Mathematical Foundations of Constraint-Based Modelling

David S. Tourigny

Columbia University Irving Medical Center

07/22/19

CONICYT Ministerio de Educación Gobierno de Chile

REDI170254

PCI

Programa de Cooperación Internacional

ESCUELA DE INGENIERÍA BIOQUÍMICA

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

Universidad de los Andes

FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

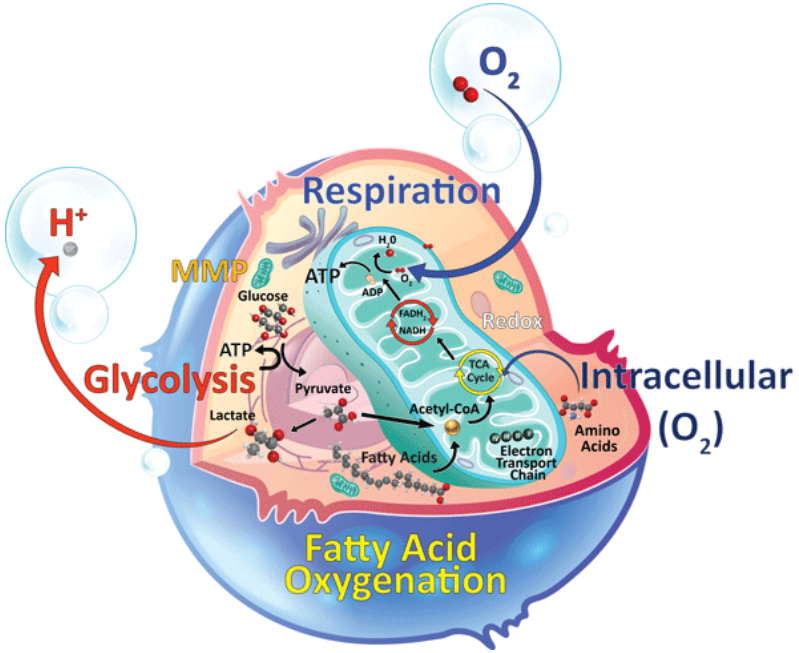
(M2S 2019)

Topics to cover

- Kinetic models of metabolism and the quasi-steady state assumption
- Constraints on metabolic flux distributions and feasibility space
- Objective functions, constrained optimization, and LP problems

Metabolism is a network of reactions

- In biology, **metabolites** are inter-converted by chemical **reactions**



Source: amsbio webpage

http://www.amsbio.com/cellular-metabolism.aspx

- Metabolites can be **intra**- or **extracellular**, and **compartmentalized**

Modelling the rate of a reaction

- The **Law of Mass Action** says that the rate of a reaction is proportional to the probability of collisions (and therefore concentrations) of the metabolites participating in the reaction

- For a reversible reaction of the form

$$\mathbf{S_1 + S_2 \rightarrow 2S_3}$$

this means the rate broken down in terms of **forward** and **reverse** directions is

$$v = v_+ - v_- = k_+ S_1 S_2 - k_- {S_3}^2$$

Modelling the rate of a reaction contd.

- The rate of reaction can be used to express the derivatives of metabolite concentrations as a system of ordinary differential equations (ODEs)

$$\frac{dS_1}{dt} = \frac{dS_2}{dt} = -v = -(k_+ S_1 S_2 - k_- {S_3}^2)$$

$$\frac{dS_3}{dt} = 2v$$

Stochiometric coefficients

- From the previous slide we have

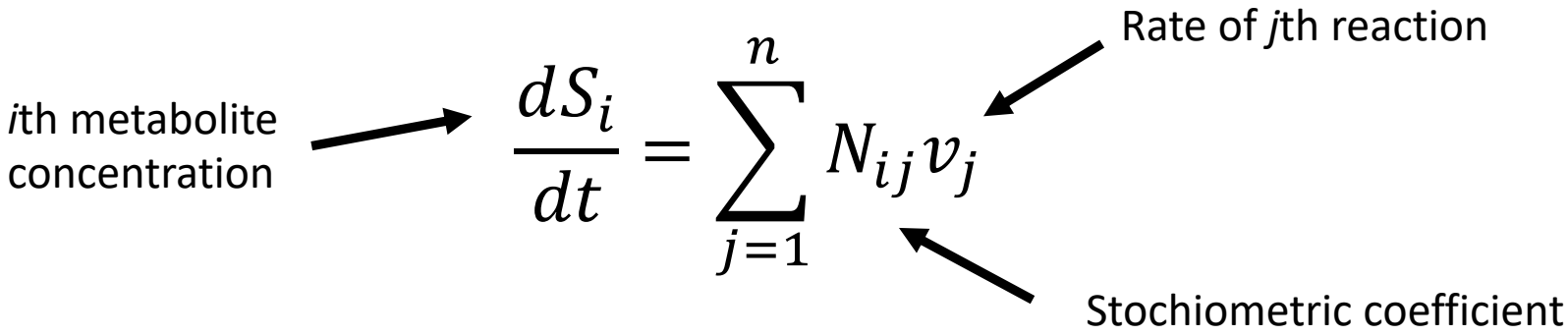
$$\frac{dS_1}{dt} = \frac{dS_2}{dt} = -v, \qquad \frac{dS_3}{dt} = 2v$$

- We say that the **stochiometric coefficient** of $\mathbf{S_1}$ and $\mathbf{S_2}$ is **-1** and that of $\mathbf{S_3}$ is **2**

- In general, stochiometric coefficients denote the proportion of each metabolite involved in a reaction

Representing a metabolic reaction network

- A **metabolic network** with n reactions can represented as



- This system of ODEs can be written in **vector form** as

$$\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v}$$

Some terminology

- In the system of ODEs

$$\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v}$$

a) $\mathbf{S}$ is called the vector of **metabolite concentrations** (of length m)

b) $\mathbf{N}$ is called the **stochiometric matrix** (of dimension $m \times n$)

c) $\mathbf{v}$ is called the **flux vector** (of length n)

Quasi-steady state assumption

- Intracellular metabolite concentrations are modelled with units per gram of biomass, X , which accumulates with growth rate μ:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v} - \mu\mathbf{S}$$

- The quasi-steady state assumption (QSSA) assumes that metabolic transients are fast compared to cell growth effects so that

$$0 = \mathbf{N} \cdot \mathbf{v}$$

Stochiometric constraints

- From the QSSA, the flux vector $\mathbf{v}$ must satisfy the **stochiometic constraints** ($\mathbf{v}$ is in the **null space** of $\mathbf{N}$)

$$\mathbf{N} \cdot \mathbf{v} = 0$$

- Solving for $\mathbf{v}$ instead of $\mathbf{S}$ (remember, $\mathbf{v} = \mathbf{v}(\mathbf{S})$ is a function of $\mathbf{S}$) means knowledge of kinetic rate parameters in not necessary

- This defines m algebraic equations for n variables. In general however, $n \gg m$ and therefore the system is **underdetermined**

The flux cone

- We will assume all reactions are irreversible so that the components of $\mathbf{v}$ are non-negative, i.e. $\mathbf{v} \geq 0$

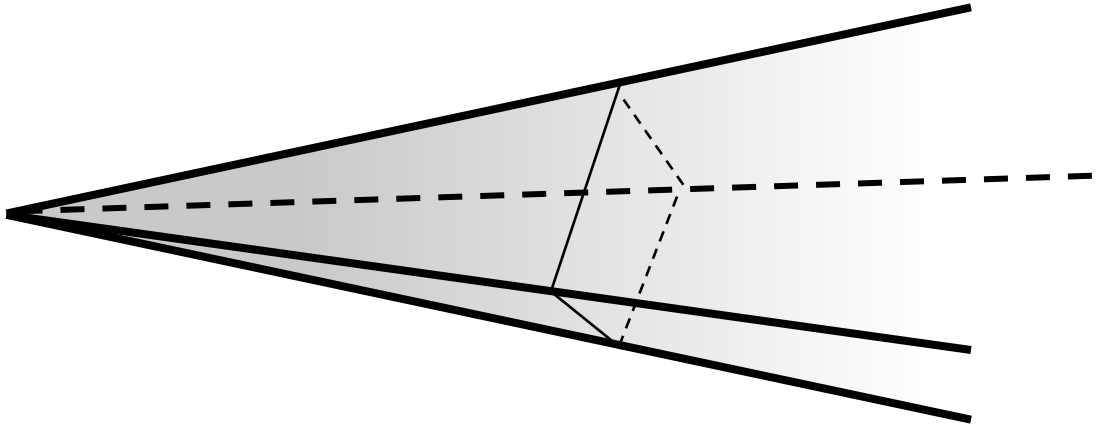
- This is always possible to achieve by splitting each reversible reaction into two: a forward and reverse reaction, both irreversible

- The flux vector $\mathbf{v}$ therefore must satisfy the constraints that define the **flux cone**:

$$\mathbf{N} \cdot \mathbf{v} = 0, \qquad \mathbf{v} \geq 0$$

The flux cone contd.

- The flux cone is really a cone:



- Each stochiometric constraint is a **linear equation of the fluxes** and defines a **hyperplane** in n-dimensional space. The flux cone is the intersection of these with the positive region of n-dimensional space

Bounds on fluxes

- The flux cone is unbounded in that infinite flux values are allowed

- In reality, we know that physical arguments prevent certain reactions proceeding at rates greater than a given upper bound, i.e.

$$\mathbf{UB} \geq \mathbf{v} \geq 0$$

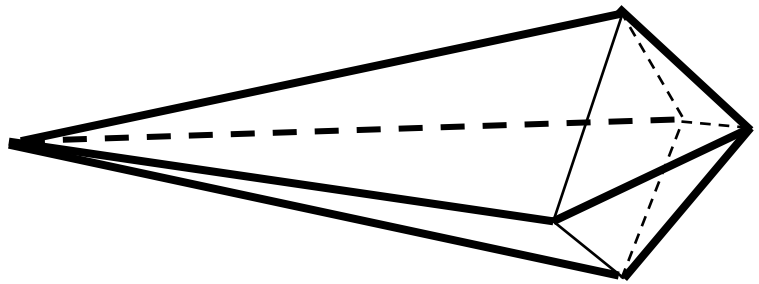
- This further constrains $\mathbf{v}$ to a **feasibility space**

Feasibility space

- The full set of constraints on $\mathbf{v}$ are

$$\mathbf{N} \cdot \mathbf{v} = 0, \qquad \mathbf{v} \geq 0, \qquad \mathbf{UB} \geq \mathbf{v} \geq 0$$

- These define a **(bounded or unbounded) convex polytope**



The objective function

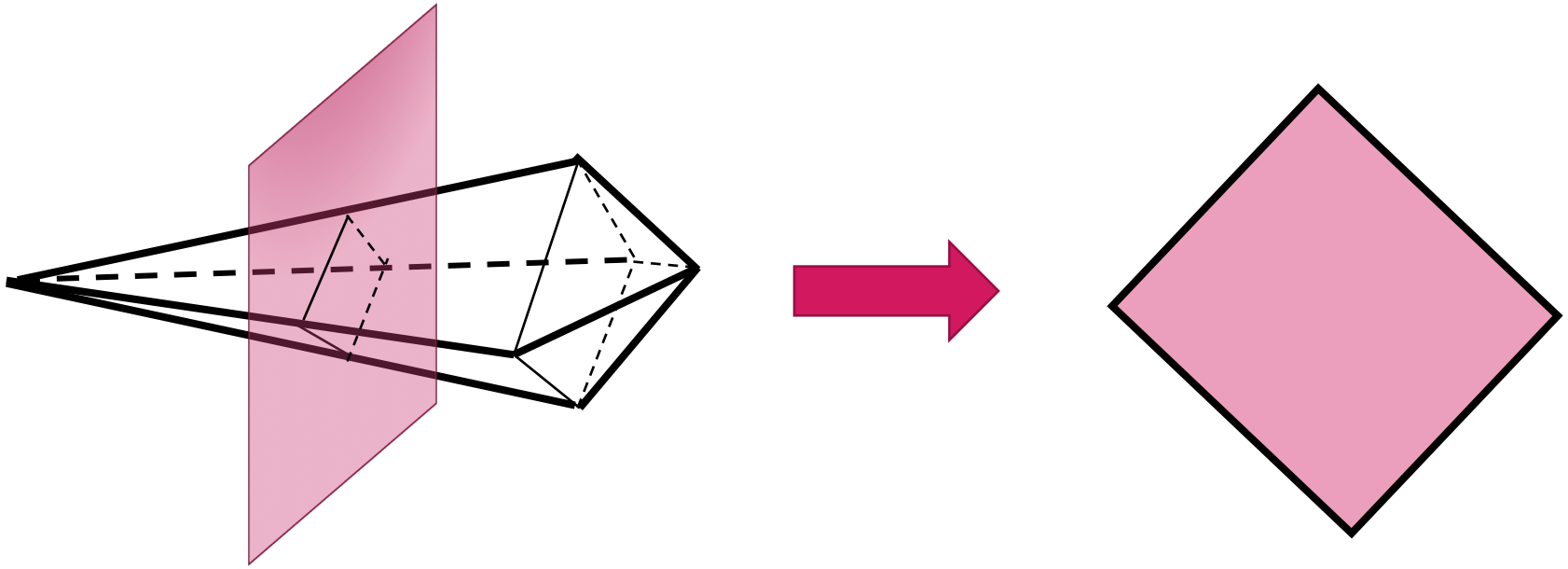
- We have avoided parameterization at a cost of underdetermination. The feasible space defines an infinite set of flux distributions, but we want just one…

- Introducing an **objective function** based on some biological assumptions can help us to find an **optimal flux distribution**

- This turns the problem of finding $\mathbf{v}$ into a **constrained optimization problem**

The objective function contd.

- An objective function will reduce the set of flux distributions



- Although the optimal value is unique, an optimal flux distribution might not be

Examples of objective functions

- By far the most common objective is **maximization of growth rate** μ. In bio-engineering, sometimes the experimentalist wants to **maximize the rate of production** of a certain metabolite (e.g. ethanol)

- In these cases, the objective function is a linear combination of the components of the flux vector and **objective coefficients**

$$f(\mathbf{v}) = \sum_{i=1}^{n} c_i v_i = \mathbf{c} \cdot \mathbf{v}$$

Examples of objective functions contd.

- Alternative objective functions sometimes considered are not linear (e.g., ratios of linear functions for yields, quadratic objectives in MOMA and for experimental data)

- **All objectives usually considered are convex** (we know how to solve convex optimization problems!)

Constrained optimization problem

- The **constrained optimization problem** for finding an **optimal flux distribution** $\mathbf{v}^*$ is

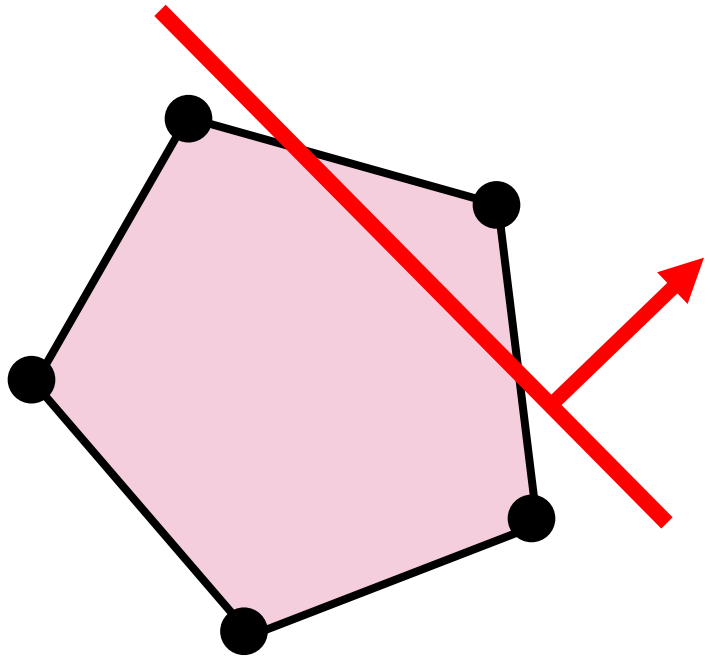
$$\text{Maximize:} \quad f(\mathbf{v})$$

$$\text{Such that:} \quad \mathbf{N} \cdot \mathbf{v} = 0,\ \mathbf{v} \geq 0,\ \mathbf{UB} \geq \mathbf{v} \geq 0$$

- $\mathbf{v}^*$ is a flux distribution that solves this problem

Linear programming

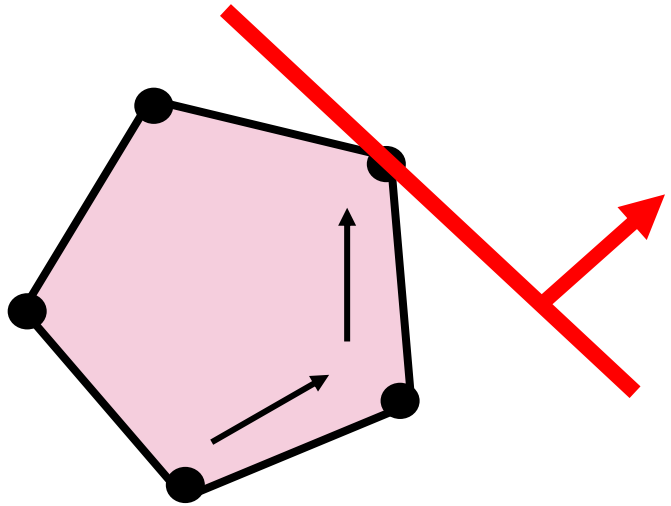
- When $f(\mathbf{v})$ is linear the problem is a **Linear Programming (LP) problem** and several methods exist for finding a solution



Maximization of $f(\mathbf{v})$

Linear programming contd.

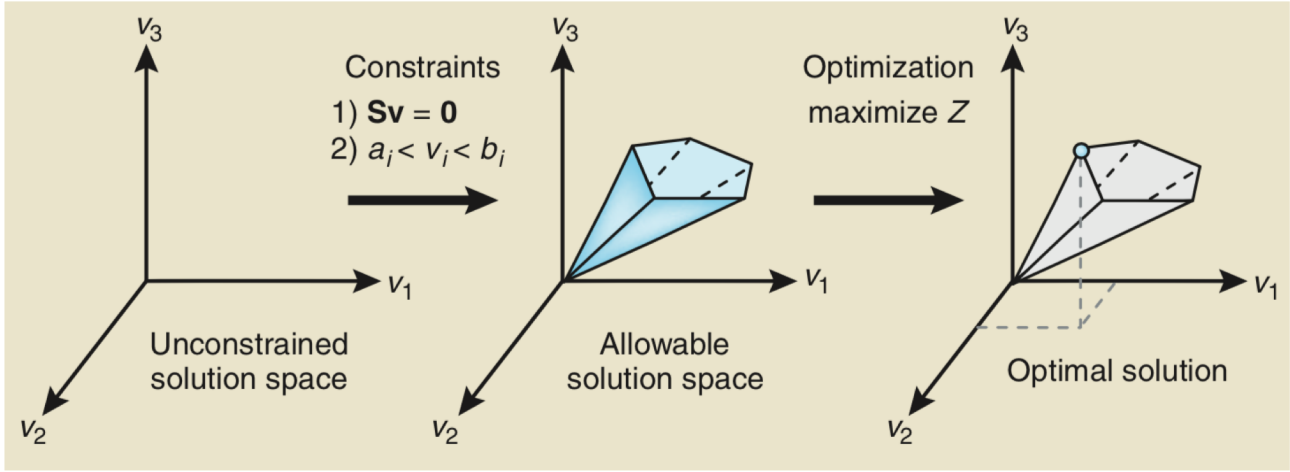
- LP problems are nice because if an (finite) optimal solution exists it can always found at a vertex



- The **simplex algorithm** for solving LP problems walks around vertices of the convex polytope looking for a solution!

Flux balance analysis

- Flux balance analysis (FBA) applies the simplex method to find an optimal flux distribution maximizing a linear objective function



Source: Orth, Thiele & Palsson (2010) *Nature Biotechnology* vol. 28

- Stochiometric constraints come from the metabolic reaction network, objective function and flux bounds from biological assumptions

Constraint-Based Reconstruction and Analysis

- FBA is just one technique falling under the umbrella of Constraint-Based Reconstruction and Analysis (COBRA)

- Other COBRA techniques are based on further biological assumptions and/or integration with experimental data

- Tools available through the open source COBRA code base:

openCOBRA

Open-source, community-developed code base for COnstraint-Based Reconstruction and Analysis.

https://opencobra.github.io/