Notes for a Systems Biology course

Biochemical reaction networks

Claudio Altafini

November 2015

Contents

4 Biochemical reaction networks

In this Chapter we study structural and dynamical properties of networks of biochemical reactions.

4.1 Reaction networks and S-R graphs

Recall for instance that for a bimolecular reaction of association

$$X_1 + X_2 \xrightarrow{k} X_3 \tag{1}$$

where k is the reaction rate constant, the mass-action ODEs are:

$$\begin{aligned} \frac{dx_1}{dt} &= -k\, x_1 x_2 \\ \frac{dx_2}{dt} &= -k\, x_1 x_2 \\ \frac{dx_3}{dt} &= k\, x_1 x_2 \end{aligned} \tag{2}$$

The ODEs are nonlinear (multilinear in this case, polynomial in general) and to have a nonambiguous representation one uses a *SR-graph* (Species-Reaction graph), i.e., a bipartite graph with two classes of nodes: molecular species and reactions, see Fig. 1. Notice in (2) that only the molecular species "upstream" of the reaction (i.e., the substrates) enter into the right hand side of the ODE. They enter with a minus sign in the equations for the substrates themselves (their concentration decreases) and with a plus sign for that of the product.

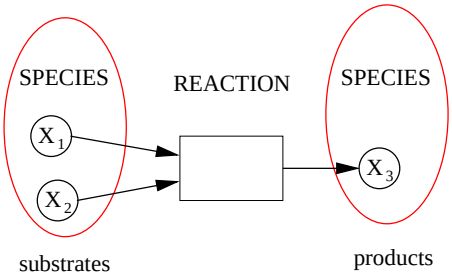

Figure 1: Species-Reactions graph for a single reaction.

When the stoichiometric coefficients are non-trivial (i.e., different from 1) then they can be indicated explicitly on the SR-graph. For instance for the reaction

$$pX_1 \; + \; X_2 \quad \underset{k_2}{\overset{k_1}{\rightleftarrows}} \quad X_3 \tag{3}$$

of ODEs:

$$\begin{aligned} \frac{dx_1}{dt} &= -p\,k_1\,x_1^p x_2 + pk_2\,x_3 \\ \frac{dx_2}{dt} &= -k_1\,x_1^p x_2 + k_2\,x_3 \\ \frac{dx_3}{dt} &= k_1\,x_1^p x_2 - k_2\,x_3 \end{aligned} \tag{4}$$

the SR-graph is shown in Fig. 2.

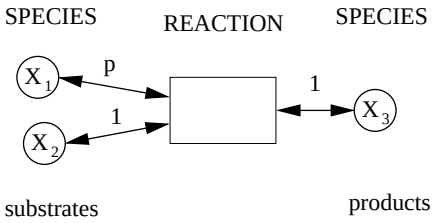

Figure 2: Species-Reactions graph with stoichiometric coefficients.

The reactions happen in a "compartment" (which for us could be anything from a tank reactor to an *in vivo* organism). The compartment is "open" when there is inflow/outflow of a specie and closed otherwise. A system can be open w.r.t. one specie and closed w.r.t. other species. Putting together multiple reactions we get a *biochemical reaction network*.

Example The "network" of reactions

$$\begin{array}{rcl} X_1 \; + \; X_2 & \underset{k_2}{\overset{k_1}{\rightleftarrows}} & X_3 \\ X_1 & \underset{k_4}{\overset{k_3}{\rightleftarrows}} & \emptyset \end{array} \tag{5}$$

is similar to (6) and in addition the system is open w.r.t. x_1 (x_1 is produced and degraded), but not w.r.t. x_2 and x_3. The ODEs become (compare with (6)):

$$\begin{aligned} \frac{dx_1}{dt} &= -k_1\, x_1 x_2 + k_2\, x_3 - k_3\, x_1 + k_4 \\ \frac{dx_2}{dt} &= -k_1\, x_1 x_2 + k_2\, x_3 \\ \frac{dx_3}{dt} &= k_1\, x_1 x_2 - k_2\, x_3 \end{aligned} \tag{6}$$

Notice in (6) that the inflow is a constant (independent of the concentration of x_1) while the outflow is a first order degradation term.

4.2 Representing biochemical networks through their stoichiometry

Consider a biochemical network involving n molecular species through r reactions. Call

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n_+$$

the vector of concentrations of the molecular species. Then $x \in \mathbb{R}^n_+$ because concentrations cannot be negative. Assume that for the r reactions the substrates and products of each reaction (with their stoichiometric coefficients) are known. These data can be though of as the entries of a *stoichiometric matrix*

$$S \in \mathbb{R}^{n \times r}.$$

Each row of S corresponds to a molecular species and each column corresponds to a reaction. If we look at S column-wise, there is a minus sign in front of the stoichiometric coefficients of the substrates (species forming the substrates of a reaction are depleted by the reaction, hence the reaction must contribute a negative term in the ODEs). The products instead have a plus sign. Row-wise, instead, the nonzero entries of the i-th row of S correspond to the reactions in which x_i is involved either as substrate (with a $-$ sign) or as a product (with a $+$ sign). For example for (3)-(4)

$$S = \begin{bmatrix} -p & p \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$$

while for (5)-(6):

$$S = \begin{bmatrix} -1 & 1 & -1 & 1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}. \tag{7}$$

Reversible reactions means linearly dependent columns in S (equal up to the sign).

To each reaction we can associate a *flux* $v_i(x,k)$, i.e., a compact expression of the kinetics action (for us mass-action) involving the species "upstream" of the reaction (and therefore which enter into the corresponding term of the ODE). Denote

$$v(x,k) = \begin{bmatrix} v_1(x,k) \\ \vdots \\ v_r(x,k) \end{bmatrix}$$

the vector of such reaction fluxes. k is a vector of parameters, representing the reaction rate constants. There is normally one rate constant per reaction, and these are nonnegative:

$$k = \begin{bmatrix} k_1 \\ \vdots \\ k_r \end{bmatrix} \in \mathbb{R}^r_+.$$

Then the ODEs for the biochemical network can be compactly expressed as

$$\frac{dx}{dt} = S\, v(x,k) \tag{8}$$

i.e., as a system of polynomial ODEs. For example, for (3)-(4), $v_1(x,k) = k_1\, x_1^p x_2$ and $v_2(x,k) = k_2\, x_3$ so that (4) can be written as

$$\frac{dx}{dt} = \begin{bmatrix} -p & p \\ -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} k_1\, x_1^p x_2 \\ k_2\, x_3 \end{bmatrix}$$

Notice that in (8) S contains the whole information about the topology of the network. Under the mass-action assumption, an arbitrarily complex network of biochemical reactions can be expressed in this way once S is given.

If in S reversible reactions are all broken down into irreversible forward and backward reactions as e.g. in (7) then $v(x,k) \geq 0$.

To unveil completely the structure of the system of ODEs (8), it is convenient to introduce a further vector $z(x)$ of "complexes", intending with that all compounds of species appearing upstream and downstream of an arrow in a reaction diagram, represented as mass-action terms. For example, in (3) the complexes are $pX_1 + X_2$ and X_3 and

$$z(x) = \begin{bmatrix} x_1^p x_2 \\ x_3 \end{bmatrix}.$$

In other words, $z(x)$ contains all basic multinomial terms appearing in $v(x,k)$, plus the "zero complex" (represented by a 1) to capture the inflow-outflow from the compartment. Assume m is the dimension of $z(x)$, $m \leq 2r$ (if all reversible reactions are split into pairs of irreversible reactions and r counts them twice, then $m \leq r$). Each reaction flux of $v(x,k)$ is obtained multiplying one of the multinomials by the corresponding reaction rate in k: $v_i(x,k) = k_i z_j(x)$. To select from the vector $z(x)$ the term $z_j(x)$ entering into $v_i(x,k)$, we need an "index matrix" $\mathcal{I} : \mathbb{R}^m \to \mathbb{R}^r$ to map the complexes $z(x)$ into the fluxes $v(x,k)$. The entire system of ODEs is then given by the composite map

species		complexes		indexed compexes		fluxes		ODEs
$\mathbb{R}^n$	$\to$	$\mathbb{R}^m$	$\to$	$\mathbb{R}^r$	$\to$	$\mathbb{R}^r$	$\to$	$\mathbb{R}^n$
x	$\mapsto$	$z(x)$	$\mapsto$	$\mathcal{I}\, z(x)$	$\mapsto$	$v(x,k) = \mathrm{diag}(k)\, \mathcal{I}\, z(x)$	$\mapsto$	$\dot{x} = S\, \mathrm{diag}(k)\, \mathcal{I}\, z(x)$

The only nonlinear step is the first, all others are linear maps.

Example Let us look at the example (5)-(6). In this case $n = 3$ and $r = 4$. The complexes are the following 4 compounds appearing in the reaction diagram (5): $\{X_1 + X_2,\, X_1,\, X_3,\, \emptyset\}$. Written in mass-action form, then,

$$z(x) = \begin{bmatrix} x_1 x_2 \\ x_1 \\ x_3 \\ 1 \end{bmatrix}$$

are all multinomial terms in the ODEs (6). If we list the 4 reactions according to the indices of the rate constants k_i given in (5), then the stoichiometric matrix is given in (7), and the index matrix is

$$\mathcal{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Computing explicitly $\dot{x} = S\, \mathrm{diag}(k)\, \mathcal{I}\, z(x)$ leads to (6). In particular the stoichiometric map is a linear map $S : \mathbb{R}^r \to \mathbb{R}^n$ and contains all the information on the topology of the system and on its dynamics.

4.3 Dynamical properties of reaction networks

Consider the biochemical reaction network

$$\frac{dx}{dt} = S\, v(x,k) = S\, \mathrm{diag}(k)\, \mathcal{I}\, z(x) \tag{9}$$

Notice that all parameters of the models (i.e., k) are concentrated in a single step of the cascade. Furthermore, since $k_i > 0$ and $\mathrm{diag}(k)$ is an invertible matrix, this step does not change any of the "structural" properties of the system. We want to study the dynamical behavior of the system *independently from the numerical value of these parameters.*

4.3.1 Invariance in $\mathbb{R}^n_+$

The vector x represents concentrations of molecular species and as such it must be and remain nonnegative.

Theorem 4.1 *If $x(0) \in \mathbb{R}^n_+$ then the solution of (9) is such that $x(t) \in \mathbb{R}^n_+$ $\forall\, t \geq 0$ and $\forall\, k \geq 0$.*

To "prove" this theorem, it is enough to observe that all negative terms in the ODEs for x_i vanish when x_i crosses the zero axis (i.e., the negative terms of the ODEs are homogeneous in x) hence $\dot{x}_i$ can never become negative if $x(0) \geq 0$. The positive terms are not required to be homogeneous in x_i. For example the inflow terms by definition are positive constants.

4.3.2 Conservation laws and left kernel of S

Consider the example (6). The ODEs are clearly redundant and we have already seen that there exist conservation laws. If we sum the first and third, or first and second equations

$$\begin{array}{lll} \dfrac{d(x_1+x_3)}{dt} = 0 & \Longrightarrow & x_1(t)+x_3(t) = \text{const} \quad \forall\, t \geq 0 \\ \dfrac{d(x_2+x_3)}{dt} = 0 & \Longrightarrow & x_2(t)+x_3(t) = \text{const} \quad \forall\, t \geq 0 \end{array} \tag{10}$$

which implies that $x_1(t)+x_3(t)$ and $x_2(t)+x_3(t)$ are constants of motion of the dynamical system (6). These constants express conservation of the total amount of a specie: x_1, by itself or bound with x_2 (in the form of the complex x_3), is conserved throughout the evolution, and similarly for x_2. Calling ξ_1 and ξ_2 the two constants in (10), by writing $x_1 = \xi_1 - x_3$ and $x_2 = \xi_2 - x_3$, the system (6) can be reduced to

$$\begin{array}{rcl} \dfrac{dx_3}{dt} & = & k_1\,(\xi_1 - x_3)(\xi_2 - x_3) - k_2\,x_3 \\ x_1 & = & \xi_1 - x_3 \\ x_2 & = & \xi_2 - x_3 \end{array}$$

i.e., each conservation law allows to replace an ODE with an algebraic equation (to be solved off-line). Notice that assigning the initial condition x_o to the system ξ_1 and ξ_2 are uniquely identified. Changing the initial conditions also the ξ_i change.

Let us look at conservation laws for the general formulation (8). the left null space of S, $\ker(S^T) = \{c \in \mathbb{R}^n \text{ s. t. } S^T c = 0\}$ is a vector subspace representing all conservation laws of the biochemical network. Assume $\text{rank}(S) = q \leq \min(n, r)$. Then $\dim(\ker(S^T)) = n - q$, i.e., the system (8) has $n - q$ constants of motion. If $c_1, \ldots, c_{n-q}$ are vectors forming a basis of $\ker(S^T)$, then $N_\ell = \begin{bmatrix} c_1^T \\ \vdots \\ c_{n-q}^T \end{bmatrix}$ is such that $N_\ell S = 0$. But then $N_\ell \dot{x} = 0$ and therefore, integrating, $N_\ell x(t) = \text{const} = \xi \in \mathbb{R}^{n-q}$ is a systematic expression of the constants of motion of the system. This can be used to reduce the dimension of (8) to q ODEs and $n - q$ algebraic equations of $x(t)$. In fact, if $N_\ell = \begin{bmatrix} N_{\ell,1} & N_{\ell,2} \end{bmatrix}$ with $(N_{\ell,2}) \in \mathbb{R}^{n-q,n-q}$ invertible, from the block splitting $\begin{bmatrix} N_{\ell,1} & N_{\ell,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \xi$ one has $x_2 = N_{\ell,2}^{-1}(\xi - N_{\ell,1}x_1)$, and the ODEs for x_2 can be dropped. Alternatively, one can say that the presence of conservation laws foliates $\mathbb{R}^n$ into *stoichiometric classes* invariant for the dynamics. Each stoichiometric class is uniquely identified by the initial condition x_o and is expressed as the affine space

$$\mathcal{SC}(x_o) = \{x_o + \text{Im}(S)\} \cap \mathbb{R}^n_+$$

Given x_o, the evolution of (8) is necessarily living in $\mathcal{SC}(x_o)$ for all times: $x(t) \in \mathcal{SC}(x_o)\ \forall\, t \geq 0$.

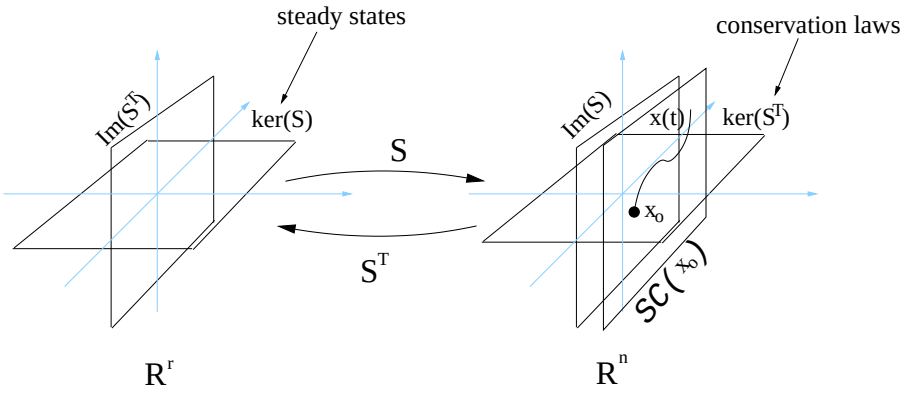

Figure 3: Stoichiometric map, with its subspaces.

4.3.3 Steady states and right kernel of S

In terms of the fluxes $v(x,k) \in \mathbb{R}^r_+$, the steady states of the system lie in the vector space $\ker(S) = \{v \in \mathbb{R}^r \text{ s. t. } Sv = 0\}$. From $\text{rank}(S) = q$, $\dim(\ker(S)) = r - q$. Since we have broken reversible reactions into pairs of irreversible reactions, we have imposed $v(x,k) \geq 0$ and our S has "twin" columns (differing only in sign, as in (7)) in correspondence of each pair of arrows "$\overrightarrow{\leftarrow}$". If all reactions are reversible, it is not complicated to show that this representation can be translated instead into an S with half columns and $v(x,k)$ that can assume any sign (i.e., the reversibility is "swapped" from the stoichiometry to the fluxes). If not all reactions are reversible, however, $\ker(S)$ should be restricted accordingly, leading in general to a convex cone in the space of fluxes. More details on this in next Chapter.

4.3.4 Equilibria and stability

We are interested in investigating the existence of (positive) equilibria, counting them, and understanding their stability properties. In particular, we want to study these properties independently of the numerical values of the parameters k. To analyze these properties, we need to introduce another class of graphs associated to the reaction network (9): the *C-graph* (Complex graph), whose nodes are the complexes and whose edges are the reactions. We must assume that the C-graphs are in "normal form" i.e., each complex label appears only once. Compare (a) and (b) of Fig. 4.

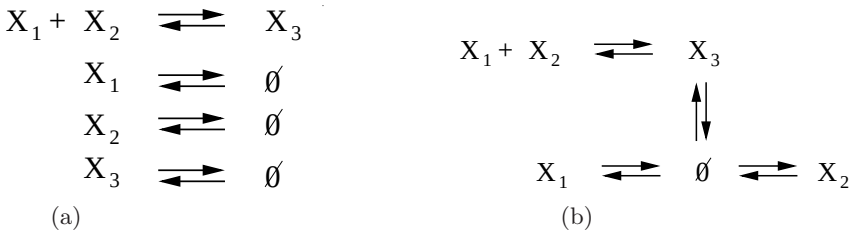

Figure 4: C graph. (a): not in normal form; (b): in normal form.

A reaction network is said weakly reversible $\exists$ a directed path of reactions connecting any two nodes of the C-graph (i.e., if the C-graph is strongly connected). For example the C-graph of Fig. 4 (b) is strongly connected, while that of (11) is not, hence the reaction network is not weakly reversible. Call ℓ (= linkage classes) the number of connected components of the C-graph.

If $q = \text{rank}(S)$, then $q = \dim(\text{Im}(S))$, i.e., the dimension of the so-called stoichiometric space $\text{Im}(S)$.

We will need an integer index, associated to the structure of a reaction network, called the *deficiency index* δ:

$$\delta := m - \ell - q$$

(recall that m = number of complexes).

Examples

- In the example of Fig. 4

$$S = \begin{bmatrix} -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

  and hence $q = 3$, $\ell = 1$ and $m = 5$, hence $\delta = m - \ell - q = 1$.

- In the example of (5)-(6), instead, $m = 4$, $\ell = 2$ and $q = 2$, hence $\delta = m - \ell - q = 0$.

The case $\delta = 0$ (*zero deficiency*) is special, because very sharp stability results are available for it. For zero deficiency networks the following theorem in fact holds.

Theorem 4.2 *For any reaction network of zero deficiency we have:*

1. *if the network is* not *weakly reversible then for arbitrary kinetics (i.e., mass-action, Michaelis-Menten, etc.) the system cannot have an equilibrium point in* $\text{int}(R^n_+)$ *and cannot have sustained oscillations.*

2. *if the network is weakly reversible, then for mass-action kinetics the network has a single positive equilibrium point* x^* *in each stoichiometric class* $\mathcal{SC}(x_o)$ *and* x^* *is "globally" asymptotically stable in* $\mathcal{SC}(x_o)$.

Meaning of 1.: lack of weak reversibility implies that one or more species will disappear asymptotically, hence x^* cannot be positive (i.e., $x^* \notin \text{int}(\mathbb{R}^n_+)$), but it must touch one or more of the axes of $\mathbb{R}^n_+$ (remember that by construction, $x(t)$ nonnegative $\forall\, t \geq 0$).

Meaning of 2.: in each leaf $\mathcal{SC}(x_o)$ in which $\mathbb{R}^n_+$ is foliated, the system has a single equilibrium point in $\text{int}(\mathbb{R}^n_+) \cap \mathcal{SC}(x_o)$ and within $\mathcal{SC}(x_o)$ this equilibrium point is globally asymptotically stable. Notice that since there is a continuum of stoichiometric classes (as we change x_o) there is also a continuum of equilibria, hence, as soon as conservation laws are present, we loose the usual notion of asymptotic stability, because every neighborhood of x^* contains infinitely many other equilibrium points. This notion is sometimes called "semistability", for example in the paper by Chellaboina et. al. mentioned at the begin. (Question: do you see any similarity with the "consensus" problem nowadays very popular??). Only when there are no conservation laws we have the "usual" asymptotic stability concept ($\mathbb{R}^n_+$ lies all in one stoichiometric class in this case). A simple way to avoid conservation laws is to have inflow/outflow for all reactions.

Example: enzyme-catalyzed reaction Consider the single substrate - single product enzyme-catalyzed reaction we saw in the first lecture, shown again in Fig. 5, whose reaction diagram is

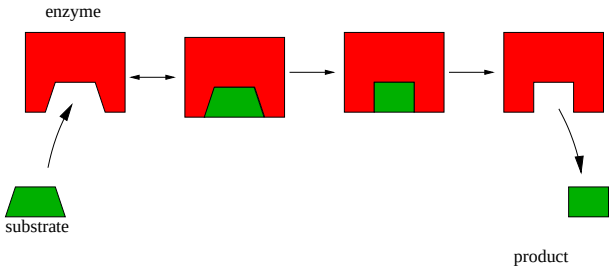

Figure 5: Sketch of an enzyme-catalyzed reaction

$$X_1 + X_2 \underset{k_2}{\overset{k_1}{\rightleftarrows}} X_3 \xrightarrow{k_3} X_2 + X_4 \tag{11}$$

The mass-action ODEs without any approximation are

$$\begin{aligned}
\frac{dx_1}{dt} &= -k_1 x_1 x_2 + k_2 x_3 \\
\frac{dx_2}{dt} &= -k_1 x_1 x_2 + (k_2 + k_3) x_3 \\
\frac{dx_3}{dt} &= k_1 x_1 x_2 - (k_2 + k_3) x_3 \\
\frac{dx_4}{dt} &= k_3 x_3
\end{aligned}$$

The stoichiometric matrix in this case is

$$S = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

Computing the deficiency index, $q = 2$, $\ell = 2$ while $m = 3$. Hence $\delta = m - \ell - q = 0$. However, the network is not weakly reversible, hence Theorem 4.2 predicts no (strictly) positive equilibrium. Let us compute the equilibria explicitly. From $\dot{x}_4 = 0 \Longrightarrow x_3 = 0$. Consequently, from $\dot{x}_1 = 0 \Longrightarrow x_1 x_2 = 0$. Hence $x^* \notin \text{int}(\mathbb{R}^4_+)$. The meaning is the following: since the substrate x_1 is transformed into product x_4 and not resupplied, $x_1 \to 0$ and consequently also $x_2 \to 0$ as $t \to \infty$. This is the concrete meaning of the reaction network not being weakly reversible and hence, from Theorem 4.2, not admitting a positive equilibrium. Notice that there are two conservation laws in the system: $x_2 + x_3 = \xi_1$ and $x_1 + x_3 + x_4 = \xi_2$.

Example (5)-(6) The network is reversible, $\delta = 0$, hence Theorem 4.2 applies and predicts that, in each stoichiometric class, the system has a unique positive equilibrium point which is asymptotically stable for all points in $\mathcal{SC}(x_o)$. Let us compute explicitly the equilibrium/a. From (6)

$$\frac{dx_3}{dt} = 0 \quad \Longrightarrow \quad x_3 = \frac{k_1}{k_2} x_1 x_2$$

Plugging into $\frac{dx_1}{dt} = 0$, we get $x_1 = \frac{k_4}{k_3}$, hence

$$x_3 = \frac{k_1 k_4}{k_2 k_3} x_2 \tag{13}$$

Eq. (13) apparently says that there is an entire ray of equilibria in the (x_2, x_3) plane, see Fig. 6. However, $q = \text{rank}(S) = 2 \Longrightarrow n - q = 3 - 2 = 1 \Longrightarrow \exists$ a conservation law. $\ker(S^T)$ is generated for example by

$$c = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

meaning that the constant of motion is determined by $c^T x = x_2 + x_3 = \xi$ or

$$x_3 = \xi - x_2 \tag{14}$$

In the plane (x_2, x_3) this constant of motion intersects (13) in a single point, see Fig. 6, meaning that on $\mathcal{SC}(x_o)$ the system has indeed a unique equilibrium point. Changing x_o means changing the

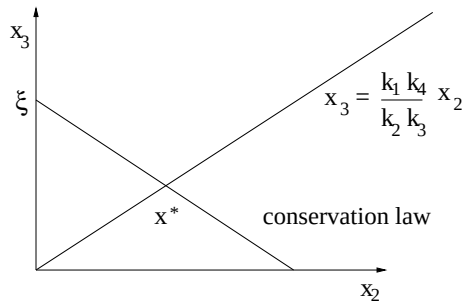

Figure 6: Steady states and conservation laws: 2D-slice of the phase plane for example (5)-(6)

value of the constant ξ, hence "sliding" the constraint (14) (i.e., passing to another stoichiometric class).

For networks of higher deficiency ($\delta > 0$) other conditions exist, although they are mostly focused on studying the "capacity for multistationarity" i.e., the possibility that for some choice of the parameters k the system may exhibit multiple equilibria in $\mathbb{R}^n_+$. As they are usually formulated as necessary but not sufficient conditions for multistationarity, they are not directly constructive, although algorithms exist to test them.