if the system is canonical.

(JN,JHvS,2008b,Th. 3)

Parametrizations

Distinguish:

- Structured system. Class of systems structured by physical laws.
- System parametrized by a canonical form.

Problem. Choice of canonical form

Consider a class of systems.
Define the i/o-equivalence relation on the class of systems such that two systems are equivalent if they have the same input-output map.
Determine a canonical form for this i/o-equivalence relation.

Solved for linear systems. Open for rational systems.

Example. Structured system

Example 2D-MM

Consider the structured biochemical reaction system.

$$
\begin{aligned}
& (X, U, f, h, x_0) \in \Sigma_{BRS,s}(P, f_{par}),\\
P &= \{p_1, p_2, p_3, p_4 \in \mathbb{R}\},\\
X &= \mathbb{R}^2, \quad U = \mathbb{R}, \quad x(0) = (1, 1),\\
f_\alpha^p &= \left(-\frac{p_1 x_1}{p_2 + x_1} + \alpha\right)\frac{\partial}{\partial x_1} + \left(\frac{p_1 x_1}{p_2 + x_1} - \frac{p_3 x_2}{p_4 + x_2}\right)\frac{\partial}{\partial x_2}, \quad \alpha \in U,\\
h_\alpha^p(x) &= \left(\frac{p_1 x_1}{p_2 + x_1}, \frac{p_3 x_2}{p_4 + x_2}\right).
\end{aligned}
$$

Structured linear system

Def. Structured linear system

$$
\begin{aligned}
dx(t)/dt &= A(p)x(t) + B(p)u(t),\ x(t_0) = x_0(p), \\
y(t) &= C(p)x(t) + D(p)u(t), \\
& P \subseteq \mathbb{R}^{n_p},\ A : P \rightarrow \mathbb{R}^{n \times n},\ B : P \rightarrow \mathbb{R}^{n \times m}, \\
& C : P \rightarrow \mathbb{R}^{k \times n},\ D : P \rightarrow \mathbb{R}^{k \times m},\ x_0 : P \rightarrow \mathbb{R}^{n}, \\
SLSP(n,m,k) &= \{(A(p), B(p), C(p), D(p), x_0(p)) \\
& \in LSP(n,m,k) | p \in P\}, \\
& f_{par} : P \rightarrow SLSP(n,m,k) \text{ parametrization map}, \\
f_{par}(p) &= (A(p), B(p), C(p), D(p), x_0(p)), \\
& (P, f_{par}) \text{ parametrization of the set.} \\
& SLSP(n,m,k).\ \Sigma_{LIN,s}(P, f_{par}).
\end{aligned}
$$

Example structured linear system

Nitrate model (J.M. van den Hof, Ph.D. thesis, 1996, Sec. 8.2).

$$\begin{aligned}
dx(t)/dt &= A(p)x(t) + Bu(t),\ x(t_0) = x_0(p), \\
y(t) &= C(p)x(t), \\
A(p) &= \begin{pmatrix} -\alpha & 0 & \delta & 0 \\ \alpha & -\beta - h_1 & 0 & 0 \\ 0 & \beta & -\gamma - \delta & 0 \\ 0 & 0 & \gamma & -\delta \end{pmatrix},\ B(p) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \\
C(p) &= \begin{pmatrix} 0 & h_2 & 0 & 0 \\ 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & \epsilon \end{pmatrix},\ x_0(p) = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{pmatrix}, \\
P &= \{(\alpha, \beta, \gamma, \delta, \epsilon, \phi_1, \phi_2, \phi_3, \phi_4) \in \mathbb{R}^9_+ \subset \mathbb{R}^9\}, \\
f_{par}(p) &= (A(p), B(p), C(p), x_0(p)).\ \Sigma_{LIN,s}(P, f_{par}).
\end{aligned}$$

Observable canonical form of linear systems

Def. Observable canonical form of a linear system

Case single-input-single-output linear system ($m = 1$, $k = 1$).

$$
\begin{aligned}
dx(t)/dt &= Ax(t) + Bu(t),\ x(t_0) = x_0, \\
y(t) &= Cx(t) + Du(t),
\end{aligned}
$$

in observable canonical form if

$$
\begin{aligned}
A &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_n \end{pmatrix},\ B = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \\
C &= \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \end{pmatrix},\ D = \delta; \\
P &= \mathbb{R}^{2n+1},\ p = (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \delta), \\
& (p, f_{par}),\ \ p \mapsto f_{par}(p) = (A(p), B(p), C(p), D(p), x_0(p)).
\end{aligned}
$$

Case $m > 1$ and $k > 1$ in the literature.

Identiability concept

Def. Identifability from the input-output map

Consider

$$\begin{array}{ll} \Sigma(P) & \text{class of parametrized systems,} \\ & (P, f_{par}),\quad f_{par} : P \to \Sigma(P),\ \text{surjective,} \\ & g : \Sigma(P) \to I/O_{maps}. \end{array}$$

Call this parametrized system
identifiable from the input-output map if

$$g \circ f_{par} : P \to I/O_{maps},\quad p \mapsto f_{par}(p) \mapsto g(f_{par}(p))$$

is injective.

Parametrization (P, f_{par}) is called continuous
if P is a connected set and the map f_{par} is continuous on P.

Structural identifiability

Def. Algebraic set and genericity

An algebraic subset of $\mathbb{R}^n$ is a subset determined by a finite set of polynomials.
A property depending on a parameter set holds generically if it holds for all parameters except for those on an algebraic subset.

Example. The determinant as a function of the entries of a matrix is generically nonzero.

Def. Structural identifiability

A structured class of systems is said to be
structurally identifiable
if it is identifiable on the parameter set P except for on an algebraic subset.

Identifiability problem

Problem. Check structural identifiability

Determine whether a parametrized class of systems is structurally identifiable.

Theorem. Identifiability of linear systems in the observable canonical form

Consider $\Sigma_{LIN}(P, f_{par})$ single-input-single-output,
with the observable canonical form.
This parametrized system is structurally identifiable
from the input-output map and it is a continuous parametrization.

Remark

The class of multi-input-multi-output linear systems does not admit a continuous canonical form. (M. Hazewinkel, R.E. Kalman, 1976).

Structural identifiability rational systems

Theorem. Structural identifiability of rational systems

$\Sigma_{RAT,s}(P, f_{par})$	$\mathcal{P} : P \rightarrow \Sigma P(P, f_{par})$, parametrization.
	Assume $\Sigma_{RAT,s}(P, f_{par})$ is structurally canonical.
	Equivalence:
(a)	parametrization $\mathcal{P}$ is structurally identifiable,
(b)	$\forall p, q \in P$ except for an algebraic subset,
	$\Sigma_{RAT}(p) \sim_{birat.eq} \Sigma_{RAT}(q)$,
	the birational map relating $\Sigma(p)$, $\Sigma(q)$,
	equals the identity and $p = q$.

J. Němcová, 2008c, Th. 4.8.

Structural identifiability

Procedure. Structural identifiability of rational systems

1. Check structural algebraic controllability and structural algebraic observability.
2. Check whether

$$
\begin{aligned}
& p, q \in P, \quad s : X_p \to X_q \text{ birational map,} \\
\Rightarrow \quad & p = q, \quad s = i.
\end{aligned}
$$

Example. Identifiability of rational system 1

Example 2D-MM.

$$
\begin{aligned}
P &= \{p_1, p_2, p_3, p_4 \in \mathbb{R}^4\}, \\
X &= \mathbb{R}^2, \quad x(0) = (1,1), \\
f_\alpha^p &= (-\frac{p_1 x_1}{p_2 + x_1} + \alpha)\frac{\partial}{\partial x_1} + (\frac{p_1 x_1}{p_2 + x_1} - \frac{p_3 x_2}{p_4 + x_2})\frac{\partial}{\partial x_2}, \quad \alpha \in U, \\
h_\alpha^p(x) &= (\frac{p_1 x_1}{p_2 + x_1}, \frac{p_3 x_2}{p_4 + x_2}).
\end{aligned}
$$

(a) The system is structurally algebraically observable, structurally algebraically controllable, hence structurally canonical, and thus a minimal rational realization.

(b) The parametrization is structurally identifiable.

(J. Němcová, 2008, Ex. 5.1).

Example. Identifiability rational system 2

Biochemical reaction system
(Margaria etal., Math. Biosc. 174 (2001), 1-26).

$$\begin{aligned}
\Sigma(p) &= (X^p, f^p, h^p, x_0^p,) \in \Sigma(P), \quad P = \{p_1, p_2, p_3, p_4 \in \mathbb{R}^4\}, \\
X^p &= \mathbb{R}^2, \quad x(0) = (1,1), \\
f^p &= (-p_1 x_1 + p_2 x_2)\frac{\partial}{\partial x_1} + (p_1 x_1 - p_2 x_2 - \frac{p_3 x_2}{p_4 + x_2})\frac{\partial}{\partial x_2}, \\
h^p(x) &= x_1, \quad x_0^p = (a, 0) \in X^p.
\end{aligned}$$

(a) The system is structurally algebraically observable, structurally algebraically controllable, hence structurally canonical, and thus a minimal rational realization outside the algebraic subset $\{p \in P | p_2 p_3 = 0\}$.

(b) The parametrization is structurally identifiable.

(J. Němcová, 2008, Ex. 5.2).

Example. Peptide chain elongation - 1

Biochemical reaction system for peptide chain elongation
(A. Heyd, D.A. Drew, Bull.Math.Bio. 65 (2003), 1095–1109.)
(E. Klipp etal, 2005; Subsec. 8.3.3).

$$
\begin{aligned}
dB(t)/dt &= -k_1 A_1(t)B(t) + k_{-1}C(t) + k_r G(t) + k_7 F(t), \\
dC(t)/dt &= k_1 A_i(t)B(t) - k_{-1}C(t) - k_2 C(t) + k_{-2}D(t), \\
dD(t)/dt &= k_2 C(t) - k_{-2}D(t) - k_3 D(t), \\
dE(t)/dt &= k_3 D(t) - k_4 E(t), \\
dF(t)/dt &= k_4 E(t) - k_5 F(t) - k_7 F(t), \\
dG(t)/dt &= k_5 F(t) - k_r G(t).
\end{aligned}
$$

Example. Peptide chain elongation - 2

B	ribosome
C	initial binding
D	condon recognition
E	GTPase activation
F	GTP hydrolysis
G	EF-Tu released peptide transfer
A1	correct aa-tRNA
A2	wrong aa-tRNA

Chemical species A1 is regarded as input.
The conclusion is that the parametrization of this structured biochemical reaction system is structurally identifiable from the input-output map except for an algebraic subset of the parameter set. (J. Němcová, 2008, Ex. 5.3).

Symbolic computations

Problem. Computational differential algebra

Algorithms needed for symbolic computation of:
structural algebraic controllability, structural algebraic observability, and structural identifiability.

- J.F. Ritt, Differential algebra (1950).
- Computational differential algebra for control and system theory of polynomial systems:
  Torkel Glad, Michel Fliess, co-workers; Maria Pia Saccomani and co-workers.
- Jana Němcová has procedures for algebraic observability and for algebraic controllability.

Needed is a long term effort to develop symbolic computations.

Identifiability from input-output time series

Def. Identifiability from an input-output time series

A class of parametrized systems $\Sigma(P)$ and a particular input-output time series, $\{\overline{y}(s),\ \overline{u}(s),\ s \in [t_0, t_1]\}$.
System called structurally identifiable from the input-output time series if

$p \mapsto_{\{\overline{u}(s), s \in T\}} \{\overline{y}(s), s \in T\},$ is injective except for on an algebraic subset.

Remarks

- Problem solved for linear systems and for Gaussian systems.
- Problem open for rational systems.

Idenfiability from input-output time series

Theorem

(J.M. van den Hof, IEEE.TAC 43 (1997), 800 – 818). If

- the system class consists of discrete-time structured linear systems $\Sigma_{LIN,s}(P, f_{par})$;
- the system is structurally weakly reachable and structurally observable;
- the system is structurally identifiable from the input-output map;
- the rank of a particular matrix with inputs and outputs is equal to a particular rank;

then the parameter vector of the structured system can be uniquely determined from the input-output time series except on algebraic subset.

- Case Gaussian systems, see reference (G. Goodwin, Payne, 1977).
- Case bilinear systems, see (E.D. Sontag, etal., IEEE.TAC 54 (2009), 195–207).

Outline

Experiment design

Procedure. Experiment design and preprocessing

Consider a system identification problem.

1. Design an input signal for the system;
2. collect the measurements of inputs and outputs, called the observed input-output time series;
3. and preprocess the observed time series.

Comments

- Condition of structural identifiability from input-output time series should be met.
- The preprocessing of the time series often consists of detrending and removing outliers.

Input design for biochemical reaction systems

Remarks. Input design

- Case of rational biochemical reaction systems under investigation.
- In practice one often supplies a piecewise-constant input signal.
- Conjecture: piecewise-constant inputs are insufficient for identifiability based on an input-output time series. (Bilinear systems E.D. Sontag etal, (2009)).
- Which class of inputs to use?
- Experimental design (C. Kreuz, J. Timmer, FEBS J. 276 (2009), 923–942).
- Two or more time scales present. Selection of the time scale. Sampling.

Outline

Approximation problem

Problem. Approximation

Consider a class of systems and an observed time series. Determine a system in the class such that the following two time series are as close together as possible according to an approximation criterion between:

- the observed input-output time series; and
- the input-output time series associated with the estimated parameter vector.

Approaches to approximation problem

1. Optimization approach.
2. Algebraic system-theoretic approach.

Approximation. Optimization approach

Example. Optimization approach to approximation problem

$$
\begin{aligned}
x(t+1) &= A(p)x(t) + B(p)u(t),\ x(t_0) = x_0(p),\\
y(t) &= C(p)x(t) + D(p)u(t),\\
&\quad \{\overline{u}(t), \overline{y}(t), t \in T = [t_0, t_1]\},\ \text{observations},\\
&\quad \text{for estimate } \hat{p} \in P \ \text{ compute with observer}\\
\hat{x}(t+1;\hat{p}) &= A(\hat{p})\hat{x}(t;\hat{p}) + B(\hat{p})\overline{u}(t) +\\
&\quad + K[\overline{y}(t) - (C(\hat{p})\hat{x}(t;\hat{p}) + D(\hat{p})\overline{u}(t))],\ \ \hat{x}(0,\hat{p}) = x_0(\hat{p}),\\
e_y(t;\hat{p}) &= \overline{y}(t) - [C(\hat{p})\hat{x}(t;\hat{p}) + D(\hat{p})\overline{u}(t)],\\
J(\hat{p}) &= \frac{1}{2(t_1 - t_0 + 1)} \left(\sum_{t=t_0}^{t_1} e_y(t;\hat{p})^T e_y(t;\hat{p}) \right)^{1/2}.
\end{aligned}
$$

Solve the optimization problem $\inf_{\hat{p} \in P} J(\hat{p})$.

Optimization approach to approximation - 1

- System identification is at its basis an approximation problem.
- Choice of approximation criterion depends in general on the class of systems considered.
- For deterministic systems, the l_2 norm is often used.
  For stochastic systems, the likelihood function is often used.
- For continuous-time system one uses sampled data

$$\{\bar{y}(t_0), \bar{y}(t_0 + \Delta t), \ldots, \bar{y}(t_0 + k\Delta t)\}.$$

- The optimization approach with partial observations requires the use of an observer.

Optimization approach to approximation - 2

Conclusion. Optimization approach

The optimization approach almost never works well!!!

Reasons:

- The optimization criterion is a nonconvex function of the parameter vector $\hat{p} \in P$ in general.
- Any optimization algorithm is likely to stop after a computation in a local minimum. There can be very many local minima.

Approximation - System-theoretic approach

Introduction. Subspace identification algorithm

- Solution to the approximation problem of a discrete-time Gaussian system.
- Based on stochastic system theory.
- Developed by H. Akaike (1976), W. Larimore (1980), P. van Overschee, B. de Moor (1996), and many others.
- Used in Matlab System Identification Toolbox (Lennart Ljung) also for initialization of the maximum likelihood method.
- Below presentation of Subspace Identification Algorithm for Gaussian systems.

Matrix approximation 1

Problem. Matrix approximation

Consider $H_1 \in \mathbb{R}^{k\times m},\ n_1 = \mathrm{rank}(H_1)$.

Solve $\inf_{H\in\mathbb{R}^{k\times m}} \|H_1 - H\|_2,\ \ \mathrm{rank}(H) = n_2 < n_1$.

$\|H_1\|_2 = \sup_{x\neq 0} \|H_1 x\|_2/\|x\|_2$.

Matrix approximation 2

Algorithm. L_2 matrix approximation (G,vL, 1983, Cor. 2.3.3)

1. Construct singular value decomposition

$$H_1 = U_1 D_1 V_1^T,\ U_1 \in \mathbb{R}^{k\times n_1},\ V_1 \in \mathbb{R}^{n_1\times m},\ D_1 \in \mathbb{R}^{n_1\times n_1},\ \text{diagonal.}$$

2. Truncate to requested size

$$D_1 = \begin{pmatrix} D_{11} & 0 \\ 0 & D_{12} \end{pmatrix},\quad D_{11} \in \mathbb{R}^{n_2\times n_2}.$$

3. Compute solution,

$$H_2 = U_1 \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix} V_1^T = \sum_{i=1}^{n_2} D_{11,i} u_i v_i^T,$$

$$D_{1,n_2+1} = \|H_1 - H_2\|_2 = \inf_{H\in\mathbb{R}^{k\times m},\text{rank}} \{H_1 - H\|_2.$$

Gaussian measure approximation

Problem. Approximate a Gaussian measure

Consider two jointly finite-dimensional Gaussian random variables,

$$(y_1, y_2) \in G(0, Q_1), \quad y_i : \Omega \to \mathbb{R}^{k_i}, \; i = 1, 2.$$

Solve,

$$\inf_{G(0,Q_2)} D(G(0,Q_1) \| G(0,Q_2)), \quad \mathrm{rank}(Q_{2,y_1,y_2}) = n_2 < n_1,$$

$$D(P_1 \| P_2) = E_2[\frac{r_1}{r_2} \ln(\frac{r_1}{r_2}) I_{\{r_2 > 0\}}],$$

if $P_1 \ll P_0, \; P_2 \ll P_0, \; r_1 = dP_1/dP_0, \; r_2 = dP_2/dP_0;$

divergence of probability measures $P_1, \; P_2$.

(A.A. Stoorvogel, JHvS, 1996).

Algorithm Gaussian measure approximation

Algorithm. Approximation finite-dimensional Gaussian measure

Consider

$$G(0, Q_1), \; Q_1 \in \mathbb{R}^{(k_1+k_2)\times(k_1+k_2)}.$$

1. Transform to canonical variable decomposition.

$$\begin{aligned} & (y_1, y_2) \mapsto \; (S_1 y_1, S_2 y_2) \in G(0, Q_{1c}), \\ Q_{1c} &= \begin{pmatrix} I & L_1 \\ L_1^T & I \end{pmatrix}, \; L_1 = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{k_1 \times k_2}, \\ D_1 &= \mathrm{Diag}(q_1, q_2, \ldots, q_{n_{11}}), \; 1 \geq q_1 \geq q_2 \geq \ldots \geq q_{n_{11}} > 0. \end{aligned}$$

Algorithm Gaussian measure approximation

Algorithm. Approximation finite-dimensional Gaussian measure

2 Truncate.

$$\begin{aligned} D_2 &= = \mathrm{Diag}(q_1, q_2, \ldots, q_{n_2}) \in \mathbb{R}^{n_2 \times n_2}, \\ L_2 &= \begin{pmatrix} D_2 & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{k_1 \times k_2}, \quad Q_2 = \begin{pmatrix} I & L_2 \\ L_2^T & I \end{pmatrix}. \end{aligned}$$

3 Compute solution.

$$G(0, Q_2), \quad y_{2,tr} = \begin{pmatrix} I_{n_2} & 0 \end{pmatrix} S_2 y_2.$$

Theorem. Optimality

The solution proceduced by the above algorithm
is the optimal solution to the above formulated problem.

(A.A. Stoorvogel, JHvS, Systems & Control Letters 35 (1998), 207–218).

Subspace identification algorithm 1

Algorithm. Subspace identification

Consider

$$\{\overline{y}(t),\ t \in T = \{t_0, \ldots, t_1\}\}, \ \text{time series.}$$

1. Estimate covariance function.

$$\hat{m} = \sum_{s=t_0}^{t_1} \overline{y}(s)/(t_1 - t_0 + 1),$$

$$\hat{W}(t) = \sum_{s=t_0}^{t_1 - t} (\overline{y}(s+t) - \hat{m})(\overline{y}(s) - \hat{m})^T/(t_1 - t - t_0 + 1).$$

2. Consider finite future and finite past of output process.

$$y^+(t) = (y(t_1), \ldots, y(t)),$$

$$y^-(t) = (y(t-1), \ldots, y(t_0)), \ (y^+, y^-) \in G(0, \hat{Q}(\hat{W})).$$

Subspace identification algorithm 2

Algorithm. Subspace identification

3 Determine low rank approximation of projection of future on past.

$$\hat{x}(t) = \begin{pmatrix} I_{n_2} & 0 \end{pmatrix} S_2 y^-(t), \quad n_2 < n_1.$$

4 Determine dynamics by regression.

$$\begin{pmatrix} \hat{x}(t+1) \\ \hat{y}(t) \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \hat{x}(t) + \begin{pmatrix} I_{n_2} & 0 \\ 0 & I_k \end{pmatrix} \hat{v}(t),$$

$$\hat{v}(t) \in G(0, Q_v).$$

5 Solution Gaussian system is,

$$\begin{aligned} x(t+1) &= Ax(t) + \begin{pmatrix} I_{n_2} & 0 \end{pmatrix} v(t), \\ y(t) &= Cx(t) + \begin{pmatrix} 0 & I_k \end{pmatrix} v(t). \end{aligned}$$

Approximation problem of biochemical reaction systems

Problem. Approximation of BRS. Existing approaches:

- Optimization approach with L_2-norm.
- Stochastic approach with likelihood function.
  (T.G. Doeswijk, Karel J. Keesman, Water Research 43 (2009), 107–116).
- Local approach.
  1. Linearize the system at a particular state.
  2. System identification of the linearized system with the Subspace Identification Algorithm.
  3. Estimation of the parameters of the biochemical reaction system from identified linear system.

  Does not always work!
- Approximation by piecewise-affine system and system identification of affine system per subset. Tested?

Research in progress (CWI): Algebraic system-theoretic algorithm.

Outline

Evaluation of determined system

Does the system identification procedure fit any system to the observed time series?

Procedure. Evaluation

1. Evaluate the estimated system in regard to the approximation of the observed input-output time series.
2. If the approximation is not so good then adjust the class of systems and repeat the procedure.

Gaussian system evaluation

Problem. Gaussian system evaluation

Estimate the dimension of the state-space.

Approaches in the literature:

- Estimation based on figure of state-space dimension against the value of the approximation criterion.
- Akaike Information Criterion (AIC).
  Proven to be incorrect for Gaussian systems,
  (R. Shibata, Biometrika 63 (1976), 117-126).
- Minimum Description Length Principle.
  (Jorma Rissanen (1985)).
  (Ted Hannan, Ann. Statist. 8 (1980), 1071-1081).

Outline

Further research - System identification

Problems of control and system theory

- Approximation algorithms and theory.
- Computational differential algebra for identification.
- Realization and identification of rational positive systems.

Problems. Mathematics

- Computational differential algebra.
- Real algebraic geometry.

Software

Software

- Matlab System Identification Toolbox
  (Lennart Ljung (U. Linköping)).
- Maple and Mathematica.
- DAISY. Identifiability of polynomial and rational systems.
  (Maria Pia Saccomani (U. Padova, Italy)).
- Mathematica toolbox for signal processing and system identification.
  (Jonas Sjöberg (Chalmers U.)).

References

Books

1. L. Ljung, System identification: Theory for the user, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.
2. T. Söderström, P. Stoica, System identification, Prentice Hall, New York, 1989.
3. Hugues Garnier and Liuping Wang, Identification of continuous-time models from sampled data, Springer, Berlin, 2008.

Research groups in system identification of biochemical reaction systems and bioreactors.

- Frank Doyle (U. of California at Santa Barbara, CA, USA).
- Frank Allgöwer (U. Stuttgart, Germany).

Acknowledgements

Colleagues

- Balázs Boros, Dirk De Vos, Jana Němcová (CWI).
- Barbara Bakker, Klaas Krab, Hans Westerhoff (VU.FALW.CellPh.).
- Frank Grosveld, Sjaak Philipsen (EUR.MC).

Financial support in part

- NWO.EW.Wisk.
- NBIC (Nijmegen, The Netherlands).
- CWI (Amsterdam, The Netherlands).

The End!