Stoichiometric network analysis

In stoichiometric analysis of metabolic networks, one concerns the effect of the network structure on the behaviour and capabilities of metabolism.
Questions that can be tackled include:

- Discovery of pathways that carry a distinct biological function (e.g. glycolysis) from the network, discovery of dead ends and futile cycles, dependent subsets of enzymes
- Identification of optimal and suboptimal operating conditions for an organism
- Analysis of network flexibility and robustness, e.g. under gene knockouts

Stoichiometric coefficients

Soitchiometric coefficients denote the proportion of substrate and product molecules involved in a reaction. For example, for a reaction

$$r : A + B \mapsto 2C,$$

the *stoichiometric coefficients* for A, B and C are $-1, -1$ and 2, respectively.

- Assignment of the coeefficients is not unique: we could as well choose $-1/2, -1/2, 1$ as the coefficients
- However, the relative sizes of the coeefficients remain in any valid choice.
- Note! We will denote both the name of a metabolite and its concentration by the same symbol.

Reaction rate and concentration vectors

- Let us assume that our metabolic network has the reactions $\mathcal{R} = \{R_1, R_2, \ldots, R_r\}$
- Let the reaction R_i operate with rate v_i
- We collect the individual reaction rates to a *rate vector* $\mathbf{v} = (v_1, \ldots, v_r)^T$
- Similarly, the *concentration vector* $X(t) = (X_1(t), \ldots, X_m(t))^T$ contains the concentration of each metabolite in the system (at time t)

Stoichiometric vector and matrix

- The stoichiometric coefficients of a reaction are collected to a vector s_r
- In s_r there is a one position for each metabolite in the metabolic system
- The stoichiometric co-efficient of the reaction are inserted to appropriate positions, e.g. for the reaction

$$r : A + B \mapsto 2C,$$

$$s_r = \begin{matrix} \cdot \\ \cdot \\ A \\ \cdot \\ \cdot \\ B \\ \cdot \\ \cdot \\ C \end{matrix} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

Stoichiometric matrix

- The stoichiometric vectors can be combined into the stoichiometric matrix S.
- In the matrix S, the is one row for each metabolite $M_1, dots, M_m$ and one column for each reaction $R_1, \ldots, R_r$.
- The coefficients s_{*j} along the j'th column are the stoichiometric coeefficients of of the reaction j.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ir} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mj} & \cdots & s_{mr} \end{bmatrix}$$

Systems equations

In a network of m metabolites and r reactions, the dynamics of the system are characterized by the systems equations

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j, \text{ for } i = 1, \ldots, m$$

- X_i is the concentration of the ith metabolite
- v_j is the rate of the jth reaction and
- s_{ij} is the stoichiometric coefficient of ith metabolite in the jth reaction.

Intuitively, each system equation states that the rate of change of concentration of a is the sum of metabolite flows to and from the metabolite.

Systems equations in matrix form

- The systems equation can be expressed in vector form as

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = S_i^T \mathbf{v},$$

where S_i contains the stoichiometric coefficients of a single metabolite, that is a row of the stoichiometric matrix

- All the systems equations of different equations together can then be expressed by a matrix equation

$$\frac{d\mathbf{X}}{dt} = S\mathbf{v},$$

- Above, the vector

$$\frac{d\mathbf{X}}{dt} = \left(\frac{d\mathbf{X_1}}{dt}, \ldots, \frac{d\mathbf{X_n}}{dt} \right)^T$$

collects the rates of concentration changes of all metabolites

Steady state analysis

- Most applications of stoichiometric matrix assume that the system is in so called steady state
- In a steady state, the concentrations of metabolites remain constant over time, thus the derivative of the concentration is zero:

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

- The requires the production to equal consumption of each metabolite, which forces the reaction rates to be invariant over time.

Steady state analysis and fluxes

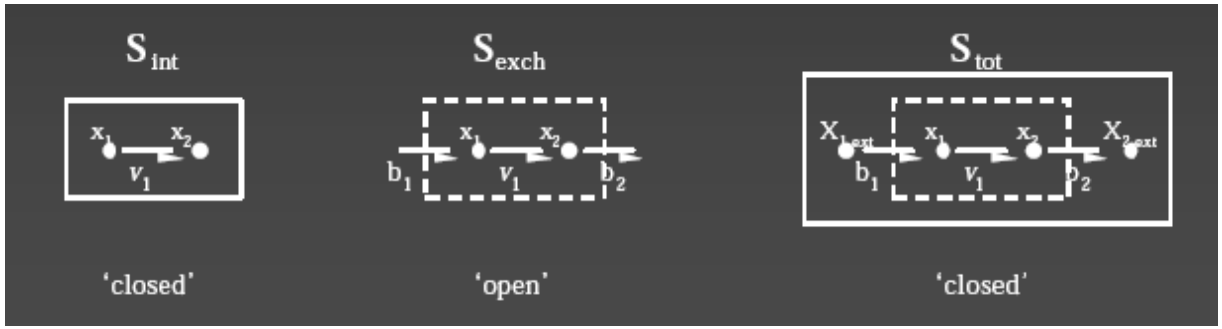
- The steady-state reaction rates $v_j, j = 1, \ldots, r$ are called the *fluxes*
- Note: Biologically, live cells do not exhibit true steady states (unless they are dead)
- In suitable conditions (e.g. continuous bioreactor cultivations) steady-state can be satisfied approximately.
- Pseudo-steady state or quasi-steady state are formally correct terms, but rarely used

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

Defining the system boundary

When analysing a metabolic system we need to consider what to include in our system
We have the following choices:

1. Metabolites and reactions internal to the cell (leftmost picture)
2. (1) + exchange reactions transporting matter accross the cell membrane (middle picture)
3. (1) + (2) + Metabolites outside the cell (rightmost picture)



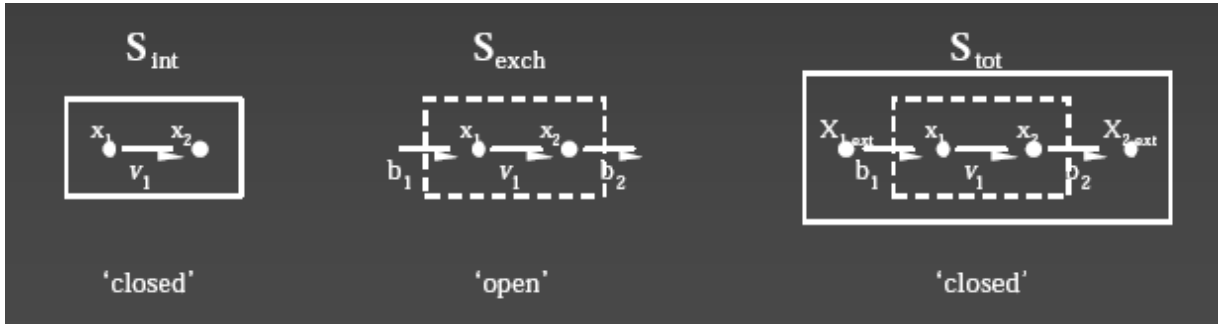
(Picture from Palsson: Systems Biology, 2006)

System boundary and the total stoichiometric matrix

The placement of the system boundary reflects in the stoichiometric matrix that will partition into four blocks:

$$\mathbf{S} = \begin{bmatrix} S_{II} & S_{IE} \\ 0 & S_{EE} \end{bmatrix}$$

- S_{II} : contains the stoichiometric coefficients of internal metabolites w.r.t internal reactions
- S_{IE} : coefficients of internal metabolites in exchange reactions i.e. reactions transporting metabolites accross the system boundary
- $S_{EI}(= 0)$: coefficients of external metabolites w.r.t internal reactions; always identically zero
- S_{EE} : coefficients of external metabolites w.r.t exchange reactions; this is a diagonal matrix.

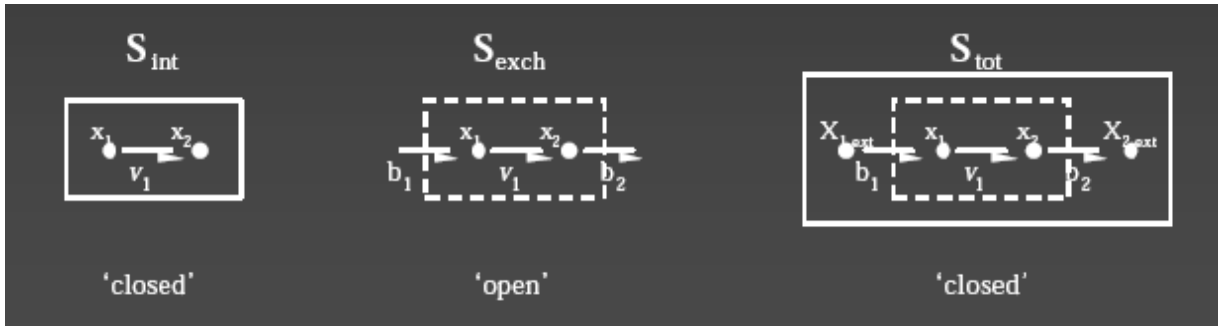


Exchange stoichiometrix matrix

In most applications handled on this course we will not consider external compounds

- The (exchange) stoichiometric matrix, containing the internal metabolites and both internal and exchange reactions, will be used
- Our metabolic system will be then open, containing exhange reactions of type $A \Rightarrow$, and $\Rightarrow B$

$$\mathbf{S} = \begin{bmatrix} S_{II} & S_{IE} \end{bmatrix}$$



System boundary and steady state analysis

- Exchange stoichiometric matrix is used for steady state analysis for a reason: it will not force the external metabolites to satisfy the steady state condition

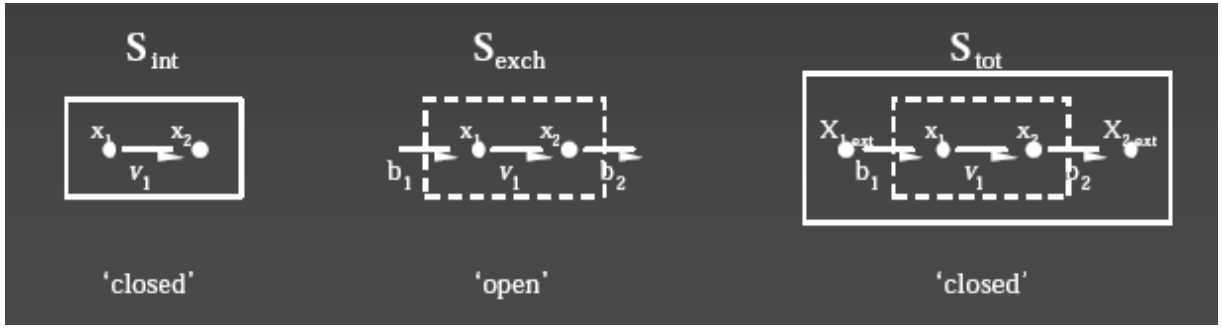
$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

- Requiring steady state for external metabolites would drive the rates of exchange reactions to zero
- That is, in steady-state, no transport of substrates into the system or out of the system would be possible!

Internal stoichiometrix matrix

- The internal stoichiometric matrix, containing only the internal metabolites and internal reactions can be used for analysis of conserved pools in the metabolic system
- The system is closed with no exchange of material to and from the system

$$\mathbf{S} = \left[S_{II} \right]$$



System boundary of our example system

- Our example system is a closed one: we do not have exchange reactions carrying to or from the system.
- We can change our system to an open one, e..g by introducing a exchange reaction $R_8 :\Rightarrow \alpha G6P$ feeding αG6P into the system and another reaction $R_9 : X5P \Rightarrow$ to push $X5P$ out of the system

R_1: $\beta\text{G6P} + \text{NADP}^+ \overset{zwf}{\Rightarrow} \text{6PGL} + \text{NADPH}$

R_2: $\text{6PGL} + \text{H}_2\text{O} \overset{pgl}{\Rightarrow} \text{6PG}$

R_3: $\text{6PG} + \text{NADP}^+ \overset{gnd}{\Rightarrow} \text{R5P} + \text{NADPH}$

R_4: $\text{R5P} \overset{rpe}{\Rightarrow} \text{X5P}$

R_5: $\alpha\text{G6P} \overset{gpi}{\Leftrightarrow} \beta\text{G6P}$

R_6: $\alpha\text{G6P} \overset{gpi}{\Leftrightarrow} \beta\text{F6P}$

R_7: $\beta\text{G6P} \overset{gpi}{\Leftrightarrow} \beta\text{F6P}$

Example

The stoichiometric matrix of our extended example contains two extra columns, corresponding to the exchange reactions $R_8 :\Rightarrow \alpha G6P$ and $R_9 : X5P \Rightarrow$

$$
\begin{array}{c}
\\
\beta G6P \\
\alpha G6P \\
\beta F6P \\
6PGL \\
6PG \\
R5P \\
X5P \\
NADP^+ \\
NADPH \\
H_2O
\end{array}
\begin{bmatrix}
-1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\
-1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Steady state analysis, continued

- The requirements of non-changing concentrations

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

  constitute a set of linear equations constraining to the reaction rates v_j.
- We can write this set of linear constraints in matrix form with the help of the stoichiometric matrix S and the reaction rate vector $\mathbf{v}$

$$\frac{d\mathbf{X}}{dt} = S\mathbf{v} = \mathbf{0},$$

- A reaction rate vector $\mathbf{v}$ satisfying the above is called the *flux* vector.

Null space of the stoichiometrix matrix

- Any flux vector $\mathbf{v}$ that the cell can maintain in a steady-state is a solution to the homogeneous system of equations

$$S\mathbf{v} = \mathbf{0}$$

- By definition, the set

$$\mathcal{N}(S) = \{\mathbf{u} | S\mathbf{u} = 0\}$$

  contains all valid flux vectors
- In linear algebra $\mathcal{N}(A)$ is referred to as the null space of the matrix A
- Studying the null space of the stoichiometric matrix can give us important information about the cell's capabilities

Null space of the stoichiometric matrix

The null space $\mathcal{N}(S)$ is a linear vector space, so all properties of linear vector spcaes follow, e.g:

- $\mathcal{N}(S)$ contains the zero vector, and closed under linear combination: $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{N}(S) \implies \alpha_1 \mathbf{v}_1 + \alpha \mathbf{v}_2 \in \mathcal{N}(S)$
- The null space has a basis $\{\mathbf{k_1}, \ldots, \mathbf{k}_q\}$, a set of $q \leq \min(n, r)$ linearly independent vectors, where r is the number of reactions and n is the number of metabolites.
- The choice of basis is not unique, but the number q of vector it contains is determined by the rank of S.

Null space and feasible steady state rate vectors

- The kernel $K = (\mathbf{k}_1, \ldots, \mathbf{k}_q)$ of the stoichiometric matrix formed by the above basis vectors has a row corresponding to each reaction. (Note: the term 'kernel' here has no relation to kernel methods and SVMs)
- K characterizes the feasible steady state reaction rate vectors: for each feasible flux vector $\mathbf{v}$, there is a vector $\mathbf{b} \in \mathbb{R}^q$ such that $K\mathbf{b} = \mathbf{v}$
- In other words, any steady state flux vector is a linear combination

$$b_1\mathbf{k}_1 + \cdots + b_q\mathbf{k}_q$$

of the basis vectors of $\mathcal{N}(S)$.

Identifying dead ends in metabolism

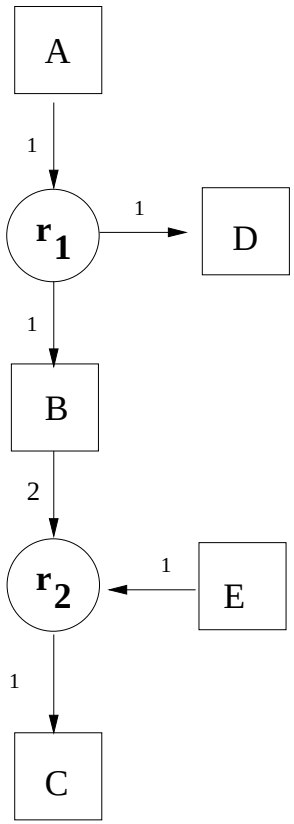
- From the matrix K, one can identify reactions that can only have zero rate in a steady state.
- Such reactions may indicate a dead end: if the reaction is not properly connected the rest of the network, the reaction cannot operate in a steady state
- Such reactions necessarily have the corresponding row K_j identically equal to zero, $K_j = 0$

Proof outline

- This can be easily proven by contradiction using the the equation $K\mathbf{b} = \mathbf{v}$:
- Assume reaction R_j is constrained to have zero rate in steady state, but assume for some i, $k_{ji} \neq 0$.
- Then we can pick the i'th basis vector of K as the feasible solution $\mathbf{v} = \mathbf{k}_i$.
- Then $v_j = k_{ji} \neq 0$ and the jth reaction has non-zero rate in a steady state.

Enzyme subsets

- An enzyme subset is a group of enzymes which, in a steady state, must always operate together so that their reaction rates have a fixed ratio.
- Consider a pair of reactions R_1 and R_2 in the metabolic network that form a linear sequence.



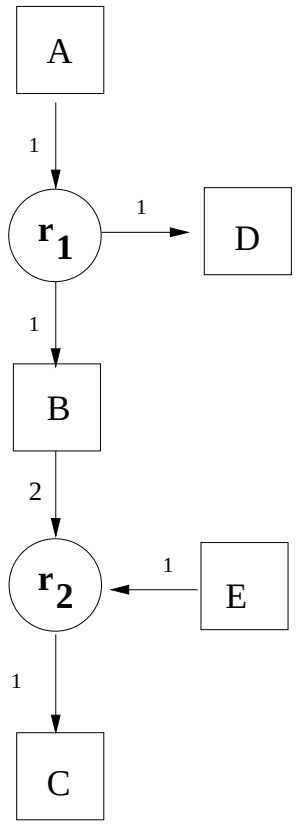
Enzyme subsets

- Let B be a metabolite that is an intermediate within the pathway produced by R_1 and consumed by R_2 for which the steady-state assumption holds. Due to the steady state assumption, it must hold true that

$$v_1 s_{i1} + v_2 s_{i2} = 0$$

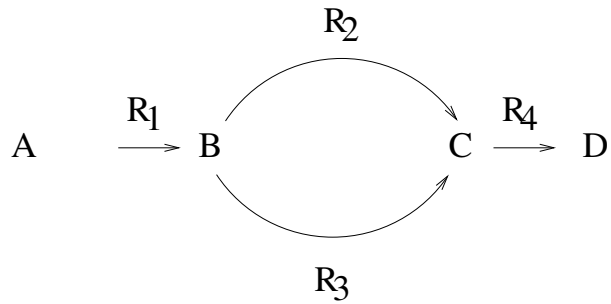
giving $v_2 = -v_1 s_{i1} / s_{i2}$.

- That is, the rates of the two reactions are linearly dependent.



Enzyme subsets

- Also other than linear pathways may be force to operate in 'lock-step'.
- In the figure, $R1$ and $R4$ form an enzyme subset, but $R2$ and $R3$ are not in that subset.



Identifying enzyme subsets

- Enzyme subsets are easy to recognize from the matrix K: the rows corresponding to an enzyme subset are scalar multiples of each other.
- That is, there is a constant α that satisfies $K_j = \alpha K_{j'}$ where K_j denotes the j'th row of the kernel matrix K
- This is again easy to see from the equation

$$K\mathbf{b} = \mathbf{v}.$$

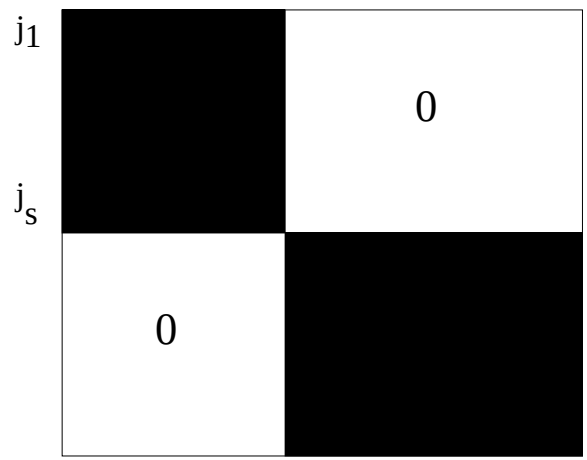
Proof outline

- Assume that reactions along rows j, j' in K correspond to an enzyme subset.
- Now assume contrary to the claim that the rows are not scalar multiples of each other. Then we can find a pair of columns i, i', where $K_{ji} = \alpha K_{j'i}$ and $K_{ji'} = \beta K_{j'i'}$ and $\alpha \neq \beta$.
- Both columns i, i' are feasible flux vectors. By the above, the rates of j and j' differ by factor α in the flux vector given by the column i and by factor β in the flux vector given by the column i'.
- Thus the ratio of reaction rates of j, j' can vary and the reactions are not force to operate with a fixed ratio, which is a contradiction.

Independent components

- Finally, the matrix K can be used to discover subnetworks that can work independently from the rest of the metabolism, in a steady state.
- Such components are characterized by a block-diagonal K: $K_{ji} \neq 0$ for a subset of rows $(j_1, \ldots, j_s)$ and a subset of columns $(i_1, \ldots, i_t)$.
- Given such a block we can change $b_{i_1}, \ldots, b_{i_t}$ freely, and that will only affect $v_{j_1}, \ldots, v_{j_s}$

$$K =$$



Example: Null space of PPP

- Consider again the set of reactions from the penthose-phospate pathway

R_1: βG6P + $NADP^+ \overset{zwf}{\Rightarrow}$ 6PGL + NADPH

R_2: 6PGL + $H_2O \overset{pgl}{\Rightarrow}$ 6PG

R_3: 6PG + $NADP^+ \overset{gnd}{\Rightarrow}$ R5P + NADPH

R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P

R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P

R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P

R_8 : $\Rightarrow \alpha G6P$

R_9 : $X5P \Rightarrow$

$$S = \begin{array}{c} \\ \beta G6P \\ \alpha G6P \\ \beta F6P \\ 6PGL \\ 6PG \\ R5P \\ X5P \\ NADP^+ \\ NADPH \\ H_2O \end{array} \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Null space of PPP

Null space of this system has only one vector

$$K = (0, 0, 0, 0, 0.5774, -0.5774, 0.5774, 0, 0, 0)^T$$

- Thus, in a steady state only reactions R_5, R_6 and R_7 can have non-zero fluxes.
- The reason for this is that there are no producers of $NADP^+$ or H_2O and no consumers of NADPH.
- Thus our PPP is effectively now a dead end!

R_1: βG6P + $NADP^+ \overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + $H_2O \overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + $NADP^+ \overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
$R_8 : \Rightarrow \alpha G6P$
$R_9 : X5P \Rightarrow$

Null space of PPP

To give our PPP non-trivial (fluxes different from zero) steady states, we need to modify our system

- We add reaction $R_{10} :\Rightarrow H_2O$ as a water source
- We add reaction R_{11}: $NADPH \Rightarrow NADP^+$ to regenerate $NADP^+$ from NADPH.
- We could also have removed the metabolites in question to get the same effect

R_1: βG6P + $NADP^+ \overset{zwf}{\Rightarrow}$ 6PGL + NADPH
R_2: 6PGL + $H_2O \overset{pgl}{\Rightarrow}$ 6PG
R_3: 6PG + $NADP^+ \overset{gnd}{\Rightarrow}$ R5P + NADPH
R_4: R5P $\overset{rpe}{\Rightarrow}$ X5P
R_5: αG6P $\overset{gpi}{\Leftrightarrow}$ βG6P
R_6: αG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
R_7: βG6P $\overset{gpi}{\Leftrightarrow}$ βF6P
$R_8 :\Rightarrow \alpha G6P$
$R_9 : X5P \Rightarrow$
R_{10}: $\Rightarrow H_2O$
R_{11}: $NADPH \Rightarrow NADP^+$

Enzyme subsets of PPP

From the kernel, we can immediately identify enzyme subsets that operate with fixed flux ratios in any steady state:

- reactions $\{R_1 - R_4, R_8 - R_{11}\}$ are one subset: R_{11} has double rate to all the others
- $\{R_6, R_7\}$ are another: R_6 has the opposite sign of R_7
- R_5 does not belong to non-trivial enzyme subsets, so it is not forced to operate in lock-step with other reactions

$$K = \begin{bmatrix} 0.2727 & 0.1066 \\ 0.2727 & 0.1066 \\ 0.2727 & 0.1066 \\ 0.2727 & 0.1066 \\ 0.3920 & -0.4667 \\ -0.1193 & 0.5733 \\ 0.1193 & -0.5733 \\ 0.2727 & 0.1066 \\ 0.2727 & 0.1066 \\ 0.2727 & 0.1066 \\ 0.5454 & 0.2132 \end{bmatrix}$$