Notes for a Systems Biology course

Metabolic Networks [Flux Balance Analysis]

Claudio Altafini

November 2015

Contents

5 Metabolic Networks [Flux Balance Analysis]

In this Chapter we study metabolic networks (i.e., of networks of biochemical reactions constituting the metabolism of an organism), making use of the Flux Balance Analysis formalism. An example of a genome-wide metabolic network is shown in Fig. 1. It is for the bacterium *E.coli* and it comprises 2383 reactions among 1668 metabolites, hence the stoichiometric matrix S is 1668×2383.

Consider the system

$$\dot{x} = Sv(x,\, k)$$

studied in the previous Chapter. In the context of metabolic networks, the idea of flux balance analysis is to disregard the dependence from x (and k) in $v(x, k)$. In this way $\frac{dx}{dt} = Sv$ is not really a system of ODEs (x no longer appears on the r.h.s.), but one can still concentrate on the properties of the stoichiometric map

$$\begin{aligned} S : \mathbb{R}^r_+ &\to \mathbb{R}^n \\ v &\mapsto \frac{dx}{dt} \end{aligned}$$

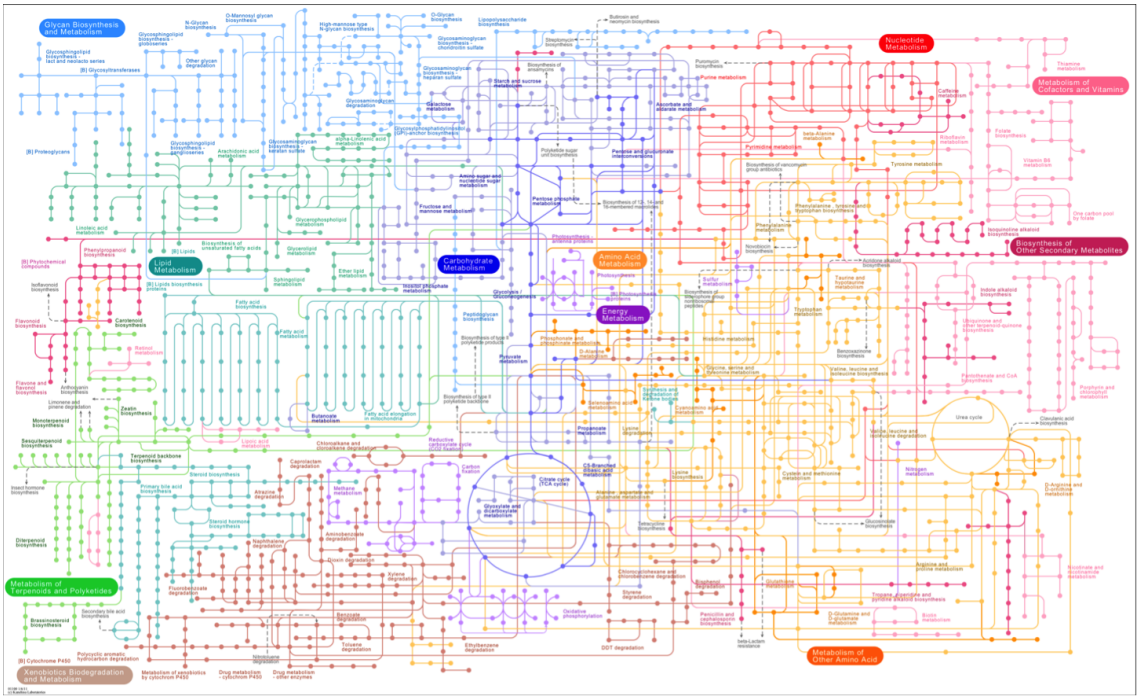


Figure 1: Full-organism metabolic network of *E.coli*.

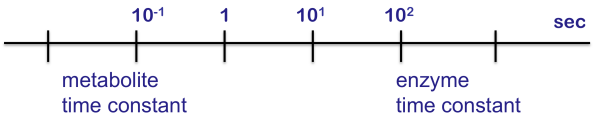


Figure 2: Time constant separation between metabolic reactions and changes in enzyme concentration (i.e., protein turover).

and in particular study the steady state flux distributions. The *rationale* behind the choice of steady states is that the time constants of the metabolic reactions are very short ($\sim 10^{-1}$ sec) when compared to most other time constants of an organism (for example transcriptional processes have time constants $\sim 10^2 - 10^4$ sec, and protein synthesis/degradation even longer, see Fig. 2), hence we can assume that the concentration of the metabolites equilibrates fast, i.e. $\frac{dx}{dt} = 0$. We can therefore limit ourselves to study the configurations of fluxes compatible with this assumption. $Sv = 0$ implies $v \in \ker(S)$.

5.1 The cone of steady state fluxes

The fact that $v \geq 0$ implies that steady state fluxes must in reality obey to the set of constraints:

$$\begin{aligned} Sv &= 0 \\ v &\geq 0 \end{aligned} \tag{1}$$

that is, the steady state fluxes must belong to a *polyhedral convex cone* given by $\ker(S) \cap \mathbb{R}^r_+$. A polyhedral convex cone in $\mathbb{R}^{r-q}_+$ is described as a nonnegative combination

$$\mathcal{C} = \{v \in \mathbb{R}^{r-q} \text{ s. t. } v = \sum_{i=1}^{d} \alpha_i w_i, \quad \alpha_i \geq 0\}$$

where w_i, $i = 1, \ldots, d$, are the generating vectors (or extreme rays), see Fig 3. Even if $\dim(\ker(S)) = r - q$ with $q = \text{rank}(S)$, the cone $\mathcal{C}$ is often described by a number of generating vectors d much larger than $r-q$. The extreme rays are called extreme pathways, as they represent pathways on the reaction graph of the network. Their calculation is a hard computational problem: for networks in which $n, r \sim 10^3$ the number of extreme pathways can be $d \sim 10^6$ or higher.

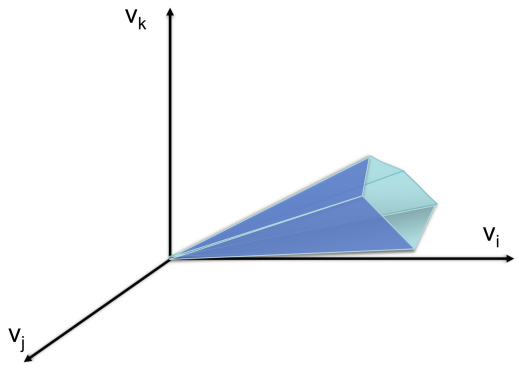


Figure 3: A cone $\mathcal{C}$ of admissible fluxes in $\mathbb{R}^3$ with $d = 5$ extreme rays.

Example Consider the network of Fig. 4. The stoichiometric matrix is

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

has $\text{rank}(S) = 3 \Longrightarrow \dim(\ker(S)) = 6 - 3 = 3$. Consider the 3 vectors $w_i \in \ker(S)$

$$w_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad w_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad w_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

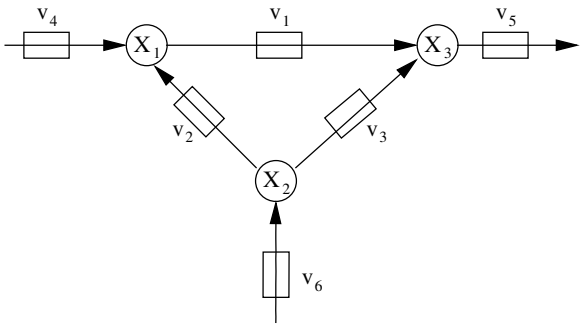
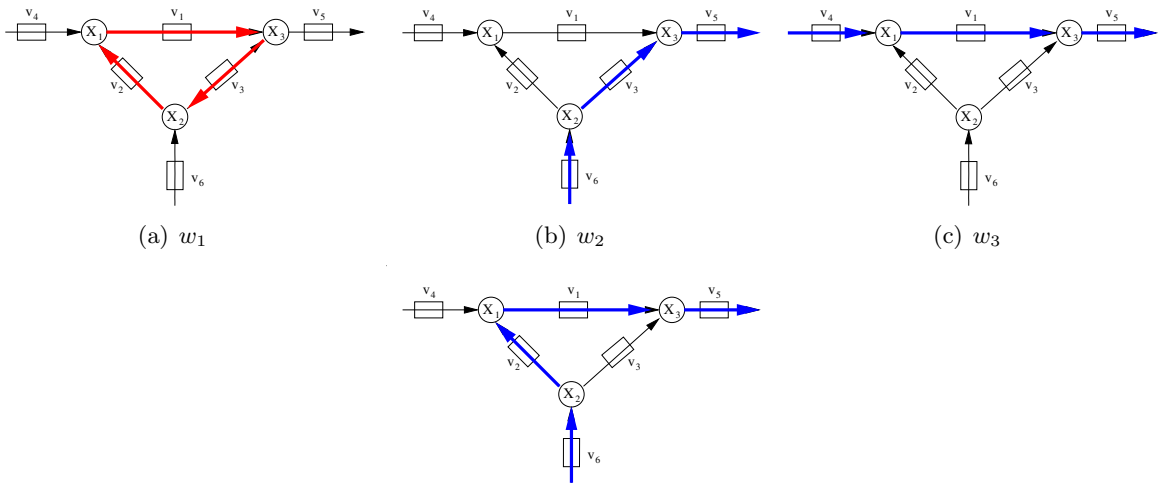


Figure 4: A basic reaction network

Clearly span(w_1, w_2, w_3) = ker(S); however the 3 vectors are not all extreme rays of the cone $\mathcal{C}$. In fact, if we look at the corresponding extreme pathways, shown in Fig. 5 (a), (b), (c), then it can be observed that w_1 is not feasible (look at the direction of the arrows), while w_2 and w_3 are.



(a) w_1 (b) w_2 (c) w_3

(d) w_4

Figure 5: Nonadmissible (red) and admissible (blue) extremal pathways.

In place of w_1 one can use instead

$$w_4 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

for which span$(w_4,\, w_2,\, w_3) = \ker(S)$ but also

$$\mathcal{C} = \left\{ v = \begin{bmatrix} w_4 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \quad \alpha_i \geq 0 \right\}.$$

In this case the cone $\mathcal{C}$ is simplicial (i. e. its generators are linearly independent in $\ker(S)$), meaning $d = r - q$. Every steady state flux is then expressed as

$$w_4 = \begin{bmatrix} \alpha_1 + \alpha_3 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_1 + \alpha_2 \end{bmatrix}, \qquad \alpha_i \geq 0$$

□

The convex cone $\mathcal{C}$ can typically be restricted to a polytope (i.e, a convex bounded polyhedral set) $\mathcal{H}$, by adding further constraints like upper bounds u on the fluxes:

$$\mathcal{H} = \mathcal{C} \cap \{0 \leq v \leq u\}$$

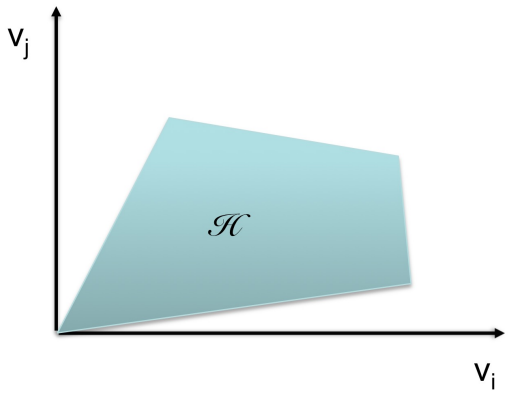


Figure 6: The polytope $\mathcal{H}$ of admissible fluxes

5.2 Thermodynamics and stoichiometric networks

In thermodynamics, the Gibbs energy is a measure of the amount of reversible work that may be performed by a system at a constant temperature and pressure. It can be defined as

$$G = E - PV - TS_e$$

where

- $E =$ energy
- $P =$ pressure
- $V =$ volume
- $T =$temperature
- $S_e =$entropy.

Here we are interested in changes of Gibbs energy in a reaction i, quantity indicated by ΔG_i. From basic thermodynamic laws, for a reaction to happen spontaneously it must be

$$v_i \Delta G_i \leq 0 \tag{2}$$

In particular, when $v_i = 0$ then also $\Delta G_i = 0$. If we associate to each molecular species X_i a chemical potential μ_i, then the Gibbs energy change for the reaction i can be expressed as

$$\Delta G_i = \sum_{j=\text{products}} S_{ji}\mu_j - \sum_{j=\text{substrates}} S_{ji}\mu_j$$

where S_{ji} are the stoichiometric coefficients. In vector form,

$$\Delta G_i = (S_{:,i})^T \mu$$

where $S_{:,i}$ is the column of S corresponding to the i-th reaction and μ is the vector of n chemical potentials. For a network of r reactions, then

$$\Delta G = S^T \mu$$

is the $r \times 1$ vector of Gibbs energy changes. From (2), if we have the restriction $v \geq 0$ then it must be $\Delta G \leq 0$.

Example If in the elementary reaction

$$X_1 \xrightarrow{v_1} X_2$$

the chemical potential of $X_1$is μ_1 and that of X_2 is μ_2, then $\Delta G_1 = \mu_2 - \mu_1$. If the reaction happens (i.e., $v_1 > 0$) then $\Delta G_1 < 0$, i.e., $\mu_1 > \mu_2$. □

Example Consider the reaction

$$pX_1 + X_2 \xrightarrow{v_1} X_3$$

with chemical potentials of X_i equal to μ_i, then

$$\Delta G_1 = S^T \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -1 & -p & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \mu_3 - p\mu_2 - \mu_1$$

Also in this case $v_1 > 0$ implies $\Delta G_1 < 0$, i.e., $\mu_1 + p\mu_2 > \mu_3$. □

5.2.1 Loop law and thermodynamically unfeasible steady states

At steady state it must be $Sv = 0$, hence also $v^T S^T = 0$. But then it must also be

$$v^T S^T \mu = v^T \Delta G = 0 \tag{3}$$

The latter expression must be valid also on subsets of the reaction network which satisfy the steady state assumption, for instance on the extreme pathways $w_1, \ldots, w_d$ which generate $\mathcal{C}$. When such subset involves only internal reactions but not inflow/outflow reactions, then the relationship (3) is called the loop law, as it forbids to have nonzero fluxes on loops (i.e., directed cycles) of reactions. Combined with (2), it can be used to discover steady state flux distributions which are thermodynamically unfeasible, and hence can be discarded.

Example Consider the reaction network shown in Fig. 7 (similar, but not identical to a previous example). Denote μ_i, the chemical potentials of X_i . The stoichiometric matrix is

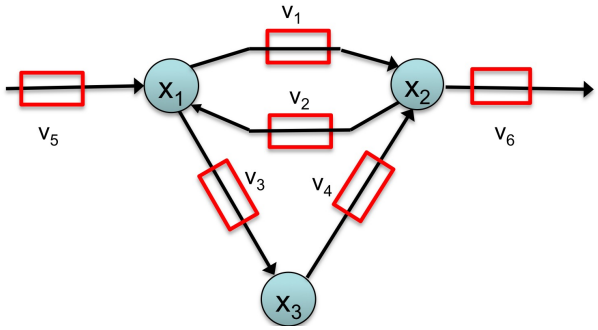


Figure 7: Example of reaction network with (unfeasible) loop.

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

and since $q = \text{rank}(S) = 3$ the system has no conservation laws. Since $r = 6$, $\dim(\ker(S)) = 6 - 3 = 3$. The following 3 vectors $w_i \in \ker(S)$

$$w_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad w_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad w_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

are linearly independent and form a basis of $\ker(S)$, admissible by the direction of the flows (hence they are admissible extreme rays of the cone $\mathcal{C}$). However, the extreme pathway w_3 corresponds to the loop shown in Fig. 8 which involves only internal fluxes. Let us shown that such pathway

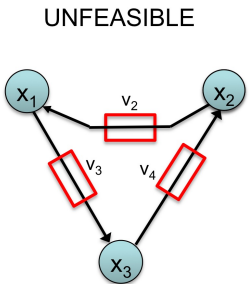


Figure 8: Extreme pathway of the reaction network of Fig. 7 representing an "internal" loop, whch is unfeasible.

is not compatible with all positive fluxes. Assume $v_3 > 0$ and $v_4 > 0$. From (2), excluding trivial cases (i.e., assuming (2) holds strictly), we have

$$v_3 > 0 \text{ and } v_3 \Delta G_3 < 0 \Longrightarrow \Delta G_3 = \mu_2 - \mu_1 < 0 \implies \mu_2 > \mu_1$$
$$v_4 > 0 \text{ and } v_4 \Delta G_4 < 0 \Longrightarrow \Delta G_4 = \mu_3 - \mu_2 < 0 \implies \mu_3 > \mu_2$$

Hence $\mu_3 > \mu_1 \Longrightarrow \Delta G_2 = \mu_1 - \mu_3 > 0$. But then, from $v_2 \Delta G_2 < 0$ it must be $v_2 < 0$, that is the only admissible reaction direction between X_1 and X_2 is v_1, not v_2! Hence the loop of Fig. 8 is thermodynamically unfeasible. We can replace w_3 e.g. with

$$w_4 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

which corresponds to the (loop-free) pathway of Fig. 9 and it is feasible. However, $w_4 = w_1 + w_2$ i.e., it is a nonnegative combination of the two already identified extreme rays, hence it is not providing a new extreme ray for $\mathcal{C}$. □

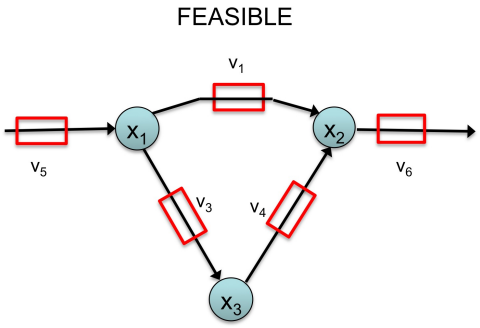


Figure 9: Feasible extreme pathway of the reaction network of Fig. 7 (no "internal" loop).

Notice that loop laws are something different from conservation laws (which are less important here because we disregard the concentrations of the metabolites).

5.3 Choosing a "preferred" steady state flux distribution [Linear Programming]

Any $v \in \mathcal{H}$ is an admissible flux distribution. How do we choose a "preferred" flux distribution within $\mathcal{H}$? In "Flux Balance Analysis", one popular choice is to optimize some cost function, for example growth rate i.e., the production of biomass of an organism (the idea is that organisms like bacteria have evolved over millions of years to optimize their growth). This is typically written as a linear cost functional $F(v) = \sum_{i=1}^{r} \beta_i v_i$, where β_i describe the (empirical) relative weights of all reactions which are crucial for growth, such as biosynthesis of nucleotides, aminoacids, fatty acids, cell-wall components, etc. The problem becomes therefore a Linear Programming problem:

$$\begin{aligned} \max_v F(v) &= \sum_{i=1}^{r} \beta_i v_i \\ \text{subject to} \quad & Sv = 0 \\ & 0 \le v \le u \end{aligned} \tag{4}$$

When solving (4), the level surfaces of the cost function $F(v)$ may hit the "upper" boundary of $\mathcal{H}$ in a single point v^{LP} as shown in Fig. 10 or may do so on a face of $\mathcal{H}$, in which case there is a degenerate solution to (4) (i.e., the optimal flux distribution is not unique, see Fig. 11).

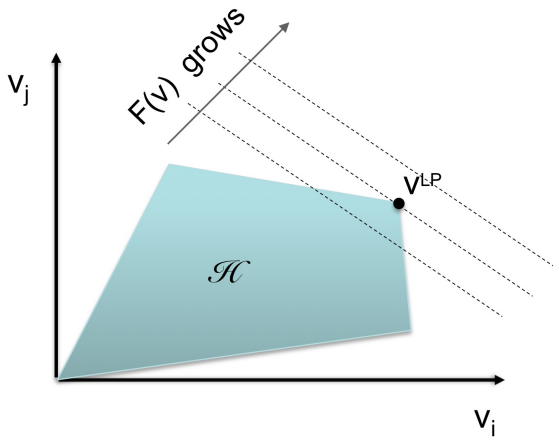


Figure 10: The level surfaces of the linear cost function $F(v)$ on the polytope $\mathcal{H}$ of admissible fluxes. The LP problem (4) has a unique solution v^{LP}.

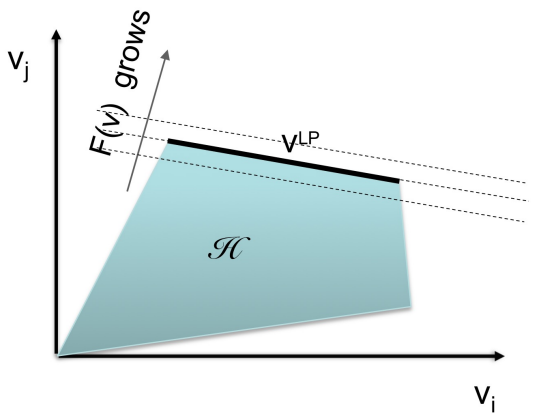


Figure 11: The level surfaces of another linear cost function $F(v)$ on the polytope $\mathcal{H}$ of admissible fluxes. The LP problem (4) has a degenerate solution in this case.

5.4 Minimization of metabolic adjustment after the knockout of a reaction [Quadratic programming]

In the optimization problem (4), taking as cost function $F(v)$ growth rate corresponds to modeling the fact that (maybe over an evolutionary horizon of millions of years) the bacteria has "shaped" its own metabolism to optimize growth (and hence "beating" other strains of bacteria). This however concerns the wild type bacterium, as found in nature.

When in the lab we do a knock-out of a gene which codes for an enzyme (and hence we suppress the flux of the corresponding reaction, for instance $v_i = 0$), then the polytope of feasible steady states restricts, because putting a reaction v_i equal to zero means collapsing $\mathcal{H}$ to the origin along the i-th axis. Call $\mathcal{H}' \subseteq \mathcal{H}$ the new polytope we obtain. The question that can be posed is what is the new optimum in correspondence of the new polytope $\mathcal{H}'$? One possibility is to re-run the LP problem (4) on the new polytope $\mathcal{H}'$. However, if we think of the regulation of metabolism as a "distributed" system, in which each enzyme has been synthetized and activated independently (and unaware) of what the other enzymes are doing, then another possibility is that the $r-1$ remaining reactions tend to behave as close as possible to the old wild type. The principle "as close as possible to the old wild type" can be modeled taking the orthogonal projection of the old optimum v^{LP} onto the new polytope $\mathcal{H}'$, see Fig. 12. The orthogonal projection is the flux distribution inside $\mathcal{H}'$ which minimizes the distance to v^{LP}. In the literature this is called *Minimization of Metabolic Adjustement* (MOMA), and it corresponds to computing the solution of following Linear-Quadratic

Problem

$$\begin{aligned} \min_v \Phi(v) = \|v - v^{\mathrm{LP}}\|_2 = \sum_{i=1}^{r} \left(v_i - v_i^{\mathrm{LP}}\right)^2 \\ \text{subject to} \quad Sv = 0 \\ 0 \le v \le u \\ v_i = 0 \end{aligned} \tag{5}$$

The extra constraint $v_i = 0$ corresponds to the knock-out of the i-th reaction. Experimentally one indeed observes that artificial mutants (i.e., knock-outs) do not seem optimal with respect to the growth rate, unlike the wild type bacteria.

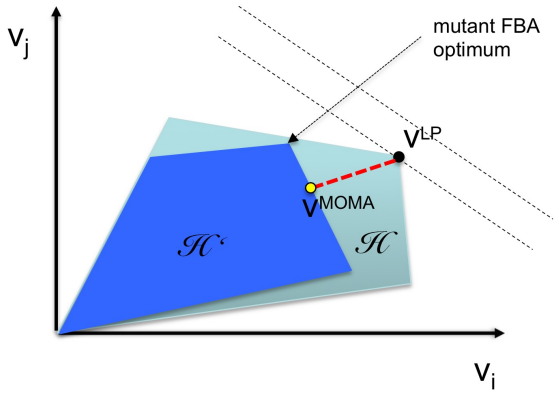


Figure 12: Minimization of metabolic adjustment after a knockout. The knock-out reduces the polytope $\mathcal{H}$ of admissible fluxes reduces to $\mathcal{H}'$. v^{MOMA} is the solution of the LQP problem (5), which is different from the solution of the LP problem (4) computed on the mutant strain (polytope $\mathcal{H}'$).