# CHEMICAL REACTION NETWORKS:
# COMPUTATION OF LINEAR FIRST INTEGRALS AND IDENTIFICATION OF THE SET OF DEPENDENT VARIABLES

**TECHNICAL INFORMATION**

**Quentin Vanhaelen** (vanhaelen@easyomics.com)

**Abstract**

*This document is associated with the code SIERRA-II designed to obtain the set of linear first integrals of MAL-CRNs, using the pivoting algorithm to compute thew reduced row echelon form allowing the identification of dependent and independent variables associated to each first integrals. The implementation also computes several matrices associated with control coefficients defined within the MCA framework. This short text describes the format and content of the different input files required to run properly the R code required to generate the source files of SIERRA-II. It provides some guidelines about how to edit these files properly. The implementation of this classic algorithm is suitable for large MAL-CRNs.*

## 1 Mathematical Background

We begin with a overview of relevant notations. The empty set is represented by the symbol $\varnothing$, the set of real numbers is noted $\mathbb{R}$, the set of integers is $\mathbb{Z}$ and the set of positive integer is $\mathbb{N}$. We represent subsets as follows. $\mathbb{R}_{<0}$ represents the set of negative real numbers, $\mathbb{R}_{\geq 0}$ represents the set of nonnegative real numbers and $\mathbb{R}_{>0}$ is the set of positive real numbers. The notation $\vec{v} \in \mathbb{R}^n_{\geq 0}$ defines a vector of $n$ components with nonnegative real values: $v_i \geq 0$, $i \in \{1, ..., n\}$. The notation $\vec{u} \in \mathbb{R}^n_{>0}$ defines a positive real vector: $u_i > 0$, $i \in \{1, .., n\}$. A $n \times m$ ($n$ rows and $m$ columns) matrix with nonnegative real entries is noted $\mathbf{A} \in \mathbb{R}^{n \times m}_{\geq 0}$. The transpose of the matrix $\mathbf{D} \in \mathbb{Z}^{n \times m}$ is $\mathbf{D}^{\mathrm{T}} \in \mathbb{Z}^{m \times n}$. The kernel of a matrix $\mathbf{A}$ is noted $\mathrm{Ker}\,\mathbf{A}$ while we use $\mathrm{Rank}\,\mathbf{A}$ and $\mathrm{Im}\,\mathbf{A}$ as notations for the rank of $\mathbf{A}$ and the image of $\mathbf{A}$ respectively. The scalar product of two vectors, $\vec{v}, \vec{w} \in \mathbb{R}^n$ for instance, is $\vec{v} \cdot \vec{w} = \sum_{k=1}^{n} v_k w_k$. The product of a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ with a vector $\vec{v} \in \mathbb{R}^m$ is noted $\mathbf{A} \cdot \vec{v}$ and each element of the resulting vector is given by $(\mathbf{A} \cdot \vec{v})_i = \sum_{k=1}^{m} A_{ik} v_k$. The product of two matrices $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{m \times l}$ is noted $\mathbf{A} \cdot \mathbf{B} = \mathbf{C} \in \mathbb{R}^{n \times l}$.

**Definition 1** A chemical reaction network (CRN) is a triplet $(\mathcal{S}, \mathcal{R}, \mathcal{C})$:

- $\mathcal{S} = (S_1, ..., S_n)$ has $n$ elements, the species. A species is a chemical substance that is involved in a reaction, either as a reactant or a product, or both.
- $\mathcal{R} = (R_1, ..., R_r)$ has $r$ elements, the reactions. The set of reactions can be represented as follows:

$$\sum_{i=1}^{n} r_{ij} S_i \rightarrow \sum_{i=1}^{n} p_{ij} S_i \quad \text{with} \quad i = 1, ..., n \quad \text{and} \quad j = 1, ..., r$$

  where the $p_{ij}$ term is the stoichiometric coefficient for the $i^{\text{th}}$ species in the $j^{\text{th}}$ reaction if appearing on the right of the reaction arrows, $r_{ij}$ if on the left. $r_{ij}$ and $p_{ij}$ are called the stoichiometric coefficients of the forward and backward reactions respectively. Using these notations, we can write a single reversible reaction as two irreversible reactions.
- $\mathcal{C} = (C_1, C_2, ..., C_l)$ has $l$ elements, the complexes. Each complex is a linear combination of species, $C_j = \alpha_{1j} S_1 + ... + \alpha_{nj} S_n \in \mathcal{C}$ where $\alpha_{ij}$ is the molecularity of species $i$ in complex $j$. Each reaction can be seen as an ordered pair of distinct complexes: $C_j \rightarrow C_i$. The complexes may be combined in the complex matrix: $\mathbf{Y} = [\alpha_{ij}] \in \mathbb{Z}^{n \times l}_{\geq 0}$.

The zero complex is a special complex with zero stoichiometric coefficients. It appears in the reaction schemes where one considers either an external source or sink of some components, i.e. , chemical reactions of the form

$$\emptyset \to \sum_{i=1}^{n} p_{ij} S_i \quad \text{or} \quad \sum_{i=1}^{n} r_{ij} S_i \to \emptyset \quad i = 1, ..., n \quad \text{and} \quad j = 1, ..., r$$

In both cases the system is thermodynamically not closed, therefore the conservation of the total mass in the system generally does not hold.

**Definition 2** One defines the stoichiometric matrix $\mathbf{N} \in \mathbb{Z}^{n \times r}$ with $\mathrm{N}_{ij} = p_{ij} - r_{ij}$. The $j^{\text{th}}$ column of the matrix $\mathbf{N}$ is the reaction vector associated with the $j^{\text{th}}$ reaction.

The vectors $p_{ij} - r_{ij}$ are the stoichiometric coefficients of the forward reactions subtracted from the stoichiometric coefficients of the backward reactions and are called the reaction vectors. They show the essential information of a reaction: the net amount of each participating species produced or used.

**Definition 3** The linear span of the reaction vectors defines the subspace of reaction space called the stoichiometric subspace. The stoichiometric subspace is the vector subspace $\mathbb{S} \subset \mathbb{R}^n$ defined as $\mathbb{S} = \operatorname{Im} \mathbf{N} = \operatorname{span}\{p_{i1} - r_{i1}, ..., p_{ir} - r_{ir}\}$. The rank of the CRN, $n_0$, is defined as $n_0 = \dim(\mathbb{S}) = \operatorname{Rank} \mathbf{N}$.

**Definition 4** Denote by $x_i$ the concentration of the species $S_i$ and by $\vec{x}(t) = (x_1(t), ..., x_n(t))$ the vector of species concentrations at time $t$.

- A rate function for a reaction $j = 1, ..., r$ is a function $v_j : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ that describes the instantaneous change in the species composition $\vec{x} = (x_1, ..., x_n)$ due to this reaction.
- A kinetics $\mathcal{V}$ is an assignment to each reaction $j = 1, ..., r$ of a rate function $v_j : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$. $\mathcal{V}$ is a set of rate functions indexed by the elements of the reaction set: $\mathcal{V} = \{v_j : j = 1, ..., r\}$. We define $\vec{v} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^r_{\geq 0} = (v_1, ..., v_r)$ as the vector of reaction rate functions (also called the reaction rate vector).

Furthermore, we assume that each component of the reaction rate vector satisfies the following monotonicity conditions:

$$\frac{\partial v_j}{\partial x_i} = \begin{cases} \geq 0 & \text{if } r_{ij} > 0 \\ = 0 & \text{if } r_{ij} = 0 \end{cases}$$

We also assume that, whenever the concentration of any of the reactants of a given reaction is 0, then, the corresponding reaction does not take place, meaning that the reaction rate is 0. These conditions can be summarized to give the definition of a well-formed reaction:

A reaction $j$ characterized by the $j^{\text{th}}$ component of reaction rate vector $v_j(\vec{x})$ is a well-formed reaction if the three following conditions hold:

1. $v_j(\vec{x})$ is a nonnegative partially differentiable function,
2. $\frac{\partial v_j}{\partial x_i} > 0$ for some value of the concentration $x_i \in \mathbb{R}_0$ of the species $S_i$ if and only if the species $S_i$ is a reactant of the reaction $R_j$,
3. $v_j(\vec{x}) = 0$ if there exists a reactant $S_i$ of the reaction $R_j$ such that its concentration $x_i = 0$.

**Definition 5** We define the function $\vec{f}(\vec{x})$, called the species formation vector, such that $f_i(\vec{x}) = \sum_{j=1}^{r} (p_{ij} - r_{ij})\, v_j$. Each $f_i(\vec{x})$ represents the contribution of the reactions acting on the species $S_i$. In vector notations $\vec{f}(\vec{x}) = \mathbf{N} \cdot \vec{v}(\vec{x})$.

**Definition 6** Let $(\mathcal{S}, \mathcal{R}, \mathcal{C})$ be a CRN. The evolution of the concentrations $\vec{x}(t) = (x_1(t), ..., x_n(t)) \in \mathbb{R}^n_{\geq 0}$ under the kinetics $\mathcal{V}$ is given by:

$$\frac{d\vec{x}}{dt} = \dot{\vec{x}} = \vec{f}(\vec{x}) = \mathbf{N} \cdot \vec{v}(\vec{x}) \quad \text{with} \quad \vec{x}(0) = \vec{x}_0 \in \mathbb{R}^n_{\geq 0} \tag{1}$$

Nonnegativity of the solutions follows from the assumption that the rate functions are non-negative.

**Definition 7** An autonomous system of ODEs $\frac{d\vec{x}}{dt} = \vec{f}(\vec{x})$ whose right hand side can be written in the form $\vec{f}(\vec{x}) = \mathbf{N} \cdot \vec{v}$ with a constant matrix $\mathbf{N} \in \mathbb{Z}^{n \times r}$ and the vector valued function $\vec{v} \in \mathbb{R}^n \to \mathbb{R}^r$ is called **pseudolinear**.

Obviously, any polynomial nonlinear system can be written in such a form, if we take as $\vec{v}$ the vector of all terms appearing on the right-hand side of the system.

We say that the quadruple $(\mathcal{S}, \mathcal{R}, \mathcal{C}, \mathcal{V})$ defines a chemical kinetic system (CKS). The system of ODEs or dynamical system $\dot{\vec{x}} = \vec{f}(\vec{x})$ is the dynamical system of the CKS.

**Definition 8** A kinetics $\mathcal{V}$ is called mass-action if the reaction rate functions introduced in Definition 5 take the form

$$v_j(k_j, \vec{x}) = k_j \phi_j(\vec{x}) \quad \text{where} \quad \phi_j(\vec{x}) = \prod_{i=1}^{n} x_i^{r_{ij}} \quad \text{with} \quad j = 1, ..., r$$

where $k_j$ is the rate constant (also referred to as kinetic parameter) of the $j^{\text{th}}$ reaction. The vector of rate constants is noted $\vec{k}$ with $k_j \in ]0, +\infty[$ for $j = 1, ..., r$. We do not attribute numerical values to the kinetic parameters. Throughout this work, kinetic parameters are considered as symbols that take positive real values (the possibility for kinetic parameters to take infinite values being unphysical and therefore excluded).

**Definition 9** A CRN defined by the triplet $(\mathcal{S}, \mathcal{R}, \mathcal{C})$ associated to a mass-action kinetics $\mathcal{V}$ is called a mass-action law chemical reaction network (MAL-CRN). The set of ODEs associated with a MAL-CRN is a sparse polynomial differential system.

# 2 Computation of first linear integrals and control coefficients

## 2.1 Deficiency in row rank and linear first integrals

MAL-CRNs often have a highly redundant stoichiometry which translates into the existence of linear dependencies between the species concentrations in the case of row deficiency and between reaction rates in the case of column deficiency. These linear dependencies are structural in the sense that they are a consequence of the structure of the stoichiometric matrix $\mathbf{N}$ only.

Linearly dependent rows within the stoichiometry matrix correspond to linear conservation relations of the system, also called linear first integrals. Each redundant row identifies a chemical species whose dynamics is completely determined by the behavior of other species in the system through a linear conservation relation. Defining $\operatorname{Ker} \mathbf{N}^{\mathrm{T}} = \{\tilde{\vec{u}} \in \mathbb{Z}^n \mid \mathbf{N}^{\mathrm{T}} \cdot \tilde{\vec{u}} = 0\}$, the linear conservation relations are $I_i := \tilde{\vec{u}}_i \cdot \vec{x} = c_i$, $(i = 1, ..., n_C)$, with $c_i$ being the components of the total-constant vector $\vec{c}$ and $n_C = \dim(\operatorname{Ker} \mathbf{N}^{\mathrm{T}}) = n - \operatorname{Rank} \mathbf{N}$.

The system $\dot{\vec{x}} = \vec{f}(\vec{x}) = \mathbf{N} \cdot \vec{v}(\vec{x})$ has flow invariant subspaces of the form $(\vec{x}_0 + \mathbb{S}) \cap \mathbb{R}^n_{\geq 0} = (\vec{x}_0 + \operatorname{Im} \mathbf{N}) \cap \mathbb{R}^n_{\geq 0}$. The affine linear spaces $\vec{x}_0 + \operatorname{Im} \mathbf{N}$ are described implicitly by the linear restrictions $I_i := \tilde{\vec{u}}_i \cdot \vec{x} = c_i$, $(i = 1, ..., n_C)$ where the $\vec{u}_i^{\mathrm{T}}$ form an orthonormal basis of the orthogonal complement of $\operatorname{Im} \mathbf{N}$. The choice $c_i = \tilde{\vec{u}}_i \cdot \vec{x}_0$ gives two equivalent descriptions $(\vec{x}_0 + \operatorname{Im} \mathbf{N}) \cap \mathbb{R}^n_{\geq 0} = \{\vec{x} \in \mathbb{R}^n \mid \tilde{\vec{u}}_i \cdot \vec{x} = c_i,\ i = 1, ..., n - \operatorname{Rank} \mathbf{N}\}$.

The rank, $n_0$, of the matrix $\mathbf{N}$ is the number of independent variables of the system. The evolution of the remaining $n - n_0$ dependent variables is a function of a set of the $n_0$ independent variables. It follows that the $n$-dimensional state enjoys only $n_0$ degrees of freedom. Thus the original description in terms of $n$ state variables is non-minimal, regardless of the form of the reaction rates. We want to partition the species concentration vector $\vec{x}$ as $\vec{x} = (\vec{x}_i^{\mathrm{T}}, \vec{x}_d^{\mathrm{T}})^{\mathrm{T}}$ where $\vec{x}_i \in \mathbb{R}_{\geq 0}^{n_0}$ is the vector of independent species and $\vec{x}_d \in \mathbb{R}_{\geq 0}^{n-n_0}$ contains the dependent species. This is done using the Gauss-Jordan algorithm on the stoichiometric matrix $\mathbf{N}$ to transform it into reduced row echelon form to distinguish dependent rows from the independent ones. The matrix $\mathbf{N}$ is then re-ordered such that its first $n_0$ rows form a basis of the whole set of rows of the matrix $\mathbf{N}$. Let $\mathbf{N_R}$ denote the matrix composed of those $n_0$ rows, then $\mathbf{N}$ can be decomposed as the product:

$$\mathbf{N} = \mathbf{L} \cdot \mathbf{N_R} \tag{2}$$

With $\mathbf{N_R}$ containing the independent rows which constitute the basis. The matrix $\mathbf{L}$ is a $n \times n_0$ matrix with the following structure: $\begin{pmatrix} \mathbf{I}_{n_0 \times n_0} \\ \mathbf{L}_0 \end{pmatrix}$ Where $\mathbf{L}_0$ is a $(n - n_0) \times n_0$ matrix whose coefficients must be computed. We perform the same re-ordering with the concentration vector. Thus, the $n_0$ first rows contain the independent variables and the last $n - n_0$ rows contain the dependent ones. Using this decomposition and following the procedure described elsewhere[1], the resulting $(n - n_0)$ independent linear conservation relationships are described by the equality $\frac{d}{dt}(\vec{x}_d - \mathbf{L}_0 \cdot \vec{x}_i) = 0$. The dependent concentrations are then expressed in terms of linear combinations of the independent concentrations.

As an example, we consider the following set of reactions:

$$* \rightarrow X_1 \tag{3}$$
$$X_1 + X_6 \rightarrow X_2 \tag{4}$$
$$X_2 + 2X_5 \rightarrow X_3 \tag{5}$$
$$X_3 \rightarrow X_4 + X_5 \tag{6}$$
$$X_4 \rightarrow X_5 + X_6 + * \tag{7}$$

The matrix $\mathbf{N}$ has the following form: $\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & -2 & 1 & 1 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}$ The first step is the computation of the rank,$n_0$, of the matrix $\mathbf{N}$. Form a computational point of view, the ran may be obtained by computing the singular value of the SVD decomposition of the matrix: the number of non-zero singular values corresponds to the rank of the matrix. The rank of the matrix is also the number of independent variables contained in the system. It means that there are $n - n_0$ dependent variables whose evolution may be obtained from the knowledge of a set of the$n_0$ independent variables.

When is rank is known, one needs to extract a subset of $n_0$ rows which constitutes a basis of the whole set of rows of matrix $\mathbf{N}$[2]. As it is always possible to change the order of the rows, we shall assume that its first $n_0$ rows form the researched basis.

Let $\mathbf{N_R}$ denote the matrix composed of those $n_0$ rows with $\mathbf{N_R} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$ containing the independent rows which constitute the basis. The matrix $\mathbf{L}$ is a $n*n_0$ matrix with the following structure: $\begin{pmatrix} \mathbf{I}_{n_0 * n_0} \\ \mathbf{L}_0 \end{pmatrix}$

[1] C. Reder, "Metabolic Control Theory: A Structural Approach", J. theor. Biol. (1988) 135, 175-201

[2] For a basis: the rows are independent, which means that there is no null linear combination of these rows, except the trivial one, where all the coefficients are zero. Every row of $\mathbf{N}$ is a linear combination of the rows of this subset. It is possible to build the basis in two steps: first computing the LU decomposition of matrix $\mathbf{N}$, then multiply it by the left part of the decomposition to obtain the row echelon form. the number of non zero rows is the rank of $\mathbf{N}$, and the labels of non zero rows give the label of the linearly independent rows of $\mathbf{N}$.

with $\mathbf{L}_0$ a $(n - n_0) * n_0$ matrix whose coefficients must be computed. For our example, we have the following system of equations for the unknown coefficients[3]:

$$\begin{pmatrix} 0 & 0 & -2 & 1 & 1 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & L_{13} & L_{14} \\ L_{21} & L_{22} & L_{23} & L_{24} \end{pmatrix} * \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Solving this system for the coefficients $Lij$, we have $\mathbf{L}_0 = \begin{pmatrix} 0 & 0 & -2 & -1 \\ 0 & -1 & -1 & -1 \end{pmatrix}$

We perform the same decomposition with the concentration vector. Thus, the $n_0$ first rows contain the independent variables the last $n - n_0$ rows contain the dependent ones.

$$\mathbf{x} = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_{n_0} \\ ---- \\ X_{n_0+1} \\ \cdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mathbf{x_R} \\ \mathbf{x_R}' \end{pmatrix}$$

We use the fact that every structural conservation relationship is a linear combination of the $(n - n_0)$ independent conservation relationships described by the equality: $\frac{d}{dt}(\mathbf{x_R}' - \mathbf{L}_0\mathbf{x_R}) = 0$ and for our example, we have $\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} - \begin{pmatrix} 0 & 0 & -2 & -1 \\ 0 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$

With $A_1, A_2$ being constant. We end up with the following two conservation relationships:

$$X_5(t) + 2X_3(t) + X_4(t) = A_1 \tag{8}$$
$$X_6(t) + X_2(t) + X_3(t) + X_4(t) = A_2 \tag{9}$$

Using these relationships, it is possible to obtain the dependent concentrations $X_6$ and $X_5$ in terns of a linear combination of independent variables:

$$X_5(t) = A_1 - 2X_3(t) - X_4(t) \tag{10}$$
$$X_6(t) = A_2 - X_2(t) - X_3(t) - X_4(t) \tag{11}$$

The initial conditions may be used to obtain a value for the constants, more precisely:

$$A_1 = X_5(0) + 2X_3(0) + X_4(0) \tag{12}$$
$$A_2 = X_6(0) + X_2(0) + X_3(0) + X_4(0) \tag{13}$$

## 2.2 Control coefficients: Definition and characterization

To begin we introduce the matrix $\mathbf{D}_x\mathbf{v}$ which is a $(r * m)$ matrix such that

$$(D_x v)_{ij} = \frac{\partial v_i}{\partial X_j} \tag{14}$$

The purpose of this section is to introduce two sets of control coefficients:

[3]From a computational point of view, the resolution of this system may done numerically by inversion of the $\mathbf{N_R}$ matrix

**Simple steady state control matrix** it is defined as

$$\Gamma = -\mathbf{L}(\mathbf{N_R}\mathbf{D}_x\mathbf{vL})^{-1}\mathbf{N_R} \tag{15}$$

**Simple steady state flux control matrix** it is defined as

$$\mathcal{C} = \mathbf{I}_r - \mathbf{D}_x\mathbf{vL}(\mathbf{N_R}\mathbf{D}_x\mathbf{vL})^{-1}\mathbf{N_R} = \mathbf{I}_r + \mathbf{D}_x\mathbf{v}\Gamma \tag{16}$$

These coefficients satisfy the following set of relationships

**Simple summation relationships**

$$\Gamma\mathbf{K} = 0 \qquad \mathcal{C}\mathbf{K} = \mathbf{K} \tag{17}$$

**Simple connectivity relationships**

$$\Gamma(\mathbf{D}_x\mathbf{vL}) = -\mathbf{L} \qquad \mathcal{C}(\mathbf{D}_x\mathbf{vL}) = 0 \tag{18}$$

It is straightforward to check (using the definition) that these relationships are satisfied by default.

One can give the following interpretation of the coefficients of the matrices $\Gamma$ and $\mathcal{C}$: suppose that for every index $j$ between 1 and $r$ there exists a parameter $\lambda_j$ (such as kinetic parameters $k_j$) that acts specifically on the rate $v_j$:

$$\frac{\partial v_j}{\partial \lambda_j} \neq 0 \qquad \frac{\partial v_i}{\partial \lambda_j} = 0 \quad \text{if} \quad i \neq j \tag{19}$$

the coefficients at row $i$ and column $j$ of the matrices $\Gamma$ represent the control of the step $j$ on the concentration $i$ and flux ($J_i = v_i$) and they do not depend on the way the specific parameter $\lambda_j$ acts on the kinetic $v_j$.

It must be noticed that there are two approaches to compute the coefficients of the matrices $\Gamma$ and $\mathcal{C}$. The first one is to directly use the definitions 15 and 16. In our case we have the detailed expression for the fluxes, thus a expression of these coefficients in terms of the kinetic parameters can be obtained. The second approach is based on the use of the set of constraints (summation and connectivity relationships). This method is used elsewhere[4] and leads to the two matrices expressed in terms of three parameters. In this last case, no knowledge of the kinetic fluxes is assumed and the derivation is based on the stoichiometric matrix only. As a consequence, the second approach does not provide any supplementary information which could be used as a constraint to determine the value of the kinetic parameter.

As an example, we consider the following system:

$$* \rightarrow X_1 \tag{20}$$

$$X_1 \rightarrow * \tag{21}$$

$$X_1 \rightarrow * \tag{22}$$

The $(n = 1, r = 3)$ matrix $\mathbf{N}$ is $\begin{pmatrix} 1 & -1 & -1 \end{pmatrix}$ The rank of the matrix is 1 and it is maximum. As a result we have $\mathbf{N_R} = \mathbf{N}$ and the matrix $\mathbf{L} = \mathbf{I}_{1*1}$. We compute $\Gamma$ using its definition, we have

$\mathbf{N_R}\mathbf{D}_x\mathbf{v} = \begin{pmatrix} 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \frac{\partial x_1}{\partial v_1} \\ \frac{\partial x_1}{\partial v_2} \\ \frac{\partial x_1}{\partial v_3} \end{pmatrix} = -k_2 - k_3$ And $\mathbf{N_R}\mathbf{D}_x\mathbf{vL} = -k_2 - k_3$. The matrix has the following form: $\Gamma = \frac{1}{k_2+k_3}\begin{pmatrix} 1 & -1 & -1 \end{pmatrix}$ This result should be compared with the expression found in the original article: $\Gamma = \begin{pmatrix} \alpha & -\alpha & -\alpha \end{pmatrix}$ with $\alpha \neq 0$. It is straightforward to see that

$$\alpha = \frac{1}{k_2 + k_3} \tag{23}$$

[4] C. Reder, "Metabolic Control Theory: A Structural Approach", J. theor. Biol. (1988) 135, 175-201

Using the definition for the matrix $\mathcal{C}$ , we have $\mathcal{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 \\ k_2 \\ k_3 \end{pmatrix} \begin{pmatrix} \alpha & -\alpha & -\alpha \end{pmatrix}$ and finally: $\mathcal{C} = \begin{pmatrix} 1 & 0 & 0 \\ k_2\alpha & 1-k_2\alpha & -k_2\alpha \\ k_3\alpha & -k_3\alpha & 1-k_3\alpha \end{pmatrix}$

The expression obtained using the constraints is $\mathcal{C} = \begin{pmatrix} a+b & 1-(a+b) & 1-(a+b) \\ a & 1-a & -a \\ b & -b & 1-b \end{pmatrix}$

The following identifications can be made:

$$a = k_2\alpha = \frac{k_2}{k_2+k_3}$$
$$b = k_3\alpha = \frac{k_3}{k_2+k_3}$$

And we have $a + b = 1$.

# 3 Generation of SIERRA II

The complete set of source codes used to perform the simulations and analysis are of two types: there are the source codes which have been written once for all and do not require any modification and/ adaptation when switching from one model to another one. The other source codes are specific to the model being simulated sometimes because they contain specific information characterizing the dimension of the system such as number of species, initial conditions,etc. or also because some parts such as the designation of the output files must be specified. These source files are generated using **R-CODE-SIERRA-II-GENERATOR**. This R script use the following files as inputs:

1. **reactions-max.csv**
2. **repressor-list-reaction.csv**
3. **initial-number-molecules.csv**
4. **rate-constants.csv**
5. **repressor-list-rate-constant.csv**

The source files used to code the chemical equations for each model obey the following general organization. Firstly, all the files are saved in the CSV (DOS) format, secondly, we have 5 separated files. However, there is never any header of any kind at the top of the file.

The syntax of input files for the R script is as follows:

- **moleculesaddition.csv**: this file should edited if molecules such as external ligands must be added to the system after the beginning of the simulation. there are three columns: the first one is the label of the species to be added. The second one is the amount of molecules to be added and the last column is the time when the species must be added.
- **rateconstants.csv**: this file contains two columns: the number of rows is equal to the number of chemical reactions of the model. the label of each row corresponds to the label of the corresponding chemical reactions. In practice, the useful information (value of the rate constant) is in the second column

- **initialnumbermolecules.csv**: this file contains one column, the number of row is equal to the number of molecular species of the model. the label of each row corresponds to the label of each chemical species. The value is the initial concentration of the given species **(automatically generated )**

- **repressorlistreaction.csv**: This file is fill with zero by default, each column corresponds to a chemical reaction. The first row of each column corresponds to the number of repressors acting of the chemical reactions. the following rows contain the label of the species acting as a repressor on this reaction. In the current version this kind of modification is allowed only for the repression of the promoter.

- **repressorlistrateconstant.csv**: This file has exactly the format than the **repressorlistreaction.csv** file but in place of the label of the species acting as repressors the corresponding value for the strength of the kinetic parameter is written. This new file allows to take into account a specific constant of repression for each repressor.

- **reactionsmax.csv**: This is the mean file of the model. It contains 17 effective columns: here is a description of the columns used:

  1. *column* 1: each row contains the name of a chemical species
  2. *column* 2: each row contains the initial concentration for the species (must be either zero or non zero values by default.)
  3. *column* 3: each row contains the value of the kinetic rate of the corresponding reaction which is written on this line, see below.
  4. *column* 5: row 1 is the number of encoded chemical reactions, row 2 is the total number of chemical species and row 3 is the number of active species (in practice, total number and number of active species are set to the same value)
  5. *column* 6: it is used to generate the XML file: it contains the official name of the compartment where the species is most likely to be found. The convention used here are as follows:

     **Nucleus** all the species inside the nucleus,

     **Cytoplasm** all the species appearing inside the cytoplasm space,

     **Membrane** it is a pseudo compartment which contains the species having most of their activities along the membrane: receptors, G-proteins, etc.

     **default** this value hold for the external environment: it is assigned to the external ligands added to activate the receptor for example.
  6. *column* 7: set to zero by default, to 1,2,3 if corresponding reactions are connected such that kinetic parameters must be adapted to ensure dynamical agreement
  7. *column* $8-11$: these $4$ columns contains the coding of the chemical reactions
  8. *column* 12: it contains the category number of the chemical reactions.
  9. *column* 13: it is the code for the rescaling procedure (label 0 is only one reactant, label 1 when there are two reactants and label 2 when there is no reactant, i.e., creation of species)
  10. *column* 14:it is the value of the delay, it is set to zero by default
  11. *column* 15: label of the kinetic rate constant
  12. *column* 16:code for the kind of delayed reactions (label 1 is for consuming and label 2 is for non consuming reactions, **label 2 is in fact for transcription and translation reaction only**)
  13. *column* 17: general labeling of the chemical reactions

The file called **reaction-max.csv** contains the core of the kinetic model. To complete it in the correct format it is easier to use the LIF model as a template. As a guideline, the first column contains the list of species. Each of them is given a label which corresponds to the row where the species is listed in the column 1. The column 2 contains the corresponding amount for each species. The column 6 contains the location of the species within the system. The columns from 8 to 11 are for the chemical reactions. The columns 8 and 9 contains the reactants and the label of the species should always be preceded by a minus sign and the columns 10 and 11 contains the products of the reaction. If there is no product (for instance for degradation) columns 10 and 11 should remain empty. Similarly,

if there is no reactant (the case of synthesis), the columns 8 and 9 will remain empty. If there is only one reactant, it should be written in column 8 only and if there is only one product, it should be written on column 10 only. Note that due to the rules and restrictions applied when building the binary tree, some restrictions apply on the kind of reactions allowed. The list of authorized discussed further in the corresponding paper. In practice, the user should keep in mind that if a reaction has two reactants, then only one product is allowed, and if the reaction releases two products, only one reactant is allowed. The value of the kinetic constant of each reaction is written on the same row but in column 6. The column 14 contains the value of the delay if any. In general, the user should ensure the coherence between the units of the different parameters (concentration, kinetic rate units, delay, etc.)