Systembiologie 551-1174-00L

Flux Balance Analysis

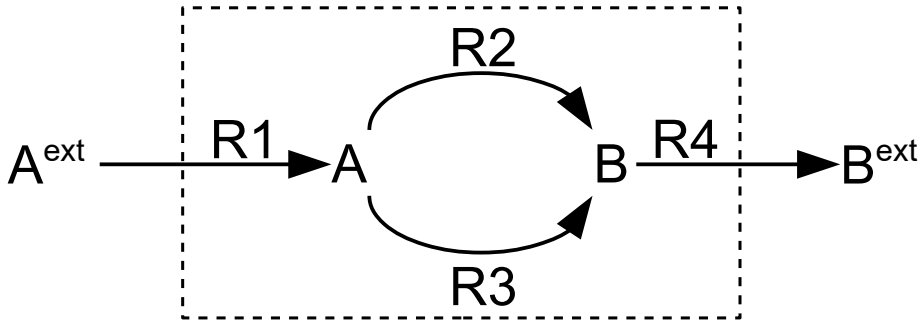
23 March, 2017

Jörg Stelling, D-BSSE

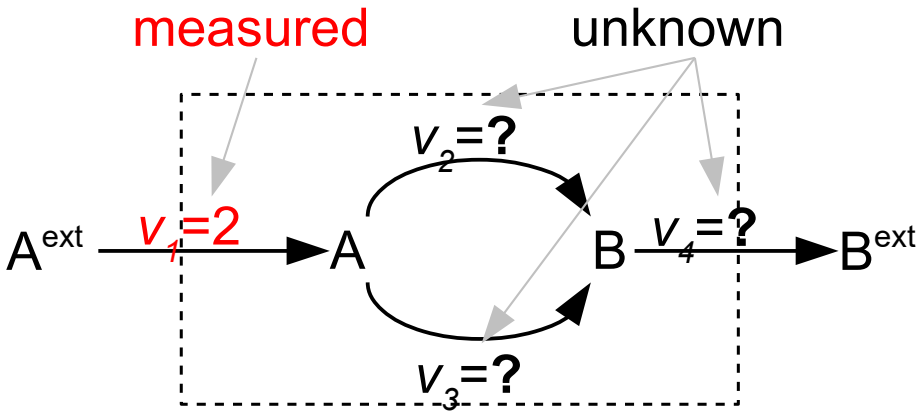
Content:

- From textbook biochemistry to genome-scale stoichiometric models
- The solution space of stoichiometric models
- Flux balance analysis

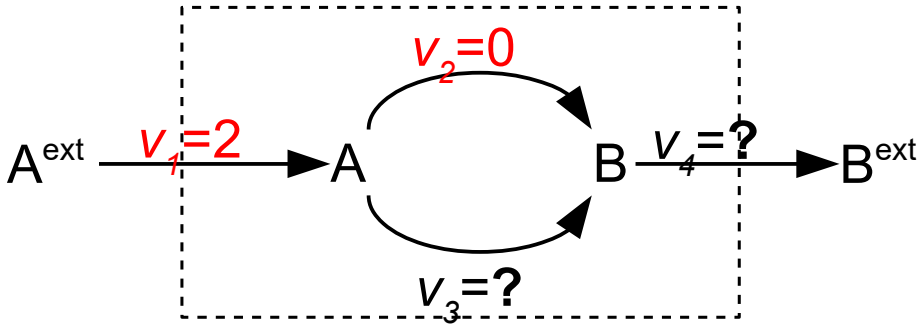
Steady-State Solution Spaces for Fluxes



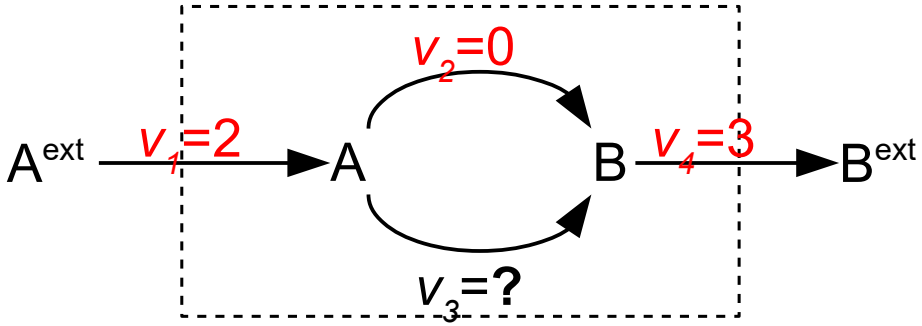
$$\boldsymbol{S} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{bmatrix}$$



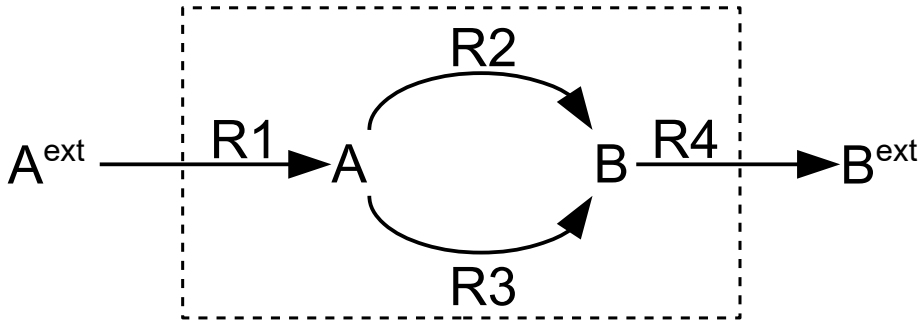
v1 = v4



v1 = v4 => v1 = v3

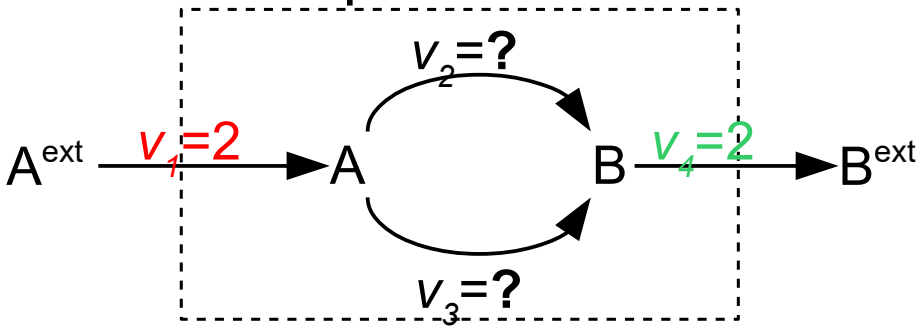


Steady-State Solution Spaces for Fluxes

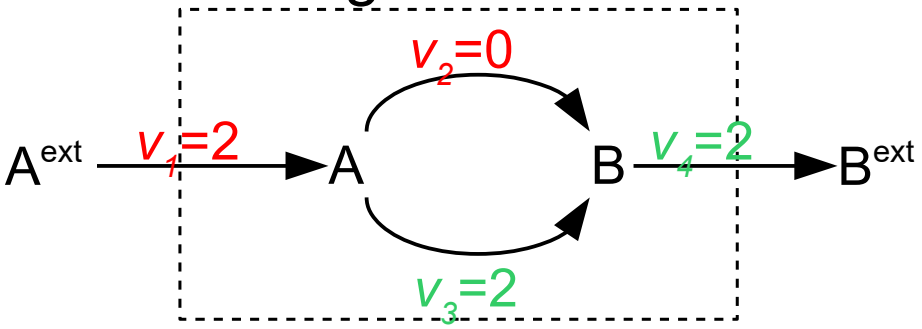


$$\boldsymbol{S} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{bmatrix}$$

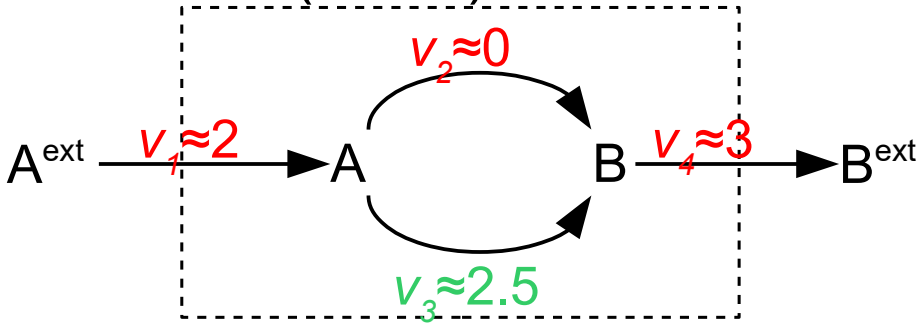
Multiple solutions



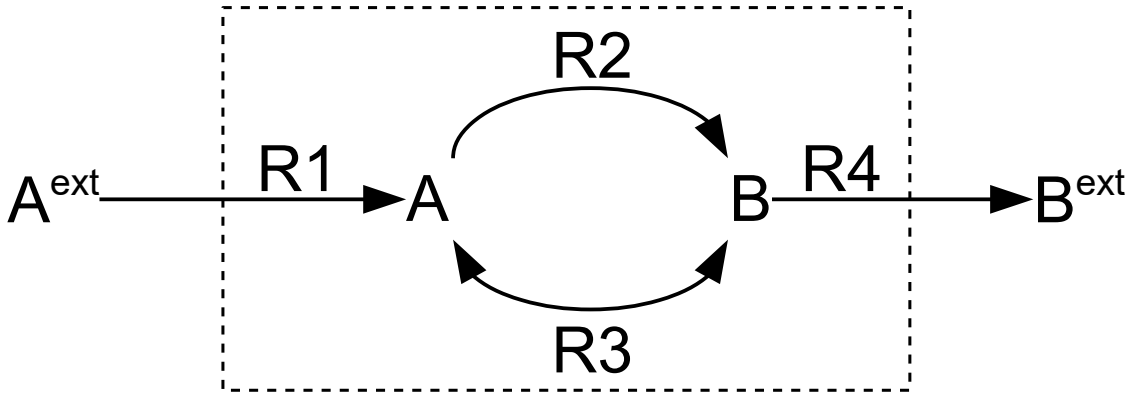
Single solution



No (exact) solution



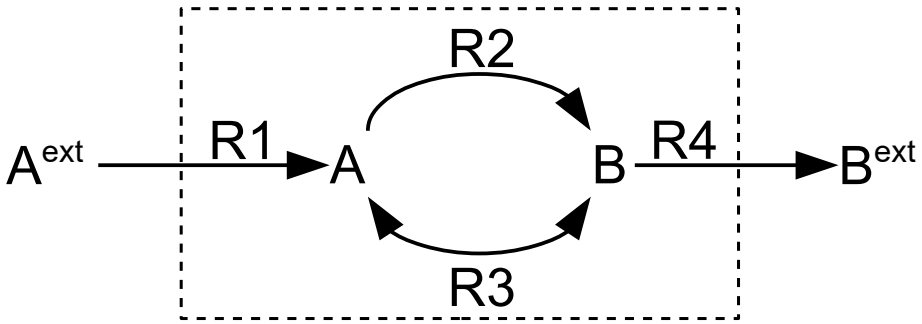
Stoichiometric Models: Dimensions



- Number of (internal) metabolites $\{A,B\}$: $m = 2$.
- Number of metabolic reactions $\{R1\text{-}R4\}$: $n = 4$.
- Sets of reversible and irreversible reactions:

  $\boldsymbol{rev} = \{R3\}$, $\boldsymbol{irrev} = \{R1,R2,R4\}$, $\boldsymbol{rev} \cap \boldsymbol{irrev} = \boldsymbol{0}$.

Stoichiometric Models: Representation



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

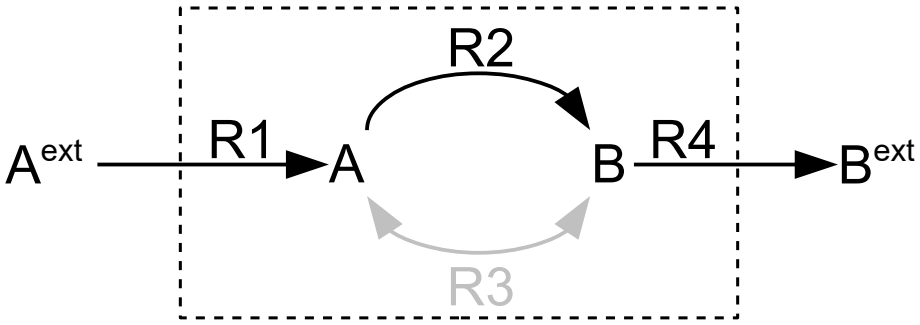
R1 R2 R3 R4 ↓ ↓ ↓ ↓

← A

← B

- Network representation: **Stoichiometric matrix** $\boldsymbol{S}$ $(m \times n)$.
- Rows → Internal metabolites i ; Columns → Reactions j.
- Elements s_{ij} : Stoichiometric coefficients (>0 for products).

Stoichiometric Models: Flux Distributions



v is feasible, v' is not

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}' = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

Which flux is feasible?

= machbar, durchführbar

- **Flux distribution:** Specification of all fluxes in the network

  → Vector $\boldsymbol{v}$ of n reaction rates that is to be determined.

- **Feasibility criterion:** $v_i \geq 0$ for all irreversible reactions.

Balanced Networks: Quasi Steady State

- Metabolic networks: Fast reactions (msec - seconds timescale) and high turnover of reactands.

- Quasi steady state → Metabolite balancing equation:

$$\frac{d\,\boldsymbol{c}(t)}{dt} = \boldsymbol{S}\cdot\boldsymbol{v}(t) \xrightarrow[const.]{\boldsymbol{c}(t),\boldsymbol{v}(t)} \mathbf{0} = \boldsymbol{S}\cdot\boldsymbol{v}$$

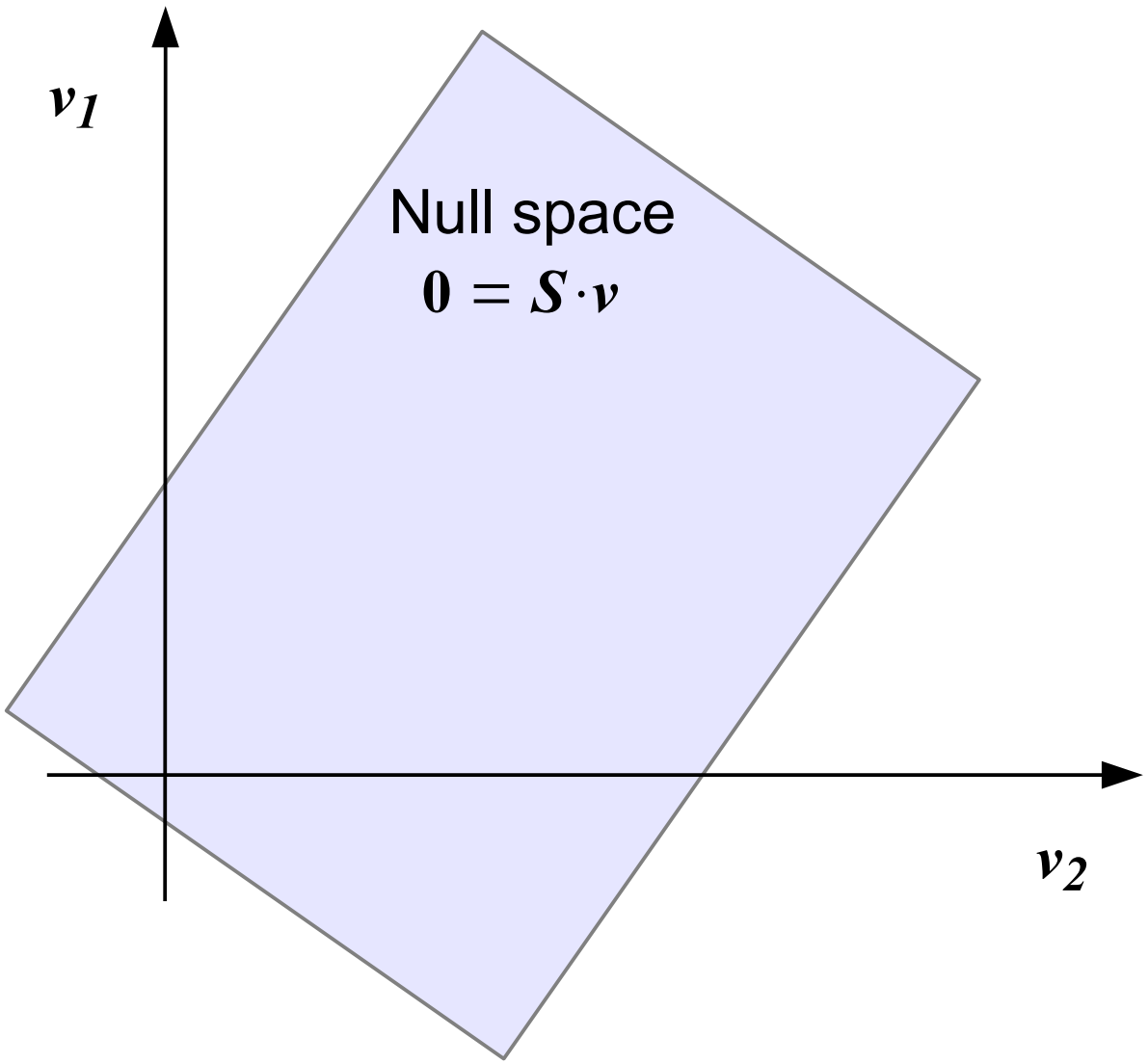
- **Homogeneous systems of linear equations:** Consumption of each metabolite equals production.

Balanced Networks: Null Space

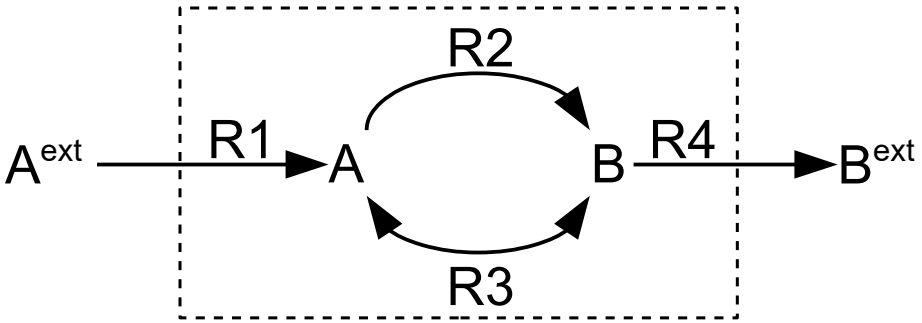
$$\mathbf{0} = \boldsymbol{S} \cdot \boldsymbol{v}$$

- Solution $\boldsymbol{v} = \boldsymbol{0} \rightarrow$ Thermodynamic equilibrium.
- Metabolic networks: $n >> m \rightarrow$ More unknowns than equations $\rightarrow$ Multiple compliant vectors $\boldsymbol{v}$.
- **Linear algebra:** All possible solutions lie in the (vector) null space (or: kernel) of $\boldsymbol{S}$ with dimension $n\text{-}rank(\boldsymbol{S})$.

Balanced Networks: Null Space



Kernel Matrix Represents the Solution Space



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

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

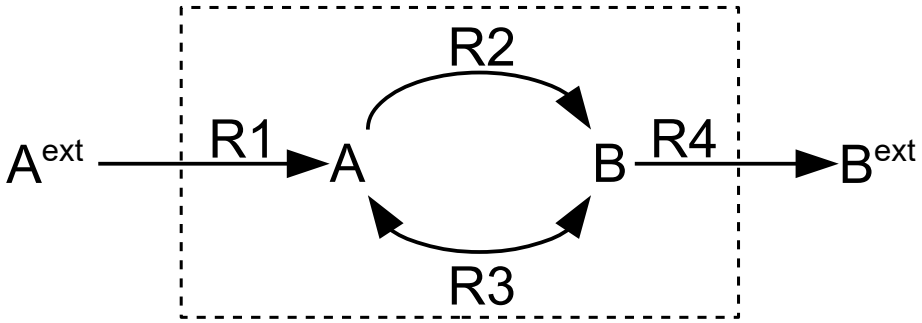
$$\boldsymbol{rev} = \{R3\}$$

$$\boldsymbol{irrev} = \{R1, R2, R4\}$$

$$n\text{-}rank(\boldsymbol{S}) = 4 - 2 = 2$$

- Basis vectors of the null space: Find $n\text{-}rank(\boldsymbol{S})$ linearly independent solutions → Arrange in a kernel matrix $\boldsymbol{K}$.
- We can reconstruct all possible $\boldsymbol{v}$ by linear combination $\boldsymbol{b}$ of the columns of the kernel matrix: $\boldsymbol{v} = \boldsymbol{K}\boldsymbol{b}$.

Kernel Matrix: Enzyme Subsets



$$\boldsymbol{S} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{bmatrix} \qquad \begin{aligned} \boldsymbol{rev} &= \{R3\} \\ \boldsymbol{irrev} &= \{R1, R2, R4\} \end{aligned}$$

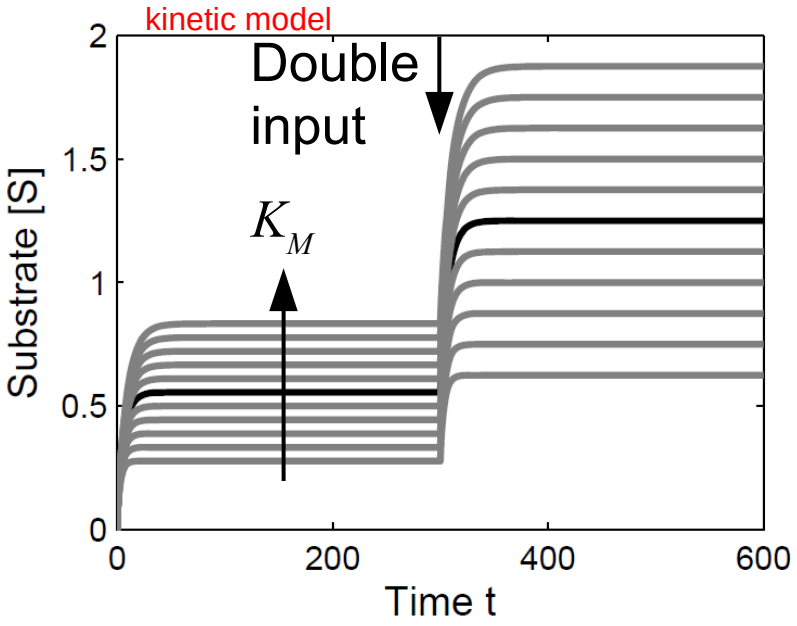
$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{matrix} \leftarrow R1 \\ \leftarrow R2 \\ \leftarrow R3 \\ \leftarrow R4 \end{matrix}$$

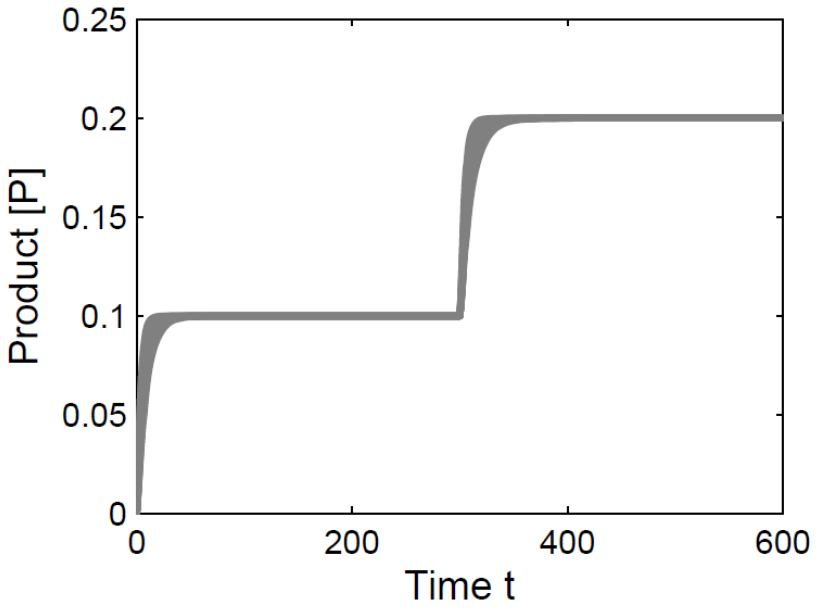
- **Enzyme subset:** Set of reactions that (in steady state) always operate together in a fixed ratio → Rows in $\boldsymbol{K}$ differ only by scalar factor.

  only influx has to be regulated and thus, the output is subsequently regulated too

- How should enzymes in a subset be regulated?

Example Revisited: Enzyme Subsets

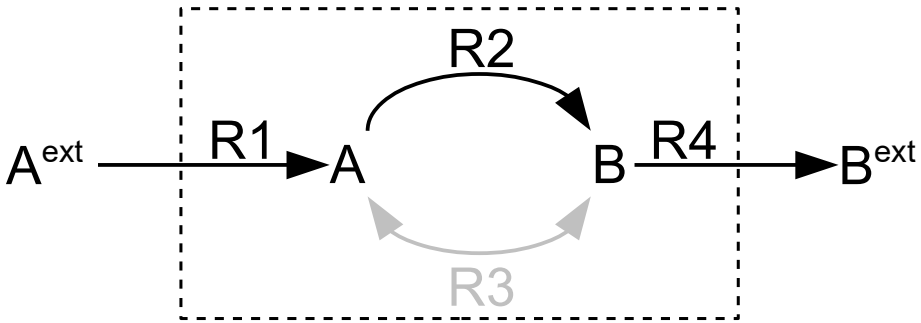




- In a linear pathway, all enzymes form a subset and all steady-state fluxes have to be equal →

$$k_1 u = k_2 [P] \Rightarrow [P] = \frac{k_1}{k_2} u$$

The Kernel Matrix Has Limitations



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

$$\boldsymbol{S} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{bmatrix} \qquad \boldsymbol{rev} = \{R3\} \qquad \boldsymbol{irrev} = \{R1, R2, R4\}$$

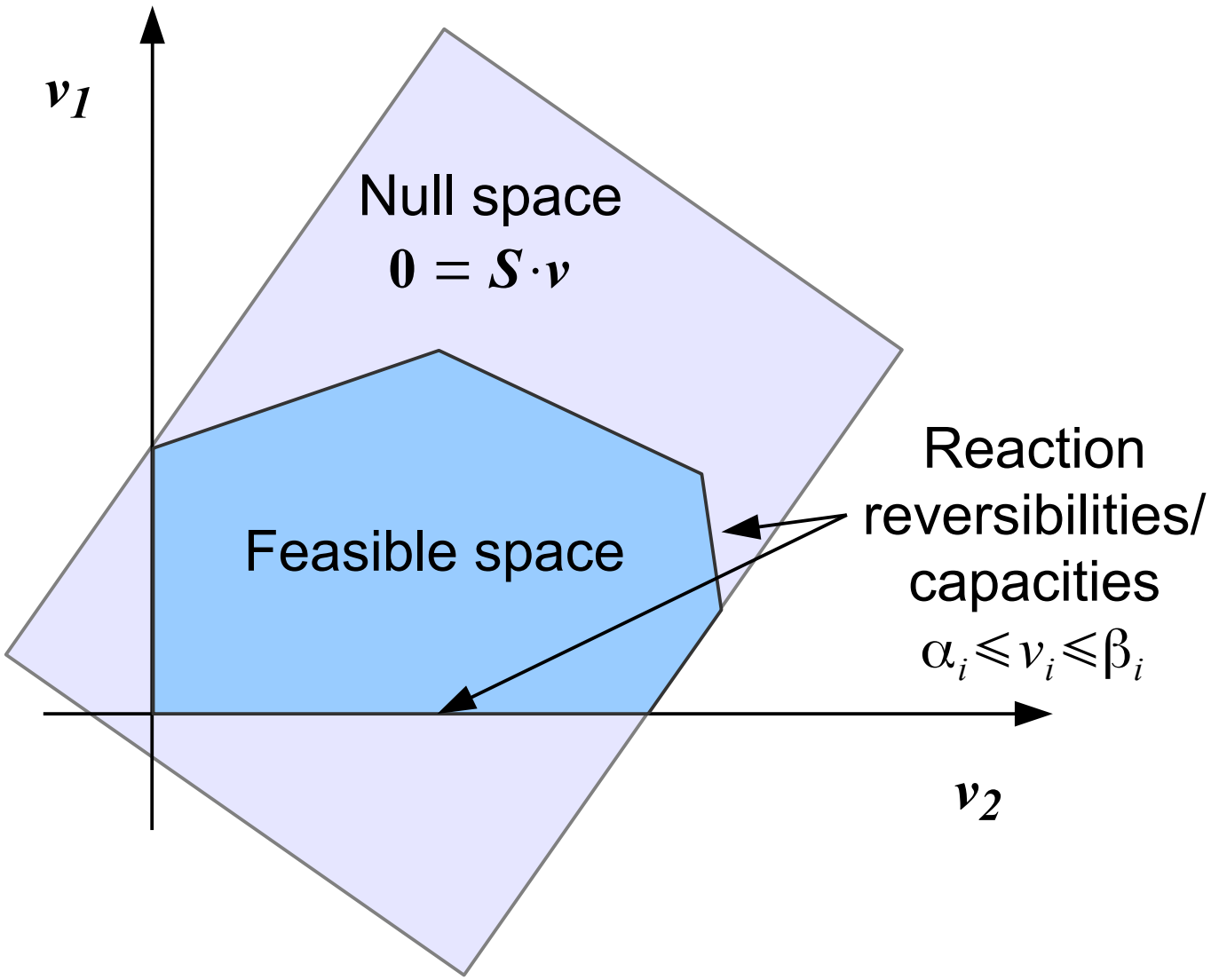
$$\boldsymbol{b} = \begin{pmatrix} 1 & -1 \end{pmatrix}$$

$$\boldsymbol{v}^{\boldsymbol{T}} = \begin{pmatrix} 1 & 1 & 0 & 1 \end{pmatrix}$$

- Limitation: Kernel matrix is not a unique representation.
- Limitation: Reaction reversibilities are **not** considered.

How to incorporate reversibilities?

Constraints Do Not Sufficiently Shrink The Solution Space



Systembiologie 551-1174-00L

Flux Balance Analysis

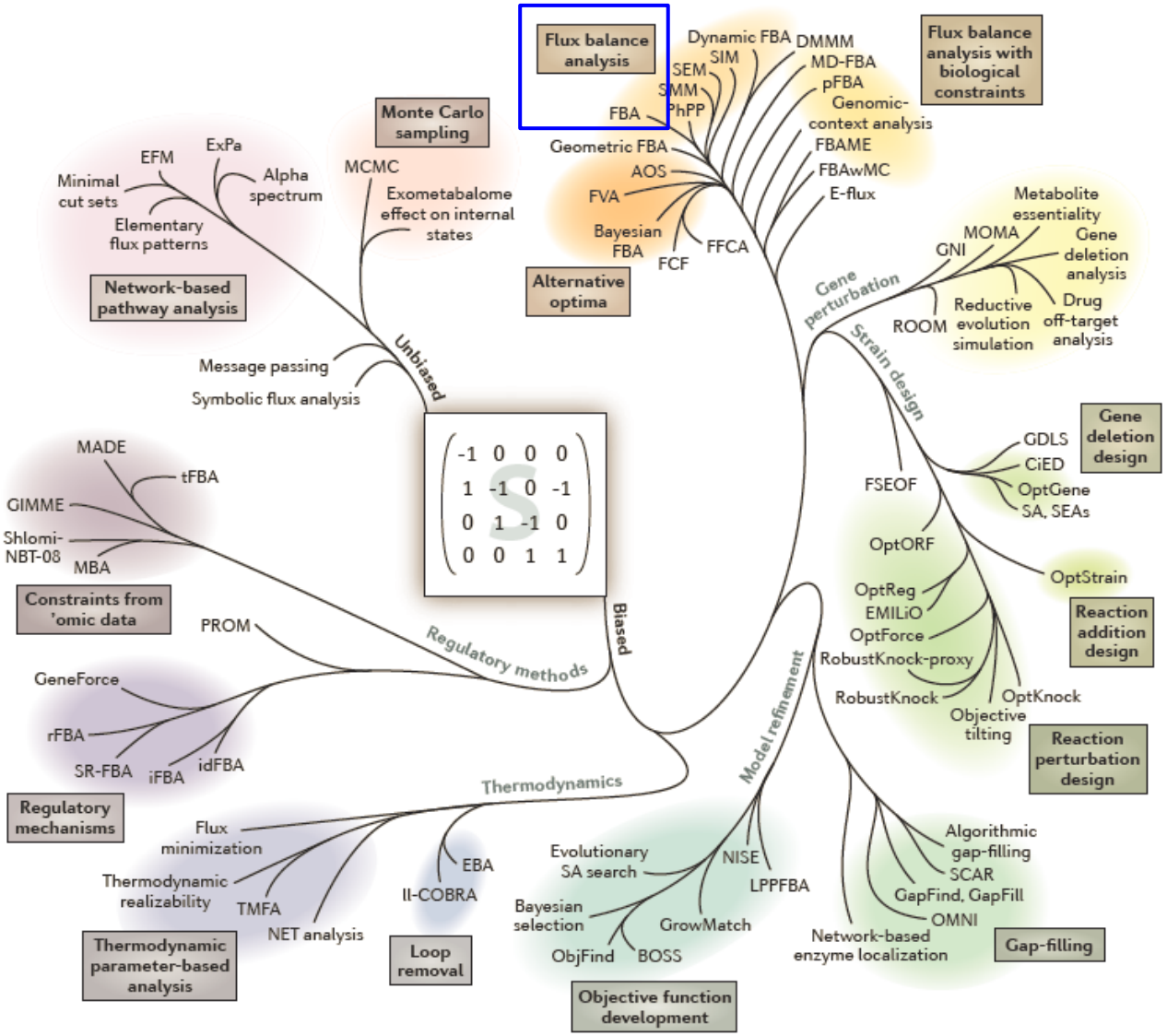
23 March, 2017

Jörg Stelling, D-BSSE

Content:

- From textbook biochemistry to genome-scale stoichiometric models
- The solution space of stoichiometric models
- Flux balance analysis

Constraint-Based Methods: Flux Balance Analysis



N.E. Lewis et al., *Nat. Rev. Microbiol.* 10: 291 (2012).

Flux Balance Analysis (FBA): Principles

- **Idea so far:** Incorporate more constraints into the model (reversibilities) → Smaller solution space → More predictive model.

- **Biological assumption of FBA:** The cell aims to achieve an **optimal** feasible steady state flux distribution.

What could be an optimal cell state?

1) maximal growth rate
2) minimal energy consumption
etc.

Flux Balance Analysis: Formally

- Quasi steady state assumption: $\mathbf{0} = \boldsymbol{S} \cdot \boldsymbol{v}$
- Reaction reversibilities / capacities: $\alpha_i \leqslant v_i \leqslant \beta_i$
- Optimal feasible steady state: $\boldsymbol{w}^T \cdot \boldsymbol{v} \rightarrow max\,!$

  where the weights $\boldsymbol{w}$ encode the assumed cellular objective function such as maximal growth rate: $w_{growth} = 1\,, \quad w_{all\ other\ fluxes} = 0$

Flux Balance Analysis: Optimization

- General optimization problem statement for FBA:

$$\Phi(\boldsymbol{v}) = \boldsymbol{w}^{T} \cdot \boldsymbol{v} \rightarrow max\,!$$

problem is linearized

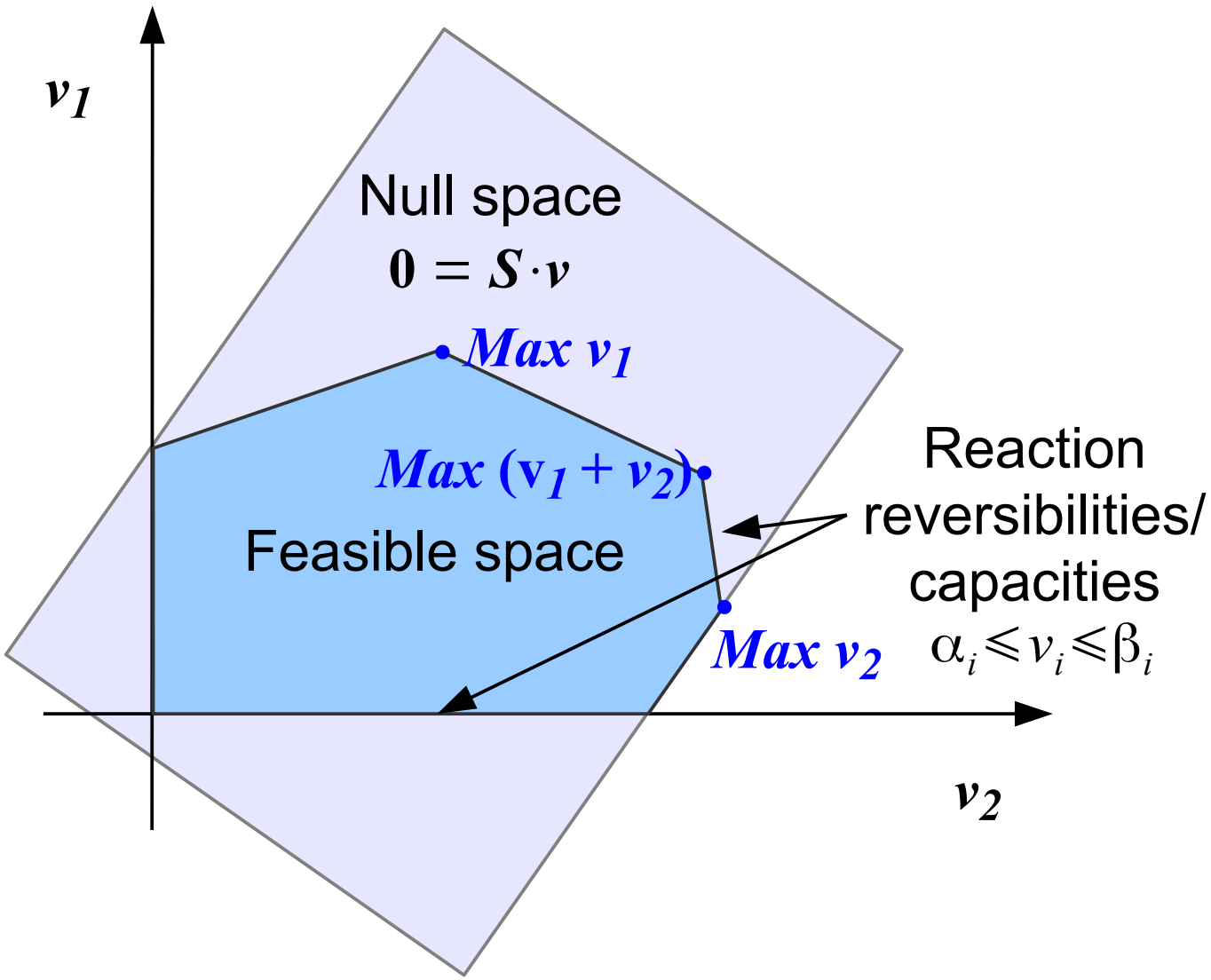
$$s.t.$$

$$\boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{0}$$

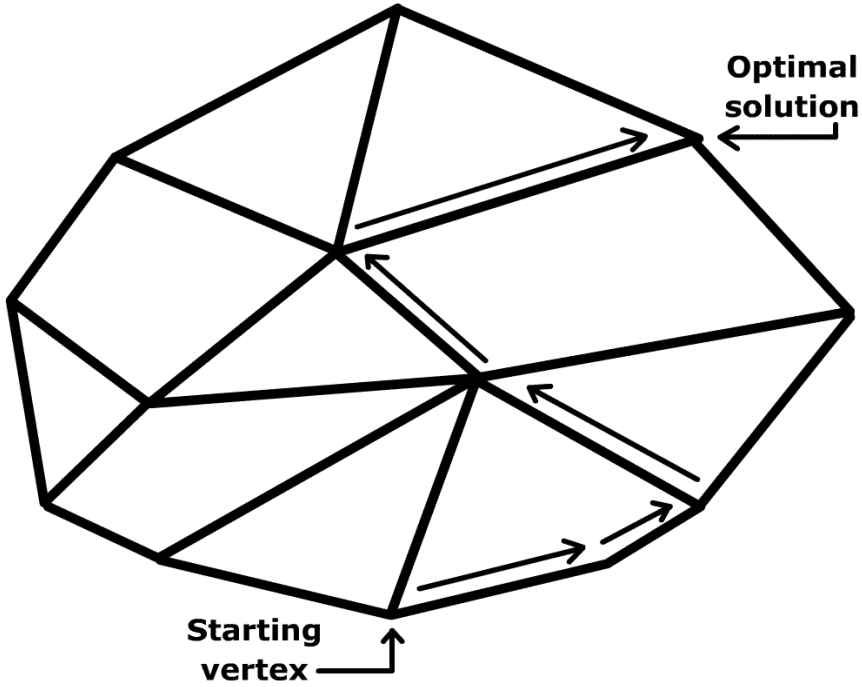
constraints: $\alpha_i \leqslant v_i \leqslant \beta_i$

- Objective function and constraints are **linear** in the unknown fluxes $\boldsymbol{v}$ → **Linear program** → Computationally efficient solution even for large models.

Flux Balance Analysis: Optimization



Flux Balance Analysis: Optimization



- **Simplex algorithm:** Solutions lie on vertices → Start on vertex → Evaluate gradients and move along edges → Continue search or stop at optimal solution.

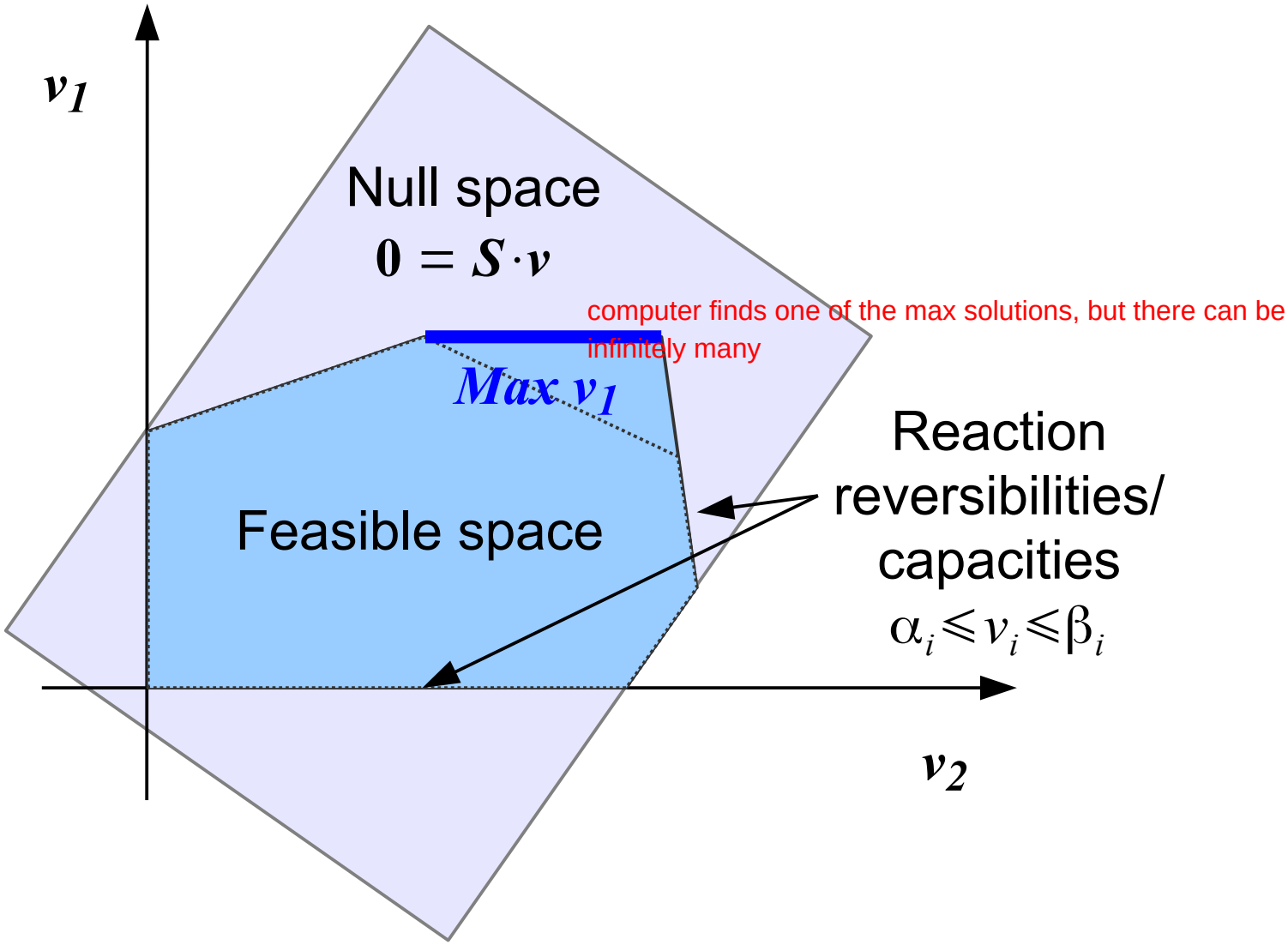
Caveat #1: Objective Function

- FBA results strongly depend on objective function.
- Use of 'natural' objective functions such as growth:
  - Not applicable to all organisms (e.g. cells in multicellular organisms → cancer ...).
  - Not applicable to all conditions (e.g. genetically engineered strain that did not evolve)..
- Alternative / conflicting approaches to optimization.

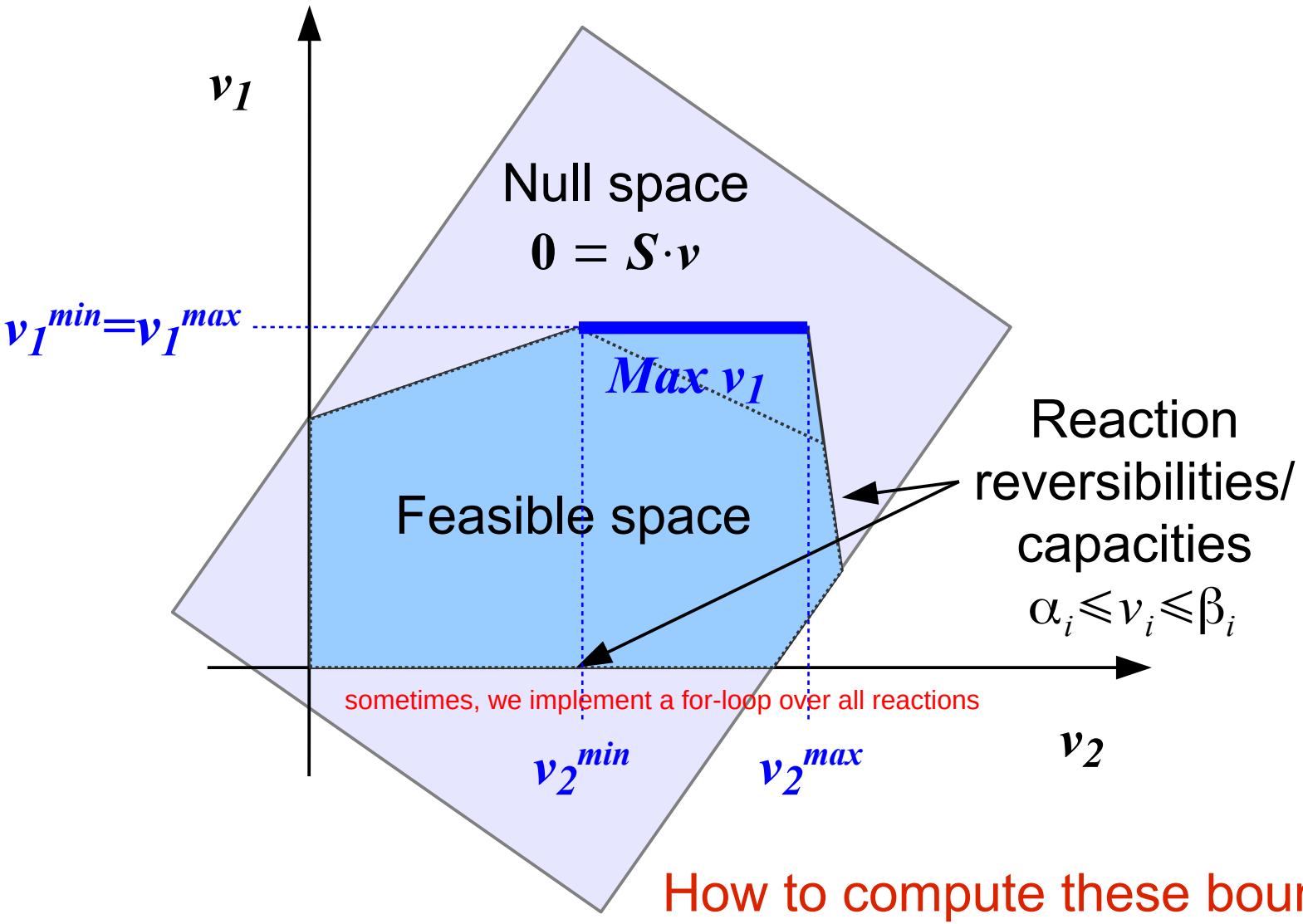
Caveat #2: Alternate Optima

- Linear programming problem implies that finding a solution can be guaranteed:
  - **Unique value** of the objective function ('growth'),
  - **But infinitely many** flux distributions with optimal value of objective function may exist.
- Without incorporating further constraints: Poor performance in predicting flux distributions.

Caveat #2: Alternate Optima



Approach: Flux Variability Analysis (FVA)



How to compute these bounds?

Flux Variability Analysis (FVA): Pseudocode

Inputs: Stoichiometric matrix $\boldsymbol{S}$ $(m \times n)$
Flux bound vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ $(n \times 1)$

Output: Matrix of minimal and maximal fluxes $\boldsymbol{F}$ $(n \times 2)$

Set $\boldsymbol{F}$ to $n \times 2$ zero matrix // initialization

For d = 1 to 2 // loop over the two directions

  For z = 1 to n // loop over all reactions

    Set $\boldsymbol{w}$ to $n \times 1$ zero vector

    Set w_z to $(-1)^d$

    Solve linear program

that's FBA I think

$$\boldsymbol{w}^T \cdot \boldsymbol{v} \rightarrow max!$$
$$s.t.$$
$$\boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{0}$$
$$\alpha_i \leqslant v_i \leqslant \beta_i$$

    Set $F_{z,d}$ to v_z

Flux Variability Analysis (FVA): Pseudocode

Inputs: Stoichiometric matrix and upper / lower bounds for the fluxes

Output: Minimal and maximal fluxes

```
For each of the reaction directions
        For each reaction as a target
                Set weights for FBA such that the target
                        reaction is the objective in the correct
                        direction
                Run FBA
                Save optimized flux for the target reaction
        End
End
```

Flux Variability Analysis (FVA): Pseudocode

that's fucking bullshit and wont give you any points in the exam

Load model
Compute something to find minimal and maximal fluxes
Plot results
Save data

This is NOT pseudocode

Summary: Teaching Goals II/III

- Network structure constrains network behavior but solutions (flux distributions) are most often not unique.
- Steady-state analysis relies on linear algebra: Null space, kernel matrix, ... → Feasible flux distributions, enzyme subsets, ...
- Flux balance analysis (FBA) adds optimality assumption in a linear (computationally efficient) optimization problem → Many applications (next week).

Exercise 6: Flux Balance Analysis

Goal

- Learning the basic ingredients of Flux Balance Analysis
- Simulate and interpret flux distributions in a toy metabolic model