Control and Cybernetics

vol. **42** (2013) No. 1

Biochemical reaction systems – system theory and decomposition*

by

Jana Němcová[1] and Jan H. van Schuppen[2]

[1]Department of Mathematics, Institute of Chemical Technology
Prague, Techniká 5, 166 28 Prague 6, Czech Republic
Jana.Nemcova@vscht.cz
[2]CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
jan.h.van.schuppen@euronet.nl

Abstract: Biochemical reaction networks may be modeled as biochemical reaction systems consisting of differential equations with rational functions. Biochemical reaction systems are defined as rational positive dynamic systems with inputs and outputs, and illustrated by examples. This formulation makes available the results from algebraic system theory for rational systems and a relation with computer algebra. It is shown how to decompose networks into subsystems and how to relate them to graphs. The realization problem for this class of systems is briefly discussed. Finally, control problems for biochemical reaction networks are formulated.

Keywords: biochemical reaction systems, rational systems, realization, decomposition, interconnection, algebra of positive real polynomials.

1. Introduction

The purpose of this paper is to present concepts, results, and problems of control and system theory for a subclass of the biochemical reaction systems which cover many examples of biochemical cell reaction networks.

The recent advances in knowledge for the genome of plants, animals, and humans now lead to increased interest in cell biology. Knowledge is needed on how a cell as a functional unit operates, how it interacts with its environment, and how the reaction network is influenced by the genome via the enzymes. Metabolic networks, signal transduction networks, and genetic networks have been analyzed by biologists and mathematicians. In principle, it is possible to model the complete biochemical reaction network of a cell though this program has so far been carried out only for small compartments of such networks.

*Submitted: June 2012; Accepted: February 2013.

Mathematical analysis for biochemical reaction systems leads to systems of ordinary or partial differential equations. Often the ordinary differential equations are of polynomial or of rational form. The number of reactions in a cell can be as high as 15000 (about half the number of estimated genomes) and the number of chemical compounds as high as 20000. A detailed mathematical analysis of a mathematical model of the complete cell reaction network may therefore not be possible in the near future. Hence, there is an interest to develop procedures to obtain from high-order mathematical models approximations in the form of low-order mathematical models. The formulation of approximate models requires understanding of the dynamics of the system, in particular of its algebraic and graph-theoretic structure and of its rate functions. It is the aim of the authors to contribute to this research effort.

In this paper attention is restricted to mathematical models for biochemical cell reaction networks in the form of biochemical reaction systems. These systems are called *positive* because the state vector is positive, it represents masses or concentrations of chemical species. Moreover, the external input representing concentrations of externally available chemical species, and the input vector of enzyme concentrations produced by the nucleus of the cell, are all positive. The dynamics of the system is often modeled as a polynomial map but in this paper attention is restricted to rational maps (each component equals a quotient of two polynomials). Such a dynamics arises in the model of Michaelis-Menten kinetics due to a singular perturbation of a bilinear system. The mathematical model of glycolysis in the unicellular organism *Trypanosoma brucei* is phrased almost entirely in terms of a biochemical reaction system and this model is regarded as realistic, see Bakker (1998), Helfert et al. (2001).

The subclass of biochemical reaction systems considered in this paper is specific due to the conditions imposed by the modeling of biochemical cell reaction networks. It is precisely because of these physically determined conditions that the subclass merits further study. The properties of such systems differ to a minor extent from those of polynomial systems considered in, for example, Sontag (2001). The graph-theoretic and the algebraic structure of biochemical reaction systems make the analysis interesting.

A summary of the main results follows. A brief formulation of the algebraic properties of positive real numbers and of rational positive functions is provided because of their major differences with respect to rational real functions (Section 2). A subclass of biochemical reaction systems is defined (Section 3). A running example is formulated and further references to models of biochemical cell reaction networks are mentioned (Section 3.3). With a biochemical reaction system is associated a directed graph. The graph is decomposed according to properties of connectivity (Section 4.1). It is established that the class of biochemical reaction systems as defined in this paper is closed with respect to a series interconnection. Decompositions of biochemical reaction systems are treated (Section 4.2). Finally, problems are formulated for the way the enzyme input controls the operation of the biochemical reaction system (Section 6).

The main contributions of the paper are:

- The formulation of the concept of biochemical reaction systems.
- The system theoretic results on the decomposition and on the interconnection of biochemical reaction systems.
- The discussion on the system theoretic properties of controllability and observability of these systems, and the realization problem.
- The formulation of control problems for biochemical reaction systems.

A preliminary version of this paper was presented in van Schuppen (2004).

2. Rational positive functions

In this section, notation for polynomials and rational functions is introduced and discussed.

Denote the set of integers by $\mathbb{Z}$, of positive integers by $\mathbb{Z}_+$, and natural numbers by $\mathbb{N} = \{0, 1, \ldots, \}$, see Birkhoff and MacLane (1997, p. 9) and Jacobson (1985, p. 15). For $n \in \mathbb{Z}_+$ denote the subsets $\mathbb{Z}_n = \{1, 2, \ldots, n\} \subset \mathbb{Z}$ and $\mathbb{N}_n = \{0, 1, 2, \ldots, n\} \subset \mathbb{N}$.

Denote the set of real numbers by $\mathbb{R}$. The set of *positive real numbers* is denoted as $\mathbb{R}_+ = [0, \infty)$ and the set of the *strictly positive real numbers* by $\mathbb{R}_{s+} = (0, \infty)$. This terminology is used in the literature and is preferred above the term of 'non-negative real numbers'. As an algebraic structure the set of positive real numbers is a *semi-ring*, it has the operations of addition and multiplication with neutral elements 0 and 1 for respectively addition and multiplication but it does not have an inverse with respect to addition though it has one with respect to multiplication when attention is restricted to the subset $(0, \infty)$. Note that $\mathbb{R}_+$ is an *integral domain*, defined by the condition that for all $a, b \in \mathbb{R}_+$, $ab = 0$ implies that either $a = 0$ or $b = 0$.

Consider for $n \in \mathbb{Z}_+$ the set of n-tuples of positive real numbers as the *positive vector space* $(\mathbb{R}_+, \mathbb{R}_+^n)$ with the understanding that the first object of this tuple is only a semi-ring as defined above and that vector addition does not have an inverse.

A positive vector space can also be defined geometrically. A *cone* $V \subseteq \mathbb{R}_+^n$ is defined to be a subset such that (1) $V + V \subseteq V$: $\forall v_1, v_2 \in V$, $v_1 + v_2 \in V$; and (2) $\mathbb{R}_+ V \subseteq V$: $\forall v \in V$ and $\forall c \in \mathbb{R}_+$, $cv \in V$. If S is a subset of $\mathbb{R}_+^n$ then there exists the smallest cone containing S, it is called the *cone generated by* S and denoted by $\text{cone}(S)$. A cone $V \subseteq \mathbb{R}_+^n$ is said to be *polyhedral cone* if it is the intersection of a finite number of half spaces. This definition is equivalent to the statement that there exists a finite set of vectors $\{v_1, \ldots, v_m\} \subset \mathbb{R}_+^n$ such that $V = \text{cone}(\{v_1, \ldots, v_m\})$. A finite set of vectors $\{v_1, \ldots, v_m\} \subset \mathbb{R}_+^n$ is said to be *positively dependent* if there exists $i \in \mathbb{Z}_n$ such that v_i is a positive linear combination of the other vectors, $v_i = \sum_{j \in \mathbb{Z}_n \setminus \{i\}} c_j v_j$ where for all $j \in \mathbb{Z}_n \setminus \{i\}$, $c_j \in \mathbb{R}_+$. It is called *positively independent* otherwise. A finite set of vectors $\{v_1, \ldots, v_m\} \subset \mathbb{R}_+^n$ is said to be a *frame* of a cone $V \subseteq \mathbb{R}_+^n$ if the cone is generated by the set and if the set is positively independent. Finally, we can state the geometric interpretation. A subset $V \subseteq \mathbb{R}_+^n$ is a finite-dimensional positive vector space if and only if it is a polyhedral cone. In this case the

space admits a representation in terms of a frame. For the theory of cones and polyhedral cones, see Gerstenhaber (1951), Rockafellar (1970).

For $n \in \mathbb{Z}_+$ denote the set of positive matrices of size $n \times n$ by $\mathbb{R}_+^{n\times n}$. As an algebraic structure this set is a *dioid* because it has neither an inverse with respect to matrix addition nor with respect to matrix multiplication even if attention is restricted to nonsingular matrices (the inverse of a nonsingular positive matrix may have negative elements). However, it is commutative with respect to addition. Note that for all $n \in \mathbb{Z}_+$ with $n > 1$, $\mathbb{R}_+^{n\times n}$ is not an integral domain as the following example shows,

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}_+^{2\times 2}, \quad A^2 = 0. \tag{1}$$

Notation and terminology for polynomial functions and rational functions in several variables follows. Fix $n \in \mathbb{Z}_+$, the dimension of the indeterminate, and denote the indeterminate by $x = (x_1, x_2, \ldots, x_n)$. Consider the multi index

$$k = \begin{pmatrix} k_1 & k_2 & \ldots & k_n \end{pmatrix}^T \in \mathbb{N}^n.$$

Note that the vector $k = 0 \in \mathbb{N}^n$ is admitted in the above definition. A *polynomial in n variables with positive coefficients* is denoted by

$$\begin{aligned} p(x) &= \sum_{k\in\mathbb{N}^n} c_p(k) \prod_{j=1}^{n} x_j^{k_j} = \sum_{k\in\mathbb{N}^n} c_p(k) x^k, \quad c_p(k) \in \mathbb{R}_+, \ \forall k \in \mathbb{N}^n, \\ & p \in \mathbb{R}_+[x_1, \ldots, x_n], \text{ abbreviated to } \mathbb{R}_+[x]. \end{aligned}$$

The understanding in the above definition is that there exists only a finite number of coefficients $\{c_p(k) \in \mathbb{R}_+, k \in \mathbb{N}^n\}$ for which $c_p(k) \neq 0$. Abuse of notation will be made because $p \in \mathbb{R}_+[x]$ denotes both a polynomial as an algebraic object and the function $p : \mathbb{R}_+^n \to \mathbb{R}_+$.

Note that $\mathbb{R}_+[x]$ as an algebraic structure is a dioid with the addition and the multiplication operation, but it has neither an inverse with respect to addition nor an inverse with respect to multiplication. The neutral element with respect to addition is the polynomial $p_z(x) = 0 \in \mathbb{R}_+$ for all $x \in X$, and the neutral element with respect to multiplication is the polynomial $p_{one}(x) = 1 \in \mathbb{R}_+$ for all $x \in X$.

Definition 1. *Consider the subset of positive polynomials,*

$$\begin{aligned} \mathbb{R}_{+,1}[x] &= \{p \in \mathbb{R}_+[x], \ c_p(0) = 1\}, \\ \mathbb{R}_{+,0}[x] &= \{p \in \mathbb{R}_+[x], \ c_p(0) = 0\}. \end{aligned}$$

For a polynomial with the above representation define the *total degree* of p as

$$\deg(p) = \max_{\{k\in\mathbb{N}^n | c_p(k)\neq 0\}} \sum_{i=1}^{n} k_i \in \mathbb{N}.$$

Define an order relation on $\mathbb{R}_{+,1}[x]$ by the relation $\text{order}(p) = \deg(p)$. The polynomials $p, q \in \mathbb{R}[x]$ are called *relatively prime* if the they do not have common zeroes.

Recall that an integral domain is a commutative ring with no proper zero divisors. Hence, $\mathbb{R}[x]$ is an integral domain because it is a commutative ring and because for all $p, q \in \mathbb{R}[x]$ it holds that $p \neq 0$ and $q \neq 0$ imply $pq \neq 0$. However, because $\mathbb{R}_+[x], \mathbb{R}_{+,1}[x]$ and $\mathbb{R}_{+,0}[x]$ are not commutative rings, neither of them is an integral domain. Nevertheless, it still holds for all $\mathbb{R}_+[x]$, $\mathbb{R}_{+,1}[x]$ and $\mathbb{R}_{+,0}[x]$ that if $p, q \in R$, where R denotes $\mathbb{R}_+[x]$, $\mathbb{R}_{+,1}[x]$ or $\mathbb{R}_{+,0}[x]$ and $pq = 0$, then either $p = 0$ or $q = 0$.

Note that the *set of units* in $\mathbb{R}_+[x]$, defined as the invertible elements within $\mathbb{R}_+[x]$, equals the set of strictly positive real numbers $\mathbb{R}_{s+} = (0, \infty)$. An element p of an integral domain R is said to be *irreducible* if (1) p is not a unit of R; and (2) $p = p_1 p_2$, where $p_1, p_2 \in R$, implies that either p_1 is a unit in R or p_2 is a unit in R. Thus, $p \in \mathbb{R}_+[x]$ is irreducible if (1) $p \notin (0, \infty)$; and (2) $p = p_1 p_2$, where $p_1, p_2 \in \mathbb{R}_+[x]$, implies that either $p_1 \in (0, \infty)$ or $p_2 \in (0, \infty)$.

An integral domain R is called *unique factorization domain* if (1) for any $p \in R$ there exists a factorization of the form,

$$p = \prod_{i=1}^{n} p_i, \text{ where } p_i \in R \text{ is irreducible;}$$

and if (2) the factorization is unique up to a reordering of the factors. Note that $\mathbb{R}[x]$ is a unique factorization domain. The subset $\mathbb{R}_{+,1}[x]$ of positive polynomials does not satisfy the conditions for being a unique factorization domain, see the following example.

Example 1. *Consider the following factorizations of the positive polynomial*

$$\begin{aligned} p(x) &= (x+2b)(x+3b)(x^2 - bx + 4b^2) \in \mathbb{R}_+[x], \\ &= (x+2b)(x^3 + 2bx^2 + b^2x + 12b^3) \\ &= (x+3b)(x^3 + bx^2 + 2b^2x + 8b^3), \quad b \in (0, \infty). \end{aligned}$$

The first factorization is a factorization over $\mathbb{R}$ *but not a factorization over* $\mathbb{R}_{+,1}$ *because of the term* $-bx$*. Moreover, the quadratic polynomial* $x^2 - bx + 4b^2$ *is irreducible over* $\mathbb{R}[x]$ *because its discriminant satisfies* $D = 4b^2 - 16b^2 = -12b^2 < 0$*. The second and the third factorizations are both factorizations over* $\mathbb{R}_+[x]$ *and, because of the first factorization, these factors are irreducible. Thus,* p *has two different factorizations into irreducible factors.*

Definition 2. *Consider a finite set of positive polynomials* $\{p_j \in \mathbb{R}_+[x], j \in \mathbb{Z}_m\}$*. Define the* common multiple *of this set as the positive polynomial* $p \in \mathbb{R}_+[x]$ *such that for all* $j \in \mathbb{Z}_m$ *there exists a positive polynomial* $q_j \in \mathbb{R}_+[x]$ *such that* $p = q_j p_j$*. Define the* least common multiple *of the finite set as the common multiple* $p \in \mathbb{R}_+[x]$ *such that for any other common multiple* $\bar{p} \in \mathbb{R}_+[x]$*,* $\text{order}(p) \leq \text{order}(\bar{p})$*. Denote then,*

$$p = \text{lcm}(\{p_j \in \mathbb{R}_+[x], j \in \mathbb{Z}_m\}, \mathbb{R}_+[x], \text{order}) = \text{lcm}(\{p_j \in \mathbb{R}_+[x], j \in \mathbb{Z}_m\}),$$

if the context is understood.

In this paper attention is restricted to a particular class of rational positive functions for which singularities cannot occur. For this purpose, define

$$\mathbb{R}_+(x) = \left\{ \frac{p(x)}{q(x)} \mid p(x), q(x) \in \mathbb{R}_+[x], q(x) \neq 0 \right\}, \tag{2}$$

$$\mathbb{R}_+(x) = \mathbb{R}_+(x_1, \ldots, x_n), \tag{3}$$

$$\mathbb{R}_{+,s}(x) = \left\{ \begin{array}{l} \frac{p(x)}{q(x)} \in \mathbb{R}_+(x) \mid p(x) = \sum_{k \in \mathbb{N}^n} c_p(k) x^k, \ c_p(0) = 0, \\ q(x) = \sum_{k \in \mathbb{N}^n} c_q(k) x^k, \ c_q(0) = 1 \end{array} \right\}. \tag{4}$$

If $p(x)/q(x) \in \mathbb{R}_{+,s}(x)$ then for all $x \in \mathbb{R}_+^n$, $q(x) \geq 1 > 0$, hence the quotient is well defined. Addition and multiplication of elements of $\mathbb{R}_{+,s}(x)$ are well defined and produce elements in $\mathbb{R}_{+,s}(x)$ as the following calculations show,

$$\begin{aligned} \frac{p_1(x)}{q_1(x)} + \frac{p_2(x)}{q_2(x)} &= \frac{p_1(x)q_2(x) + p_2(x)q_1(x)}{q_1(x)q_2(x)}, \\ q_1(x)q_2(x) &= \sum_{k \in \mathbb{N}^n} c_{q_1 q_2}(k) x^k, \quad c_{q_1 q_2}(0) = c_{q_1}(0) c_{q_2}(0) = 1, \\ p_1(x)q_2(x) + p_2(x)q_1(x) &= \sum_{k \in \mathbb{N}^n} c_{p_1 q_2 + p_1 q_2}(k) x^k, \\ & \quad c_{p_1 q_2 + p_2 q_2}(0) = c_{p_1}(0) c_{q_1}(0) + c_{p_2}(0) c_{q_2}(0) = 0. \end{aligned}$$

In the remainder of the paper rational functions $p(x, x_{ex})/q(x, x_{ex})$ are considered for two sets of indeterminates $(x_1, \ldots, x_n)$ and $(x_{ex,1}, \ldots, x_{ex,n_{ex}})$. In this case decompose the $k \in \mathbb{N}^{n+n_{ex}}$ vector as $k = (k_x, \ k_{x_{ex}})$ with $k_x \in \mathbb{N}^n$ and $k_{x_{ex}} \in \mathbb{N}^{n_{ex}}$. The following notation will be used,

$$\begin{aligned} p(x, x_{ex}) &= \sum_{k \in \mathbb{N}^n} c_p(k) \prod_{j=1}^{n} x_j^{(k_x)_j} \prod_{m=1}^{n_{ex}} x_{ex,m}^{(k_{x_{ex}})_m}, \\ \deg_{x,ex}(p) &= \max_{\{k \in \mathbb{N}^n \mid c_p(k) \neq 0\}} [\sum_{i=1}^{n} (k_x)_i + \sum_{j=1}^{n_{ex}} (k_{x_{ex}})_j] \in \mathbb{N}. \end{aligned}$$

Rational functions on a variety are treated in Cox, Little and O'Shea (1992) while rings of quotients are treated in Lam (1999, Chapter 10).

3. Biochemical reaction systems

3.1. Example of a biochemical reaction system

An example is presented to introduce the definition of a rational positive system for a biochemical reaction network. The reader is referred to the very nice papers by M. Feinberg on modeling chemical reaction networks by ordinary differential equations, see Feinberg and Horn (1974, 1977), Feinberg (1978, 1988).

EXAMPLE 2. *Consider the academic artificial biochemical reaction network consisting of two reactions only. There are five chemical species (A, C, D, E, F) in the network, one external chemical species (B), and one chemical species is available to the outside of the network. The chemical reactions follow.*

$$A + 4B \quad \leftrightarrow \quad 2C + 2D, \tag{5}$$
$$2D + E \quad \rightarrow \quad F. \tag{6}$$

Define the state components, the external variable, and the output variable to be the concentrations of the respective chemical species. Namely,

$$\begin{aligned} & x_1 = [A],\ x_2 = [C],\ x_3 = [D],\ x_4 = [E],\ x_5 = [F], \\ & x_{ex} = [B],\ z = [F] = x_3, \\ x \quad = \quad & \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix},\ x_{ex},\ z. \end{aligned}$$

The concentrations are of the dimension grams per cubic centimeter.

Reaction (5) is called Reaction 1*, it is a reversible reaction, and Reaction (6) is called* Reaction 2*, it is an irreversible reaction. Each reaction has a reaction rate defined below.*

The stoichiometric matrix relates the reaction rates to the differentials of the concentrations. For the example, the system is displayed below in which the symbol N denotes the stoichiometric matrix,

$$\begin{aligned} dx(t)/dt \quad &= \quad N\mathrm{Diag}(r(x(t), x_{ex}(t)))u(t) \\ &= \quad \begin{pmatrix} -1 & 0 \\ +2 & 0 \\ +2 & -2 \\ 0 & -1 \\ 0 & +1 \end{pmatrix} \begin{pmatrix} r_1(x(t), x_{ex}(t))u_1(t) \\ r_2(x(t), x_{ex}(t))u_2(t) \end{pmatrix}, \end{aligned}$$

$$N \quad = \quad N^+ - N^- = \begin{pmatrix} 0 & 0 \\ +2 & 0 \\ +2 & 0 \\ 0 & 0 \\ 0 & +1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$\begin{aligned} y(t) \quad &= \quad Cx(t) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix} x(t), \\ z(t) \quad &= \quad H\mathrm{Diag}(r(x(t), x_{ex}(t)))u(t) \\ &= \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} \mathrm{Diag}(r(x(t), x_{ex}(t))))u(t), \end{aligned}$$

$$
\begin{aligned}
r_1^+(x, x_{ex}) &= \frac{x_1 x_{ex}^4}{1 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_{ex}} = \frac{[A][B]^4}{1 + c_1[A] + c_2[C] + c_3[D] + c_4[B]}, \\
r_1^-(x, x_{ex}) &= \frac{x_2^2 x_3^2}{1 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_{ex}}, \\
r_1(x, x_{ex}) &= r_1^+(x, x_{ex}) - r_1^-(x, x_{ex}), \\
r_2^+(x, x_{ex}) &= \frac{x_3^2 x_4}{1 + c_5 x_3 + c_6 x_4}, \quad r_2^-(x, x_{ex}) = 0, \quad r_2(x, x_{ex}) = r_2^+(x, x_{ex}).
\end{aligned}
$$

Because Reaction 1 is reversible, it has both a forward rate function, $r_1^+(x, x_{ex})$, and a backward rate function, $r_1^-(x, x_{ex})$. Below, the rate functions are multiplied by the enzyme concentrations to obtain the actual reaction rates. If both the forward and the backward reaction rates of one reaction are rational functions, then the denominators are usually identical. For an irreversible reaction the backward reaction rate is zero to make the notation consistent.

The differential equations for the concentrations of the chemical species $[A]$ and $[D]$ are then

$$
\begin{aligned}
dx_1(t)/dt &= N_{11} r_1^+(x(t), x_{ex}(t)) u_1(t) - N_{11} r_1^-(x(t), x_{ex}(t) u_1(t) \\
&= N_{11} r_1(x(t), x_{ex}(t)) u_1(t) \\
&= -\left[\frac{x_1(t) x_{ex}(t)^4}{1 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_{ex}} - \frac{x_2(t)^2 x_3(t)^2}{1 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_{ex}} \right], \\
dx_3(t)/dt &= N_{31} r_1^+(x(t), x_{ex}(t)) u_1(t) - N_{31} r_1^-(x(t), x_{ex}(t)) u_1(t) + \\
&\quad + N_{32} r_2^+(x(t), x_{ex}(t)) u_2(t), \\
&= 2 \left[\frac{x_1(t) x_{ex}(t)^4 - x_2(t)^2 x_3(t)^2}{1 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_{ex}} \right] u_1(t) + \\
&\quad -2 \left[\frac{x_3(t)^2 x_4(t)}{1 + c_5 x_3(t) + c_6 x_4(t)} \right] u_2(t), \\
z(t) &= H_5 r_2^+(x(t), x_{ex}(t)) u_2(t) = r_2(x(t), x_{ex}(t)) u_2(t) \\
&= H \mathrm{Diag}(r(x(t), x_{ex}(t)) u(t).
\end{aligned}
$$

To summarize, the biochemical reaction system of this example is described by the following vector differential equation and output equation,

$$
\begin{aligned}
dx(t)/dt &= N \mathrm{Diag}(r(x(t), x_{ex}(t)) u(t), & (7) \\
z(t) &= H \mathrm{Diag}(r(x(t), x_{ex}(t)) u(t). & (8)
\end{aligned}
$$

3.2. Concepts

Biochemical reaction systems are defined formally below.

Definition 3. *A* biochemical reaction system *for a biochemical reaction network is defined as a control system, as understood in system theory, defined by*

the system of differential equations

$$
\begin{aligned}
dx(t)/dt &= N\mathrm{Diag}(r(x(t),x_{ex}(t)))u(t),\ x(t_0)=x_0\in X=\mathbb{R}^n_+, && (9)\\
y(t) &= Cx(t), && (10)\\
z(t) &= H\mathrm{Diag}(r(x(t),x_{ex}))u(t), && (11)\\
&\textit{or, per component } i\in\mathbb{Z}_n,\\
dx_i(t)/dt &= \sum_{j=1}^{m} N_{i,j}r_j(x(t),x_{ex}(t))u_j(t) && (12)\\
&= \sum_{j=1}^{m}(N^+_{i,j}-N^-_{i,j})[r^+_j(x(t),x_{ex}(t))-r^-_j(x(t),x_{ex}(t))]u_j(t) && (13)\\
&= f_i(x(t),x_{ex}(t),u(t)),\ x_i(t_0)=x_{i,0},
\end{aligned}
$$

with the definitions, $n,\ m\in\mathbb{Z}_+,\ n_{ex},\ n_z\in\mathbb{N}$,

$$
\begin{aligned}
T &= [t_0,\infty),\ \textit{the time index set,}\\
X &= \mathbb{R}^n_+,\ \textit{the state set,}\\
X_{ex} &= \mathbb{R}^{n_{ex}}_+,\ \textit{the set of external concentrations,}\\
U &= \mathbb{R}^m_+,\ \textit{the set of enzyme concentrations,}\\
N\in\mathbb{Z}^{n\times m} &\quad \textit{called the } \text{stoichiometric matrix,}\\
&\quad \textit{with decomposition, } N=N^+-N^-,\quad N^+,N^-\in\mathbb{N}^{n\times m},\\
&\quad u:T\to U,\quad \textit{an input function,}\\
&\quad r:X\times X_{ex}\to\mathbb{R}^m,\ \forall j\in\mathbb{Z}_m,\\
r_j(x,x_{ex}) &= r^+_j(x,x_{ex})-r^-_j(x,x_{ex})=\frac{p^+_j(x,x_{ex})}{q^+_j(x,x_{ex})}-\frac{p^-_j(x,x_{ex})}{q^-_j(x,x_{ex})}, && (14)\\
&\quad \frac{p^+_j(x,x_{ex})}{q^+_j(x,x_{ex})},\frac{p^-_j(x,x_{ex})}{q^-_j(x,x_{ex})}\in\mathbb{R}_{+,s}(x,x_{ex}),\ q^+_j(x,x_{ex})=q^-_j(x,x_{ex}),\\
\mathrm{Diag}(r(x,x_{ex})) &= \mathrm{Diag}((r_1(x,x_{ex})),\ldots,(r_m(x,x_{ex})))\in\mathbb{R}^{m\times m}_+, && (15)\\
&\quad \mathrm{Diag}(w),\ \textit{denotes a diagonal matrix with the elements of}\\
&\quad \textit{the vector } w \textit{ on the diagonal,}\\
&\quad z:T\to\mathbb{R}^{n_z},\ H\in\mathbb{N}^{n_z\times m},\\
&\quad \textit{where } z \textit{ represents the outflow rate of the system.}
\end{aligned}
$$

The following conditions are assumed to hold:

1. *For all* $j\in\mathbb{Z}_m$, *the two tuples* (p^+_j,q^+_j) *and* (p^-_j,q^-_j) *are each assumed to be relatively prime polynomials.*
2. *For all* $i\in\mathbb{Z}_n$, $j\in\mathbb{Z}_m$, $\forall x\in X$ *and* $\forall x_{ex}\in X_{ex}$,
$$
\begin{aligned}
x_i=0\ \wedge\ (N^+_{i,j}-N^-_{i,j}>0) &\ \Rightarrow\ r^-_j(x,x_{ex})=0;\\
x_i=0\ \wedge\ (N^+_{i,j}-N^-_{i,j}<0) &\ \Rightarrow\ r^+_j(x,x_{ex})=0.
\end{aligned}
$$

3. *For all* $i \in \mathbb{Z}_{n_z}$ *and* $j \in \mathbb{Z}_m$, $\forall x \in X$ *and* $\forall x_{ex} \in X_{ex}$,

$$H_{i,j} > 0 \;\Rightarrow\; p_j^-(x, x_{ex}) = 0.$$

4. *The components of* r, *thus* $\{r_j(.), j \in \mathbb{Z}_m\}$, *are linearly independent functions.*
5. *For* $T = [t_0, \infty)$, *for any initial condition* $x_0 \in \mathbb{R}^n$, *any external concentration function* $x_{ex} : T \to \mathbb{R}_+^{n_{ex}}$, *and any continuous input function* $u : T \to \mathbb{R}_+^m$ *there exists a unique solution* $x : T \to \mathbb{R}$ *to the system of ordinary differential equations (9).*

The reader should clearly distinguish between the state function and the vector of *external concentrations* $x_{ex} \in \mathbb{R}_+^{n_{ex}}$. The external concentrations represent masses of concentrations or chemical species which are not part of the model and whose dynamics is not part of the system. They may represent chemical species which are available in abundance and whose values may not change over time depending on the other state variables. Such an external concentration could be assumed to be constant over time during the interval considered or could be assumed to be time varying in a prescribed way. In the example of Subsection 3.3, the external concentrations are present in the model. Distinguish also the outflow rate z and the state x.

Comments on the conditions of Definition 3 follow. The first condition is to obtain a mathematically economical expression for the rate functions. If the condition is not met then it can be obtained by canceling common factors. The second condition is necessary and sufficient for the positive orthant $\mathbb{R}_+^n$ to be a positively invariant set of the system, see the next section. The third condition is to enforce that the outflow rate refer to an outflow only, there is no inflow into the systems. To relax the condition requires further modeling of the inflows into the biochemical reaction system also from the outflows. The fourth condition is to obtain a nonredundant set of reactions. If the condition is not met then the corresponding enzyme inputs can be combined so that a system with one reaction less is obtained. The last condition is needed for mathematical reasons.

For references on positive systems see Berman, Neumann and Stern (1989), Farina and Rinaldi (2000), van den Hof (1996).

3.3. Examples

Biochemical reaction networks are used to represent metabolic networks, but also signal transduction networks (with the purpose of controlling to communicate signals), and genetic networks (with the purpose of controlling the operation of a cell from the nucleus).

Other references on dynamic systems for biochemical cell reaction networks are Fall et al. (2002), Heinrich and Schuster (1996). For models of polynomial systems see Sontag (2001). General references on nonlinear systems as considered in control and system theory include Isidori (1989, 1999), Sontag (1998). Rational systems without the positivity condition have been treated in Bartosiewicz (1987b), Wang, Sontag (1992).

There follows an example of a biochemical reaction system.

EXAMPLE 3. Glycolysis in the bacterium Trypanosoma brucei. *The bacterium Trypanosoma brucei is a parasite of humans. They live in the blood and tissue fluids of mammals. In Africa, the parasites are transmitted by the tsetse flies, which occur only in the sub-Saharan region. A subspieces of T. brucei causes the African Sleeping Sickness in humans; approximately 200 000 new infections occur every year. The infection is lethal unless treated. The few existing drugs have severe side effects. The parasites also have a significant economic impact through infection of lifestock. There is thus a need for medicine to counter the effects of this parasite.*

The bacterium gets its energy within a mammal by the well known chemical process called glycolysis, meaning the splitting of sugar. Under aerobic conditions, most glucose is metabolized to pyruvate although about 10% is converted into glycerol. A mathematical model has been formulated for the dynamic behavior of the biochemical reaction network of glycolysis. With the model, biologists predict the effect of adjusting the external conditions on the network. Think for this of adjusting the amounts of enzymes. The effects noticed can then be used to select chemical compounds for medical drugs. The model has been formulated by P. Michels (Brussels) and B.M. Bakker (Groningen) and co-workers, see Bakker (1998) and Helfert et al. (2001). The readers who are not biologists may want to read in Campbell, Reece, Mitchell (1999, Ch. 9) about the operation of glycolysis in a cell.

A specification of the model follows. The model has 28 chemical compounds, 20 reactions, one external input, and two output flows. The model consists of a differential-algebraic system, that is, a set of differential equations and algebraic equations for the state variables. The model takes too much space for this paper. A part of the model follows below. For the full model see the references quoted above.

Notation. The states represent chemical compounds in the cell, denoted by x_i for $i \in \mathbb{Z}_{28}$, the enzyme inputs are denoted by u_j for $j \in \mathbb{Z}_{20}$, the reaction rates are denoted by r_k, for $k \in \mathbb{Z}_{20}$. Of the reactions, those numbered 3, 6, 9, 13, 15, 17, 19, and 20 are treated as in equilibrium and the corresponding equations are not listed in this paper.

The states and the chemical compounds which they represent are denoted by,

$$
\begin{aligned}
&x_1 = GLC_{in}, && x_2 = [hexose - P], && x_3 = [Fru - 1,6 - BP]_g,\\
&x_4 = triose - P, && x_5 = [1,3 - BPGA]_g, && x_6 = N,\\
&x_7 = [PYR]_c, && x_8 = [NADH]_g, && x_9 = P_g,\\
&x_{10} = P_c, && x_{11} = ADP, && x_{12} = ATP,\\
&x_{13} = 3 - PGA, && x_{14} = DHAP, && x_{15} = Gly - 3 - P,\\
&x_{16} = NAD^+, && x_{17} = GA - 3 - P, && x_{18} = Gly,\\
&x_{ex} = GLC_{out}.
\end{aligned}
$$

The notations for the input components and the corresponding enzymes are,

u_1	*transport of glucose through the plasma and glycolysis membrane,*
u_2	*HK,* u_4 *PFK,* u_5 *ALD,* u_7 *GAPDH,* u_8 *PGK,* u_{10} *PYK,*
u_{11}	*transport of pyruvate across the plasma membrane,*
u_{12}	*GDH,* u_{14} *GPO,* u_{16} *GK,* u_{18} *ATP utilization.*

The rate functions are only partly described and the algebraic relations are not listed at all,

$$\begin{aligned}
r_1 &= r_1^+ - r_1^- = c_2 \frac{c_{21}x_{ex} - c_{22}x_1}{c_{23} + x_1 + x_{ex} + c_{24}x_1x_{ex}},\\
r_2 &= r_2^+ = c_1 \frac{c_3x_{13}c_4x_1}{(1 + c_3x_{12} + c_5x_{11})(1 + c_4x_1)},\\
r_7 &= r_7^+ - r_7^- = c_1 \frac{[c_9x_{17}c_{10}x_{16} - c_6c_7x_5c_8x_8]}{(1 + c_9x_{17} + c_7x_5)(1 + c_{10}x_{16} + c_8x_8)},\\
r_8 &= r_8^+ - r_8^- = c_1 \frac{[c_{11}x_5c_{12}x_{11} - c_{13}c_{14}x_{13}c_{15}x_{12}]}{(1 + c_{11}x_5 + c_{14}x_{13})(1 + c_{12}x_{11} + c_{15}x_{12})},\\
r_{12} &= r_{12}^+ - r_{12}^- = c_1 \frac{[c_{16}x_{14}c_{18}x_8 - c_{18}c_{19}x_{15}c_{20}x_{16}]}{(1 + c_{16}x_{14} + c_{19}x_5)(1 + c_{17}x_8 + c_{20}x_{16})}.
\end{aligned}$$

The differential equations can then be expressed in terms of the rates,

$$\begin{aligned}
\dot{x}_1(t) &= c_{38}[r_1(x(t)) - r_2(x(t))],\\
\dot{x}_2(t) &= c_{40}[r_2(x(t)) - r_4(x(t))],\\
\dot{x}_3(t) &= c_{40}[r_4(x(t)) - r_5(x(t))],\\
\dot{x}_4(t) &= c_{38}[2r_5(x(t)) - r_7(x(t)) - r_{12}(x(t)) - r_{14}(x(t))],\\
\dot{x}_5(t) &= c_{40}[r_7(x(t)) - r_8(x(t))],\\
\dot{x}_6(t) &= c_{38}[r_8(x(t)) - r_{10}(x(t))],\\
\dot{x}_7(t) &= c_{39}[r_{10}(x(t)) - r_{11}(x(t))],\\
\dot{x}_8(t) &= c_{40}[r_7(x(t)) - r_{12}(x(t))],\\
\dot{x}_9(t) &= c_{40}[r_8(x(t)) + r_{16}(x(t)) - r_2(x(t)) - r_4(x(t))],\\
\dot{x}_{10}(t) &= c_{39}[r_{10}(x(t)) - r_{18}(x(t))].
\end{aligned}$$

The full system is then,

$$\begin{aligned}
dx(t)/dt &= NM\mathrm{Diag}(r(x(t), x_{ex}(t))u(t),\\
M &= \mathrm{Diag}(c_{38}, c_{40}, c_{40}, c_{38}, c_{40}, c_{38}, c_{39}, c_{40}, c_{40}, c_{39}).
\end{aligned}$$

Due to the particular representation of the model, there is an extra matrix M *which is not stated in the formal definition.*

Example 4. *The following small example illustrates the transformation of a mathematical model of the biochemical processes of a cell to a dynamic system.*

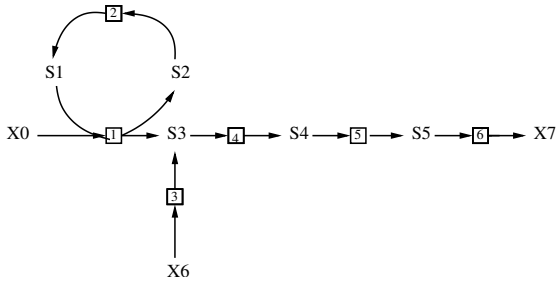

Figure 1. The biochemical network of the example

The model is derived from the example described by J.M. Rohwer in Rohwer (1997, pp. 32, 37). See Fig. 1. The inputs and state variables are, in terms of the notation used in that reference:

$$\begin{aligned} n &= 5,\ n_{ex} = 2,\ n_z = 1,\ m = 6, \\ x_1 &= S_1,\ \ldots,\ x_5 = S_5,\ u_1 = e_1,\ \ldots,\ u_6 = e_6,\ \ x_{ex,1} = X0,\ x_{ex,2} = X6, \\ z &= X7. \end{aligned}$$

The stoichiometric matrix and the rate functions are:

$$\begin{aligned} N &= \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \\ r_1(x, x_{ex}) &= \frac{10 x_{ex,1} x_1}{1 + x_{ex,1} x_1 + x_2 x_3} - \frac{x_2 x_3}{1 + x_{ex,1} x_1 + x_2 x_3} \\ &= \frac{p_1^+(x, x_{ex})}{q_1^+(x, x_{ex})} - \frac{p_1^-(x, x_{ex})}{q_1^-(x, x_{ex})}, \\ r_2 &= \frac{10 x_2}{1 + x_1 + x_2} - \frac{x_1}{1 + x_1 + x_2}, \\ r_3 &= \frac{5 x_{ex,2}}{1 + x_3 + x_{ex,2}} - \frac{x_3}{1 + x_3 + x_{ex,2}}, \\ r_4 &= \frac{10 x_3}{1 + x_3 + x_4} - \frac{x_4}{1 + x_3 + x_4}, \\ r_5 &= \frac{10 x_4}{1 + x_4 + x_5} - \frac{x_5}{1 + x_4 + x_5}, \quad r_6 = \frac{10 x_5}{1 + x_5}, \\ H &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{Z}^{1 \times m}. \end{aligned}$$

The resulting dynamic system then has the form

$$\begin{aligned} \dot{x}(t) &= N\ \mathrm{Diag}(r(x(t), x_{ex}))u(t),\ x(t_0) = x_0, \\ z(t) &= H\mathrm{Diag}(r(x(t), x_{ex}))u(t). \end{aligned}$$

Note that for each $i \in Z_6$, in the function $r_i : \mathbb{R}^n_+ \times \mathbb{R}^{n_{ex}}_+ \to \mathbb{R}$ each of the terms in the difference is a rational function in the indeterminates $(x_1, ..., x_n, x_{ex,1}, ..., x_{ex,n_{ex}})$ of which the numerator and the denominator degrees are equal. Note that for $i = 1$,

$$\begin{aligned} x_1 &= 0\ \wedge\ N_{11} = -1 \Rightarrow p_1^+(x, x_{ex}) = 0; \\ x_1 &= 0\ \wedge\ N_{12} = +1 \Rightarrow p_2^-(x, n_{ex}) = 0; \textit{ etc. for } i = 2, \ldots, 5. \end{aligned}$$

4. System decomposition, interconnection, and interaction

The motivation for the decomposition of high-dimensional biochemical reaction networks is primarily to be able to evaluate the functioning of small networks, to investigate the system reduction of subsystems of networks, and to combine the reduced subsystems in a larger system.

4.1. Rational positive systems and their graphs

The dynamics of a biochemical reaction system is determined by a graph and by the reaction rates as specified by the rate function $r : X \times X_{ex} \to \mathbb{R}^m$. The graph relates directly to the chemical reaction network of the system, because the system is a model for such a network. How one can recover the graph from a specification of the system is described below. Of course, the graph can also be determined from the reaction network but typically this is not done by biochemists. The resulting graph can be used to analyze the dynamics of the system as illustrated in the following subsections.

At the interface of system theory and graph theory there have appeared publications on how to use graphs to analyze the underlying network of system states. For sources see Murota (2000), Reinschke (1988). For concepts and theorems on graphs see Gondran and Minoux (1984).

A *directed graph* is a tuple (V, E) consisting of a *set of vertices* V and a *set of edges* $E \subset V \times V$. Denote the vertices by $V = \{v_1, v_2, \ldots, v_n\}$ and an edge $e = (v_i, v_j) \in E$ denotes that there is an arrow from vertex v_i to vertex v_j. A *path* from vertex v_i to vertex v_j is a sequence of edges

$$\{(v_i, v_1), (v_1, v_2), \ldots, (v_m, v_j)\} \subset E.$$

Denote this by $v_i \mapsto v_j$. It is called as *elementary path* if in the path no vertex occurs twice except possibly at the start and terminal vertex of the path. An *elementary cycle* is an elementary path from a vertex back to the same vertex. Define the *strong connection relation* on vertices as $(v_i, v_j) \in R_{sc} \subset V \times V$ if

there exists a path from v_i to v_j and there exists a path from v_j to v_i. Denote by $C = \{C_1, C_2, \ldots, C_p\}$ the *set of equivalence classes* of the strong connection relation and call $p \in \mathbb{Z}_+$ the *connectivity number*. The graph is said to be *strongly connected* if the connectivity number p equals 1.

Of interest to biochemical reaction systems is the graph induced by (V, E) on the set of equivalence classes induced by the strong connection relation.

Define the directed *graph of strongly connected classes* as the tuple (C, E_C) where C is the set of equivalence classes defined above and the set of oriented edges is defined by $(C_1, C_2) \in E_C$ if there exist vertices $v_i \in C_1$ and $v_j \in C_2$ such that there is a path from vertex v_i to vertex v_j. From the definition of the strong connection relation follows that the graph (C, E_C) does not contain a cycle.

DEFINITION 4. *Consider a biochemical reaction system as defined in Definition 3,*

$$\begin{aligned} \dot{x}_i(t) &= \sum_{j=1}^{m} N_{i,j} r_j(x(t), x_{ex}(t)) u_j(t) = f_i(x(t), x_{ex}(t), u(t)), \; x(t_0) = x_0, \\ z(t) &= H \mathrm{Diag}(r(x(t), x_{ex}(t))) u(t). \end{aligned}$$

Define the system graph (V, E) *by the* vertex set $V = \mathbb{Z}_{n+n_z+n_{ex}}$*, where the vertex* $i \in V$ *corresponds: (1) the* state nodes*: for* $i \in \mathbb{Z}_n$*, to the* i*-the component* x_i *of the state vector* x*; (2) to the* output nodes*, for* $i = n+1, \ldots, n+n_z$*, to the output flows* z_{i-n}*; and (3) to the* external concentration nodes*, for* $i = n+n_z+1, \ldots, n+n_z+n_{ex}$*, to* $x_{ex,i-(n+n_z)}$*; and by the edge set* $E \subset V \times V$ *where* $(i,j) \in E$ *if (1) for* $j = 1, \ldots, n$ *the function* f_i *depends on the state component* x_j*; (2) for* $j = n+n_z+1, \ldots, n+n_z+n_v$ *the function* f_i *depends on the external concentrations* $x_{ex,j-(n+n_z)}$*; and (3) for* $i = n+1, \ldots, n+n_z$ *the function* $h_{i-n}(x)$ *depends on* x_j*. Associate with the graph the* graph matrix

$$A \in \{0,1\}^{(n+n_z)\times(n+n_z+n_{ex})}, \quad A_{i,j} = \begin{cases} 1, & \text{if } (j,i) \in E, \\ 0, & = \text{otherwise.} \end{cases}$$

The graph can be extended with nodes for the external input rate.

ALGORITHM 1. Computation of the system graph. *Consider a biochemical reaction system.*
Input data: $n \in \mathbb{Z}_+$ *and* $n_{ex}, n_z \in \mathbb{N}$ *and the functions* f *and* H*.*

1. *Set* $V = \mathbb{Z}_{n+n_z+n_{ex}}$*.*
2. *For all* $j \in \mathbb{Z}_n$ *determine the* index set of reactions affecting the rate of component j *and the outflow component* $g \in \mathbb{Z}_{n_z}$*, as*

$$K_j = \{k \in \mathbb{Z}_m | (N^+_{j,k} - N^-_{j,k}) \neq 0\}, \quad K^H_g = \{k \in \mathbb{Z}_m | H_{g,k} \neq 0\}.$$

3. *For all* $i, j \in V$ *with* $j = 1, \ldots, n$ *and for all* $k \in K_j$*, include* $(i,j) \in E$ *if the positive component* $r^+_k(x, x_{ex})$ *of the rate function* $r_k(x, x_{ex})$ *depends directly on the variable* x_i *or if that component depends on* $x_{ex,i-(n+n_z)}$*.*

For all $i = n+1, \ldots, n_z$, $j \in V$, and $k \in K_j^H$ include $(i,j) \in E$ if the rate function $r_k(x, x_{ex})$ depends on x_i or on $x_{ex,i-n}$.

Example 5. *Consider Example 4. Recall that*

$$N = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \ n = 5, \ n_{ex} = 2, \ n_z = 1, \ m = 6.$$

Algorithm 1 is applied:

$$\begin{aligned} & n + n_z + n_{ex} = 5 + 1 + 2 = 8, \\ V &= \{1,2,3,4,5,6,7,8\}, \quad E \subset V \times V, \text{ is computed as follows,} \\ j &= 1, \ K_1 = \{1,2\}, \quad \text{as follows from the first row of the matrix } N, \\ & (1,1), \ (2,1), \ (3,1), \ (1,7) \in E; \\ j &= 2, \ K_2 = \{1,2\}, \quad (1,2), \ (2,2), \ (3,2), \ (2,7) \in E; \\ j &= 3, \ K_3 = \{1,3,4\}, \quad (1,3), \ (2,3), \ (3,3), \ (4,3), (3,7), (3,8) \in E; \\ j &= 4, \ K_4 = \{4,5\}, \quad (3,4), \ (4,4), \ (5,4) \in E; \\ j &= 5, \ K_5 = \{5,6\}, \quad (4,5), \ (5,5) \in E; \\ j &= 6, \ K_6 = \{5\}, \quad (5,6) \in E. \end{aligned}$$

Another way to see this is to note that in the description of Example 3 the rate function r_1 depends on the external variable $x_{ex,1}$. The components of the right-hand sides of the differential equations, f_1, f_2, and f_3 depend on the rate function r_1 according to the first column of the stoichiometric matrix N. Hence, the seventh column of the matrix A has three one's in the first three rows, see (16). Similarly, the rate function r_3 depends on the external variable $x_{ex,2}$, the function f_3 depends on r_3 according to the stoichiometric matrix N, and consequently the eighth column of the matrix A has a one in the third row.

The graph matrix of the system is then

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{16}$$

Note that the graph, restricted to the state nodes $1, \ldots 5$, is strongly connected and that there exist paths from the two external concentration Nodes $7, 8$ to the output Node 6. The diagram of the graph is displayed in Fig. 2 except that self-loops have been omitted. A self-loop is an edge going from a node back to the same node.

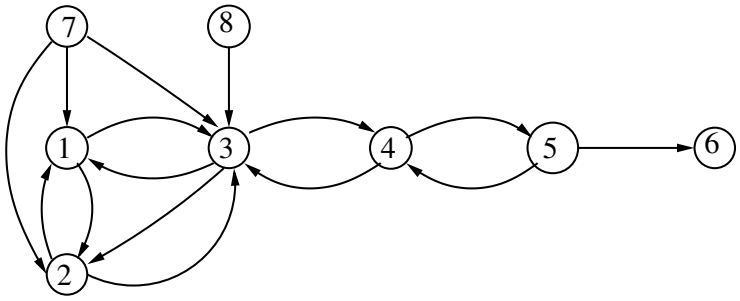

Figure 2. The diagram of the system graph with the state nodes $1, \ldots, 5$, the external input nodes $7, 8$, and the outflow node 6. The self-loops have been omitted

4.2. Decomposition of biochemical reaction systems

A biochemical cell reaction network may be considered as an interconnected network of irreducible systems. It is therefore of interest to decompose the network into irreducible systems as is regularly done for algebraic systems.

But a biochemical reaction system is an *open system* rather than a *closed system*, it interacts with the outside world. The interaction with its environment takes place through the external concentration and the outflow rate. Because of this, the following concepts of the theory of compartmental systems are useful, see Fife (1972), Jacquez (1985), Jacquez and Simon (1993):

Definition 5. *Consider a biochemical reaction system as defined in Definition 3.*

(a) A subsystem *is another biochemical reaction system of which the state set is a subset of the states of the original system, say indexed by $I \subseteq \mathbb{Z}_n$, the dynamics is that of the original system restricted to those states, and where the interconnections to the other states, those indexed by the state components $\mathbb{Z}_n \backslash I$, are formulated as external input rates and outflow rates. For a subsystem indexed by $I \subseteq \mathbb{Z}_n$ call the corresponding subsystem, indexed by $\mathbb{Z}_n \backslash I$, the* complementary subsystem.

(b) A trap *of the biochemical reaction system considered is a subsystem from which there is no outflow, neither to the outside environment nor to the complementary subsystem. A* simple trap *is a trap which does not strictly contain a nontrivial trap.*

(c) An internal source *of the biochemical reaction system considered is a subsystem that does neither have an external input rate nor an inflow from the complementary subsystem. A* simple internal source *is an internal source which does not strictly contain a nontrivial internal source.*

Theory is needed on the characterization of biochemical reaction systems with traps or internal sources and on the decomposition of the graph of a bio-

chemical reaction system. The theory may be based on that for compartmental systems, see Fife (1972), Foster and Jacquez (1975), Jacquez and Simon (1993). In particular, characterizations are needed of a trap and of an internal source in terms of dynamics, and a decomposition of the network is useful. These results will have an impact on the realization problem formulated elsewhere in the paper.

4.3. Interconnections of biochemical reaction systems

In biochemical reaction networks one may combine networks in various ways. It is therefore of interest to investigate the question whether the class of biochemical reaction systems is closed with respect to interconnection operations. The modular formulation of such networks is important.

A preliminary classification of the possible ways in which biochemical reaction systems interact follows:

- The external concentration of a system is determined by another system, hence there exists a chemical species, which is common to the two systems, but there are no common reactions of the two systems.
- There are both common chemical species and common reactions between the subsystems.

In this paper only the first case described above is discussed. The other case is left for a future publication.

Theorem 1. *Consider two biochemical reaction systems as defined in Definition 3.*

$$dx_1(t)/dt = N_1\mathrm{Diag}(r_1(x_1(t), x_{ex,1}(t)))u_1(t),\ x_1(t_0) = x_{1,0}, \tag{17}$$

$$dx_2(t)/dt = N_2\mathrm{Diag}(r_2(x_2(t), x_{ex,2}(t)))u_2(t),\ x_2(t_0) = x_{2,0}. \tag{18}$$

Consider the series connection defined by the relation

$$x_{ex,2}(t) = Cx_1(t) + x_{ex,4}(t), \tag{19}$$

where the external concentration $x_{ex,4}$ is a new external concentration of the interconnected system. The interconnection of the two subsystems is then again a biochemical reaction system with the following representation and the corresponding formulas:

$$\begin{aligned} dx_3(t)/dt &= N_3\mathrm{Diag}(r_3(x_3(t), x_{ex,3}(t)))u_3(t),\ x_3(t_0) = x_{3,0}, \\ &\quad n_3 = n_1 + n_2,\ m_3 = m_1 + m_2,\ n_{ex,3} = n_{ex,1} + n_{ex,2}, \end{aligned} \tag{20}$$

$$\begin{aligned} x_3 &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},\ x_{ex,3} = \begin{pmatrix} x_{ex,3,1} \\ x_{ex,3,2} \end{pmatrix},\ u_3 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \\ r_3(x_3, x_{ex,3}) &= \begin{pmatrix} r_1(x_1, x_{ex,1}) \\ r_2(x_2, Cx_1 + x_{ex,4}) \end{pmatrix} = \begin{pmatrix} r_{3,1}(x_1, x_{ex,3,1}) \\ r_{3,2}(x_2, x_{ex,3,2}) \end{pmatrix}, \\ N_3 &= \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix}. \end{aligned} \tag{21}$$

Proof With the notation introduced above, one obtains,

$$\begin{aligned}
& r_1(x_1, x_{ex,1}) = r_1(x_1, x_{ex,3,1}) = r_{3,1}(x, x_{ex,3}), \\
dx_1(t)/dt &= N_1 \mathrm{Diag}(r_1(x_1(t), x_{ex,1}(t))u_1(t) \\
&= \begin{pmatrix} N_1 & 0 \end{pmatrix} \mathrm{Diag}(r_3(x_3(t), x_{ex,3}(t))u_3(t), \\
& r_2(x_2, x_{ex,2}) = r_2(x_2, Cx_1 + x_{ex,4,2}) = r_{3,2}(x_3, x_{ex,4,2}), \\
dx_2(t)/dt &= N_2 \mathrm{Diag}(r_2(x_2(t), x_{ex,2}(t))u_2(t) \\
&= \begin{pmatrix} 0 & N_2 \end{pmatrix} \mathrm{Diag}(r_3(x_3(t), x_{ex,3}(t)))u_3(t), \\
dx_3(t)/dt &= N_3 \mathrm{Diag}(r_3(x_3(t), x_{ex,3}(t)))u_3(t).
\end{aligned}$$

The representation of a biochemical reaction system is then obtained. It remains to prove that the interconnected system satisfies the conditions (1–5) of Definition 3. By assumption of Condition 5 for the Subsystems 1 and 2 and by interconnection, Condition 5 is met for System 3. Because the rate function of the interconnected system is a composition of those of Subsystem 1 and Subsystem 2, it meets Condition 1. For Condition 2, consider $i \in \mathbb{Z}_{n_1}$ and $j \in \mathbb{Z}_{m_1}$. From Condition 1 for Subsystem 1 it follows that $x_{3,i} = x_{1,i} = 0$ implies that $r_3(x_3, x_{ex,3}) = r_1^-(x_1, x_{ex,1}) = 0$. Likewise, one proceeds to prove that Condition 3 holds for System 3. Condition 3 follows directly from that of Subsystem 2. Condition 4 follows, because the rate function of the interconnected system is that of the Subsystems 1 and 2 combined, while their state variables are disjoint. □

5. Realization of rational positive systems

The realization problem of system theory aims at studying dynamic systems as relations between input and output functions. The conditions for the existence and uniqueness of a realization reappear as conditions for the existence of control laws or of observers for such systems. The results of the realization problem are also relevant for identifiability and for system reduction.

In this section we will deal with the realization problem for the class of positive rational systems modeling biochemical reaction networks. However, we do not provide a complete solution to it. We expound the realization theory for rational systems as it is introduced in Němcová and van Schuppen (2009, 2010) and discuss several generalizations and open problems. The applied approach has been motivated by the results on discrete-time polynomial systems, Sontag (1979), continuous-time polynomial systems, Bartosiewicz (1985, 1988), and by the framework for rational systems introduced in Bartosiewicz (1987b).

5.1. Rational systems

Throughout the section we assume without loss of generality that the enzyme concentrations $u(t)$ and the external concentrations $x_{ex}(t)$, which appear in rational systems introduced in Definition 3 as models of biochemical reaction

networks, are piecewise constant functions. Despite the fact that the assumption on inputs to be piecewise constant is not very realistic, it is still sufficient to consider such inputs for modeling purposes. Note that any continuous input (continuous functions are sufficiently general to describe any real behavior of concentrations) can be approximated by piecewise constant functions. Then, when the topology of uniform convergence is considered on inputs, the outputs depend continuously on inputs. The proof can be found, for example, in Sussmann (1979). The way one proceeds with modeling is as follows:

Since the measurements are in practice usually very noisy and inprecise, it is not necessary to require that the outputs (measurements) be modeled by systems with identical behavior, reasonable approximations are acceptable. Therefore, if one models the outputs corresponding to continuous inputs by rational systems, then it is sufficient to find a rational system with outputs which (1) are 'close' to the measured outputs and (2) correspond to piecewise constant inputs which approximate the continuous ones. It is possible owing to the assumption that one can model the measurements by rational systems and owing to the continuous dependence of outputs on inputs. Then, the problem whether such rational system exists is a topic of realization theory. How to choose the right approximative output/input to be modeled by a rational system is another difficult problem.

Let $\gamma_i = (\alpha_i, \beta_i) \in X_{ex} \times U \subseteq \mathbb{R}_+^{n_{ex}+m}$, $i = 1, \ldots, k$, denote the constant values of the external and enzyme concentrations such that

$$v = (\gamma_1, t_1)(\gamma_2, t_2) \cdots (\gamma_k, t_k), \quad \text{i.e.} \quad v(t) = \gamma_j \text{ for } t \in (\sum_{l=0}^{j-1} t_l, \sum_{l=0}^{j} t_l),$$

is a piecewise-constant input to a rational system Σ. The length of the time interval on which v is defined is denoted by $t_v = \sum_{l=0}^{k} t_l$. For the input v, the dynamics of rational systems we consider, see Definition 3, is given as $\dot{x}(t) = f(x(t), v(t))$, where $f : \mathbb{R}^n \times \mathbb{R}_+^{n_{ex}+m} \to \mathbb{R}^n$ is such that for every γ_j, $j = 1, \ldots, k$, the components $f_{\gamma_j,i}$, $i = 1 \ldots n$ of the vector field $f_{\gamma_j} : \mathbb{R}^n \ni x \mapsto f(x, \gamma_j) \in \mathbb{R}^n$ are rational functions. The output functions are given either as measurements of concentrations $y(t) = Cx(t)$ or as outflows of the systems $y(t) = H\mathrm{Diag}(r(x(t), x_{ex}))u(t)$.

To sum up, to deal with the realization problem for rational systems modeling biochemical reaction networks we restrict attention to rational systems Σ,

- defined on irreducible real affine varieties $X \subseteq \mathbb{R}^n$ (state spaces are the sets of zero points of finitely many polynomials),
- with the dynamics given by the family of rational vector fields $f_\gamma = \sum_{i=1}^{n} f_i(x, \gamma) \frac{\partial}{\partial x_i} = \sum_{i=1}^{n} f_{\gamma,i}(x) \frac{\partial}{\partial x_i}$, where $\gamma \in X_{ex} \times U$,
- with an output map $y = h(x)$, having components, which are rational functions,
- with an initial state $x_0 \in X$.

We use the notation $\Sigma = (X, f, h, x_0)$. Note that in case of measurements of

concentrations taken as outputs, the output map h is linear in x. In case outflows of the system are considered outputs (with the assumption that $x_{ex}(t)$ and $u(t)$ are constant in time), h is rational in x. Further, note that these rational systems we consider to deal with realization problem are not necessarily positive.

5.2. Controllability

There are several notions of controllability introduced for nonlinear systems. For the class of rational systems we adopt the notion introduced in Bartosiewicz (1987b), i.e. controllability refers to the density in terms of the Zariski topology of reachable sets in state spaces (in Zariski topology, closed sets are varieties). This concept seems more natural for the systems with such obvious algebraic structure as rational systems.

Let us recall that the state trajectory of a rational system $\Sigma = (X, f, h, x_0)$, corresponding to an input $v = (\gamma_1, t_1)(\gamma_2, t_2) \cdots (\gamma_k, t_k)$ is a continuous piecewise-differentiable function $x_{\Sigma,v}(\cdot) : [0, t_v] \to X$ such that $x_{\Sigma,v}(0) = x_0$ and $\frac{d}{dt} x_{\Sigma,v}(t) = f(x_{\Sigma,v}(t), v(t))$ for $t \in [0, t_v]$. The set of all piecewise-constant inputs for which there exists a trajectory of Σ is denoted by $\mathcal{U}(\Sigma)$.

DEFINITION 6. *A rational system $\Sigma = (X, f, h, x_0)$ is called* controllable *if the reachable set $\mathcal{R}(x_0) = \{x_{\Sigma,v}(t_v) \mid v \in \mathcal{U}(\Sigma)\}$ is Zariski-dense in X, i.e. the smallest variety which contains $\mathcal{R}(x_0)$ equals X.*

It is possible to provide an algebraic characterization of such concept of controllability. For polynomial systems it was derived in Bartosiewicz (1988). The extension for rational systems, presented in Němcová and van Schuppen (2010), is as follows:

PROPOSITION 1. *Consider a rational system $\Sigma = (X, f, h, x_0)$. If Σ is controllable then there are no ideals $(0) \neq I \subseteq A$, where A denotes all polynomials on X, such that for every $\varphi \in I$ it holds that $\varphi(x_0) = 0$ and $\{$ (numerator of $f_\gamma \varphi$) $\in A \mid \gamma \in X_{ex} \times U\} \subseteq I$.*

With stronger assumptions on the inputs (see the condition (iii) in Section 5.4), the condition above is also sufficient for controllability. However, checking this condition can be very difficult. Some relations with other concepts of controllability, see for example Isidori (2001), Nijmeijer and van der Schaft (1990), are determined in Němcová and van Schuppen (2010). Moreover, if the external concentration is assumed constant, the considered rational system is a nonlinear system affine in inputs and the Lie algebraic approach to check controllability can be applied directly.

5.3. Observability

Observability refers to the property of being able to determine the state of a system from its past inputs and past outputs. There are several approaches to study observability introduced in the literature. In case of rational systems

we focus on algebraic approaches. There are two slightly different concepts of algebraic observability known. The first one, which strongly relies on the notions of differential algebra, was introduced by Diop and Fliess in (1991). It refers to the existence of algebraic relations between state variables and successive derivatives of inputs and outputs. According to the second notion, developed by Sontag in (1979) for discrete-time polynomial systems, a system is algebraically observable if the initial state can be expressed as a polynomial function of finitely many observations. Sontag's notion of algebraic observability was adopted and further developed by Bartosiewicz in Bartosiewicz (1985, 1988) for continuous-time polynomial systems and in Bartosiewicz (1987b) for continuous-time rational systems. Let us recall this notion below. A similar approach to study observability of nonlinear systems was used in Hermann (1977), Baillieul (1980, 1981) and Bartosiewicz (1995).

Definition 7. *Let $\Sigma = (X, f, h, x_0)$ be a rational system. Let Q denote the field of rational functions on X. The* observation algebra $A_{obs}(\Sigma)$ *of Σ is the smallest subalgebra of Q which contains the components of h and which is closed with respect to the derivatives determined by the vector fields f_γ, $\gamma \in X_{ex} \times U$. In particular, $A_{obs}(\Sigma) = \mathbb{R}[\{h, f_\gamma h \mid \gamma \in X_{ex} \times U\}]$. The* observation field $Q_{obs}(\Sigma)$ *of Σ is the field of quotients of $A_{obs}(\Sigma)$, i.e. $Q_{obs}(\Sigma) = \mathbb{R}(\{h, f_\gamma h \mid \gamma \in X_{ex} \times U\})$.*

Definition 8. *A rational system Σ is called* observable *if $Q_{obs}(\Sigma) = Q$, where Q denotes the field of rational functions on the state space of Σ.*

5.4. Problem formulation, response maps

In order to formulate the realization problem for the class of rational systems specified in Section 5.1, we have to define the notion of a response map. To this end, we have to recall the notion of an admissible input set $\mathcal{U}$.

Let $\gamma_i = (\alpha_i, \beta_i) \in X_{ex} \times U \subseteq \mathbb{R}_+^{n_{ex}+m}$, $i = 1, \ldots, I$ and let $v = (\gamma_1, t_1)(\gamma_2, t_2) \cdots (\gamma_I, t_I)$ be a piecewise-constant input to a rational system Σ such that the corresponding trajectory exists on the interval $[0, t_v]$. A set $\mathcal{U}$ of such piecewise-constant inputs is called a *set of admissible inputs* if:

(i) $\forall v \in \mathcal{U}\ \forall t \in [0, t_v] : v_{[0,t]} \in \mathcal{U}$, ($v_{[0,t]}$ denotes the input v restricted to the time-domain $[0, t]$),

(ii) $\forall v \in \mathcal{U}\ \forall \gamma \in X_{ex} \times U\ \exists t > 0 : (v)(\gamma, t) \in \mathcal{U}$,

(iii) $\forall v = (\gamma_1, t_1) \cdots (\gamma_k, t_k) \in \mathcal{U}\ \exists \delta > 0\ \forall\, \overline{t_i} \in [0, t_i + \delta], i = 1, \ldots, k :$
$$\overline{v} = (\gamma_1, \overline{t_1}) \cdots (\gamma_k, \overline{t_k}) \in \mathcal{U}.$$

These conditions are technical, but necessary for the development of realization theory for rational systems.

Let us assume for the rest of this section that $\mathcal{U}$ is a set of admissible inputs.

We say that $p : \mathcal{U} \to \mathbb{R}^r$ is a *response map* if for every component $p_i, i = 1, \ldots, r$ of p and for every sequence of input values $\gamma_1, \ldots, \gamma_k \in X_{ex} \times U$, $k > 0$, the function $p_{i;\gamma_1,\ldots,\gamma_k}$ is analytic. The definition of $p_{i;\gamma_1,\ldots,\gamma_k}$ is as follows:

Denote by $\mathbb{T}_{\gamma_1,\ldots,\gamma_k}$ the set of all k-tuples $(t_1,\ldots,t_k) \in [0,+\infty)^k$ such that the input $(\gamma_1,t_1)\cdots(\gamma_k,t_k)$ belongs to $\mathcal{U}$. Then, for all $(t_1,\ldots,t_k) \in \mathbb{T}_{\gamma_1,\ldots,\gamma_k}$, the map $p_{i;\gamma_1,\ldots,\gamma_k} : \mathbb{T}_{\gamma_1,\ldots,\gamma_k} \to \mathbb{R}$ is defined by $p_{i;\gamma_1,\ldots,\gamma_k}(t_1,\ldots,t_k) = p_i((\gamma_1,t_1)\cdots(\gamma_k,t_k))$.

One can prove that the set $\mathcal{A}(\mathcal{U}\to\mathbb{R})$ of all real functions defined on $\mathcal{U}$, which are analytic with respect to switching times of the inputs of $\mathcal{U}$, is an integral domain (for this proof the conditions (i)-(iii) above are necessary). This allows for taking the quotient field of $\mathcal{A}(\mathcal{U}\to\mathbb{R})$, denoted by $\mathcal{Q}(\mathcal{U}\to\mathbb{R})$.

Let us formulate the realization problem.

PROBLEM 1. *Let $p : \mathcal{U} \to \mathbb{R}^r$ be a given response map. The realization problem for the class of rational systems consists of three sub-problems:*

- *find a rational system Σ, which is a realization of p. In particular, find a rational system $\Sigma = (X, f, h, x_0)$ such that $\mathcal{U} \subseteq \mathcal{U}(\Sigma)$ and $p(u) = h(x_{\Sigma,v}(t_v))$ for all $v \in \mathcal{U}$;*
- *find a rational realization which is controllable, observable, minimal;*
- *provide constructive procedures for the first two problems.*

The results on the realization problem for rational systems (without positivity constraint) are presented in Němcová and van Schuppen (2009, 2010). Note that the approach applied in those papers is similar to the approach in Bartosiewicz (1988), Jakubczyk (1980), Sontag (1979), Bartosiewicz and Pawluszewicz (2008), where the response maps are considered in the realization problem. One could also consider input-output maps, equations of higher orders, power series instead of response maps. However, for the class of rational systems, except of the approach based on response maps, we are aware of only one other approach which is the approach based on formal power series presented in Wang and Sontag (1992). In that paper only the existence part of the realization problem is considered.

To point out the differences between Němcová and van Schuppen (2009, 2010) and Bartosiewicz (1988), Jakubczyk (1980), Sontag (1979), Bartosiewicz and Pawluszewicz (2008), let us mention the following: in Bartosiewicz (1988) and Sontag (1979) the realization problem for polynomial discrete-time systems and polynomial continuous time systems is considered, with Bartosiewicz (1988) generalizing the results of Sontag (1979). The existence of a polynomial realization for a response map due to Bartosiewicz (1988) is implied by the existence of a rational realization due to Němcová and van Schuppen (2009) since polynomial systems are a subclass of rational systems. However, the opposite implication does not hold. Further, in Němcová and van Schuppen (2009, 2010) a slightly different notion of observability was introduced and the respective relations with minimal realizations were derived.

Compared to Němcová and van Schuppen (2009, 2010), in Bartosiewicz and Pawluszewicz (2008) only the existence part of the realization problem is considered. On the other hand, in Bartosiewicz and Pawluszewicz (2008) more general classes of systems are considered (continuous of order k, smooth and analytic

systems on time scales). The proofs of the existence of realizations in Bartosiewicz and Pawluszewicz (2008) and in Němcová and van Schuppen (2009) are similar since they are both related to Bartosiewicz (1988). Nevertheless, the results are not equivalent since the system classes are different. If the conditions for the existence of rational realizations stated in Němcová and van Schuppen (2009) are fulfilled, then the conditions in Bartosiewicz and Pawluszewicz (2008) hold and a realization within k-continuous, smooth, analytic systems exists. The opposite implication does not hold.

5.5. Existence and minimality of realizations

We will provide an overview of the main results on the existence and minimality of rational realizations according to Němcová and van Schuppen (2009, 2010).

To state the characterization of the existence of rational realizations, let us recall the notion of the observation field $Q_{obs}(p)$ for a response map p. It is derived in a similar way as the observation field of a system in Definition 7. Specifically, it is the smallest field over $\mathbb{R}$, which contains the components of p and is closed with respect to D_γ derivatives. By D_γ derivative of $\varphi \in \mathcal{A}(\mathcal{U} \to \mathbb{R})$, where $\gamma \in X_{ex} \times U$, we refer to the derivative $(D_\gamma\varphi)(v) = \frac{d}{dt}\varphi((v)(\gamma,t))|_{t=0+}$, where $(v)(\gamma,t) \in \mathcal{U}$. Note that for $D_\gamma\varphi$ to be well-defined, the assumptions (i), (ii) on $\mathcal{U}$ stated in Section 5.4 are necessary.

Theorem 2. *Let $p : \mathcal{U} \to \mathbb{R}^r$ be a response map. There exists a rational realization of p if and only if $Q_{obs}(p)$ is a finite field extension of $\mathbb{R}$.*

The proof of this theorem is constructive. To implement the proposed procedure one has to overcome several obstacles of computer algebra. For example, how to determine the generators for the observation field of the response map if one knows only the rule how to construct it by recursive differentiation, how to find a rational combination of the generators of the observation field such that it equals a given function, etc.

In a constructive way it is also proven that the existence of rational realization of a response map is equivalent to the existence of observable rational realization and even to the existence of realization which is both observable and controllable.

Another important property of realizations is their dimension. By the dimension dim Σ of a rational system Σ we refer to the dimension of its state space X. Because X is an irreducible real affine variety, dim Σ = dim X equals the transcendence degree of the field Q of rational functions defined on X, i.e. it equals the maximal number of algebraically independent elements of Q over $\mathbb{R}$.

Definition 9. *A rational realization Σ of a response map p is called* minimal *if its dimension is minimal within dimensions of all rational realizations of p, i.e. if there does not exist a rational realization Σ' of p such that* dim Σ' < dim Σ.

In Němcová and van Schuppen (2010) the following characterization of minimal realizations was derived:

THEOREM 3. *A rational realization* Σ *of a response map* p *is minimal if and only if* $\dim\ X = \mathrm{trdeg}\ Q_{obs}(p)$.

For linear realizations the property of being minimal and the property of being canonical (both controllable and observable) are equivalent. For rational systems one obtains a weaker relation.

THEOREM 4. *Let* Σ *be a rational realization of a response map* p *such that* $Q \setminus Q_{obs}(\Sigma)$ *are not algebraic over* $Q_{obs}(\Sigma)$. *Then,* Σ *is a minimal realization of* p *if and only if* Σ *is controllable and observable.*

5.6. Nash systems and positivity

The positivity was not treated in this section. It creates extra complications for the formulation of the realization problem for rational systems and its solution.

Regular rational systems are a subclass of so-called Nash systems. These systems are continuous-time dynamic systems defined on analytic manifolds (which can be described by finitely many polynomial equalities and inequalities) and with dynamics and outputs determined by analytic functions satisfying an algebraic equation. This class of systems, unlike rational systems, includes the systems with the state space being $(0, \infty)^n$, which is a natural domain for models of biochemical processes. Therefore, it seems reasonable to assume that the class of Nash systems is more suitable to deal with positivity. Preliminary results on realization problem for such systems are derived in Němcová, Petreczky and van Schuppen (2009).

5.7. System identification

By applying the results of realization theory for rational systems it was possible to derive results on identifiability of parametrizations of rational systems with parameters, see Němcová (2010). Identifiability refers to the fact that the parameter values of a model can be determined uniquely from input-output data. There are many concepts of identifiability introduced in the literature and many approaches to study them, for example the approach based on power series expansions of outputs (Pohjanpalo, 1978), differential-algebraic approach (Ljung and Glad, 1994), generating series approach (Walter, 1982), and similarity transformation method (Travis and Haddock, 1981). Identifiability for a class of nonlinear systems which includes the class of rational systems, is studied also in Xia and Moog (2003), but in a linear algebraic setting, which is related to the differential-algebraic approach. Other papers dealing with identifiability of polynomial and rational systems, but without inputs, are Denisvidal, Jolyblanchard and Noiret (2001), Evans et al. (2002).

The characterization of identifiability of parametrizations for rational systems is provided in Němcová (2010) by specifying the properties of certain isomorphisms of systems. In particular, it is a direct consequence of the fact that realizations (of the same response map), which are both controllable and observable, are isomorphic.

Identifiability of parametrizations of rational positive systems is a topic of future research. It is expected that once the realization theory for these systems is developed, the characterization of identifiability will be derived in the same way as for general rational systems.

6. Control of biochemical reaction systems

Cell biologists and biomedical engineers want to know how a cell controls its biochemical reaction network. Biologists and biochemists have a theory that the nucleus produces RNA, RNA produces enzymes, and that enzymes catalyze the individual chemical reactions. The control of the reaction network by the nucleus is therefore represented in the biochemical reaction system by the input signal of the enzyme concentration.

A biochemical reaction network uses a large number of enzymes. Many of these enzymes are present in the cell at nontrivial concentrations. The corresponding reactions can therefore take place continuously. If a particular enzyme is present in a rather low concentration or not at all, then the corresponding reaction will take place at a low rate or not at all, respectively. It is therefore of interest for the understanding of control of the biochemical reaction network to develop a mathematical model and control theory for the enzyme control of such a network.

There is another reason for the interest in control of these networks. The effect of a medicine on the biochemical network is often the inhibition of a particular reaction. *Inhibition* is the name for the following phenomenon. For certain enzymes, a particular chemical species injected into the cell takes the binding place on the enzyme where otherwise a particular reaction takes place in the form of an assembly of a new molecule. Moreover, once the binding site is occupied by the molecule, it never goes away. The effect of this binding is then that the regular reaction cannot take place, is inhibited. It is therefore of interest for the use of medicine and drugs to determine which reactions, when inhibited partly or completely, have as effect that components of the outflow rate of the network become zero or at least small. When these reactions are known then the biochemists can search for appropriate chemical compounds which are suitable for binding to the enzymes at the binding places.

There are several ways to study the effect of enzyme control on a biochemical reaction network: (1) Metabolic control coefficients; (2) Metabolic pathway analysis with the concept of elementary flux modes; and (3) A control theoretic way, formulated below. The first two approaches are discussed first.

In the research group of Hans Westerhoff of the Vrije Universiteit, a particular method for the sensitivity of reaction networks to inhibition has been developed, called *metabolic control analysis*. Control theorists describe this method as sensitivity analysis. The sensitivity is a local way of determining the effect of inhibition. For a control theoretic way of formulating sensitivity coefficients, see Sontag (2002).

The method of metabolic pathway analysis with the concept of elementary

flux modes is treated in Klamt and Stelling (2003), Schuster et al. (2002) and Stelling et al. (2002).

In this paper a different method is proposed, namely to evaluate the effect on the network of enzyme concentrations which are zero or almost zero.

PROBLEM 2. Control of a biochemical reaction system. *Consider a biochemical reaction system. Determine how control of this system by enzyme inputs can be effectively executed. Subproblems are:*

(a) Determine the effect of a zero enzyme input component on the dynamics of the biochemical reaction network and on the continuous dynamics.

(b) Determine which input components (enzymes) most effectively control the network at low concentrations.

(c) What are the controllability and the accessibility properties of these systems, see Sontag (1998).

PROBLEM 3. Effect of zero enzyme input on the steady outflow rate. *Consider a positive rational system as defined in Definition 3. Consider an index set* $I \subseteq \mathbb{Z}_m$ *and a constant input function,*

$$\begin{aligned} & u : T \to U = \mathbb{R}_+^m, \text{ such that,} \\ u_i(t) &\equiv 0, \ \forall i \in I, \ \forall t \in T, \\ u_j(t) &= u_{s_j} \in U, \ j \in \mathbb{Z}_m \backslash I, \ \forall t \in T. \end{aligned}$$

Determine the effect of this input function on the steady state $x_s \in X$ *and the steady outflow rate* $z_s \in \mathbb{R}_+^{n_z}$ *associated with this input. Is the outflow rate identically zero?*

The problem is motivated by inhibition of a biochemical reaction network as described above. Almost always only one component of the input is zero, corresponding to one enzyme. It is of interest and relatively novel to consider the problem also for constant input vectors with two or more components equal to zero corresponding to a medical drug with two chemical species.

Consider first the case in which only one input component is set to zero. This problem can be answered by first constructing the graph matrix of the system obtained from the original biochemical reaction system by setting the enzyme input component u_i to zero. The result is displayed for the example.

EXAMPLE 6. *Consider Example 4. Assume that the enzyme input component* $u_4 = 0$. *A computation then yields that the graph matrix of the remaining system equals,*

$$A = \left(\begin{array}{ccc|ccc|cc} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right). \tag{22}$$

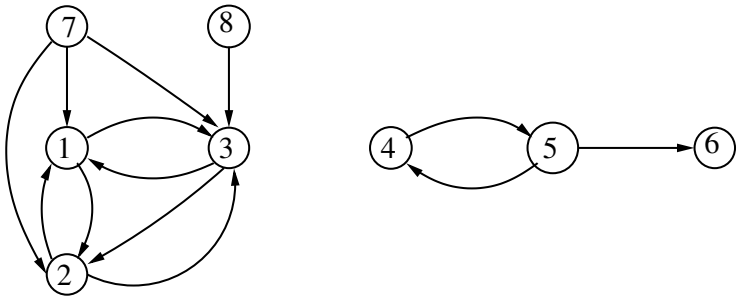

Figure 3. The diagram of the system graph in case of $u_4 = 0$

The conclusion is that the graph has become disconnected with two components. There is no path leading to the output node 6 *from either of the external concentration nodes,* $7, 8$*. Thus, the production of the network stops when reaction 4 is blocked. The diagram of the graph for this case is displayed in Fig. 3.*

Proposition 2. *Consider a biochemical reaction system as defined in Definition 3 with* $n_z = 1$ *and assume that all external exogeneous inputs are required for the single outflow component. If for* $i \in \mathbb{Z}_m$ *the enzyme input components* $u_i(t) = 0$ *for all* $t \in T$ *and if for the biochemical reaction system with this enzyme input there does not exist a path from each of the required external input nodes to the considered outflow node, then the steady outflow rate is identically zero.*

Proof Follows directly by consideration of the differential equation for this input. □

If the graph has not become disconnected, then one has to determine the steady state and the steady state outflow rate for the enzyme input considered to check whether $z_s = 0$. From the example it appears that the outflow rate may not be zero hence in general one has to compute the outflow rate for the equilibrium state.

Consider Problem 3 of determining a subset of systems which, when disabled, produces a zero outflow rate after a transient period. This problem with the above analysis can be mathematically formulated as a cut set problem of graph theory or of combinatorial optimization.

Consider a graph with one or more inflow nodes or origins and one or more outflow nodes or destinations. A *cutset* of a graph is a subset of the edges which, when deleted from the graph, result in a graph with two or more disconnected components with one component containing all outflow nodes and no inflow nodes, and another component containing inflow nodes but no outflow nodes.

Problem 4. *The* minimal cut set problem *for a graph is to determine a cut set with the minimal number of cuts, corresponding to the minimal number of reactions to be disabled.*

In the literature there are algorithms to compute minimal cut sets of a graph.

The second control problem for a biochemical reaction system aims at clarifying how the cell controls the outflow rate of a biochemical reaction network.

DEFINITION 10. *Consider a biochemical reaction system as defined in Definition 3. Assume that for every $u_s \in U$ there exists a unique outflow rate $z_s \in \mathbb{R}_+^{n_z}$, denoted by $z_s(u_s)$. Consider a* required outflow rate *$z_r \in \mathbb{R}_+^{n_z}$. Define the set of* required inputs *for this outflow rate,*

$$U(z_r) = \{u \in U | z_s(u) \geq z_r\}. \tag{23}$$

Thus, $U(z_r)$ consists of all input vectors $u \in U$ such that a constant input function $u(t) = u_s \in U$ for all $t \in T$, produces in steady state an outflow rate $z_s(u_s)$ which is larger than or equal to z_r. Define the infimal input *which is minimally necessary to obtain the outflow rate z_r by, for all $i \in \mathbb{Z}_m$,*

$$u_{min,i}(z_r) = \inf_{u \in U(z_r)} u_i, \tag{24}$$

$$u_{min}(z_r) = (u_{min,1}(z_r), \ldots, u_{min,m}(z_r)) \in \mathbb{R}_+^m. \tag{25}$$

For any particular input $u \in U(z_r)$ define the index set of limiting input components

$$I(z_r, u) = \{i \in \mathbb{Z}_m | u_i = u_{min,i}(z_r)\}. \tag{26}$$

PROBLEM 5. Outflow rate limiting enzymes. *For a biochemical reaction system and a required outflow rate $z_r \in \mathbb{R}_+^{n_z}$, determine the set of required inputs $U(z_r)$, the infimal input, and, for a range of inputs $U_r \subseteq U(z_r)$, the index set of limiting input components $I(z_r, u)$.*

The motivation for this problem is the study of how the nucleus controls a biochemical reaction network. Can the biochemical reaction system provide information about the index set of limiting input components? If so, then the enzymes corresponding to those components can be used by the nucleus for control. Note that several enzymes are available in abundance in the cell or in part of the cell, hence in normal circumstances those components are never limiting the outflow rate.

Note that in general $u_{min} \notin U(z_r)$. A question is whether the map $u \mapsto z_s(u)$ is *monotone*: if $u, \bar{u} \in U$ with $u \leq \bar{u}$ is then $z_s(u) \leq z_s(\bar{u})$, where the inequalities for vectors hold only if they hold for all components? Is $U(z_r)$ a convex set? Further research on these problems is needed.

7. Concluding remarks

The paper presents the class of biochemical reaction systems as mathematical models of biochemical cell reaction networks. Results are presented for: the positive invariance of the positive orthant for the differential equation; the computation and interpretation of the graphs of such networks; the closure of the

class of biochemical reaction systems with respect to an interconnection relation; and for control of such systems by enzyme inputs.

Further research is required for:

- System theory of biochemical reaction systems. Computational procedures for determination of controllability and of observability of rational systems; and of controllable and observable subsystems.
- Decomposition of biochemical reaction systems in particular, concepts and algorithms.
- System identification. A characterization of identifiability of biochemical reaction systems based on the reference Němcová (2010). Algorithms for system identification.
- System reduction of biochemical reaction systems.
- Control of a biochemical reaction system with as input external concentrations or as input the enzyme concentration. Characterization of controllability and control laws.

Acknowledgements

The authors acknowledge the stimulation and advice provided by Hans V. Westerhoff and Barbara M. Bakker both of the Department of Cell Physiology, of the Faculty of Earth and Life Sciences, of the Vrije Universiteit in Amsterdam. They also acknowledge discussions on the topic of this paper with Mr. Siddhartha Jha during an internship of the latter at CWI in the Summer of 2001, with Prof. André C.M. Ran of VU University Amsterdam, and with Ms. Ilona Verburg during her master project. They also thank several unnamed researchers for comments on drafts of the paper.

References

Baillieul, J. (1980) The geometry of homogeneous polynomial dynamical systems. *Nonlinear Anal. – Theory, Meth. and Appl.* **4**, 879–900.

Baillieul, J. (1981) Controllability and observability of polynomial dynamical systems. *Nonlinear Anal. – Theory, Meth. and Appl.* **5**, 543–552.

Bakker, B.M. (1998) *Control and regulation of glycolis in Trypanosoma brucei.* PhD thesis, Vrije Universiteit, Amsterdam.

Bartosiewicz, Z. (1985) A new setting for polynomial continuous-time systems, and a realization theorem. *IMA J. Mathematical Control Inf.* **2**.

Bartosiewicz, Z. (1987) Rational systems and observation fields. *Systems & Control Lett.* **9**, 379–386.

Bartosiewicz, Z. (1988) Minimal polynomial realizations. *Math. Control Signals Systems* **1**, 227 – 237.

Bartosiewicz, Z. (1995) Local observability of nonlinear systems. *Systems & Control Lett.* **25**, 295 – 298.

Bartosiewicz, Z. and Pawluszewicz, E. (2008) Realizations of nonlinear control systems on time scales. *IEEE Trans. Automatic Control* **53**, 571–575.

Berman, A., Neumann, M. and Stern, R.J. (1989) *Nonnegative Matrices in Dynamic Systems*. John Wiley & Sons, New York.

Birkhoff, G. and MacLane, A. (1977) *A Survey of Modern Algebra, fourth edition*. MacMillan Publ. Co. Inc., New York.

Campbell, N.A., Reece, J.B. and Mitchell, L.G. (1999) *Biology (Fifth Ed.)*. Addison Wesley Longoman Inc., Menlo Park.

Cox, D.A., Little, D.A. and O'Shea, D. (1992) *Ideals, Varieties, Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer, Berlin.

Denis-Vidal, L., Joly-Blanchard, G. and Noiret, C. (2001) Some effective approaches to check the identifiability of uncontrolled nonlinear systems. *Mathematics and Computers in Simulation* **57**, 35–44.

Diop, S. and Fliess, M. (1991) On nonlinear observability. In: *Proc. 1st European Control Conf.*, Paris. EUCA, Hermès, 152–157.

Evans, N.D., Chapman, M.J., Chappell, M.J. & Godfrey, K.R. (2002) Identifiability of uncontrolled nonlinear rational systems. *Automatica* **38**, 1799–1805.

Fall, C., Marland, E., Wagner, J. and Tyson, J. (2002) *Computational Cell Biology. Interdisciplinary Mathematics*. Springer, New York, **20**.

Farina, L. and Rinaldi, S. (2000) *Positive Linear Systems: Theory and Applications*. Pure and Applied Mathematics. John Wiley & Sons, New York.

Feinberg, M. (1987) Chemical reaction network structure and the stability of complex isothermal reactors I. The deficiency zero and the deficiency one theorems. *Chemical Engineering Science* **42**, 2229–2268.

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

Feinberg, M. and Horn, F.J.M. (1974) Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Engrg. Sci.* **29**, 775–787.

Feinberg, M. and Horn, F.J.M. (1977) Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Arch. Rational Mech. Anal.* **66**, 83–97.

Fife, D. (1972) Which compartmental systems contain traps? *Math. Biosciences* **14**, 311–315 .

Foster, D. and Jacquez, J.A. (1975) Multiple zeroes for eigenvalues and the multiplicity of traps in a linear compartmental system. *Math. Biosciences* **26**, 89–97.

Gerstenhaber, M. (1951) Theory of convex polyhedral cones. In: Tj.C. Koopmans, ed., *Activity Analysis of Production and Allocation*. Wiley & Sons, New York, 298–316.

Gondran, M. and Minoux, M. (1984) *Graphs and Algorithms*. John Wiley & Sons, Chichester.

Heinrich, R. and Schuster, S. (1996) *The Regulation of Cellular Systems.* Chapman and Hall, New York.

Helfert S., Estévez A., Bakker B., Michels P. & Clayton Ch. (2001) Roles of triosephosphate isomerase and aerobic metabolism in Trypanosoma brucei. *Biochem. J.* **357**, 117–125.

Hermann, R. (1977) *Algebro-geometric and Lie-theoretic Techniques in Systems Theory – Part A. Interdisciplinary Mathematics* **XIII**, Math. Sci Press.

Isidori, A. (1989) *Nonlinear Control Systems - An Introduction, 2nd Edition.* Springer-Verlag, Berlin.

Isidori, A. (1999) *Nonlinear Control Systems II.* Communications and Control Engineering Series. Springer-Verlag, London.

Isidori, A. (2001) *Nonlinear Control Systems (3rd. ed.).* Springer, Berlin.

Jacobson, N. (1985) *Basic Algebra, Volume 1 (2nd edition).* W.H. Freeman and Company, New York.

Jacquez, J.A. (1985) *Compartmental Analysis in Biology and Medicine, 2nd Ed.* The University of Michigan Press, Ann Arbor.

Jacquez, J.A. and Simon, C.P. (1993) Qualitative theory of compartmental systems. *SIAM Rev.* **35**, 43–79.

Jakubczyk, B. (1980) Existence and uniqueness of realizations of nonlinear systems. *SIAM J. Control & Opt.* **18**, 455–471.

Klamt, S. and Stelling, J. (2003) Two approaches for metabolic pathway analysis. *Trends in Biotechnology* **21**, 64–69.

Lam, T.Y. (1999) *Lectures on Modules and Rings. Graduate Texts in Mathematics* **189**, Springer, Berlin.

Ljung, L. and Glad, T. (1994) *Modeling of Dynamic Systems.* PTR Prentice Hall, Englewood Cliffs.

Murota, K. (2000) *Matrices and matroids for systems analysis. Algorithmics and Combinatorics* **20**. Springer-Verlag, Berlin.

Němcová, J., Petreczky, M. and van Schuppen, J.H. (2009) Realization theory of Nash systems. In: *Proc. 48th IEEE Conference on Decision and Control (CDC.2009)*, New York. IEEE Press, 5935–5940.

Němcová, (2010) Structural identifiability of polynomial and rational systems. *Math. Biosci.* **223**, 83–96.

Němcová, J. and van Schuppen, J.H. (2010) Realization theory for rational systems: Minimal rational realizations. *Acta Applicandae Mathematicae* **110**, 605–626.

Němcová, J. and van Schuppen, J. (2009) Realization theory for rational systems: The existence of rational realizations. *SIAM J. Control & Opt.* **48**, 2840 – 2856.

Nijmeijer, H. and van der Schaft, A.J. (1990) *Nonlinear Dynamical Control Systems.* Springer-Verlag, Berlin.

Pohjanpalo, H. (1978) System identifiability based on the power series expansion of the solution. *Math. Biosci.* **41**, 21–33.

Reinschke, K.J. (1988) *Multivariable Control - A Graph-theoretic Approach, Lecture Notes in Control and Informations Sciences* **108**. Springer-Verlag, Berlin.

Rockafellar, R.T. (1970) *Convex Analysis.* Princeton University Press, Princeton.

Rohwer, J.M. (1997) *Interaction of Functional Units in Metabolism.* PhD thesis, Vrije Universiteit, Amsterdam.

Schuster, S., Hilgetag, C., Woods, J.H. & Fell, D.A. (2002) Reaction routes in biochemical reacton systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* **45**, 153–181.

Sontag, E.D.(1979) *Polynomial Response Maps. Lecture Notes in Control and Information Sciences* **13**, Springer-Verlag, Berlin.

Sontag, E.D. (1979) Realization theory of discrete-time nonlinear systems: I. The bounded case. *IEEE Trans. Circuits & Systems* **26**, 342–356.

Sontag, E.D. (1998) *Mathematical Control Theory: Deterministic Finite Dimensional Systems (2nd. Ed.). Graduate Texts in Applied Mathematics* **6**, Springer, New York.

Sontag, E.D. (2001) Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. *IEEE Trans. Automatic Control* **46**, 1028–1047.

Sontag, E.D. (2002) Asymptotic amplitutes and Cauchy gains: A small-gain principle and an application to inhibitory biological feedback. *Systems & Control Lett.* **47**, 167–179.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193.

Sussmann, H.J.(1979) Single-input observability of continuous-time systems. *Math. Systems Theory* **12**, 371–393.

Travis, C.C. and Haddock, G.(1981) On structural identification. *Math. Biosci.* **56**, 157–173.

van den Hof, J.M.(1996) *System theory and system identification of compartmental systems.* PhD thesis, University of Groningen, Groningen.

van Schuppen, J. H.(2004) System theory of rational positive systems for cell reaction networks. In: Bart De Moor et al., ed., *Proc. MTNS.2004*, Leuven. Katholieke Universiteit Leuven.

Walter, E. (1982) *Identifiability of State Space Models. Lecture Notes in Biomathematics* **46**, Springer-Verlag, Berlin.

Wang, Y. and Sontag, E.D.(1992) Algebraic differential equations and rational control systems. *SIAM J. Control & Opt.* **30**, 1126–1149.

Xia, X. and Moog, C. (2003) Identifiability of nonlinear systems with application to hiv/aids models. *IEEE Trans. Automatic Control* **48**, 330–336.