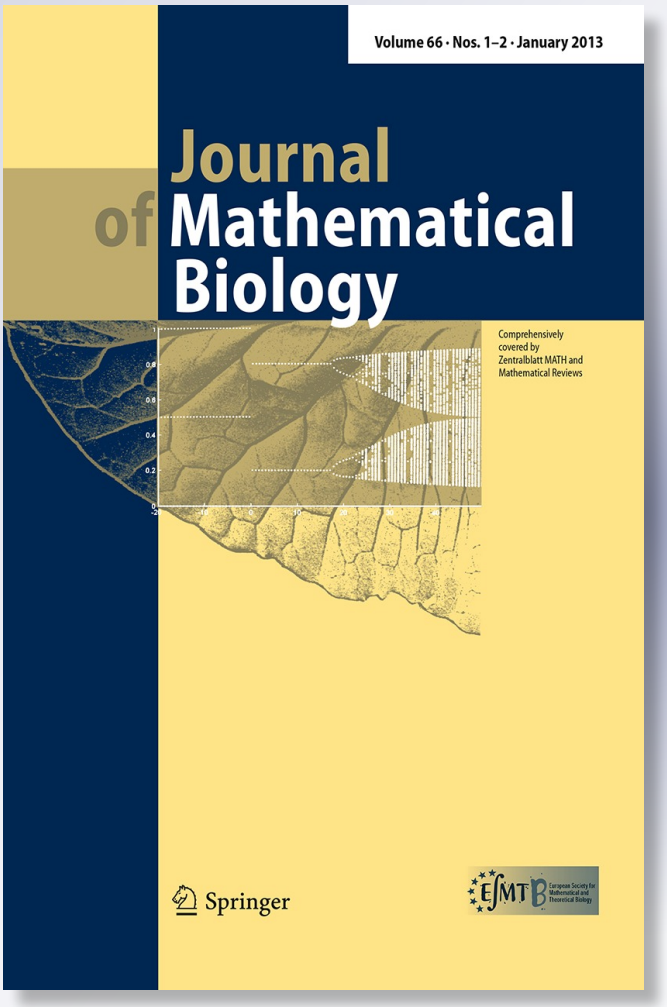
Variable elimination in post-translational modification reaction networks with mass-action kinetics

Elisenda Feliu & Carsten Wiuf

Journal of Mathematical Biology

ISSN 0303-6812
Volume 66
Combined 1-2

J. Math. Biol. (2013) 66:281-310
DOI 10.1007/s00285-012-0510-4



Springer

J. Math. Biol. (2013) 66:281–310
DOI 10.1007/s00285-012-0510-4

Mathematical Biology

Variable elimination in post-translational modification reaction networks with mass-action kinetics

Elisenda Feliu · Carsten Wiuf

Received: 29 July 2011 / Revised: 1 November 2011 / Published online: 5 February 2012

Abstract We define a subclass of chemical reaction networks called post-translational modification systems. Important biological examples of such systems include MAPK cascades and two-component systems which are well-studied experimentally as well as theoretically. The steady states of such a system are solutions to a system of polynomial equations. Even for small systems the task of finding the solutions is daunting. We develop a mathematical framework based on the notion of a *cut* (a particular subset of species in the system), which provides a linear elimination procedure to reduce the number of variables in the system to a set of *core* variables. The steady states are parameterized algebraically by the core variables, and graphical conditions for when steady states with positive core variables imply positivity of all variables are given. Further, *minimal* cuts are the connected components of the species graph and provide conservation laws. A criterion for when a (maximal) set of independent conservation laws can be derived from cuts is given.

Keywords Polynomial equations · Species graph · MAPK cascade · Rational functions · Chemical reaction networks

Mathematics Subject Classification (2000) 92C42 · 80A30

1 Introduction

Signaling systems play an important role in regulation of cellular processes and are essential for cellular decision making. Typical signaling systems react to stimulus in the (cellular) environment and transmit a signal through connected layers of biochemical

E. Feliu (✉) · C. Wiuf
Department of Mathematics, University of Copenhagen,
Universitetsparken 5, Copenhagen 2100, Denmark
e-mail: efeliu@math.ku.dk

species. The layers provide means to adjust the response according to the stimulus. A common form of signaling systems is Post-Translational Modification (PTM) systems where species are activated in chemical reactions in order to propagate the signal through the system.

PTM systems have attracted considerable theoretical attention due to their abundance in nature (Huang and Ferrell 1996) and regular form (Thomson and Gunawardena 2009). They are a special type of chemical reaction networks and their dynamics can be modeled as $\frac{dx(t)}{dt} = p(x)$, where $x = (x_1, \ldots, x_n)$ are the variables (concentrations of species) of the system and $p(x)$ is a vector of polynomials in x (assuming mass-action kinetics). In particular, small specific systems have been scrutinized, focusing on the dynamical behavior and the steady states of the systems. Examples include the biologically important MAPK cascades (Huang and Ferrell 1996; Kholodenko and Birtwistle 2009; Markevich et al. 2004), as well as simpler signaling cascades systems (Feliu et al. 2012; Feliu and Wiuf 2011a; Heinrich et al. 2002; Ventura et al. 2008).

We focus on the steady states of a PTM system (defined formally in the next section) and how to determine them. Taken with mass-action kinetics, the system's steady states are solutions to a set of polynomial equations in the species and with coefficients given by unknown kinetic rates (i.e. unspecified parameters). In particular, the number of equations to be solved is equal to the number of species. Even small systems might have many variables such that analytical solutions are difficult to obtain and numerical solutions are prone to errors. It is therefore of interest to develop formal procedures to eliminate variables and reduce the number of equations.

In addition, many PTM systems admit multistationarity (the existence of more than one steady state under particular biological conditions) which can be seen as a mechanism for cellular decision making (Markevich et al. 2004). Several non-necessary conditions are known to infer monostationarity (at most one positive steady state) (Angeli et al. 2010; Craciun and Feinberg 2005; Feinberg 1987). However, when these fail, multistationarity is difficult to determine and often decided based on a random parameter search. Procedures to eliminate variables not only assist in reducing the number of equations to solve but also in reducing the number of effective parameters of the system. In this way further mathematical as well as computational manipulation of the steady-state equations can be facilitated.

Our work is inspired by previous work by Thomson and Gunawardena (2009) (see also King and Altman 1956 for an early development of closely related techniques), which we extend to embrace a range of important PTM systems such as signaling cascades, including the MAPK cascade, and two-component systems with phosphorelays and phosphotransfer (Krell et al. 2010), as well as systems with self-interactions. A PTM system has two types of species: substrates ($\mathscr{S}$) and intermediate complexes ($\mathscr{Y}$). We show that the intermediate complexes always can be eliminated from the steady-state equations. Further, we develop the idea of a *cut* $\mathscr{S}_\alpha$, a subset of $\mathscr{S}$ with certain properties that allow us to express the steady-state equations as rational functions in the *core* variables $\mathscr{S}\backslash\mathscr{S}_\alpha$, providing an algebraic parameterization of the steady states in terms of the core variables only.

Elimination of the intermediate complexes is closely linked to the procedure known as the *quasi-steady state approximation* which is often employed to simplify the modeling equations (King and Altman 1956; Cornish-Bowden 2004). Under the quasi-steady

state approximation some reactions are assumed to occur at a much faster rate than other reactions (that is, there is a separation of time scales) such that a steady state effectively has been reached for the fast reactions. For a special class of PTM systems this procedure has been formalized by Thomson and Gunawardena (2009). We show that this elimination can be done generally for PTM systems.

The idea of a cut $\mathscr{S}_\alpha$ enables us to write the steady-state equations as a linear system of equations in the variables in $\mathscr{S}_\alpha$ with coefficients that depend on the reaction rate constants and the core variables $\mathscr{S}\backslash\mathscr{S}_\alpha$. This simplification allows us to relate the steady-state equations to the *Laplacian* of a directed graph that is uniquely determined by the cut and the reactions in the system. Using primarily Tutte's Matrix-Tree theorem (Thomson and Gunawardena 2009; Tutte 1948), we show that properties of the graph determine properties of the steady-state solutions. For example, if the core variables take positive values at steady state then all other concentrations (including those of the intermediate complexes) are either zero or positive as well.

Further, we show that cuts relate to *conservation laws* (conserved quantities that imply that the dynamics takes place in an affine invariant subspace of $\mathbb{R}^n$). These arise as connected components in the species graph and are often used as a first step to reduce the dimensionality of the system. In our approach, conservation laws come into play after elimination of variables from the steady-state equations. In this way, we allow for a larger reduction in the number of core variables and a control of the positivity of the solutions.

The outline of the paper is as follows. In Sect. 2 we review some basic graph properties and in Sect. 3 we define PTM systems, introduce the notion of a cut and a non-interacting graph, and relate the concepts to conservation laws in the system. Section 4 is devoted to variable elimination in the steady-state equations. In Sect. 5 we provide examples of the elimination procedure, focusing mainly on signaling cascades.

2 Preliminaries

2.1 Graphs and the Matrix-Tree theorem

Consider a graph G with set of nodes V and set of edges E. Two nodes v, w are *linked* if there exists a path between them. Being linked is an equivalence relation which decomposes G into subgraphs called *connected components*. In particular, a graph is *connected* if there is a path between every pair of nodes v and w.

Given a directed graph G, one says that G is *strongly connected* if for any pair of nodes v, w there is a directed path from v to w. A *spanning tree* τ of G is a directed subgraph with the same node set as G and such that the corresponding undirected graph is connected and acyclic (that is, contains no cycles). There is a unique undirected path between any two nodes in a spanning tree (Diestel 2005). A spanning tree τ is said to be *rooted* at a node v if the unique path between any node w and v is directed from w toward the root node v. It follows that v is the only node with no out-edges, that is, there is no edge of the form $v \to w$ in τ. In addition, there cannot be a node with two out-edges in τ. Any directed path from v to w in a strongly connected graph can be extended to a spanning tree rooted at w. Some general references for graph theory are Diestel (2005) and Gross and Yellen (2006).

If G is labeled, that is, each edge is assigned a non-zero element (label) in some ring R, then any spanning tree τ inherits a labeling from G and we define

$$\pi(\tau) = \prod_{x \xrightarrow{a} y \in \tau} a.$$

Assume that G has no self-loops. Order the node set $\{v_1, \ldots, v_n\}$ of G and denote by $a_{i,j}$ the label of the edge $v_i \to v_j$. We set $a_{i,j} = 0$ if there is no edge from v_i to v_j (thus $a_{i,i} = 0$). Let $\mathscr{L}(G) = \{\alpha_{i,j}\}$ be the *Laplacian* of G, that is the matrix with

$$\alpha_{i,j} = \begin{cases} a_{j,i} & \text{if } i \neq j \\ -\sum_{k=1}^{n} a_{i,k} & \text{if } i = j, \end{cases}$$

such that the column sums are zero. For each node v_j, let $\Theta(v_j)$ be the set of spanning trees of G rooted at v_j. Then, the Matrix-Tree theorem states that the maximal minor $\mathscr{L}(G)_{(ij)}$ (the determinant of the minor obtained by removing the ith row and the jth column of $\mathscr{L}(G)$) is:

$$\mathscr{L}(G)_{(ij)} = (-1)^{n-1+i+j} \sum_{\tau \in \Theta(v_j)} \pi(\tau).$$

Note that for notational simplicity we have defined the Laplacian as the transpose of how it is usually defined and the Matrix-Tree theorem has been adapted consequently.

3 Post-translational modification systems

In this work we focus on chemical reaction networks that arise in connection with post-translational modification (PTM) mechanisms. The majority of these mechanisms relate to post-translational modification of proteins catalyzed by enzymes, and relay and transfer of modifier groups:

$$E + S \rightleftarrows Y \longrightarrow E + S^* \qquad P^* + S \rightleftarrows Y \rightleftarrows P + S^* \qquad S \to S^*.$$

Here S^*, P^* are modified proteins (substrates), S, P their corresponding unmodified forms, E a substrate acting as an enzyme, and Y an intermediate complex. In the first case, the attachment of the modifier group is catalyzed by the enzyme E and follows a Michaelis–Menten mechanism. In the second case, a modifier group is transferred from P^* to S, whereas in the third case the modification of the substrate does not involve other species. Additionally, intermediate complexes may undergo some transformation before dissociation. For instance, the modification catalyzed by an enzyme can be modeled with the formation of two (or more) intermediate complexes:

$$E + S \rightleftarrows Y_1 \rightleftarrows Y_2 \longrightarrow E + S^*.$$

PTM systems thus consist of two types of species: *substrates* that interact with each other and *intermediate complexes* that are formed by interaction of substrates. Further, a framework for the study of PTM systems needs to incorporate reactions that account for the formation and dissociation of intermediate complexes, as well as the modification of substrates and intermediate complexes without the involvement of other species.

In the next subsection we formally define PTM systems. We proceed in the subsequent subsections to discuss the associated mass-action differential equations and the existence of conserved amounts.

3.1 PTM system

A *post-translation modification (PTM) system* consists of two non-empty sets of species, $\mathscr{S} = \{S_1, \ldots, S_N\}$ (the *substrates*) and $\mathscr{Y} = \{Y_1, \ldots, Y_P\}$ (the *intermediate complexes*) with $\mathscr{S} \cap \mathscr{Y} = \emptyset$, and a set of reactions $\mathrm{Rct} = R_a \cup R_b \cup R_c \cup R_d$ with associated positive reaction rate constants:

$$R_a = \{S_i + S_j \xrightarrow{a^k_{i,j}} Y_k | (i,j,k) \in I_a\} \qquad R_c = \{Y_i \xrightarrow{c_{i,j}} Y_j | (i,j) \in I_c, i \neq j\}$$
$$R_b = \{Y_k \xrightarrow{b^k_{i,j}} S_i + S_j | (i,j,k) \in I_b\} \qquad R_d = \{S_i \xrightarrow{d_{i,j}} S_j | (i,j) \in I_d, i \neq j\}$$

for $I_a, I_b \subseteq \{1, \ldots, N\}^2 \times \{1, \ldots, P\}$, $I_c \subseteq \{1, \ldots, P\}^2$ and $I_d \subseteq \{1, \ldots, N\}^2$. To fix the notation, we assume that any $(i, j, k) \in I_a \cup I_b$ satisfies $i \leq j$, so that self-interactions a priori are allowed. If the rate constants are not required, we put an arrow to indicate a reaction and omit the rates. Further:

(i) All chemical species are involved in at least one reaction.
(ii) For every intermediate complex Y_k there exist $i \leq j$, indices $k_1, \ldots, k_r$ and a chain of reactions $Y_k \to Y_{k_1} \to \cdots \to Y_{k_r} \to S_i + S_j$.

Assumption (ii) ensures that Y_k *ultimately dissociates into two substrates*. Also, we allow that there are more than one Y_k such that $S_i + S_j \to Y_k$ or $Y_k \to S_i + S_j$ for given S_i, S_j.

For convenience, we put $c_{i,j} = 0, d_{i,j} = 0$ if $(i, j) \notin I_c$ or I_d respectively, and similarly $a^k_{i,j} = 0$ and $b^k_{i,j} = 0$ if $(i, j, k) \notin I_a$ or I_b, respectively. For $i \leq j$ and k, we define $a^k_{j,i} = a^k_{i,j}$ and $b^k_{j,i} = b^k_{i,j}$. For later use, we define

$$\mathscr{S}_{\circlearrowleft} = \{S_i \in \mathscr{S} | (i,i,k) \in I_a \cup I_b \text{ for some } k\}$$

to be the set of *self-interacting substrates*, which contains the substrates for which reactions $2S_i = S_i + S_i \to Y_k$ or $Y_k \to 2S_i$ exist for some k.

As an example consider the PTM system with $\mathscr{S} = \{S_1, S_2, S_3, S_4, S_5\}$, $\mathscr{Y} = \{Y_1, Y_2, Y_3\}$ and reactions

$$S_1 \xrightarrow{d_{1,2}} S_2 \qquad S_2 + S_5 \underset{b^1_{2,5}}{\overset{a^1_{2,5}}{\rightleftarrows}} Y_1 \underset{c_{2,1}}{\overset{c_{1,2}}{\rightleftarrows}} Y_2 \xrightarrow{b^2_{1,4}} S_1 + S_4 \qquad S_3 + S_4 \underset{b^3_{3,4}}{\overset{a^3_{3,4}}{\rightleftarrows}} Y_3 \xrightarrow{b^3_{3,5}} S_3 + S_5 \tag{1}$$

One interpretation is that S_1 is modified to S_2. The modifier group is then transferred from S_2 to S_5 with the formation of two intermediate complexes Y_1, Y_2, causing the modification of S_5 to S_4 and the demodification of S_2 to S_1. Finally, S_4 is demodified via a Michaelis–Menten mechanism catalyzed by an enzyme S_3.

Nomenclature We introduce a few concepts that will be used in the following, some of which are taken from Chemical Reaction Network Theory (CRNT) (Feinberg 1980; Feinberg and Horn 1977). Consider the set of *complexes* of the reaction system:

$$\mathscr{C} = \mathscr{Y} \cup \{S_i,\ S_j |\ (i, j) \in I_d\} \cup \{S_i + S_j |\ (i, j, k) \in I_a \cup I_b \text{ for some } k\}.$$

Then:

- $A \in \mathscr{C}$ *reacts to* $B \in \mathscr{C}$ if there exists a reaction $A \to B$.
- $A \in \mathscr{C}$ *ultimately reacts to* $B \in \mathscr{C}$ if there exists a sequence of reactions $A \to A_1 \to \cdots \to A_r \to B$ with $A_m \in \mathscr{C}$. If $A_m \in \widetilde{\mathscr{Y}} \subseteq \mathscr{Y}$ for all m, then A *ultimately reacts to* B *via* $\widetilde{\mathscr{Y}}$.
- S_i and S_j *interact* if for some Y_k either $S_i + S_j$ reacts to Y_k or vice versa.
- S_i, S_j are *1-linked* if $d_{i,j}$ or $d_{j,i} \neq 0$. Y_k, Y_v are 1-linked if $c_{k,v}$ or $c_{v,k} \neq 0$. S_i and Y_k are 1-linked if for some j, $S_i + S_j$ reacts to Y_k or vice versa ($j = i$ is allowed).
- A reaction $A \to B$ is *reversible* if $B \to A$ is also a reaction. It is called *irreversible* otherwise.

Assumption (ii) of a PTM system ensures that all intermediate complexes ultimately react to some $S_i + S_j$ via $\mathscr{Y}$.

3.2 Mass-action kinetics

The set of reactions together with their associated rate constants give rise to a polynomial system of ordinary differential equations taken with *mass-action kinetics*:

$$\dot{Y}_k = \sum_{j=1}^{N}\sum_{i=1}^{j}(a^k_{i,j}S_iS_j - b^k_{i,j}Y_k) + \sum_{v=1}^{P}(c_{v,k}Y_v - c_{k,v}Y_k), \qquad k = 1, \ldots, P,$$

$$\dot{S}_i = \sum_{j=1}^{N}\sum_{k=1}^{P}\epsilon_{i,j}(-a^k_{i,j}S_iS_j + b^k_{i,j}Y_k) + \sum_{j=1}^{N}(d_{j,i}S_j - d_{i,j}S_i), \quad i = 1, \ldots, N,$$

where $\epsilon_{i,j} = 1$ if $i \neq j$ and 2 if $i = j$ and where $\dot{x}$ denotes dx/dt for $x = x(t)$. Here we abuse notation and let S_i, Y_k denote the concentrations of the species S_i, Y_k as well. The steady states are the solutions to the polynomial system obtained by setting the derivatives to zero, i.e. $\dot{Y}_k = 0$ and $\dot{S}_i = 0$:

$$0 = \sum_{j=1}^{N}\sum_{i=1}^{j}(a_{i,j}^k S_i S_j - b_{i,j}^k Y_k) + \sum_{v=1}^{P}(c_{v,k}Y_v - c_{k,v}Y_k), \qquad k = 1, \ldots, P, \quad (2)$$

$$0 = \sum_{j=1}^{N}\sum_{k=1}^{P}\epsilon_{i,j}(-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j=1}^{N}(d_{j,i}S_j - d_{i,j}S_i), \quad i = 1, \ldots, N. \quad (3)$$

This system is quadratic in the variables Y_k, S_i, but the only quadratic terms are of the form $S_i S_j$. It is linear in Y_k.

It is convenient to treat the reaction rate constants as parameters with unspecified (positive) values and view $a_{i,j}^k, b_{l,m}^r, c_{u,v}, d_{w,t}$ as symbols. For that, let

$$\begin{aligned}\mathrm{Con} = \{a_{i,j}^k | (i,j,k) \in I_a\} \cup \{b_{i,j}^k | (i,j,k) \in I_b\} \\ \cup \{c_{k,v} | (k,v) \in I_c\} \cup \{d_{k,v} | (k,v) \in I_d\}\end{aligned}$$

be the set of the non-zero parameters (symbols). Then, the system (2)–(3) is quadratic in $\mathscr{S} \cup \mathscr{Y}$ with coefficients in the field $\mathbb{R}(\mathrm{Con})$. Further, if all S_i are considered part of the coefficient field, then the system is linear with coefficients in $\mathbb{R}(\mathrm{Con} \cup \mathscr{S})$ and variables $Y_1, \ldots, Y_P$.

Only *non-negative solutions* of the steady-state equations are biologically meaningful. To study positivity of solutions, we use the concept of *S-positivity*. Let $X = \{x_1, \ldots, x_r\}$ be a finite set. A non-zero polynomial in $\mathbb{R}[X]$ with non-negative coefficients is called *S-positive*. Similarly, a rational function f is S-positive if it is a quotient of two S-positive polynomials. If $x_1, \ldots, x_r$ are substituted by positive real numbers in f, we obtain a positive real number. In general, a rational function $f = p/q$ in $z_1, \ldots, z_s$ and coefficients in $\mathbb{R}(X)$ is S-positive if the non-zero coefficients of p and q are S-positive rational functions in $x_1, \ldots, x_r$. If f is a rational function in $x_1, \ldots, x_r$ and $x_i = g(x_1, \ldots, \widehat{x_i}, \ldots, x_r)$ with g a rational function, then substituting g into f gives f as a rational function in $x_1, \ldots, \widehat{x_i}, \ldots, x_r$.

The differential equations of Example (1) are:

$$\begin{aligned}
\dot{Y}_1 &= a_{2,5}^1 S_2 S_5 - (b_{2,5}^1 + c_{1,2})Y_1 + c_{2,1}Y_2 & \dot{S}_1 &= -d_{1,2}S_1 + b_{1,4}^2 Y_2 \\
\dot{Y}_2 &= c_{1,2}Y_1 - (b_{1,4}^2 + c_{2,1})Y_2 & \dot{S}_2 &= d_{1,2}S_1 - a_{2,5}^1 S_2 S_5 + b_{2,5}^1 Y_1 \\
\dot{Y}_3 &= a_{3,4}^3 S_3 S_4 - (b_{3,4}^3 + b_{3,5}^3)Y_3 & \dot{S}_3 &= -a_{3,4}^3 S_3 S_4 + (b_{3,4}^3 + b_{3,5}^3)Y_3 \\
\dot{S}_5 &= -a_{2,5}^1 S_2 S_5 + b_{2,5}^1 Y_1 + b_{3,5}^3 Y_3 & \dot{S}_4 &= -a_{3,4}^3 S_3 S_4 + b_{1,4}^2 Y_2 + b_{3,4}^3 Y_3.
\end{aligned} \quad (4)$$

To compute the steady states, we can use $\dot{Y}_3 = 0$ to eliminate Y_3 as a function of the substrates. Also Y_1, Y_2 can be eliminated by solving the linear system $\dot{Y}_1 = \dot{Y}_2 = 0$. This is a general feature of PTM systems which is covered in Sect. 4.1.

Further, observe that $\dot{S}_3 + \dot{Y}_3 = 0$, which implies that the sum $S_3 + Y_3$ is independent of time and thus conserved. In fact, it implies that one of the equations $\dot{Y}_3 = 0$

and $\dot{S}_3 = 0$ is redundant. Removing one of them leaves a polynomial system with 7 equations in 8 variables, and thus the solutions to the steady-state equations form an algebraic variety of dimension at least one (over the complex numbers). This redundancy can be compensated for by fixing the value $S_3 + Y_3 = \overline{A}$ and adding this relation to the steady-state equations.

In general, conserved amount equations, called *conservation laws*, complement the steady-state equations. Their existence plays an important role in the elimination procedure below. Therefore, the next section is devoted to understand conservation laws and how they arise. Further, we characterize those that are relevant for the elimination procedure by a graphical condition on a certain graph.

3.3 Conservation laws

We consider systems where inflow of species is not allowed and species are not degraded or able to diffuse out. Such systems are "entrapped" in contrast to open systems (so-called "continuous flow stirred tank reactors") as considered for instance by Craciun and Feinberg (2005). PTM systems are entrapped and, as we will see below, always have conservation laws that reflect that the total amount of species remains constant either in free form S_i or in bounded form Y_j. These laws follow from the system of differential equations and appear as linear combinations of species (e.g. $S_3 + Y_3 = \overline{A}$ in the example above).

The existence of conservation laws implies that the dynamics of the system takes place in a proper invariant subspace of $\mathbb{R}^{N+P}$. We identify $\mathbb{R}^{N+P}$ with the real vector space generated by $\mathscr{S} \cup \mathscr{Y}$ so that $\mathbb{R}^{N+P} \equiv \langle S_1, \ldots, S_N, Y_1, \ldots, Y_P \rangle$. The species S_i and Y_k are unit vectors with a one in the ith and $(N+k)$th entry, respectively, and all other entries being zero. A vector $v = (\lambda_1, \ldots, \lambda_N, \mu_1, \ldots, \mu_P)$ is identified with the linear combination of species $\sum_i \lambda_i S_i + \sum_k \mu_k Y_k$.

Consider the *stoichiometric subspace* of $\mathbb{R}^{N+P}$ of a PTM system:

$$\begin{aligned} \Gamma = &\langle S_i + S_j - Y_k \mid (i,j,k) \in I_a \cup I_b \rangle \\ &+ \langle Y_k - Y_v \mid (k,v) \in I_c \rangle + \langle S_i - S_j \mid (i,j) \in I_d \rangle. \end{aligned}$$

If $(\lambda_1, \ldots, \lambda_N, \mu_1, \ldots, \mu_P) \in \Gamma^\perp$, then $\sum_i \lambda_i \dot{S}_i + \sum_k \mu_k \dot{Y}_k = 0$ and thus $\sum_i \lambda_i S_i + \sum_k \mu_k Y_k$ is independent of time. The vector ω with $\lambda_i = 1$ and $\mu_j = 2$ belongs to $\Gamma^\perp$ for any PTM system, showing that these systems always have conserved amounts. As shown by Feinberg and Horn (1977), not all conserved amounts are derived from vectors in $\Gamma^\perp$.

It follows that any basis $\{\omega^1, \ldots, \omega^d\}$ of $\Gamma^\perp$ provides a set of independent conserved quantities $\sum_{i=1}^N \lambda_i^l S_i + \sum_{k=1}^P \mu_k^l Y_k$ if $\omega^l = (\lambda_1^l, \ldots, \lambda_N^l, \mu_1^l, \ldots, \mu_P^l)$. Therefore, if *total amounts* $\overline{A}_1, \ldots, \overline{A}_d \in \mathbb{R}_+$ are provided, we require the steady-state solutions to satisfy:

$$\overline{A}_l = \sum_{i=1}^N \lambda_i^l S_i + \sum_{k=1}^P \mu_k^l Y_k \qquad l = 1, \ldots, d. \tag{5}$$

Total amounts are fixed by the initial concentrations of the species. We say that equations (5) are independent if the system has maximal rank, or equivalently, if the corresponding vectors of $\Gamma^{\perp}$ are independent.

Remark 1 By definition, the stoichiometric subspace Γ, and hence the conservation laws, do not depend on the reactions being reversible or irreversible.

We will show that some conservation laws can be determined by a graphical condition together with a property of the complexes. For that, we introduce the concepts of a cut and a non-interacting graph. These concepts do not depend on the reactions being reversible or irreversible, because neither do the conservation laws.

Definition 1 Let a non-empty set $\mathscr{S}_\alpha \subseteq \mathscr{S}$ be given and let the associated set $\mathscr{Y}_\alpha \subseteq \mathscr{Y}$ be the smallest set such that $Y_k \in \mathscr{Y}_\alpha$ if Y_k is 1-linked to some $S_i \in \mathscr{S}_\alpha$ or to $Y_m \in \mathscr{Y}_\alpha$. Then:

(i) $\mathscr{S}_\alpha$ is *closed* if S_j belongs to $\mathscr{S}_\alpha$ whenever $S_i \in \mathscr{S}_\alpha$ is 1-linked to S_j, and if S_i and S_j interact and are 1-linked to $Y_k \in \mathscr{Y}_\alpha$, then S_i or S_j are in $\mathscr{S}_\alpha$.
(ii) $\mathscr{S}_\alpha$ is a *cut* if (a) $S_i, S_j \in \mathscr{S}_\alpha$ do not interact for any i, j, and (b) $\mathscr{S}_\alpha$ is closed.
(iii) A cut $\mathscr{S}_\alpha$ is *minimal* if it has no proper closed subsets.

Condition (ii) implies that a self-interacting substrate $S \in \mathscr{S}_\circlearrowleft$ cannot belong to any cut, that is, $\mathscr{S}_\alpha \cap \mathscr{S}_\circlearrowleft = \emptyset$ for any cut $\mathscr{S}_\alpha$. Note that a closed subset $\mathscr{S}'$ of a cut is also a cut. The union of two disjoint cuts $\mathscr{S}_\alpha, \mathscr{S}'_\alpha$ is a cut if $\mathscr{Y}_\alpha \cap \mathscr{Y}'_\alpha = \emptyset$.

In the PTM system with reactions $S_1 + S_4 \rightleftarrows Y_2 \rightleftarrows S_2 + S_4$ and $Y_1 \rightleftarrows S_2 + S_3$, the set $\{S_1, S_2\}$ is a cut, while $\{S_1, S_3\}$ is not. There are no proper closed subsets of $\{S_1, S_2\}$ and thus the cut is minimal.

Definition 2 Let a non-empty set $\mathscr{S}_\alpha \subseteq \mathscr{S}$ be given and let $\mathscr{Y}_\alpha \subseteq \mathscr{Y}$ be as in Definition 1. Let $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$ be the undirected graph with node set $\mathscr{S}_\alpha \cup \mathscr{Y}_\alpha$ and edges between 1-linked nodes. The graph is *non-interacting* if it is connected and $\mathscr{S}_\alpha$ is a cut.

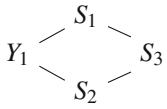
If $\mathscr{S}_\alpha = \mathscr{S}$, then $\mathscr{Y}_\alpha = \mathscr{Y}$. All graphs $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$ are naturally subgraphs of $G_{\mathscr{S}, \mathscr{Y}}$. Without proof we state the following:

Lemma 1 *Let $\mathscr{S}_\alpha$ be a cut and G' be a connected subgraph of $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$ with node set $\mathscr{S}' \cup \mathscr{Y}'$, $\mathscr{S}' \subseteq \mathscr{S}_\alpha$ and $\mathscr{Y}' \subseteq \mathscr{Y}_\alpha$. The following statements are equivalent:*

(i) *$\mathscr{S}'$ is closed with associated set $\mathscr{Y}'$.*
(ii) *G' is a connected component of $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$.*
(iii) *G' is non-interacting and contains only species in $\mathscr{S}_\alpha \cup \mathscr{Y}_\alpha$.*

If either is the case, then $\mathscr{S}'$ is a minimal cut and $G' = G_{\mathscr{S}', \mathscr{Y}'}$.

Thus, the non-interacting graphs containing substrates only in a cut $\mathscr{S}_\alpha$ are exactly the connected components of $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$. All non-interacting graphs contain some node from $\mathscr{S}$ (condition (ii) of a PTM system). However, such a graph might not exist. Consider for example the system with reactions $S_1 \rightleftarrows S_3$, $S_2 \rightleftarrows S_3$, and $S_1 + S_2 \rightleftarrows Y_1$. The graph $G_{\mathscr{S}, \mathscr{Y}}$ is



Condition (B) Of Definition 1(ii) implies that any non-interacting graph must contain all four species, which contradicts condition (a) of the same definition. Note that $S_1 + S_2 + S_3 + 2Y_1$ is conserved.

In the following proposition we show that each non-interacting graph gives rise to a conservation law.

Proposition 1 *Let $H_1, \ldots, H_n$ be the non-interacting graphs of a PTM system, C_l the node set of H_l, $\mathscr{S}_l = \mathscr{S} \cap C_l$ and $\mathscr{Y}_l = \mathscr{Y} \cap C_l$. Then, $\dot{\omega}_l = 0$ for*

$$\omega_l = \sum_{S \in \mathscr{S}_l} S + \sum_{Y \in \mathscr{Y}_l} Y \qquad l = 1, \ldots, n.$$

That is, H_l corresponds to a conservation law and ω_l is determined by the initial concentrations.

Proof Substrates in C_l interact only with substrates in $\mathscr{S} \backslash \mathscr{S}_l$ and thus, by definition of $\mathscr{Y}_l$, if $a^k_{i,j} \neq 0$ or $b^k_{i,j} \neq 0$ for $i \neq j$ then: (a) if S_i (resp. S_j) is in $\mathscr{S}_l$, then S_j (resp. S_i) belongs to $\mathscr{S} \backslash \mathscr{S}_l$, and $Y_k \in \mathscr{Y}_l$; (b) if $Y_k \in \mathscr{Y}_l$, then either S_i or S_j, but not both, belongs to $\mathscr{S}_l$. If $c_{v,k} \neq 0$ or $c_{k,v} \neq 0$, then Y_k, Y_v belong to the same non-interacting graph (if any); if $d_{i,j} \neq 0$ or $d_{j,i} \neq 0$, then S_i, S_j belong to the same non-interacting graph (if any). Since $\mathscr{S}_l \cap \mathscr{S}_\circlearrowleft = \emptyset$ for $Y_k \in \mathscr{Y}_l$ and $S_i \in \mathscr{S}_l$ we have:

$$\dot{Y}_k = \sum_{i | S_i \in \mathscr{S}_l} \sum_{j | S_j \in \mathscr{S} \backslash \mathscr{S}_l} (a^k_{i,j} S_i S_j - b^k_{i,j} Y_k) + \sum_{v | Y_v \in \mathscr{Y}_l} (c_{v,k} Y_v - c_{k,v} Y_k)$$

$$\dot{S}_i = \sum_{k | Y_k \in \mathscr{Y}_l} \sum_{j | S_j \in \mathscr{S} \backslash \mathscr{S}_l} (-a^k_{i,j} S_i S_j + b^k_{i,j} Y_k) + \sum_{j | S_j \in \mathscr{S}_l} (d_{j,i} S_j - d_{i,j} S_i).$$

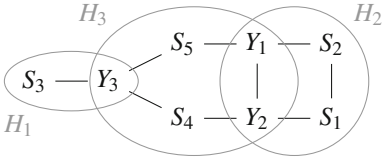
It follows that

$$\sum_{k | Y_k \in \mathscr{Y}_l} \sum_{v | Y_v \in \mathscr{Y}_l} (c_{v,k} Y_v - c_{k,v} Y_k) = 0 \quad \text{and} \quad \sum_{i | S_i \in \mathscr{S}_l} \sum_{j | S_j \in \mathscr{S}_l} (d_{j,i} S_j - d_{i,j} S_i) = 0.$$

Similarly, the remaining terms in $\dot{\omega}_l$ cancel. Thus, $\dot{\omega}_l = 0$. □

Thus, each non-interacting graph gives rise to a conserved amount. If each non-interacting graph contains a species that only belongs to that graph, then the ω_l's are independent. In particular, conservation laws derived from the connected components of $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$ for some cut $\mathscr{S}_\alpha$ are independent. In general, the set of conservation laws found from Proposition 1 can be reduced to a set of independent conservation laws.

In Example (1), the graph $G_{\mathscr{S}, \mathscr{Y}}$ is



The non-interacting graphs H_1, H_2, H_3 are circled. If total amounts $\overline{A}_1, \overline{A}_2, \overline{A}_3 \in \mathbb{R}_+$ are provided then the steady-state solutions must satisfy:

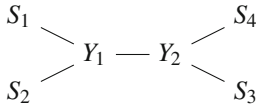
$$\overline{A}_1 = S_3 + Y_3, \qquad \overline{A}_2 = S_1 + S_2 + Y_1 + Y_2, \qquad \overline{A}_3 = S_4 + S_5 + Y_1 + Y_2 + Y_3. \tag{6}$$

These conserved total amounts are easily verified by differentiation using (4).

This procedure provides an easy construction of conservation laws. In the example above, the conservation laws obtained from the graph are independent and, additionally, determine all conservation laws arising from $\Gamma^\perp$ (dim $\Gamma^\perp = 3$). However, this is not always the case. Consider for instance the reaction system

$$S_1 + S_2 \rightleftarrows Y_1 \rightleftarrows Y_2 \rightleftarrows S_3 + S_4. \tag{7}$$

The graph $G_{\mathscr{S},\mathscr{Y}}$ is

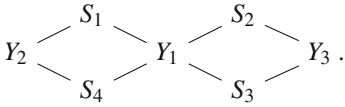


There are 4 non-interacting graphs that give the conserved total amounts $\overline{A}_1 = S_1 + S_3 + Y_1 + Y_2, \overline{A}_2 = S_1 + S_4 + Y_1 + Y_2, \overline{A}_3 = S_2 + S_3 + Y_1 + Y_2$, and $\overline{A}_4 = S_2 + S_4 + Y_1 + Y_2$. This system has rank 3, implying that one of the relations is redundant. In this case the procedure still gives all conservation laws, because the dimension of $\Gamma^\perp$ is 3.

Consider the following reaction system:

$$S_1 + S_3 \rightleftarrows Y_1 \rightleftarrows S_2 + S_4 \qquad S_1 + S_4 \rightleftarrows Y_2 \qquad S_2 + S_3 \rightleftarrows Y_3. \tag{8}$$

The graph $G_{\mathscr{S},\mathscr{Y}}$ is



There are 2 non-interacting graphs that give the conserved total amounts $\overline{A}_1 = S_1 + S_2 + Y_1 + Y_2 + Y_3$, and $\overline{A}_2 = S_3 + S_4 + Y_1 + Y_2 + Y_3$. However, $\dim \Gamma^\perp = 3$ and the procedure fails to provide three independent conservation laws. A third conservation law is $\overline{A}_3 = S_1 + S_4 + Y_1 + 2Y_2$, and the coefficient 2 of Y_2 cannot be obtained from a non-interacting graph.

3.4 Cuts of $\mathscr{S}$ and conservation laws

The previous example shows that a maximal set of independent conservation laws cannot always be obtained from non-interacting graphs. In this subsection, we provide a criterion to guarantee that there are $\dim \Gamma^\perp$ independent conservation laws derived from non-interacting graphs. The criterion will be used in Sect. 5.

In the following we make use of Lemma 1 without further reference. Let $\mathscr{S}_\alpha$ be a cut with associated set $\mathscr{Y}_\alpha$. Define $\mathscr{S}_\alpha^c = \mathscr{S}\backslash\mathscr{S}_\alpha$ and $\mathscr{Y}_\alpha^c = \mathscr{Y}\backslash\mathscr{Y}_\alpha$, and let N_α, P_α (resp. N_α^c, P_α^c) be the cardinality of $\mathscr{S}_\alpha$, $\mathscr{Y}_\alpha$ (resp. $\mathscr{S}_\alpha^c$, $\mathscr{Y}_\alpha^c$). Extend the set of conservation laws derived from the n_α connected components of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ to a maximal set of n independent conservation laws derived from *other* non-interacting graphs (thus containing species in $\mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c$). Let $n_\alpha^c = n - n_\alpha$.

Lemma 2 *Let $\mathscr{S}_\alpha$ be a cut and keep the notation introduced above. Then, we have that* $\dim\left(\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle \cap \Gamma\right) \leq N_\alpha^c + P_\alpha^c - n_\alpha^c$ *and* $\dim \Gamma^\perp = n$ *if and only if*

$$dim\left(\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle \cap \Gamma\right) = N_\alpha^c + P_\alpha^c - n_\alpha^c.$$

Proof Without loss of generality we can assume that $\mathscr{Y}_\alpha = \{Y_1, \ldots, Y_{P_\alpha}\}$ and $\mathscr{S}_\alpha = \{S_1, \ldots, S_{N_\alpha}\}$. Identify $\mathbb{R}^{N+P}$ with $\mathbb{R}^{N_\alpha} \times \mathbb{R}^{P_\alpha} \times \mathbb{R}^{N_\alpha^c} \times \mathbb{R}^{P_\alpha^c}$ and let

$$\Gamma_\alpha = \langle A - B | \text{ for each edge between } A \text{ and } B \text{ in } G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}\rangle.$$

The space $\Gamma_\alpha^\perp$ is generated by the vectors which are sums of species in each connected component of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ and hence $\dim \Gamma_\alpha^\perp = n_\alpha$. We have $\dim \Gamma^\perp \geq n = n_\alpha + n_\alpha^c$ and we want to determine when equality holds. Equivalently, we want to determine when $\dim \Gamma = N + P - n$. If this is not the case, then $\dim \Gamma < N + P - n$. Note that $N = N_\alpha + N_\alpha^c$ and $P = P_\alpha + P_\alpha^c$.

Note that $\dim \Gamma_\alpha = N_\alpha + P_\alpha - n_\alpha$. Let $\pi : \mathbb{R}^{N+P} \to \mathbb{R}^{N_\alpha+P_\alpha}$ denote the projection onto the first $N_\alpha + P_\alpha$ coordinates and $\pi_\alpha : \Gamma \to \Gamma_\alpha$ its restriction to Γ (π_α a surjective map). Then, $\dim \Gamma = \dim \Gamma_\alpha + \dim \ker \pi_\alpha$ and so $\dim \ker \pi_\alpha \leq N_\alpha^c + P_\alpha^c - n_\alpha^c$. Further, $\dim \Gamma^\perp = n$ if and only if $\dim \ker \pi_\alpha = N_\alpha^c + P_\alpha^c - n_\alpha^c$. Finally, note that $\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle \cap \Gamma = \ker \pi_\alpha$. Indeed, let $i : \Gamma \hookrightarrow \mathbb{R}^{N+P}$ and $i_\alpha : \Gamma_\alpha \hookrightarrow \mathbb{R}^{N_\alpha+P_\alpha}$ denote the natural inclusions. We have that $i_\alpha \circ \pi_\alpha = \pi \circ i$. The kernel of π is clearly $\mathbb{R}^{N_\alpha^c+P_\alpha^c} = \langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle$ from where it follows that the kernel of π_α is $\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle \cap \Gamma$.

Therefore, we have that $\dim\left(\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c\rangle \cap \Gamma\right) = \dim \ker \pi_\alpha = N_\alpha^c + P_\alpha^c - n_\alpha^c$ if and only if $\dim \Gamma^\perp = n$. □

As each non-interacting graph corresponds to a minimal cut (Lemma 1), the lemma above provides a condition for when all conservation laws can be recovered from cuts.

Remark 2 An easy way to construct elements of $\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c \rangle \cap \Gamma$ is by considering:

(i) Vectors $S_i - S_j$ for pairs $S_i, S_j \in \mathscr{S}_\alpha^c$ for which there are $S_m \in \mathscr{S}$ and reactions $S_m + S_i \text{ — } B_1 \text{ — } \cdots \text{ — } B_r \text{ — } S_m + S_j$ with $B_u \in \mathscr{C}$ and — being $\leftarrow$ or $\rightarrow$.

(ii) Vectors $S_i + S_j - Y_k$, $S_i - S_j$ or $Y_k - Y_v$ corresponding to reactions with $S_i, S_j \in \mathscr{S}_\alpha^c$ and $Y_k, Y_v \in \mathscr{Y}_\alpha^c$.

If we can construct $N_\alpha^c + P_\alpha^c - n_\alpha^c$ independent elements of $\langle \mathscr{S}_\alpha^c \cup \mathscr{Y}_\alpha^c \rangle \cap \Gamma$ of the previous type, then the previous lemma holds.

In Example (1) consider the cut $\mathscr{S}_\alpha = \{S_1, S_2, S_3\}$ with $\mathscr{S}_\alpha^c = \{S_4, S_5\}$ and the conservation laws given in (6) ($n = 3$). We have $N_\alpha^c = 2$ and $n_\alpha^c = 1$. Further, $\mathscr{Y}_\alpha = \mathscr{Y}$ so that $P_\alpha^c = 0$. The element $S_5 - S_4 = (S_3 + S_5 - Y_3) - (S_3 + S_4 - Y_3)$ belongs to $\langle \mathscr{S}_\alpha^c \rangle \cap \Gamma$. In addition, $N_\alpha^c + P_\alpha^c - n_\alpha^c = 1$ and thus $\dim(\langle \mathscr{S}_\alpha^c \rangle \cap \Gamma) = N_\alpha^c + P_\alpha^c - n_\alpha^c$, implying that a set of independent conservation laws is found from non-interacting graphs.

In Example (7), consider the cut $\mathscr{S}_\alpha = \{S_1, S_3\}$ with $\mathscr{S}_\alpha^c = \{S_2, S_4\}$ so that $N_\alpha^c = 2$, $P_\alpha^c = 0$. There is only one conservation law in $\mathscr{S}_\alpha \cup \mathscr{Y}_\alpha$, $S_3 + S_1 + Y_1 + Y_2$, and since $n = 3$, then $n_\alpha^c = 2$. It follows that $N_\alpha^c + P_\alpha^c - n_\alpha^c = 0$, and we are guaranteed that the dimension of $\langle S_2, S_4 \rangle \cap \Gamma$ is zero.

In Example (8), consider the cut $\mathscr{S}_\alpha = \{S_1, S_2\}$ with $\mathscr{S}_\alpha^c = \{S_3, S_4\}$ and $N_\alpha^c = 2$, $P_\alpha^c = 0$. We have $n_\alpha^c = 1$ and $N_\alpha^c + P_\alpha^c - n_\alpha^c = 1$. However, $\langle S_3, S_4 \rangle \cap \Gamma$ has dimension zero and thus extra conservation laws are required.

4 Variable elimination

In this section we show that the intermediate complexes can always be eliminated and expressed as polynomials in the substrates with coefficients in $\mathbb{R}(\mathrm{Con})$ (Sect. 4.1). After choosing a cut $\mathscr{S}_\alpha$, the substrates in $\mathscr{S}_\alpha$ can be expressed in terms of those in $\mathscr{S}_\alpha^c = \mathscr{S} \backslash \mathscr{S}_\alpha$ (Sect. 4.3).

4.1 Elimination of intermediate complexes

Consider the system $\dot{Y}_i = 0$ in (2) as a linear system of P polynomial equations with coefficients in $\mathbb{R}[\mathrm{Con} \cup \mathscr{S}]$ and P variables $Y_1, \ldots, Y_P$. If the system has maximal rank, then there is a unique solution in $\mathbb{R}(\mathrm{Con} \cup \mathscr{S})$.

Specifically, we have a linear system $MY = z$ where $Y = (Y_1, \ldots, Y_P)^t$ and $M = \{\lambda_{k,v}\}$ is a $P \times P$ matrix with coefficients in $\mathbb{R}[\mathrm{Con}]$,

$$\lambda_{k,v} = \begin{cases} -c_{v,k} & \text{if } k \neq v \\ \sum_{j=1}^{N} \sum_{i=1}^{j} b_{i,j}^k + \sum_{u=1}^{P} c_{k,u} & \text{if } k = v. \end{cases}$$

The independent term $z = (z_1, \ldots, z_P)^t$ is in $\mathbb{R}[\mathrm{Con} \cup \mathscr{S}]$: $z_k = \sum_{i \leq j} a_{i,j}^k S_i S_j$.

Assume that M has maximal rank P in $\mathbb{R}(\mathrm{Con})$. Then, using Cramer's rule to solve linear systems of equations, we obtain that $Y_k = \rho_k / \rho$ with $\rho = \det(M) \neq 0$ and ρ_k

the determinant of M with the kth column substituted by z. Since the determinant is a homogeneous polynomial in the entries of the matrix, it follows that $\rho \in \mathbb{R}[\mathrm{Con}]$ and $\rho_k \in \mathbb{R}[\mathrm{Con} \cup \mathscr{S}]$. Therefore, Y_k takes the form

$$Y_k = \sum_{i \leq j} \mu_{i,j}^k S_i S_j$$

with $\mu_{i,j}^k \in \mathbb{R}(\mathrm{Con})$ and hence Y_k is a polynomial in $\mathbb{R}(\mathrm{Con})[\mathscr{S}]$. If both ρ, ρ^k are S-positive elements of $\mathbb{R}[\mathrm{Con}]$ and $\mathbb{R}[\mathrm{Con} \cup \mathscr{S}]$, respectively, then $\mu_{i,j}^k$ is S-positive in $\mathbb{R}(\mathrm{Con})$ and for positive rate constants and non-negative values of S_i, the steady-state value of Y_k is non-negative as well.

S-positivity of ρ, ρ^k is proven in the next section using the *Matrix-Tree theorem* (Tutte 1948).

4.2 Decomposition of the system

The matrix M is not a Laplacian, since the column sums $\sum_{j=1}^{N} \sum_{i=1}^{j} b_{i,j}^k$ are not zero. Therefore, the Matrix-Tree theorem cannot be directly applied. However, M can be extended such that its determinant is a maximal minor of a Laplacian.

Let $G_{\mathscr{Y}}$ be the directed graph with node set $\mathscr{Y}$ and a directed edge $Y_k \to Y_v$ if $(k, v) \in I_c$. That is, $G_{\mathscr{Y}}$ is the graph obtained by restricting the set of reactions to those among the intermediate complexes. The node sets of the connected components of $G_{\mathscr{Y}}$ determine a partition of $\mathscr{Y}$: $\mathscr{Y} = \mathscr{Y}_1 \cup \cdots \cup \mathscr{Y}_s$. Let P_l be the cardinality of $\mathscr{Y}_l$ and rename the intermediate complexes such that $\mathscr{Y}_l = \{Y_{P_1+\cdots+P_{l-1}+1}, \ldots, Y_{P_1+\cdots+P_l}\}$.

If $Y_k \in \mathscr{Y}_l$ for some l, then $c_{k,v} = c_{v,k} = 0$ for any v such that $Y_v \notin \mathscr{Y}_l$. It follows, that M is a block diagonal matrix $\mathrm{diag}(M_1, \ldots, M_s)$ with M_l being a $P_l \times P_l$ matrix. Solving $MY = z$ is thus equivalent to solving s "smaller" systems with matrices M_l. Further, M has maximal rank P if and only if M_l has maximal rank P_l for all l.

Consider the connected component $G_{\mathscr{Y}_l}$ corresponding to $\mathscr{Y}_l$. We construct an extended labeled directed graph $\widehat{G}_{\mathscr{Y}_l}$ as follows:

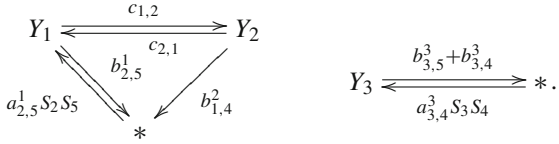
Definition 3 Order the set $\mathscr{Y}_l \cup \{*\}$ such that $Y_{P_1+\cdots+P_{l-1}+k}$ is the kth node and $*$ the $(P_l + 1)$th node. Let

$$b^k = \sum_{i \leq j} b_{i,j}^k, \qquad \text{and} \qquad a^k = \sum_{i \leq j} a_{i,j}^k S_i S_j.$$

The graph $\widehat{G}_{\mathscr{Y}_l}$ has node set $\mathscr{Y}_l \cup \{*\}$ and labeled directed edges:

$$Y_k \xrightarrow{c_{k,v}} Y_v, \text{ if } (k, v) \in I_c, \qquad Y_k \xrightarrow{b^k} *, \text{ if } b^k \neq 0, \qquad * \xrightarrow{a^k} Y_k \text{ if } a^k \neq 0.$$

In Example (1), the graph $G_{\mathscr{Y}}$ has two connected components $\mathscr{Y}_1 = Y_1 \rightleftarrows Y_2$ and $\mathscr{Y}_2 = Y_3$. The graphs $\widehat{G}_{\mathscr{Y}_1}$ and $\widehat{G}_{\mathscr{Y}_2}$ are



Let $\mathscr{L} = \{\alpha_{k,v}\}$ be the Laplacian of $\widehat{G}_{\mathscr{Y}_l}$. If $k, v \le P_l$, then $\alpha_{k,v} = -\lambda_{k,v}$. The entries of the last row are $\alpha_{P_l+1,k} = b^k$ for $k \le P_l$ and the entries of the last column are $\alpha_{k,P_l+1} = a^k$ $(= z_k)$ for $k \le P_l$. We conclude that the (P_l+1, P_l+1) principal minor of $\mathscr{L}$ is exactly $-M_l$ and thus, by the Matrix-Tree theorem, we have

$$\det(M_l) = (-1)^{P_l}\mathscr{L}_{(P_l+1,P_l+1)} = \sum_{\tau\in\Theta(*)} \pi(\tau).$$

Assumption (ii) of a PTM system ensures that each Y_k ultimately reacts to some $S_i + S_j$ via $\mathscr{Y}$, and hence there exists at least one spanning tree rooted at $*$. Thus, $\det(M_l) \neq 0$ and $\det(M_l)$ is an S-positive element of $\mathbb{R}[\mathrm{Con}]$.

By the definition of ρ_k and the Matrix-Tree theorem,

$$\rho_k = (-1)^{k+1}\mathscr{L}_{(P_l+1,k)} = \sum_{\tau\in\Theta(Y_{P_{l-1}+k})} \pi(\tau),$$

and hence ρ_k is either zero or an S-positive element of $\mathbb{R}[\mathrm{Con}\cup\mathscr{S}]$.

If there exists at least one spanning tree rooted at $v_k = Y_{P_{l-1}+k}$, then $\rho_k \neq 0$. A necessary condition for this to happen is the existence of at least one in-edge to v_k. Otherwise the concentration at steady state of v_k is zero, which is expected if v_k is only consumed and never produced. Similarly, if there is no reaction of the form $S_i + S_j \to Y_{P_{l-1}+m}$ for any m (that is, a directed edge $* \to v_m$), then $\rho_k = 0$ for all k.

The term ρ_k is a homogeneous polynomial of degree 2 in $\mathscr{S}$ with coefficients in $\mathbb{R}[\mathrm{Con}]$, because any spanning tree rooted at a node v_k has exactly one edge of the form $* \to v_m$ for some m. Further, a monomial S_iS_j appears in ρ_k only if $S_i + S_j$ ultimately reacts to v_k via $\mathscr{Y}_l$. If $\widehat{G}_{\mathscr{Y}_l}$ is strongly connected, then this condition is both sufficient and necessary. Indeed, if $S_i + S_j$ ultimately reacts to v_k via $\mathscr{Y}_l$, then there is a spanning tree rooted at v_k containing this path.

The next proposition summarizes the discussion above:

Proposition 2 *Consider a PTM system with intermediate complexes $\mathscr{Y}$ and substrates $\mathscr{S}$. Then, $\dot{Y}_k = 0$ for all k, if and only if*

$$Y_k = \sum_{i\le j} \mu^k_{i,j} S_iS_j \tag{9}$$

with $\mu^k_{i,j} \in \mathbb{R}(\mathrm{Con})$ being either zero or S-positive. Further:

(i) *If $S_i + S_j$ does not ultimately react to Y_k via $\mathscr{Y}$, then $\mu^k_{i,j} = 0$.*
(ii) *If $\widehat{G}_{\mathscr{Y}_l}$ is strongly connected and $Y_k \in \mathscr{Y}_l$, then $\mu^k_{i,j} \neq 0$ if and only if $S_i + S_j$ ultimately reacts to Y_k via $\mathscr{Y}_l$.*
(iii) *$\widehat{G}_{\mathscr{Y}_l}$ is strongly connected if and only if Y_k in (9) is a non-zero polynomial in $\mathbb{R}(\mathrm{Con})[\mathscr{S}]$ for all $Y_k \in \mathscr{Y}_l$.*

Remark 3 The condition that $\widehat{G}_{\mathscr{Y}_l}$ is strongly connected is biochemically reasonable: The intermediate complexes are not the initial or final products of the system and should eventually be broken up into parts.

In Example (1), the graph $\widehat{G}_{\mathscr{Y}_1}$ has three spanning trees rooted at $*$ so that $\det(M_1) = b^2_{1,4}c_{1,2} + b^1_{2,5}c_{2,1} + b^2_{1,4}b^1_{2,5}$. There is one spanning tree rooted at Y_2, giving $\rho_2 = c_{1,2}a^1_{2,5}S_2S_5$, and two spanning trees rooted at Y_1, giving $\rho_1 = (b^2_{1,4} + c_{2,1})a^1_{2,5}S_2S_5$. The graph $\widehat{G}_{\mathscr{Y}_2}$ has one spanning tree rooted at $*$ so that $\det(M_2) = b^3_{3,5} + b^3_{3,4}$, and one spanning tree rooted at Y_3, giving $\rho_3 = a^3_{3,4}S_3S_4$. Thus:

$$Y_1 = \mu^1_{2,5}S_2S_5, \quad Y_2 = \mu^2_{2,5}S_2S_5, \quad Y_3 = \mu^3_{3,4}S_3S_4,$$

with $\mu^1_{2,5} = \frac{(b^2_{1,4}+c_{2,1})a^1_{2,5}}{\det(M_1)}$, $\mu^2_{2,5} = \frac{c_{1,2}a^1_{2,5}}{\det(M_1)}$, and $\mu^3_{3,4} = \frac{a^3_{3,4}}{\det(M_2)}$.

Lemma 3 *Let $\widehat{G}_{\mathscr{Y}} = \cup_l \widehat{G}_{\mathscr{Y}_l}$. The graphs $\widehat{G}_{\mathscr{Y}_l}, l = 1, \ldots, s$, are strongly connected if and only if the graph $\widehat{G}_{\mathscr{Y}}$ is.*

Proof Assume that the graphs $\widehat{G}_{\mathscr{Y}_l}$ are strongly connected. Then, for any $v \in \widehat{G}_{\mathscr{Y}_l}$ and $\omega \in \widehat{G}_{\mathscr{Y}_j}$, there are directed paths $v \to *$ in $\widehat{G}_{\mathscr{Y}_l}$ and $* \to v$ in $\widehat{G}_{\mathscr{Y}_j}$, which by composition give a directed path between v and w.

For the reverse implication, let v, ω be two elements of $\mathscr{Y}_l$. Since $\widehat{G}_{\mathscr{Y}}$ is strongly connected, there exists a directed path $\alpha : v \to w$ in $\widehat{G}_{\mathscr{Y}}$. We can assume that $v, w \neq *$. A path connecting an intermediate complex in $\mathscr{Y}_l$ to one in $\mathscr{Y}_j$ for $j \neq l$ must pass through $*$. If a path α goes through $\widetilde{v} \in \mathscr{Y}_j$, for $j \neq l$, then it must go through $*$, first in and then out, potentially many times until it goes back to $\mathscr{Y}_l$ and to w. Therefore, α has the form $v \xrightarrow{\alpha_1} * \xrightarrow{\beta} * \xrightarrow{\alpha_2} w$ with α_1 and α_2 being paths in $\widehat{G}_{\mathscr{Y}_l}$. It follows that the path $v \xrightarrow{\alpha_1} * \xrightarrow{\alpha_2} w$ is a directed path from v to w in $\widehat{G}_{\mathscr{Y}_l}$. □

4.3 Elimination of substrates

Equation (9) shows that at steady state the intermediate complexes are given as polynomials in $\mathscr{S}$ with coefficients in $\mathbb{R}(\mathrm{Con})$. Insertion of (9) into the (time dependent) differential equations for the substrates is the procedure known as the quasi-steady state approximation. The rationale is that intermediate complexes tend to reach steady state much faster than substrates and thus some variables in the dynamical system can be eliminated. We have shown here that PTM systems mathematically enable this simplification although biochemical justification is required in concrete examples.

We now use the steady-state equation (3) to further eliminate some of the substrates in terms of others. Recall equation (3), that is $\dot{S}_i = 0$,

$$0 = \sum_{j=1}^{N}\sum_{k=1}^{P} \epsilon_{i,j}(-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j=1}^{N}(d_{j,i}S_j - d_{i,j}S_i) \tag{10}$$

for $i = 1, \ldots, N$. After substitution of the values for Y_k, equation $\dot{S}_i = 0$ corresponds to

$$0 = \sum_{u=1}^{N}\sum_{k=1}^{P}\sum_{j\le t} \epsilon_{i,j} b_{i,u}^k \mu_{j,t}^k S_j S_t - \sum_{j=1}^{N}\sum_{k=1}^{P} \epsilon_{i,j} a_{i,j}^k S_i S_j + \sum_{j=1}^{N}(d_{j,i}S_j - d_{i,j}S_i), \tag{11}$$

These equations are quadratic in $\mathscr{S}$. To proceed with linear elimination it is necessary to decide which variables are to be eliminated and which will be taken as part of the coefficient field. Since a monomial $S_i S_j$ appears only if S_i and S_j interact, we can proceed as long as $\mathscr{S}$ can be partitioned in an appropriate way.

Lemma 4 *Given a cut $\mathscr{S}_\alpha$, equations (11) for $S_i \in \mathscr{S}_\alpha$ form a homogeneous linear system of equations in the set of substrates $\mathscr{S}_\alpha$ and with coefficients in $\mathbb{R}(\mathrm{Con} \cup \mathscr{S}_\alpha^c)$.*

Proof For the three sums in (11) we make the following observations: If $S_i \in \mathscr{S}_\alpha$ and $d_{j,i} \neq 0$ or $d_{i,j} \neq 0$, then also $S_j \in \mathscr{S}_\alpha$. If $S_i \in \mathscr{S}_\alpha$ and $a_{i,j}^k \neq 0$, then $S_j \notin \mathscr{S}_\alpha$, otherwise S_i and S_j would interact. Finally, if $\mu_{j,t}^k \neq 0$, then according to Proposition 2, $S_j + S_t$ ultimately reacts to Y_k via the set $\mathscr{Y}_\alpha$ associated to $\mathscr{S}_\alpha$. Since $\mathscr{S}_\alpha$ is a cut, one of S_j and S_t (but not both) belongs to $\mathscr{S}_\alpha$. Thus equations (11) for the substrates in $\mathscr{S}_\alpha$ form a homogeneous linear system of equations in the species in $\mathscr{S}_\alpha$. □

Assume that there exists a cut $\mathscr{S}_\alpha$, let $\mathscr{Y}_\alpha$ be the associated set of intermediate complexes (Definition 1) and reorder the substrates so that $\mathscr{S}_\alpha = \{S_1, \ldots, S_{N_\alpha}\}$. Equations (11) for $S_i \in \mathscr{S}_\alpha$ form an $N_\alpha \times N_\alpha$ homogeneous linear system of equations with variables $\mathscr{S}_\alpha$ and coefficients in $\mathbb{R}(\mathrm{Con} \cup \mathscr{S}_\alpha^c)$. Further, $\epsilon_{i,j} = 1$ if $S_i \in \mathscr{S}_\alpha$.

Let B be the matrix with entries $\widetilde{b}_{i,j}$ for $i \neq j$ and $\widetilde{b}_{i,i} - \widetilde{a}_i$ for $i = j$, where

$$\widetilde{a}_i = \sum_{j=1}^{N_\alpha} d_{i,j} + \sum_{j=N_\alpha+1}^{N}\sum_{k=1}^{P} a_{i,j}^k S_j, \quad \widetilde{b}_{i,j} = d_{j,i} + \sum_{t=N_\alpha+1}^{N}\sum_{k=1}^{P} b_i^k \mu_{j,t}^k S_t,$$

$$b_i^k = \sum_{u=N_\alpha+1}^{N} b_{i,u}^k,$$

so that (11) becomes

$$0 = \sum_{j=1, j\neq i}^{N_\alpha} \widetilde{b}_{i,j} S_j + (\widetilde{b}_{i,i} - \widetilde{a}_i) S_i. \tag{12}$$

Consider Example (1) and the cut $\mathscr{S}_\alpha = \{S_1, S_2, S_3\}$. Then the equations (11) corresponding to $\dot{S}_1 = 0$, $\dot{S}_2 = 0$ and $\dot{S}_3 = 0$ are

$$0 = -d_{1,2}S_1 + b^2_{1,4}\mu^2_{2,5}S_2S_5, \qquad 0 = d_{1,2}S_1 + (b^1_{2,5}\mu^1_{2,5} - a^1_{2,5})S_2S_5,$$
$$0 = ((b^3_{3,4} + b^3_{3,5})\mu^3_{3,4} - a^3_{3,4})S_3S_4.$$

Therefore,

$$\widetilde{a}_1 = d_{1,2}, \qquad \widetilde{a}_2 = a^1_{2,5}S_5, \qquad \widetilde{a}_3 = a^3_{3,4}S_4$$
$$\widetilde{b}_{1,2} = b^2_{1,4}\mu^2_{2,5}S_5, \quad \widetilde{b}_{2,1} = d_{1,2}, \qquad \widetilde{b}_{2,2} = b^1_{2,5}\mu^1_{2,5}S_5, \quad \widetilde{b}_{3,3} = (b^3_{3,4} + b^3_{3,5})\mu^3_{3,4}S_4.$$

while the rest of the coefficients are zero. Note that indeed the coefficients of the system belong to $\mathbb{R}(\mathrm{Con} \cup \{S_4, S_5\})$. The equation $\dot{S}_3 = 0$ is trivial since $(b^3_{3,4} + b^3_{3,5})\mu^3_{3,4} - a^3_{3,4} = 0$. Observe also that the equations corresponding to $\dot{S}_1$ and $\dot{S}_2$ sum to zero since $b^1_{2,5}\mu^1_{2,5} + b^2_{1,4}\mu^2_{2,5} - a^1_{2,5} = 0$. As we show next, this is a general feature due to the conservation laws $\dot{Y}_3 + \dot{S}_3 = 0$ and $\dot{S}_1 + \dot{S}_2 + \dot{Y}_1 + \dot{Y}_2 = 0$.

Proposition 1 ensures that there is a conservation law for each connected component of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$. Let $C^\alpha_1, \ldots, C^\alpha_{n_\alpha}$ be the node sets of the connected components and define $\mathscr{S}_{\alpha,l} = \mathscr{S}_\alpha \cap C^\alpha_l$ and $\mathscr{Y}_{\alpha,l} = \mathscr{Y}_\alpha \cap C^\alpha_l$ so that $\sum_{S_i \in \mathscr{S}_{\alpha,l}} \dot{S}_i + \sum_{Y_k \in \mathscr{Y}_{\alpha,l}} \dot{Y}_k = 0$ for $l = 1, \ldots, n_\alpha$, are conservation laws. Imposing only that the intermediate complexes are at steady state, that is $\dot{Y}_k = 0$ for all k, we obtain

$$\sum_{S_i \in \mathscr{S}_{\alpha,l}} \dot{S}_i = 0, \qquad l = 1, \ldots, n_\alpha. \tag{13}$$

Here the equations for $\dot{S}_i$ depend only on the substrates and the intermediate complexes being substituted by their steady-state expressions (9). It follows that the column sums of the matrix B restricted to the rows corresponding to $\mathscr{S}_{\alpha,l}$ are all zero. Consequently, the matrix B has rank at most $N_\alpha - n_\alpha$.

Definition 4 Let $G_{\mathscr{S}_\alpha}$ be the labeled directed graph with node set $\mathscr{S}_\alpha$ and an edge from S_j to S_i labeled $\widetilde{b}_{i,j}$ whenever $\widetilde{b}_{i,j} \neq 0$, and $i \neq j$.

Recall that $\widetilde{b}_{i,j} \in \mathbb{R}(\mathrm{Con})[\mathscr{S}^c_\alpha]$ is S-positive. Let $G_{\mathscr{Y}_\alpha}$ be $G_{\mathscr{Y}}$ restricted to the nodes $\mathscr{Y}_\alpha$ and define $\widehat{G}_{\mathscr{Y}_\alpha}$ similarly (cf. Lemma 3). It follows from the definition of $\mathscr{Y}_\alpha$ (Definition 1) that $G_{\mathscr{Y}_\alpha}$ is a union of connected components of $G_{\mathscr{Y}}$. Let $N_{\alpha,l}$ be the cardinality of $S_{\alpha,l}$.

Lemma 5 *After reordering of the substrates in $\mathscr{S}_\alpha$, B is a block diagonal matrix,* $\mathrm{diag}(B_1, \ldots, B_{n_\alpha})$ *with B_l an $N_{\alpha,l} \times N_{\alpha,l}$ matrix. If there is a reaction $S_j \to S_i$, then $\widetilde{b}_{i,j} \neq 0$. If this is not the case and $\widetilde{b}_{i,j} \neq 0$ for $i \neq j$, then $S_j + S_t$ ultimately reacts to $S_i + S_u$ via $\mathscr{Y}_\alpha$ for some $S_u, S_t \in \mathscr{S}^c_\alpha$. If in addition $\widehat{G}_{\mathscr{Y}_\alpha}$ is strongly connected, then the reverse is true.*

Proof It follows from Lemma 3 and Proposition 2(i) that if $\mu^k_{j,t} \neq 0$ then $S_j + S_t$ ultimately reacts to Y_k. By definition, $b^k_i \neq 0$ if and only if there exists a reaction

$Y_k \to S_i + S_u$ for some $S_u \in \mathscr{S}_\alpha^c$. We have $\widetilde{b}_{i,j} \neq 0$ if and only if $d_{j,i} \neq 0$ or $b_i^k \mu_{j,t}^k \neq 0$ for some k and t, and hence either there is a reaction $S_j \to S_i$ or there exist $S_u, S_t \in \mathscr{S}_\alpha^c$, so that $S_j + S_t$ ultimately reacts to $S_i + S_u$ via $\mathscr{Y}_\alpha$. If $\widehat{G}_{\mathscr{Y}_\alpha}$ is strongly connected then by Proposition 2(ii) the existence of these reactions is a sufficient condition. It follows, after reordering of the species in $\mathscr{S}_\alpha$, that B is a block diagonal matrix with blocks given by the species in each connected component of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$. Indeed, if S_i, S_j are in different components, then $\widetilde{b}_{i,j} = \widetilde{b}_{j,i} = 0$. □

It follows from the lemma that a necessary condition for $\widetilde{b}_{i,j} \neq 0$ is that S_i can be "produced" from S_j. Also, a consequence of the lemma is that the graph $G_{\mathscr{S}_\alpha}$ has at least l connected components. If two species S_i, S_j belong to the same connected component of $G_{\mathscr{S}_\alpha}$ then they belong to the same connected component of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$. If $\widehat{G}_{\mathscr{Y}_\alpha}$ is strongly connected then the reverse is true. In the following we restrict the study to the case where $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ is connected and note that the results apply to every connected component individually. However, the propositions to be derived below are stated in full generality, that is, without the assumption that $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ is connected.

If $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ is connected then using (13), the column sums of B are zero and B is the Laplacian of the labeled directed graph $G_{\mathscr{S}_\alpha}$. By the Matrix-Tree theorem, the principal minors $B_{(i,j)}$ of $B = \mathscr{L}(G_{\mathscr{S}_\alpha})$ are

$$B_{(i,j)} = (-1)^{N_\alpha - 1 + i + j} \sum_{\tau \in \Theta(S_j)} \pi(\tau).$$

Thus, B has rank $N_\alpha - 1$ if and only if there exists at least one spanning tree in $G_{\mathscr{S}_\alpha}$ rooted at some S_j with $j \in \{1, \ldots, N_\alpha\}$. A necessary condition for this to happen is that $G_{\mathscr{S}_\alpha}$ is connected as well. Therefore, from now on we assume that $G_{\mathscr{S}_\alpha}$ is connected as well. Otherwise the elimination procedure does not apply. For a general PTM system with a selected cut $\mathscr{S}_\alpha$, we obtain the following proposition:

Proposition 3 *The graph $G_{\mathscr{S}_{\alpha,l}}$ has at least one rooted spanning tree if and only if there is only one independent conservation law involving the species in $\mathscr{S}_{\alpha,l} \cup \mathscr{Y}_{\alpha,l}$. Consequently, if this is the case for all l, then the non-interacting graphs provide a maximal set of independent conservation laws involving the species in $\mathscr{S}_\alpha \cup \mathscr{Y}_\alpha$.*

Proof For a fixed l, the non-interacting graphs of $G_{\mathscr{S}_{\alpha,l},\mathscr{Y}_{\alpha,l}}$ provide all conservation laws involving only $\mathscr{S}_{\alpha,l} \cup \mathscr{Y}_{\alpha,l}$ if and only if all conservation laws are multiples of $\sum_{S_i \in \mathscr{S}_{\alpha,l}} S_i + \sum_{Y_k \in \mathscr{Y}_{\alpha,l}} Y_k$, which is the case if and only if the rank of B_l is $N_{\alpha,l} - 1$. As stated above this is equivalent to the existence of a rooted spanning tree in $G_{\mathscr{S}_{\alpha,l}}$. □

Remark 4 In particular, the proposition holds if $G_{\mathscr{S}_\alpha}$ is strongly connected. If $\widehat{G}_{\mathscr{Y}_\alpha}$ is strongly connected, then to check that $G_{\mathscr{S}_\alpha}$ is strongly connected we do not need to calculate the labels of $G_{\mathscr{S}_\alpha}$. Whether there is an edge or not between two nodes follows from the set of reactions, cf. Lemma 5.

For simplicity we assume that there exists a spanning tree of $G_{\mathscr{S}_\alpha}$ rooted at S_1. Then, the variables $S_2, \ldots, S_{N_\alpha}$ can be solved in the coefficient field $\mathbb{R}(\mathrm{Con} \cup \mathscr{S}_\alpha^c \cup \{S_1\})$.

In particular, using Cramer's rule and the Matrix-Tree theorem, we obtain

$$S_j = \frac{(-1)^{j+1} B_{(1,j)}}{B_{(1,1)}} = \frac{\sigma_j(\mathscr{S}_\alpha^c)}{\sigma_1(\mathscr{S}_\alpha^c)} S_1 = r_j^S(\mathscr{S}_\alpha^c) S_1, \text{ where } \sigma_j(\mathscr{S}_\alpha^c) = \sum_{\tau \in \Theta(S_j)} \pi(\tau) \tag{14}$$

for $j = 2, \ldots, N_\alpha$. By assumption, $\sigma_1(\mathscr{S}_\alpha^c) \neq 0$ is S-positive and $\sigma_j(\mathscr{S}_\alpha^c)$ is either a zero or S-positive element of $\mathbb{R}(\mathrm{Con})[\mathscr{S}_\alpha^c]$. If the graph $G_{\mathscr{S}_\alpha}$ is strongly connected, then $\sigma_j(\mathscr{S}_\alpha^c) \neq 0$ for all j and any choice of S_j could be used instead of S_1. Further:

Proposition 4 *The subgraph $G_{\mathscr{S}_{\alpha,l}}$ of the graph $G_{\mathscr{S}_\alpha}$ is strongly connected if and only if $\sigma_j(\mathscr{S}_\alpha^c)$ is a non-zero polynomial in $\mathbb{R}(\mathrm{Con})[\mathscr{S}_\alpha^c]$ for all $S_j \in \mathscr{S}_{\alpha,l}$.*

The results shown above provide a proof of the following lemma.

Lemma 6 *If a substrate $S_{t_1} \in \mathscr{S}_\alpha^c$ is a variable in $\sigma_j(\mathscr{S}_\alpha^c)$ for some $S_j \in \mathscr{S}_\alpha$, then there exist $S_{i_1}, \ldots, S_{i_m} \in \mathscr{S}_\alpha$ with $S_{i_m} = S_j$ such that $S_{i_1} + S_{t_1} \overset{\mathscr{Y}_\alpha}{\Rightarrow} S_{i_2} + S_{t_2}$ for some $S_{t_2} \in \mathscr{S}_\alpha^c$ and for $l = 2, \ldots, m-1$, either S_{i_l} reacts to $S_{i_{l+1}}$ or $S_{i_l} + S_{t_l} \overset{\mathscr{Y}_\alpha}{\Rightarrow} S_{i_{l+1}} + S_{t_{l+1}}$ for some $S_{t_l}, S_{t_{l+1}} \in \mathscr{S}_\alpha^c$.*

After substitution of the value of S_j given in (14) into Y_k (9) we obtain

$$Y_k = r_k^Y(\mathscr{S}_\alpha^c) S_1, \tag{15}$$

where r_k^Y is either zero or an S-positive rational function in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con})$. If $\widehat{G}_{\mathscr{Y}_\alpha}$ is strongly connected, $\sigma_j(\mathscr{S}_\alpha^c) \neq 0$, $\sigma_i(\mathscr{S}_\alpha^c) \neq 0$, and $\mu_{i,j}^k \neq 0$ for some i, j, then this function is non-zero.

Conservation laws The sum of the species concentrations in $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$ is conserved. If the total amount $\overline{A}_1 = S_1 + \cdots + S_{N_\alpha} + Y_1 + \cdots + Y_{P_\alpha}$ is given, we obtain

$$\overline{A}_1 = (1 + r_2^S(\mathscr{S}_\alpha^c) + \cdots + r_{N_\alpha}^S(\mathscr{S}_\alpha^c) + r_1^Y(\mathscr{S}_\alpha^c) + \cdots + r_{P_\alpha}^Y(\mathscr{S}_\alpha^c)) S_1.$$

The coefficient of S_1 is an S-positive element of $\mathbb{R}(\mathrm{Con} \cup \mathscr{S}_\alpha^c)$ and thus,

$$S_1 = \overline{r}_1^S(\mathscr{S}_\alpha^c) := \frac{\overline{A}_1}{1 + r_2^S(\mathscr{S}_\alpha^c) + \cdots + r_{N_\alpha}^S(\mathscr{S}_\alpha^c) + r_1^Y(\mathscr{S}_\alpha^c) + \cdots + r_{P_\alpha}^Y(\mathscr{S}_\alpha^c)}$$

with $\overline{r}_1^S$ an S-positive rational function in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con} \cup \{\overline{A}_1\})$.

If $\sigma_1(\mathscr{S}_\alpha^c) \neq 0$ when evaluated at some non-negative values for the substrates in $\mathscr{S}_\alpha^c$, and $\overline{A}_1 > 0$, then $S_1 > 0$. Further, $\sigma_1(\mathscr{S}_\alpha^c) = 0$ only if some substrates in $\mathscr{S}_\alpha^c$ are zero. This remark and Proposition 4 imply:

Proposition 5 *Assume that the graph $G_{\mathscr{S}_{\alpha,l}}$ has a rooted spanning tree and that there exists a non-negative steady state with positive values for the substrates in $\mathscr{S}_\alpha^c$. Then, $G_{\mathscr{S}_{\alpha,l}}$ is strongly connected if and only if for any total amount $\overline{A}_l > 0$, the corresponding steady-state value satisfies $S_j \neq 0$ for all $S_j \in \mathscr{S}_{\alpha,l}$.*

By substitution of S_1 by $\overline{r}_1^S$, we obtain

$$Y_k = \overline{r}_k^Y(\mathscr{S}_\alpha^c) := r_k^Y(\mathscr{S}_\alpha^c)\overline{r}_1^S(\mathscr{S}_\alpha^c), \qquad S_j = \overline{r}_j^S(\mathscr{S}_\alpha^c) := r_j^S(\mathscr{S}_\alpha^c)\overline{r}_1^S(\mathscr{S}_\alpha^c) \quad (16)$$

with $\overline{r}_k^Y, \overline{r}_j^S$ either zero or S-positive rational functions in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con} \cup \{\overline{A}_1\})$.

Proposition 6 *Assume that for $l = 1, \ldots, n_\alpha$, there is a spanning tree of $G_{\mathscr{S}_{\alpha,l}}$ rooted at some species S_{i_l}. Then, there exists a zero or S-positive rational function r_j^S in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con})$, such that equation (11) for $S_j \in \mathscr{S}_{\alpha,l}$ is satisfied in $\mathbb{R}(\mathrm{Con} \cup \mathscr{S}_\alpha^c)$ if and only if*

$$S_j = r_j^S(\mathscr{S}_\alpha^c)S_{i_l}, \qquad S_j \in \mathscr{S}_{\alpha,l}.$$

Further, there exists an S-positive rational function $\overline{r}_{i_l}^S$ in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con} \cup \{\overline{A}_l\})$, such that the conservation law $\overline{A}_l = \sum_{S \in \mathscr{S}_{\alpha,l}} S + \sum_{Y \in \mathscr{Y}_{\alpha,l}} Y$ is fulfilled if and only if

$$S_{i_l} = \overline{r}_{i_l}^S(\mathscr{S}_\alpha^c). \quad (17)$$

In Example (1), considering the cut $\mathscr{S}_\alpha = \{S_1, S_2, S_3\}$, the graph $G_{\mathscr{S}_\alpha}$ has two connected components: S_3, which does not allow further eliminations, and $S_1 \underset{\widetilde{b}_{1,2}}{\overset{\widetilde{b}_{2,1}}{\rightleftharpoons}} S_2$, which is strongly connected. Selecting S_1 as the non-eliminated substrate we obtain

$$S_2 = \frac{\widetilde{b}_{2,1}S_1}{\widetilde{b}_{1,2}} = \frac{d_{1,2}S_1}{b_{1,4}^2\mu_{2,5}^2S_5}, \quad Y_1 = \frac{d_{1,2}\mu_{2,5}^1S_1}{b_{1,4}^2\mu_{2,5}^2} = \frac{d_{1,2}(b_{1,4}^2 + c_{2,1})S_1}{b_{1,4}^2c_{1,2}}, \quad Y_2 = \frac{d_{1,2}S_1}{b_{1,4}^2}.$$

The equations for the total amounts $\overline{A}_1 = S_3 + Y_3$ and $\overline{A}_2 = S_1 + S_2 + Y_1 + Y_2$ give:

$$\overline{A}_1 = S_3(1 + \mu_{3,4}^3S_4), \quad \overline{A}_2 = \frac{d_{1,2}}{b_{1,4}^2}\left(\frac{b_{1,4}^2}{d_{1,2}} + \frac{1}{\mu_{2,5}^2S_5} + \frac{b_{1,4}^2 + c_{2,1}}{c_{1,2}} + 1\right)S_1.$$

Let $\widetilde{r}_1^S(S_4, S_5) = \overline{A}_2\left(\frac{b_{1,4}^2}{d_{1,2}} + \frac{1}{\mu_{2,5}^2S_5} + \frac{b_{1,4}^2+c_{2,1}}{c_{1,2}} + 1\right)^{-1}$. Then,

$$S_1 = \frac{b_{1,4}^2}{d_{1,2}}\widetilde{r}_1^S(S_4, S_5), \qquad S_2 = \frac{\widetilde{r}_1^S(S_4, S_5)}{\mu_{2,5}^2S_5}, \qquad S_3 = \frac{\overline{A}_1}{1 + \mu_{3,4}^3S_4},$$
$$Y_1 = \frac{(b_{1,4}^2 + c_{2,1})}{c_{1,2}}\widetilde{r}_1^S(S_4, S_5), \quad Y_2 = \widetilde{r}_1^S(S_4, S_5), \quad Y_3 = \frac{\mu_{3,4}^3\overline{A}_1S_4}{1 + \mu_{3,4}^3S_4}. \quad (18)$$

Thus, all species are given as S-positive rational functions of S_4, S_5 in the coefficient field $\mathbb{R}(\mathrm{Con} \cup \{\overline{A}_1, \overline{A}_2\})$.

4.4 Steady-state equations

To summarize, at steady state the intermediate complexes $\mathscr{Y}$ can be expressed as rational functions of the substrates $\mathscr{S}$ and therefore eliminated. Further, provided a cut $\mathscr{S}_\alpha$ exists (and some spanning trees), the variables $\mathscr{S}_\alpha$ can be expressed as functions of $\mathscr{S}_\alpha^c = \mathscr{S}\backslash\mathscr{S}_\alpha$ and therefore also eliminated. For the latter statement, we make use of the conservation laws (with given total amounts) for the species in $\mathscr{S}_\alpha$ determined by the connected components of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$. Assume that the graph $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ has n_α connected components, each admitting a rooted spanning tree. By Proposition 3, each component provides one conservation law.

Specifically, consider the steady-state equations (11) for $\mathscr{S}_\alpha^c$. Substituting the expressions in (15) and (14) for $\mathscr{Y}$ and $\mathscr{S}_\alpha$ provides the steady-state equations in terms of $\mathscr{S}_\alpha^c$ and the selected variables S_{i_l} (one for each connected component of $G_{\mathscr{S}_\alpha}$). Using the conservation law, equation (16) provides steady-state equations in terms of $\mathscr{S}_\alpha^c$ only. Let $\mathscr{S}_\alpha^c = \{S_{N_\alpha+1}, \ldots, S_N\}$. Since $\mathscr{S}_\alpha$ is a cut, if $S_i \in \mathscr{S}_\alpha^c$ and $d_{i,j} \neq 0$ or $d_{j,i} \neq 0$, then $S_j \in \mathscr{S}_\alpha^c$. Using Eqs. (16) and (11), the equation $\dot{S}_i = 0$ for $i = N_\alpha + 1, \ldots, N$ can be written as:

$$0 = \sum_{k=1}^{P}\left(\sum_{j=1}^{N_\alpha}\epsilon_{i,j}a_{i,j}^k\overline{r}_j^S(\mathscr{S}_\alpha^c)S_i + \sum_{j=N_\alpha+1}^{N}\epsilon_{i,j}a_{i,j}^k S_i S_j\right) + \sum_{j=1}^{N}\sum_{k=1}^{P}\epsilon_{i,j}b_{i,j}^k\overline{r}_k^Y(\mathscr{S}_\alpha^c) + \sum_{j=N_\alpha+1}^{N}(d_{j,i}S_j - d_{i,j}S_i).$$

The right-hand side of the equation is a rational function in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con}\cup\{\overline{A}_1, \ldots, \overline{A}_{n_\alpha}\})$. Since the non-vanishing $\overline{r}_j^S(\mathscr{S}_\alpha^c)$ and $\overline{r}_k^Y(\mathscr{S}_\alpha^c)$ are S-positive rational functions, their denominators can be chosen to be S-positive. Thus, we can multiply the expressions by the denominators without changing the *positive* solutions and a polynomial equation is obtained. Therefore, we define $\Phi_u(\mathscr{S}_\alpha^c) = 0$ to be the equation obtained from $\dot{S}_u = 0$ after elimination of $\mathscr{Y}$ and $\mathscr{S}_\alpha$ and removal of denominators in this way.

Extend the set of n_α conservation laws arising from $\mathscr{S}_\alpha$ to a maximal set of $\dim(\Gamma^\perp)$ laws. Following a similar procedure, each additional equation for a conserved amount $\overline{A}_l = \sum_i \lambda_i^l S_i + \sum_k \mu_k^l Y_k$ for $l = n_\alpha + 1, \ldots, \dim(\Gamma^\perp)$ is transformed into an equation $\overline{A}_l = \varphi_l(\mathscr{S}_\alpha^c)$. The function φ_l is a rational function in $\mathscr{S}_\alpha^c$ with coefficients in $\mathbb{R}(\mathrm{Con}\cup\{\overline{A}_1, \ldots, \overline{A}_{n_\alpha}\})$ and its denominator can be chosen to be S-positive.

Theorem 1 *Consider a PTM system for which there exists a cut $\mathscr{S}_\alpha$ and let $\mathscr{S}_{\alpha,l}$, $l = 1, \ldots, n_\alpha$, be the node sets of the connected components of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$. Further, assume that each subgraph $G_{\mathscr{S}_{\alpha,l}}$ admits a rooted spanning tree. If total amounts $\overline{A}_l$ are given for the n_α connected components of $G_{\mathscr{S}_\alpha,\mathscr{Y}_\alpha}$ and the $\dim(\Gamma^\perp) - n_\alpha$ additional conservation laws, then the non-negative steady states of the system with positive values for the substrates in $\mathscr{S}_\alpha^c$ are in one-to-one correspondence with the positive*

solutions to

$$\Phi_u(\mathscr{S}_\alpha^c) = 0, \qquad \overline{A}_l = \varphi_l(\mathscr{S}_\alpha^c)$$

for $u = N_\alpha + 1, \ldots, N$ *and* $l = n_\alpha + 1, \ldots, \dim(\Gamma^\perp)$. *Here* $\Phi_u(\mathscr{S}_\alpha^c)$ *is a polynomial and* $\varphi_l(\mathscr{S}_\alpha^c)$ *is a rational function in* $\mathscr{S}_\alpha^c$ *with coefficients in* $\mathbb{R}(\mathrm{Con} \cup \{\overline{A}_1, \ldots, \overline{A}_{n_\alpha}\})$.

Proof We have shown that any non-negative steady state with positive values for all substrates in $\mathscr{S}_\alpha^c$ must satisfy these equations. For the reverse, consider a positive solution $s = (s_{N_\alpha+1}, \ldots, s_N)$ to the equations $\Phi_u(\mathscr{S}_\alpha^c) = 0$ and $\overline{A}_l = \varphi_l(\mathscr{S}_\alpha^c)$. For $i = 1, \ldots, N_\alpha$, define s_i through equation (17) and $y_k, k = 1, \ldots, P$, through equation (9). For positive rate constants and positive total amounts, s_i, y_k are non-negative (because the rational functions defining them are S-positive). By construction these definitions automatically ensure that the conservation laws with total amounts $\overline{A}_l, l = 1, \ldots, N_\alpha$, are satisfied (cf. Proposition 6).

By Proposition 2, the values $y_1, \ldots, y_P$ satisfy (2) for all k and hence the steady-state equations of the intermediate complexes are satisfied. By Proposition 6 the values $s_1, \ldots, s_{N_\alpha}$ satisfy (11). Since the latter is just (3) after substitution of (9), we see that (3) holds as well. Since $\Phi_u(\mathscr{S}_\alpha^c) = 0$ is the steady-state equation $\dot{S}_u = 0$ for $u = N_\alpha + 1, \ldots, N$, after substitution of (9) and (17), this equation is also satisfied. The same reasoning applies to the equation $\overline{A}_l = \varphi_l(\mathscr{S}_\alpha^c), l > n_\alpha$. Thus, $S_i = s_i, i = 1, \ldots, N$, and $Y_k = y_k, k = 1, \ldots, P$, is a solution to the steady-state equations and satisfy the conservation laws with total amounts $\overline{A}_l$. □

If each $G_{\mathscr{S}_{\alpha,l}}$ is strongly connected, then this theorem together with Proposition 2(iii) and Proposition 4 provide a statement about *positive* steady states, and not only about *non-negative* steady states.

Corollary 1 *With the notation of Theorem 1, assume that* $\widehat{G}_{\mathscr{Y}}$ *and each* $G_{\mathscr{S}_{\alpha,l}}$ *are strongly connected. Then, the positive steady states of the system are in one-to-one correspondence with the positive solutions to*

$$\Phi_u(\mathscr{S}_\alpha^c) = 0, \qquad \overline{A}_l = \varphi_l(\mathscr{S}_\alpha^c)$$

for $u = N_\alpha + 1, \ldots, N$ *and* $l = n_\alpha + 1, \ldots, \dim(\Gamma^\perp)$. *Further, if at steady state* $S_i = 0, S_i \in \mathscr{S}_\alpha$, *or* $Y_k = 0, Y_k \in \mathscr{Y}$, *then there exists some* $S_j \in \mathscr{S}_\alpha^c$ *such that also* $S_j = 0$.

In Example (1), $\mathscr{S}_\alpha^c = \{S_4, S_5\}$, $\dim(\Gamma^\perp) - n_\alpha = 1$ and there is only one additional conservation law, $\overline{A}_3 = S_4 + S_5 + Y_1 + Y_2 + Y_3$. Using Eq. (18), the elimination procedure leads to the steady-state equations consisting of $\dot{S}_4 = 0$ (that is, equation $\Phi_4(S_4, S_5) = 0$) and $\overline{A}_3$ (that is, equation $\varphi_3(S_4, S_5) = \overline{A}_3$):

$$0 = \Phi_4(S_4, S_5) = -b^3_{3,5}\mu^3_{3,4}\overline{A}_1 S_4 \widetilde{d}^S_1(S_4, S_5) + b^2_{1,4}(1 + \mu^3_{3,4}S_4)\widetilde{n}^S_1(S_4, S_5),$$

$$\overline{A}_3 = \varphi_3(S_4, S_5) = S_4 + S_5 + \frac{(b^2_{1,4} + c_{2,1} + c_{1,2})}{c_{1,2}}\widetilde{r}^S_1(S_4, S_5) + \frac{\mu^3_{3,4}\overline{A}_1 S_4}{1 + \mu^3_{3,4}S_4},$$

where $\widetilde{r}_1^S(S_4, S_5) = \widetilde{n}_1^S(S_4, S_5)/\widetilde{d}_1^S(S_4, S_5)$. Since the conditions of Corollary 1 are fulfilled, any positive solution of this reduced system provides a positive steady state of the PTM system. The steady states of the other species, $S_1, S_2, S_3, Y_1, Y_2, Y_3$, are found from (18). In this specific example, the first equation is linear in S_4 and S_5 separately, and hence either S_4 or S_5 can be eliminated as well. This provides a polynomial equation in the remaining variables. However, S-positivity is no longer guaranteed.

In the example we intentionally selected $\mathscr{S}_\alpha$ to have the highest possible number of elements, because all variables in $\mathscr{S}_\alpha$ are subsequently eliminated. In some systems (see e.g. Sect. 5.2) there two different cuts $\mathscr{S}_\alpha, \mathscr{S}'_\alpha$ might exist, such that the union is not a cut, but still all variables in $\mathscr{S}_\alpha \cup \mathscr{S}'_\alpha$ can be eliminated. Thus, more species might be eliminated if different cuts are considered.

5 Examples

5.1 TG framework

Thomson and Gunawardena (2009) provide a linear elimination procedure for the special case in which the set of substrates is partitioned into two distinct sets. In their context, a *PTM system* (here called *TG system*) consists of three non-empty and disjoint sets of species called enzymes, substrates, and intermediate complexes:

$$\text{Enz} = \{E_1, \ldots, E_L\},\ \text{Sub} = \{S_1, \ldots, S_N\},\ \mathscr{Y} = \{Y_1, \ldots, Y_P\},$$

and a set of reactions $\text{Rct} = R_a \cup R_b \cup R_c$ with $R_a = \{E_i + S_j \xrightarrow{a_{i,j}^k} Y_k | (i, j, k) \in I_a\}$, $R_b = \{Y_k \xrightarrow{b_{i,j}^k} E_i + S_j | (i, j, k) \in I_b\}$ and $R_c = \{Y_i \xrightarrow{c_{i,j}} Y_j | (i, j) \in I_c\}$, for $I_a, I_b \subseteq \{1, \ldots, L\} \times \{1, \ldots, N\} \times \{1, \ldots, P\}$ and $I_c \subseteq \{1, \ldots, P\}^2$, such that

(i) All chemical species are involved in at least one reaction.
(ii) For every intermediate complex Y_k there is at most one enzyme $E_{\eta(k)}$, such that $(\eta(k), j, k) \in R_a \cup R_b$ for some j.
(iii) If two intermediate complexes Y_k, Y_v are 1-linked, then $E_{\eta(k)} = E_{\eta(v)}$.

The graph $\widehat{G}_{\mathscr{Y}}$ and each connected component of the graph G_{Sub} are required to be strongly connected. In particular, the assumption that $\widehat{G}_{\mathscr{Y}}$ is strongly connected implies that any Y_k ultimately reacts to $S_i + S_j$ for some i, j. If we let $\mathscr{S} = \text{Sub} \cup \text{Enz}$, it follows that a TG system is a special case of a PTM system.

Essentially, they consider PTM systems in which enzymes are not allowed to be modified and substrates are not allowed to modify. Let $\mathscr{S}_\alpha = \text{Sub}$, $\mathscr{S}_\alpha^c = \text{Enz}$ and note that $\mathscr{Y}_\alpha = \mathscr{Y}_\alpha^c = \mathscr{Y}$. Properties (i)-(iii) imply that $\mathscr{S}_\alpha$ and $\mathscr{S}_\alpha^c$ are cuts. Thus our framework applies to TG systems and is in fact an extension.

By assumption (iii) the graph $G_{\mathscr{S}_\alpha^c, \mathscr{Y}}$ has L connected components that give L conservation laws for the enzymes: $\overline{E}_i = E_i + \sum_{k|\eta(k)=i} Y_k$, for $i = 1, \ldots, L$. With the notation of Lemma 2, $N_\alpha^c = n_\alpha^c = L$, $P_\alpha^c = 0$, so that $N_\alpha^c + P_\alpha^c - n_\alpha^c = 0$. Thus a set of independent conservation laws for a TG system can be derived from the non-interacting graphs of $G_{\mathscr{S}, \mathscr{Y}}$. Further, the form of R_a and R_b ensures that

any non-interacting graph contains species from Enz or Sub, but not both. Thus, all conservation laws are associated with a connected component of $G_{\mathrm{Enz},\mathscr{Y}}$ or $G_{\mathrm{Sub},\mathscr{Y}}$.

If all intermediate complexes ultimately dissociate into an enzyme and a substrate, and each connected component of $G_{\mathscr{S}_\alpha}$ has a rooted spanning tree, then the variables in $\mathscr{S}_\alpha \cup \mathscr{Y}$ can be eliminated. In this case, the steady-state equations are reduced to L equations derived from the total amount of enzymes. Observe that these conditions are milder than requiring $\widehat{G}_{\mathscr{Y}}$ and $G_{\mathscr{S}_\alpha}$ to be strongly connected.

5.2 Signaling cascades

Our setting is well-suited to study elimination of variables in signaling cascades. Signaling cascades form a special type of PTM systems where some substrates also act as enzymes (Feliu et al. 2012; Huang and Ferrell 1996) and thus extends TG systems.

Definition 5 A *signaling cascade* is a collection of TG systems $R^1, \ldots, R^n$, with corresponding sets of species

$$\mathrm{Enz}^i = \{E_1^i, \ldots, E_{L_i}^i\}, \quad \mathrm{Sub}^i = \{S_1^i, \ldots, S_{N_i}^i\}, \quad \mathscr{Y}^i = \{Y_1^i, \ldots, Y_{P_i}^i\}$$

and sets of reactions $\mathrm{Rct}^i = R_a^i \cup R_b^i \cup R_c^i$, for $i = 1, \ldots, n$, satisfying the following conditions:

(i) $(\mathrm{Enz}^i \cup \mathrm{Sub}^i \cup \mathscr{Y}^i) \cap (\mathrm{Enz}^j \cup \mathrm{Sub}^j \cup \mathscr{Y}^j) = \{E_1^{i+1}\} = \{S_{N_i}^i\}$ for $j = i+1$ and $i < n$. The intersection is empty otherwise.
(ii) For $i = 1, \ldots, n$, the graph G_{Sub^i} admits a spanning tree rooted at $S_{N_i}^i$.
(iii) All intermediate complexes ultimately dissociate into two substrates.

Condition (i) implies that a signaling cascade consists of independent TG systems joined by a single substrate acting as an enzyme in the layer below. Condition (iii) ensures that the intermediate complexes can be eliminated.

Let $N = N_1 + \cdots + N_n$, $L = L_1 + \cdots + L_n$ and $\mathscr{S} = \bigcup_i \mathrm{Enz}^i \cup \mathrm{Sub}^i$. Consider the closed subset $\mathrm{Sub}^i \subset \mathscr{S}$ with associated set $\mathscr{Y}_{\mathrm{Sub}^i} = \mathscr{Y}^i \cup \{Y_k \in \mathscr{Y}^{i+1} \mid E_{\eta(k)} = S_{N_i}^i\}$ for $i = 1, \ldots, n$. By definition, substrates in Sub^i do not interact and thus Sub^i is a cut.

Proposition 6 ensures that the variables in Sub^i can be eliminated and that

$$S_j^i = r_j^i(\mathrm{Enz}^i) S_{N_i}^i, \qquad S_j^i \in \mathrm{Sub}^i \setminus \{S_{N_i}^i\}.$$

By Lemma 6, r_j^i depends on the species in Enz^i only: if $S_u^i + S_t$ ultimately reacts to $S_j^i + S_r$ via $\mathscr{Y}$ for some species S_u^i in Sub^i and $S_r \in \mathscr{S} \setminus \mathrm{Sub}^i$, then since $S_j^i \neq S_{N_i}^i$, $S_t = S_r = E_\eta^i$ for some $E_\eta^i \in \mathrm{Enz}^i$. Further, if $Y_k \in \mathscr{Y}_{\mathrm{Sub}^i}$, we let $Y_k = r_k^Y(\mathrm{Enz}^i) S_{N_i}^i$ be the corresponding rational function.

5.2.0.1 Conservation laws. Since G_{Sub^i} is connected and admits a rooted spanning tree, there is only one conservation law among the species in $\mathrm{Sub}^i \cup \mathscr{Y}_{\mathrm{Sub}^i}$,

$$\overline{S}^i = \sum_{S^i_j \in \mathrm{Sub}^i} r^i_j(\mathrm{Enz}^i) S^i_{N_i} + \sum_{Y_k \in \mathscr{Y}_{\mathrm{Sub}^i}} r^Y_k(\mathrm{Enz}^i) S^i_{N_i}, \tag{19}$$

where $\overline{S}^i$ is the total amount. For $i = n$, $S^n_{N_n} \notin \mathrm{Enz}^n$, and therefore $S^n_{N_n}$ can be isolated from (19) and expressed as a rational function in Enz^n.

Thus, if we let $\mathrm{Enz} = \bigcup_i \mathrm{Enz}^i$, then the species in $\mathscr{S} \setminus \mathrm{Enz}$ are given as rational functions in Enz with coefficients in $\mathbb{R}(\mathrm{Con})$. Condition (iii) implies that for $E^i_\eta \in \mathrm{Enz}^i \setminus \{S^i_{N_i}\}$, $\{E^i_\eta\}$ is a cut with associated (connected) graph $G_{E^i_\eta, \mathscr{Y}_{E^i_\eta}}$. Thus, if the total amount $\overline{E}^i_\eta$ is provided, the steady states must fulfill the equality

$$\overline{E}^i_\eta = E^i_\eta + \sum_{k|\eta=\eta(k)} Y_k = E^i_\eta + \sum_{k|\eta=\eta(k)} r^Y_k(\mathrm{Enz}^i) S^i_{N_i}. \tag{20}$$

We conclude that the non-negative steady states of a signaling cascade are solutions to L equations in Enz with coefficients in $\mathbb{R}(\mathrm{Con})$, provided that total amounts for Enz are given: $\overline{S}^1, \ldots, \overline{S}^{n-1}$ in (19) for the enzymes $S^i_{N_i}$ and $\overline{E}^i_\eta$ in (20) for $E^i_\eta \in \mathrm{Enz} \setminus \{S^1_{N_1}, \ldots, S^{n-1}_{N_{n-1}}\}$.

The number of conservation laws obtained in this way is $m = \sum_i L_i + 1 = L + 1$ (remember $\overline{S}^n$). Let $\epsilon = 1$ if n is even and 0 otherwise, and define $\epsilon^c = 1 - \epsilon$. The cuts provide all conservation laws. In fact the graph associated to the cut

$$\mathscr{S}_\alpha = \bigcup_{i \text{ even}} \mathrm{Sub}^i \cup \bigcup_{i \text{ odd}} \mathrm{Enz}^i$$

has $n_\alpha = \epsilon + \sum_{i \text{ odd}} L_i$ connected components and thus, $n^c_\alpha = \epsilon^c + \sum_{i \text{ even}} L_i$. We have $N_\alpha = \sum_{i \text{ odd}} L_i + \sum_{i \text{ even}} (N_i - 1) + \epsilon$, and $N^c_\alpha = \sum_{i \text{ even}} L_i + \sum_{i \text{ odd}} (N_i - 1) + \epsilon^c$. Further, $\mathscr{Y}_\alpha = \mathscr{Y}$, so that $P^c_\alpha = 0$.

Let $\Gamma = N^c_\alpha - n^c_\alpha = \sum_{i \text{ odd}} (N_i - 1)$. By Lemma 2, if there are Γ independent terms in $\mathscr{S}^c_\alpha \cap \Gamma$, then all conservation laws come from non-interacting graphs. By assumption, the graph G_{Sub^i} has a spanning tree rooted at $S^i_{N_i}$. This implies that for every $S_u \neq S^i_{N_i}$ in Sub^i, there exists a directed path $S_u \to S_{k_1} \to \cdots \to S_{k_r} \to S^i_{N_i}$. The conditions of a TG system and Lemma 5 ensure that for any edge $S_{k_v} \to S_{k_s}$ in G_{Sub^i}, $E + S_{k_v}$ ultimately reacts to $E + S_{k_s}$ via $\mathscr{Y}$ for some E. We see that $S_u - S^i_{N_i} \in \mathscr{S}^c_\alpha \cap \Gamma$ for all $S_u \neq S^i_{N_i}$ in Sub^i, implying that indeed there are Γ independent vectors in $\mathscr{S}^c_\alpha \cap \Gamma$.

5.3 Biological examples

MAPK signaling cascade We consider the first two layers of the MAPK cascade with a one-site modification in the first layer and a two-site modification in the second layer. In the latter, dephosphorylation is considered sequential whereas this is not the case for phosphorylation (Markevich et al. 2004).

The reactions of the system in the first layer are

$$E + S_0^1 \rightleftarrows Y_1^1 \longrightarrow E + S_1^1 \qquad F_1 + S_1^1 \rightleftarrows Y_2^1 \longrightarrow F_1 + S_0^1,$$

accounting for phosphorylation and dephosphorylation, respectively, via a Michaelis–Menten mechanism. In the second layer we have the phosphorylation reactions

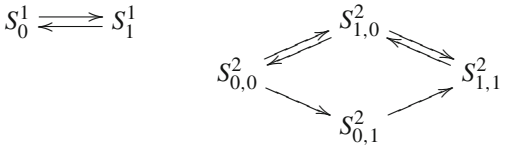
$$S_1^1 + S_{0,0}^2 \rightleftarrows Y_1^2 \longrightarrow S_1^1 + S_{1,0}^2 \qquad S_1^1 + S_{0,0}^2 \rightleftarrows Y_2^2 \longrightarrow S_1^1 + S_{0,1}^2$$
$$S_1^1 + S_{1,0}^2 \rightleftarrows Y_3^2 \longrightarrow S_1^1 + S_{1,1}^2 \qquad S_1^1 + S_{0,1}^2 \rightleftarrows Y_4^2 \longrightarrow S_1^1 + S_{1,1}^2.$$

Dephosphorylation proceeds sequentially in the following way:

$$F_2 + S_{1,1}^2 \rightleftarrows Y_5^2 \longrightarrow F_2 + S_{1,0}^2 \qquad F_2 + S_{1,0}^2 \rightleftarrows Y_6^2 \longrightarrow F_2 + S_{0,0}^2.$$

The sets of enzymes are $\mathrm{Enz}^1 = \{E, F_1\}$ and $\mathrm{Enz}^2 = \{S_1^1, F_2\}$. The sets of substrates are $\mathrm{Sub}^1 = \{S_0^1, S_1^1\}$ and $\mathrm{Sub}^2 = \{S_{0,0}^2, S_{1,0}^2, S_{0,1}^2, S_{1,1}^2\}$. The sets of intermediate complexes are $\mathscr{Y}^1 = \{Y_1^1, Y_2^1\}$ and $\mathscr{Y}^2 = \{Y_1^2, Y_2^2, Y_3^2, Y_4^2, Y_5^2, Y_6^2\}$. We have $\mathrm{Enz}^2 \cap \mathrm{Sub}^1 = \{S_1^1\}$, so that the modified substrate in the first layer is a kinase of the next layer. The superindex denotes the layer, while the subindex denotes phosphorylation state (the presence of the phosphate group is represented by 1).

Each connected component of $\widehat{G}_{\mathscr{Y}}$ consists of a single intermediate complex and is thus strongly connected. The graphs $\widehat{G}_{\mathrm{Sub}^1}$ and $\widehat{G}_{\mathrm{Sub}^2}$ are



which are also strongly connected. The conservations laws (all derived from non-interacting graphs) are $\overline{E} = E + Y_1^1, \overline{F}_1 = F_1 + Y_2^1, \overline{F}_2 = F_2 + Y_5^2 + Y_6^2, \overline{S}^1 = S_0^1 + S_1^1 + Y_1^1 + Y_2^1 + Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2$ and $\overline{S}^2 = S_{0,0}^2 + S_{1,0}^2 + S_{0,1}^2 + S_{1,1}^2 + Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 + Y_5^2$. Therefore, if total amounts are provided, then the steady states of the two-layer cascade are found as solutions to a system of four polynomial equations in four variables, namely E, F_1, F_2, S_1^1.

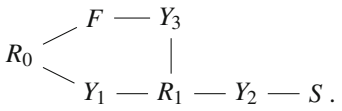
Receptor protein-tyrosine kinase Receptor protein-tyrosine kinases (RPTK) are cell surface receptors linked to enzymes that phosphorylate their substrate proteins in tyrosine residues. The common mechanism for activation is autophosphorylation following ligand-induced dimerization (Cooper and Hausman 2009, §15). The phosphorylated receptor serves as binding site to downstream signaling molecules, such as SH2 domain containing proteins. Further, the receptor can be dephosphorylated by several protein tyrosine phosphatases (PTPs) (Ostman and Bohmer 2001).

A simple model describing the phosphorylation state of an RPTK is:

$$2R_0 \rightleftarrows Y_1 \longrightarrow 2R_1 \qquad S + R_1 \rightleftarrows Y_2 \qquad F + R_1 \rightleftarrows Y_3 \longrightarrow F + R_0$$

where R_0, R_1 stand for the unphosphorylated and phosphorylated RPTK, respectively, S is a protein binding R_1, and F is a PTP.

We have $\mathscr{S} = \{R_0, R_1, S, F\}$ and $\mathscr{Y} = \{Y_1, Y_2, Y_3\}$. Note that $\mathscr{S}_\circlearrowright = \{R_0, R_1\}$ are the self-interacting substrates and thus cannot be part of a cut. First of all, the intermediate complexes can be eliminated in terms of $\mathscr{S}$. The graph $G_{\mathscr{S},\mathscr{Y}}$ is



The non-interacting graphs provide two conservation laws: $\overline{F} = F + Y_3$, and $\overline{S} = S + Y_2$, associated to the cut $\mathscr{S}_\alpha = \{F, S\}$. Thus, the substrates F, S can be eliminated. We conclude that at steady state all species are described as rational functions in R_0, R_1 and the non-negative steady states are in one-to-one correspondence with the non-negative solutions to the equations corresponding to $\dot{R}_0$ and the remaining conservation law $\overline{R} = R_0 + R_1 + 2Y_1 + Y_2 + Y_3$.

6 Discussion and perspectives

In this paper we have shown that the steady states of a PTM system are parameterized algebraically by a set of core variables selected among all substrates. The eliminated variables cannot be any set but must fulfil certain criteria. Further, the intermediate complexes can always be eliminated and expressed as polynomials in the substrates, or as rational functions in the core variables. The intermediate complexes are transient molecules involved in enzymatic catalytic activities and eventually dissociate into substrates. Elimination of intermediate complexes is a standard procedure known as the quasi-steady state approximation and is often applied to simplify the dynamics of the system. Here we have justified this procedure mathematically at steady state for general PTM systems.

An interesting consequence of the variable elimination procedure is that conditions for a steady state to have zero concentration(s) are derived. Several results on chemical reaction networks apply to strictly positive steady-state solutions and do not consider solutions on the boundary, e.g. Craciun and Feinberg (2005). Our conditions allow

the identification of systems with species for which the steady-state concentrations always are on the boundary.

Cuts can be identified algorithmically and thus a fully automated procedure for identifying sets of variables that can be eliminated from the steady-state equations can be implemented. However, such a procedure might be computationally expensive. Given a cut $\mathscr{S}_\alpha$, the minimal cuts $\mathscr{S}' \subseteq \mathscr{S}_\alpha$ can easily be identified, e.g. as the connected components of the graph $G_{\mathscr{S}_\alpha, \mathscr{Y}_\alpha}$. Importantly, elimination of the variables in $\mathscr{S}'$ from the steady-state equations and calculation of the corresponding rational expressions can also be done efficiently (algorithmically) using symbolic programming packages such as `Mathematica` or `Singular`.

The framework and the results derived here are in the setting of PTM systems, which are particularly useful for understanding cellular signaling. The specific structure of PTM systems ensures that the set of intermediate complexes can be fully eliminated and that the structure between the set of substrates and the set of intermediate complexes can be explored. However, we are aware that the notion of a cut and the graphical techniques apply more generally to chemical reaction networks taken with mass-action (Feliu and Wiuf 2011b). The special peculiarities and structure of the PTM systems and signaling cascades are however lost in the general exposition.

Acknowledgments E.F. is supported by a postdoctoral grant from the "Ministerio de Educación" of Spain and the project MTM2009-14163-C02-01 from the "Ministerio de Ciencia e Innovación". C.W. is supported by the Lundbeck Foundation, Denmark and the Leverhulme Trust, UK. Part of this work was done while E.F. and C.W. were visiting the Imperial College London in the fall 2011.

References

Angeli D, De Leenheer P, Sontag E (2010) Graph-theoretic characterizations of monotonicity of chemical networks in reaction coordinates. J Math Biol 61:581–616

Cooper GM, Hausman RE (2009) The cell, 5th edn. ASM Press, Washington

Cornish-Bowden A (2004) Fundamentals of enzyme kinetics, 3rd edn. Portland Press, London

Craciun G, Feinberg M (2005) Multiple equilibria in complex chemical reaction networks. I. The injectivity property. SIAM J Appl Math 65(5):1526–1546

Diestel R (2005) Graph theory, graduate texts in mathematics, vol 173, 3rd edn. Springer, Berlin

Feinberg M (1980) Lectures on chemical reaction networks 1. pp 1–18. http://www.che.eng.ohio-state.edu/~feinberg/LecturesOnReactionNetworks/

Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors I. The deficiency zero and deficiency one theorems. Chem Eng Sci 42(10):2229–2268

Feinberg M, Horn FJM (1977) Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. Arch Ration Mech Anal 66(1):83–97

Feliu E, Wiuf C (2011a) Enzyme sharing as a cause of multistationarity in signaling systems. J R Soc Interf doi:10.1098/rsif.2011.0664 (epub ahead of print, Nov 2)

Feliu E, Wiuf C (2011b) Variable elimination in chemical reaction networks with mass action kinetics. arXiv 1109.1505

Feliu E, Knudsen M, Andersen LN, Wiuf C (2012) An algebraic approach to signaling cascades with n layers. Bull Math Biol 74(1):45–72

Gross JL, Yellen J (2006) Graph theory and its applications, 2nd edn. Discrete mathematics and its applications. Chapman and Hall/CRC, Boca Raton

Heinrich R, Neel BG, Rapoport TA (2002) Mathematical models of protein kinase signal transduction. Mol Cell 9:957–970

Huang CY, Ferrell JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93:10078–10083

Kholodenko BN, Birtwistle MR (2009) Four-dimensional dynamics of MAPK information processing systems. Wiley Interdiscip Rev Syst Biol Med 1:28–44

King E, Altman C (1956) A schematic method of deriving the rate laws for enzyme-catalyzed reactions. J Phys Chem 60:1375–1378

Krell T, Lacal J, Busch A, Silva-Jimenez H, Guazzaroni ME, Ramos JL (2010) Bacterial sensor kinases: diversity in the recognition of environmental signals. Annu Rev Microbiol 64:539–559

Markevich NI, Hoek JB, Kholodenko BN (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. J Cell Biol 164:353–359

Ostman A, Bohmer FD (2001) Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatases. Trends Cell Biol 11:258–266

Thomson M, Gunawardena J (2009) The rational parameterization theorem for multisite post-translational modification systems. J Theor Biol 261:626–636

Tutte WT (1948) The dissection of equilateral triangles into equilateral triangles. Proc Camb Philos Soc 44:463–482

Ventura AC, Sepulchre JA, Merajver SD (2008) A hidden feedback in signaling cascades is revealed. PLoS Comput Biol 4:e1000041