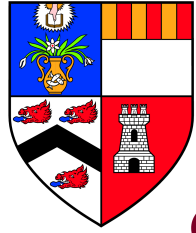


1 4 9 5



UNIVERSITY
OF ABERDEEN



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

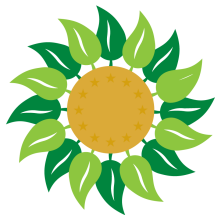


CEPLAS

Cluster of Excellence on Plant Sciences

Linear programming applied to genome-scale metabolic network models

Oliver Ebenhöf



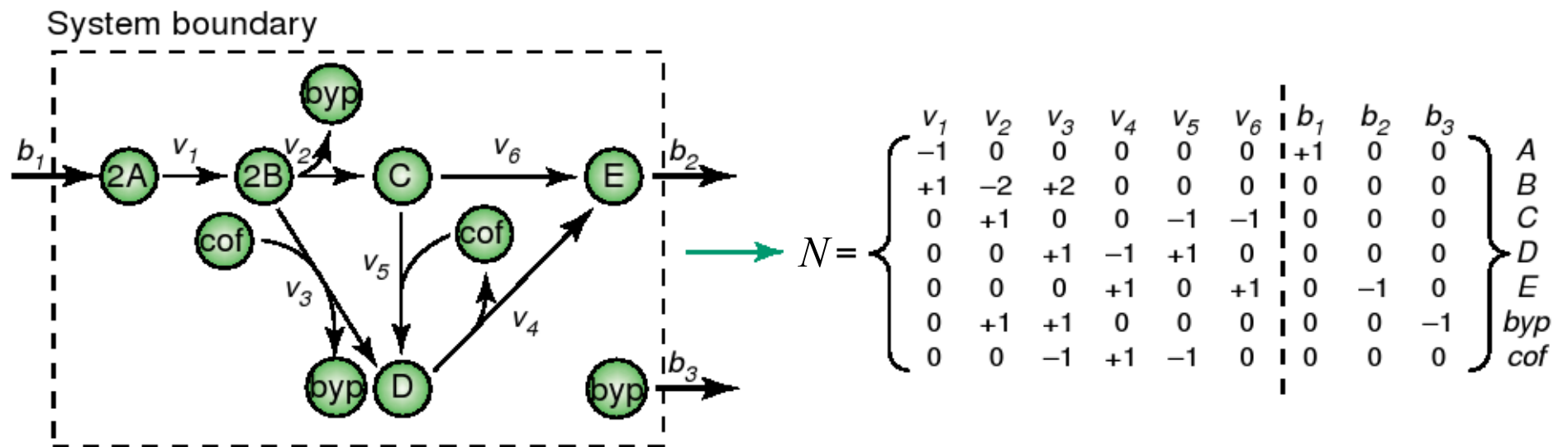
AccliPhot

www.accliphot.eu

Rigi Workshop, January 19, 2015

Stoichiometric modelling

A metabolic system is defined by **internal reactions** and **exchange fluxes**



The temporal change of the concentrations is given by

$$\frac{dX}{dt} = N \cdot v$$

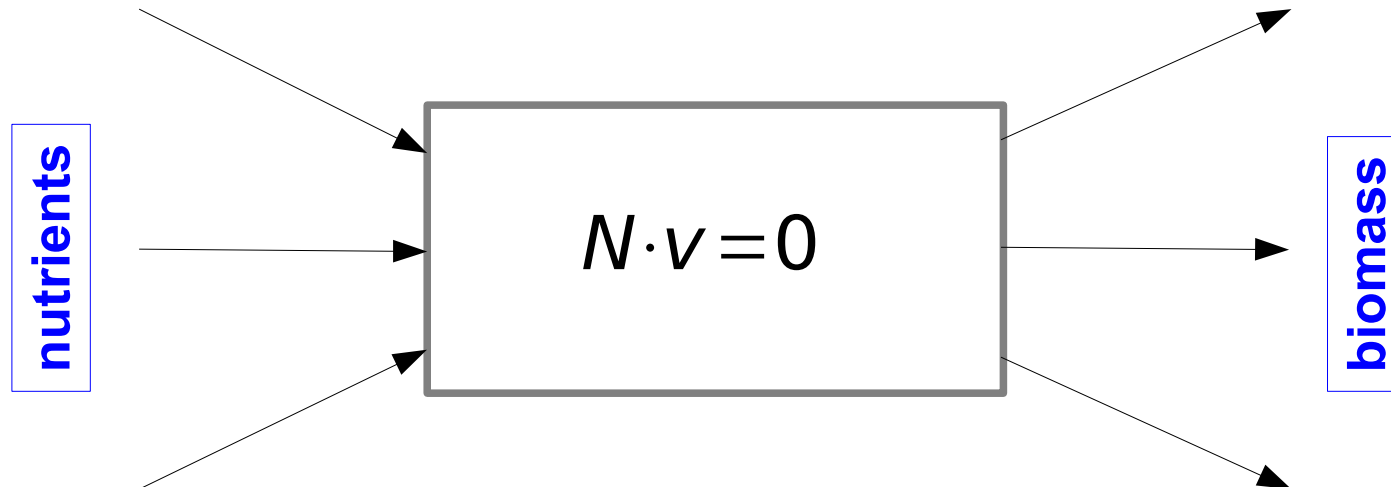
Steady state is characterised by

$$N \cdot v = 0$$

$$N \cdot v = 0$$

Problem: This equation has many solutions!

Which one is correct?



- We know what goes in
- We know what goes out
- We do not know what is happening inside!

Constraints allow to reduce the number of possible solutions

Optimisation allows to find special fluxes (and answer 'what if...?')

Linear Programming

Linear Programming (LP) is an optimisation technique

Identify variable values which result in a maximal (or minimal) value of a function which linearly depends on the parameters under given constraints

Introduction

General idea: optimise a linear function under inequality constraints

Variables: $x_i, \quad i=1 \dots N$

Constraints: $l_i \leq x_i \leq u_i$

Objective: $\Omega = \sum_i^N c_i \cdot x_i$

EXAMPLES

Example 1: Maximising profit

A farmer has 10 acres to plant in wheat and rye. He has to plant at least 7 acres. However, he has only \$1200 to spend and each acre of wheat costs \$200 to plant and each acre of rye costs \$100 to plant. Moreover, the farmer has to get the planting done in 12 hours and it takes an hour to plant an acre of wheat and 2 hours to plant an acre of rye. If the profit is \$500 per acre of wheat and \$300 per acre of rye how many acres of each should be planted to maximize profits?

Steps to solve the problem:

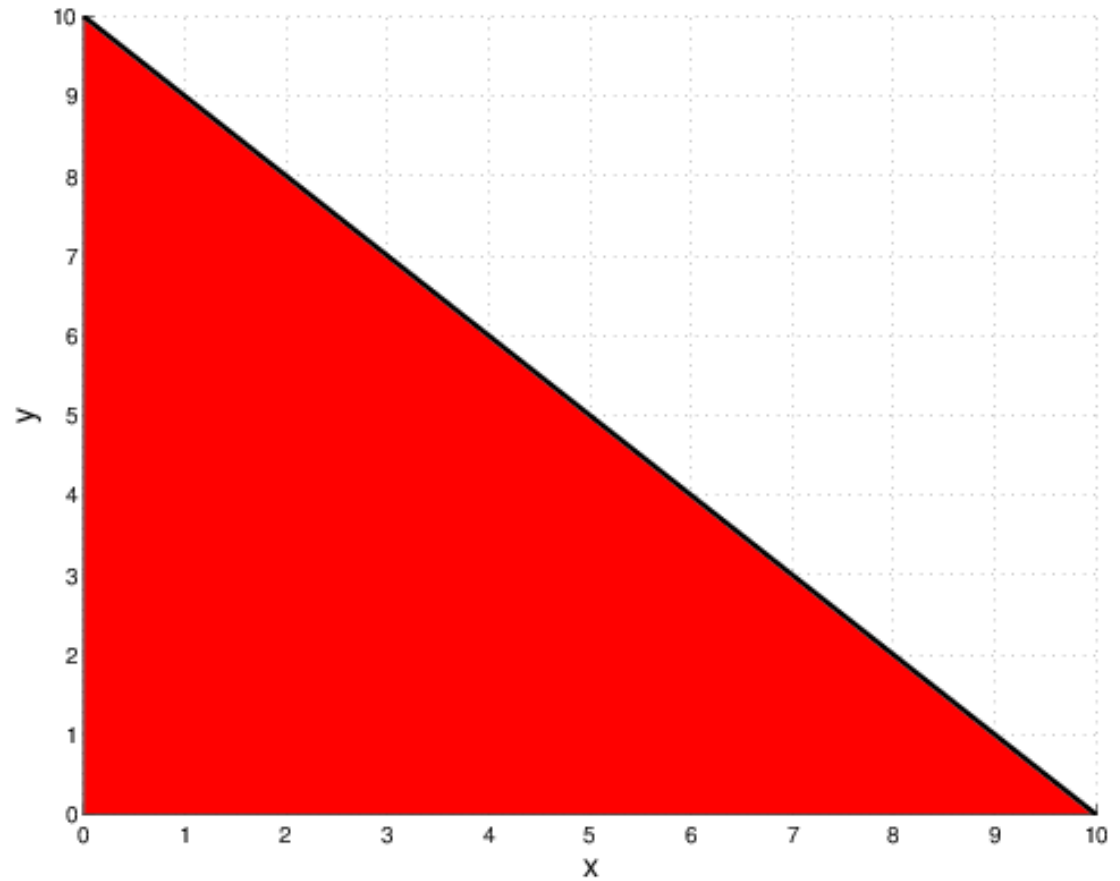
0. Read the whole problem.
1. Define your unknowns.
2. Express the objective function and the constraints.
3. Graph the constraints.
4. Find the corner points to the region of feasible solutions.
5. Evaluate the objective function at all the feasible corner points.

see http://www.sonoma.edu/users/w/wilsonst/courses/math_131/lp/

Example 1: Maximising profit

Total land:

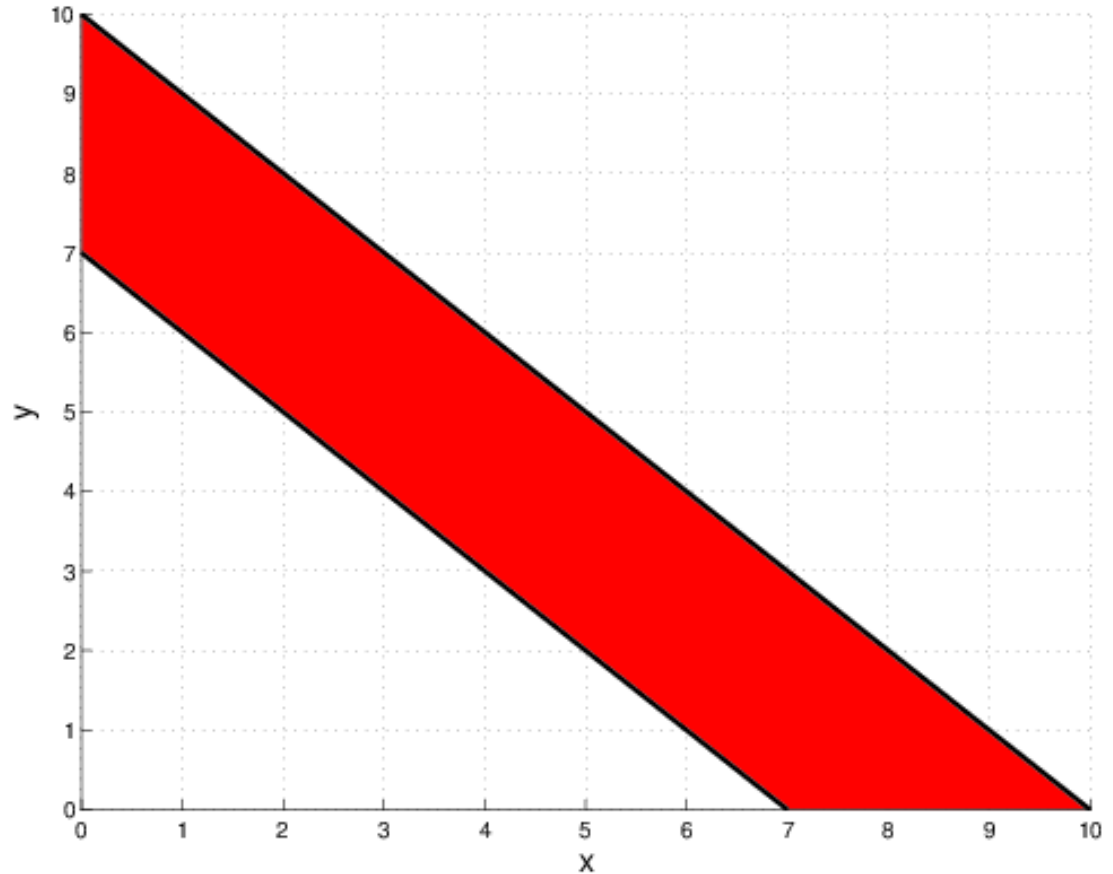
$$x + y \leq 10$$



Example 1: Maximising profit

Total land: $x + y \leq 10$

Minimum planting: $x + y \geq 7$

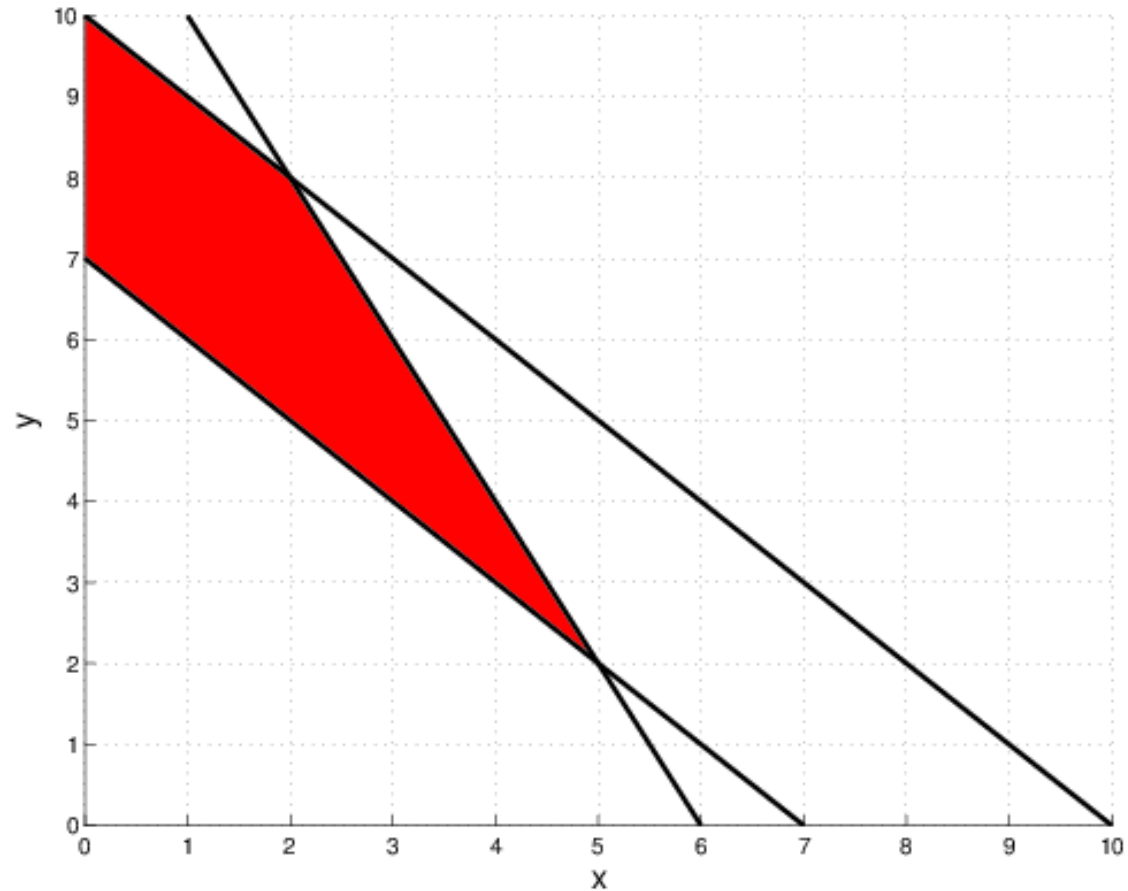


Example 1: Maximising profit

Total land: $x + y \leq 10$

Minimum planting: $x + y \geq 7$

Limited funds: $200x + 100y \leq 1200$



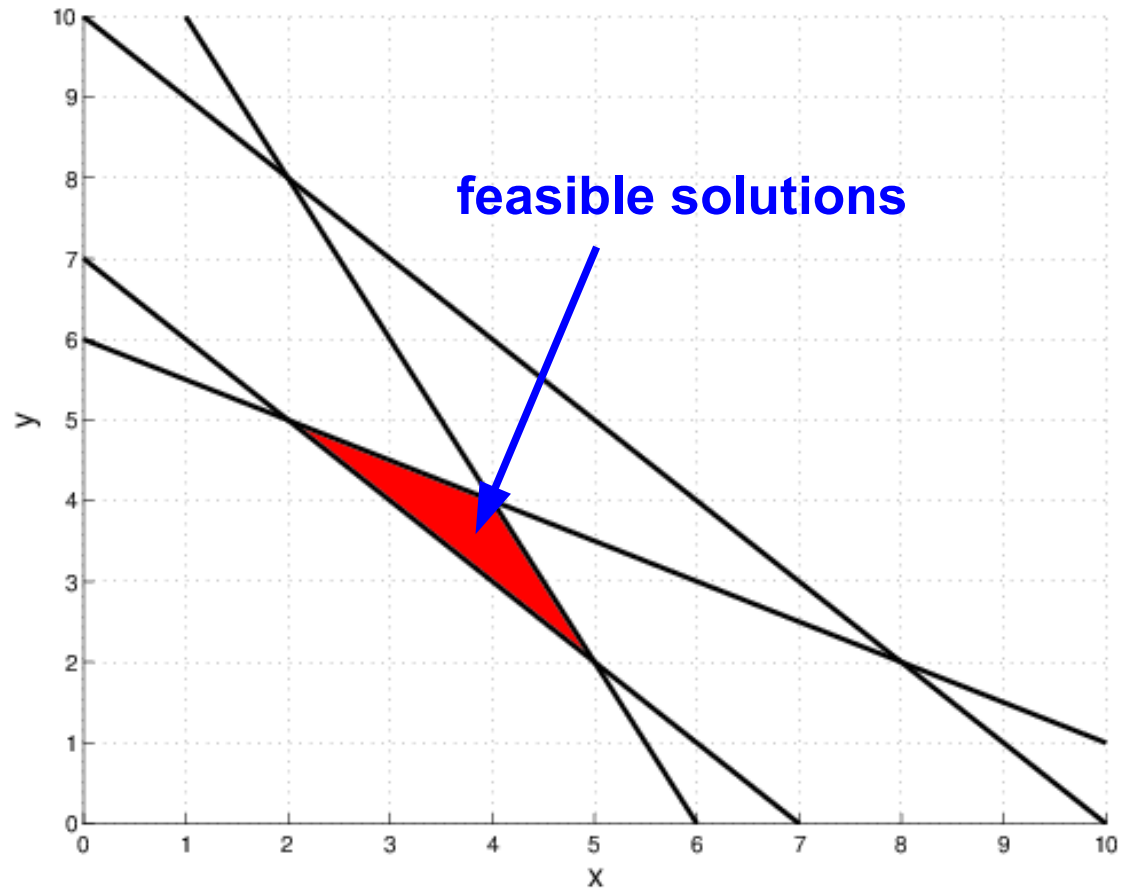
Example 1: Maximising profit

Total land: $x + y \leq 10$

Minimum planting: $x + y \geq 7$

Limited funds: $200x + 100y \leq 1200$

Limited time: $x + 2y \leq 12$



Which one of the feasible solutions gives the most profit?

Profit: $\Omega = 500x + 300y$

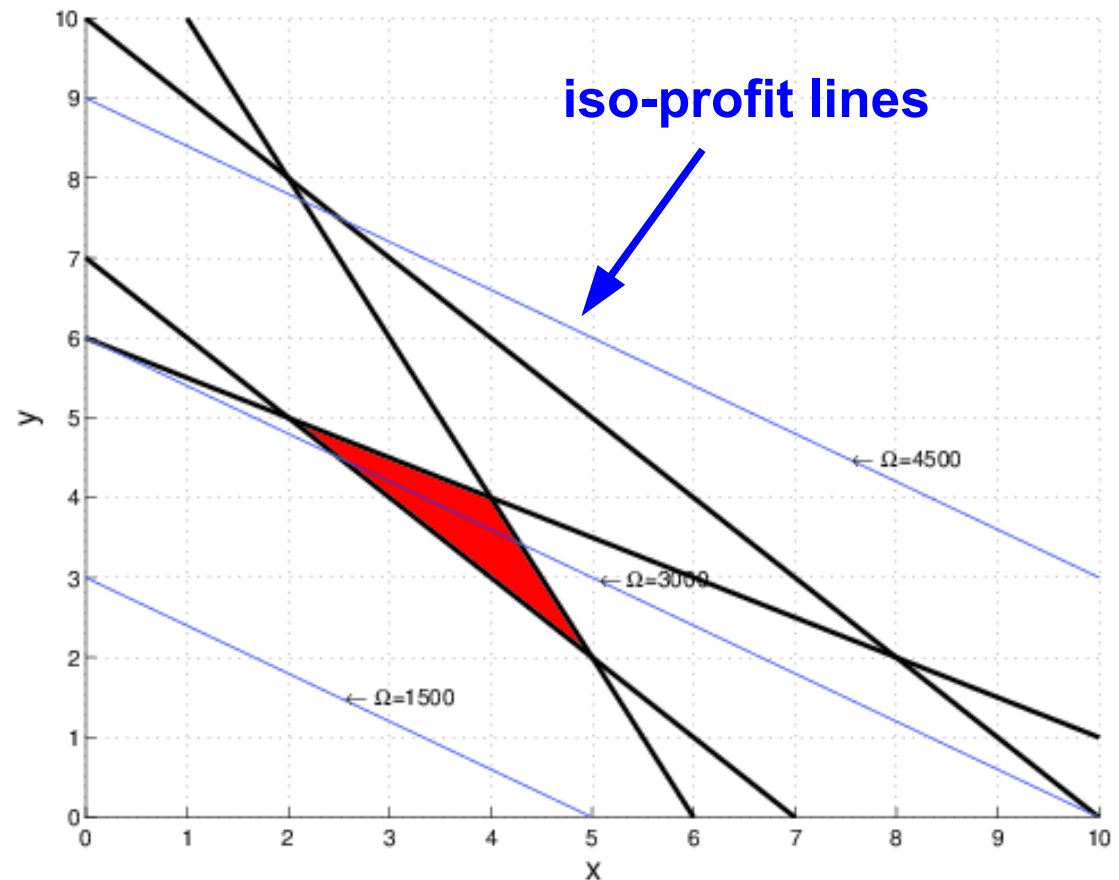
Example 1: Maximising profit

Total land: $x + y \leq 10$

Minimum planting: $x + y \geq 7$

Limited funds: $200x + 100y \leq 1200$

Limited time: $x + 2y \leq 12$



Which one of the feasible solutions gives the most profit?

Profit: $\Omega = 500x + 300y$

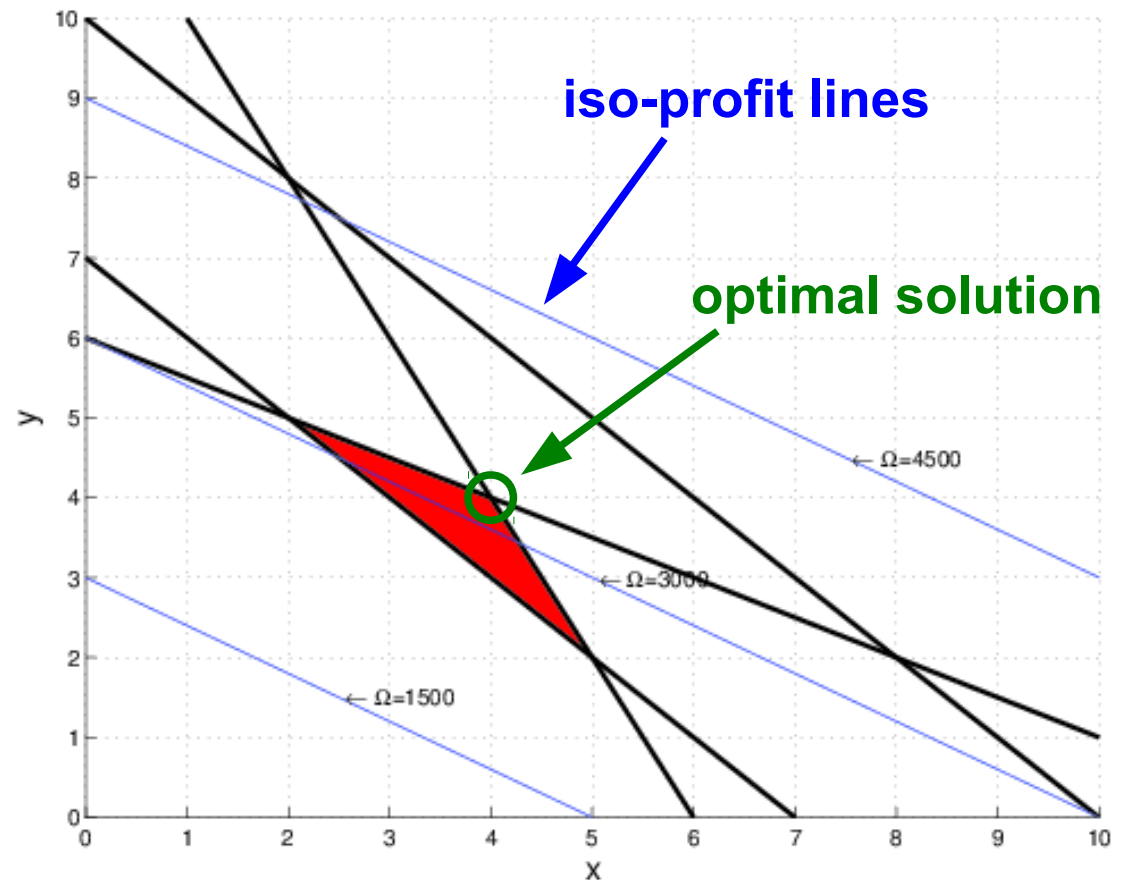
Example 1: Maximising profit

Total land: $x + y \leq 10$

Minimum planting: $x + y \geq 7$

Limited funds: $200x + 100y \leq 1200$

Limited time: $x + 2y \leq 12$



Maximal profit at: $x=4, y=4$

- In general:
- space of feasible solution is a convex polyhedron
 - optimal solution is always at a vertex

Example: Optimising happyness

A week has 168 hours

- We need time to study (S), party (P) and for everything else (E – incl. sleep, eat)
- To survive, we need at least 8h rest per day: $E \geq 56$
- To maintain sanity, we need to party or rest a bit more: $P+E \geq 70$
- To pass exams, we need to study at least 60h/week: $S \geq 60$
- But longer if we don't sleep enough or party too much: $2S+E-3P \geq 150$

(this means, for every missed hour of sleep, we need to study 30 min longer and for every hour partying, we need to study 1.5h longer because of hangovers)

Objective: Maximise happyness, expressed by $\Omega = 2P+E$
(extra rest makes happy, partying makes twice as happy)

Example: Optimising happyness

The problem:

maximise $\Omega = 2P + E$

under the constraints that

(week) $S + P + E = 168$

(survival) $E \geq 56$

(min. study) $S \geq 60$

(hangovers) $2S + E - 3P \geq 150$

(sanity) $P + E \geq 70$

(no negative times) $P \geq 0$

Eliminate S

$$S = 168 - P - E$$

$$168 - P - E \geq 60 \Leftrightarrow P + E \leq 108$$

$$2(168 - P - E) + E - 3P \geq 150 \Leftrightarrow E + 5P \leq 186$$

$$P + E \geq 70$$

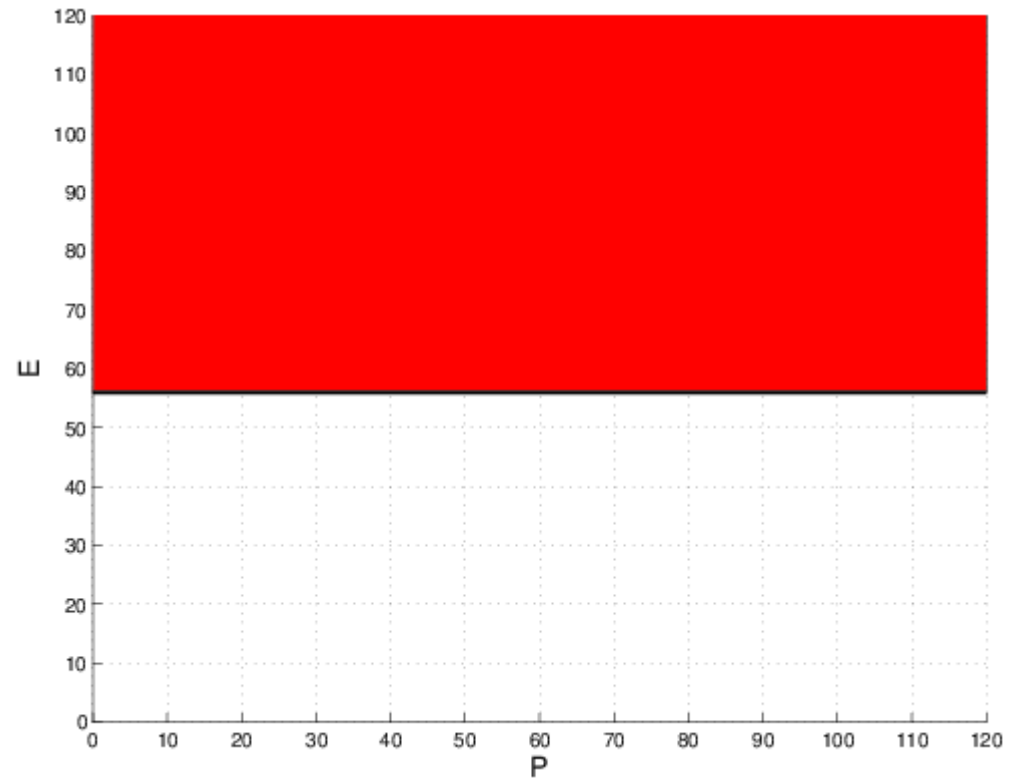
$$P \geq 0$$

$$E \geq 56$$

Example: Optimising happiness

Survival:

$$E \geq 56$$



Example: Optimising happiness

Survival: $E \geq 56$

minimal study: $P + E \leq 108$

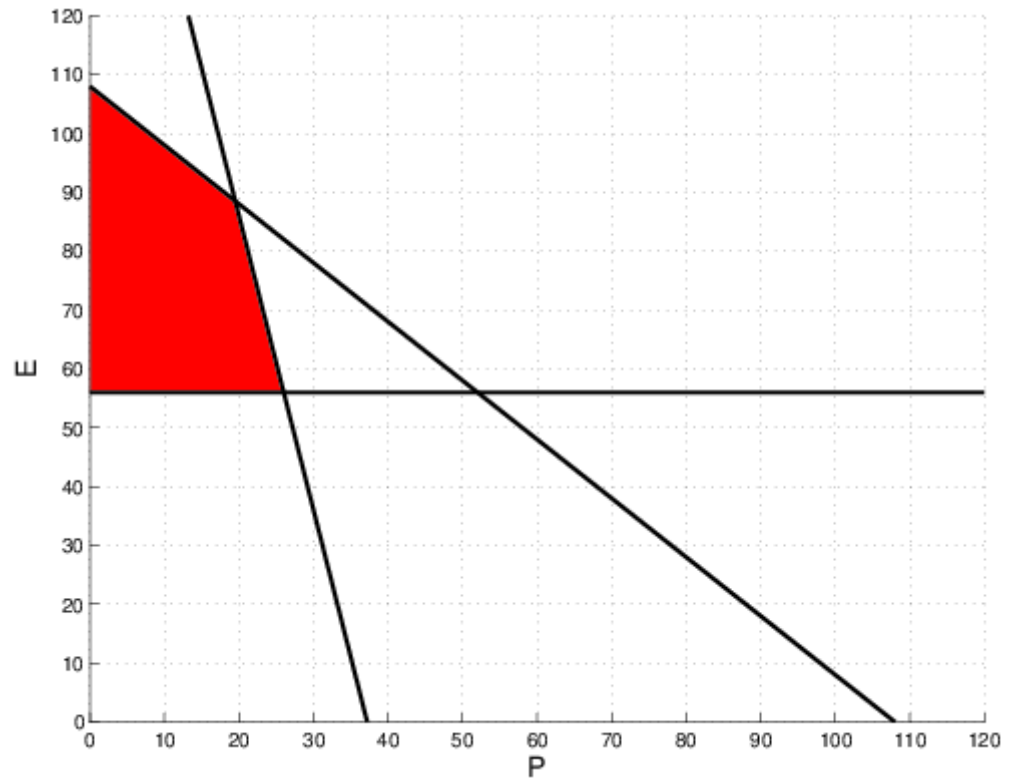


Example: Optimising happyness

Survival: $E \geq 56$

minimal study: $P + E \leq 108$

hangovers: $E + 5P \leq 186$



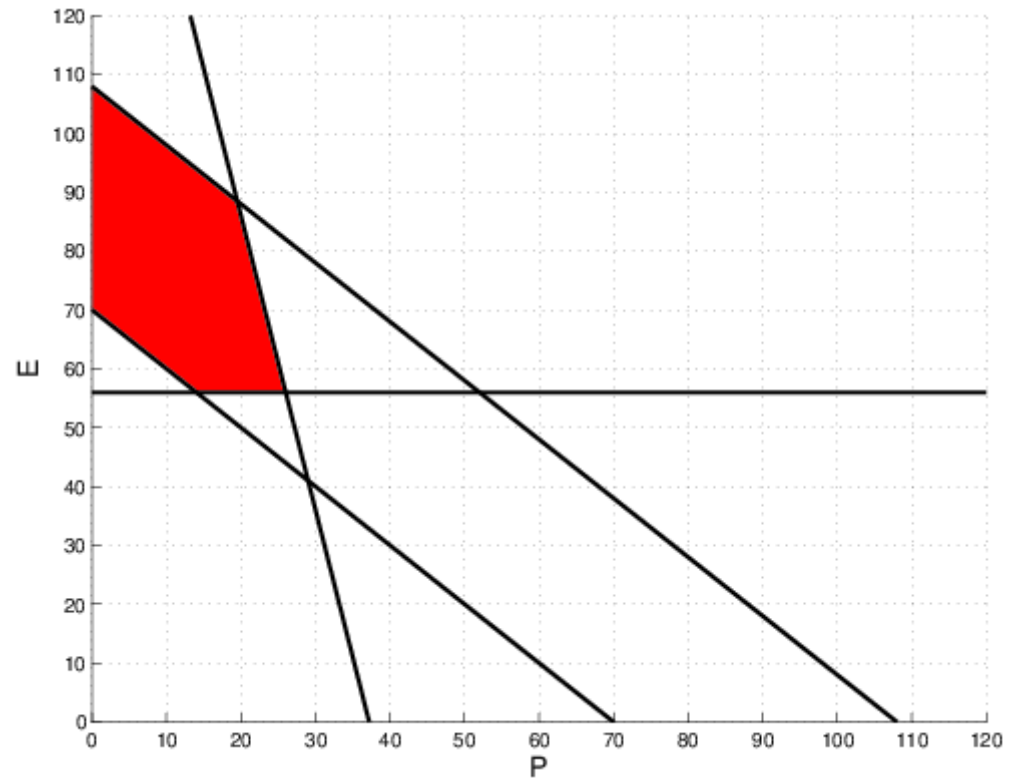
Example: Optimising happyness

Survival: $E \geq 56$

minimal study: $P + E \leq 108$

hangovers: $E + 5P \leq 186$

sanity: $P + E \geq 70$



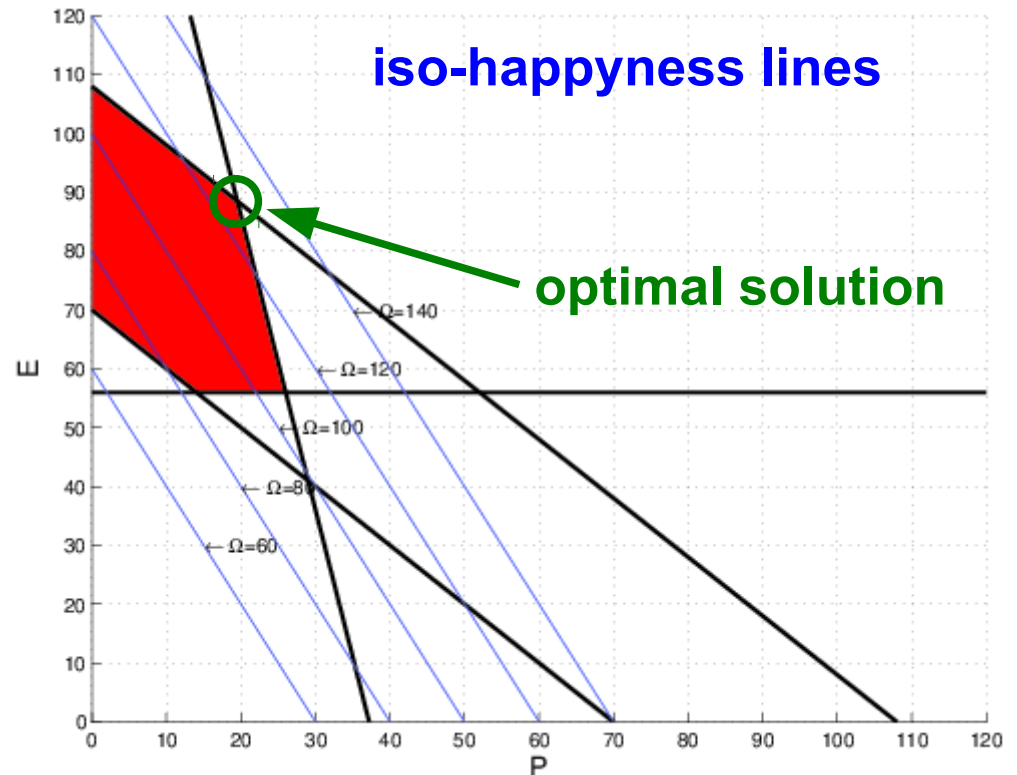
Example: Optimising happiness

Survival: $E \geq 56$

minimal study: $P + E \leq 108$

hangovers: $E + 5P \leq 186$

sanity: $P + E \geq 70$



Calculate optimal solution by finding intersection of two lines:

$$186 - 5P = 108 - P \Leftrightarrow P = 19.5$$

$$E = 108 - P \Rightarrow E = 88.5$$

$$S = 168 - P - E \Rightarrow S = 60$$

Application to metabolic networks

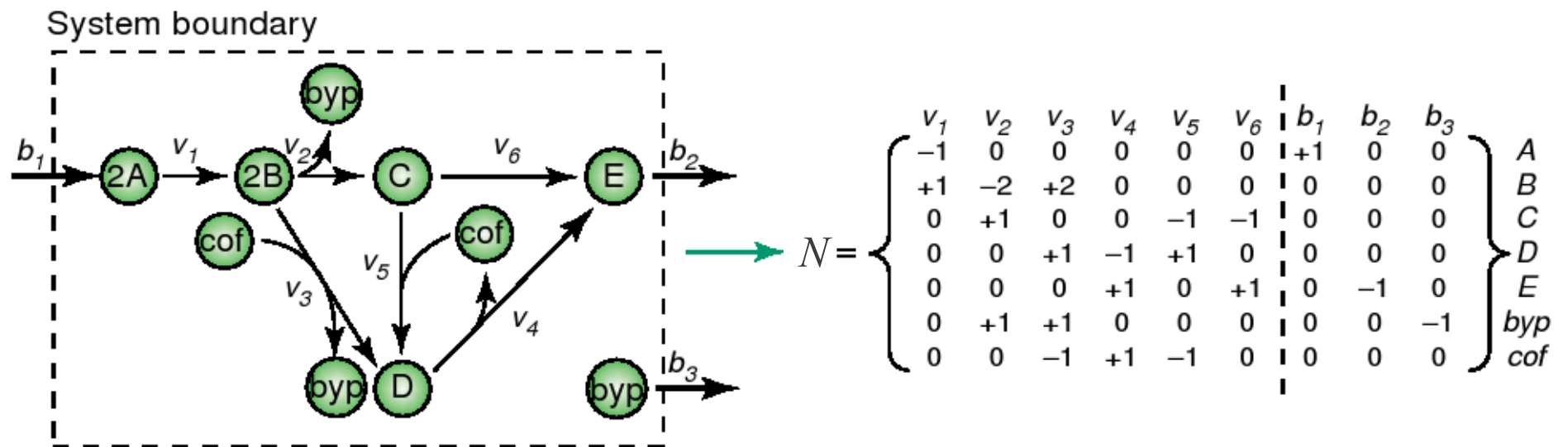
Variables: *FLUXES* $v_i, \quad i=1 \dots R$

Constraints: ?

Objective: ?

Application to metabolic networks

A metabolic system is defined by **internal reactions** and **exchange fluxes**



The temporal change of the concentrations is given by $\frac{dX}{dt} = N \cdot v$

Steady state is characterised by

$$N \cdot v = 0$$

Constraint #1

Other constraints

Because of thermodynamic reasons, some reactions can only proceed in one direction



$$K_{\text{eq}} = \frac{[\text{ADP}]_{\text{eq}} \cdot [\text{G6P}]_{\text{eq}}}{[\text{ATP}]_{\text{eq}} \cdot [\text{Glc}]_{\text{eq}}} = e^{-\Delta G^0/RT} = e^{10.05} = 23000$$

With $[\text{ATP}]/[\text{ADP}]=3$ and $[\text{Glc}]=1 \text{ mM}$ the reaction runs in reverse if

$$[\text{G6P}] > K_{\text{eq}} \cdot [\text{Glc}] \cdot \frac{[\text{ATP}]}{[\text{ADP}]} = 69000 \text{ mM} = 69 \text{ M} \quad !!!$$

(pure water has 55.5 M)

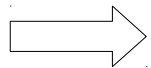


directionality implies $v_j \geq 0$ **Constraints #2**

Other constraints

Some process have upper bounds

- maximal uptake rates
- known maximal enzyme activities



limitation implies

$$v_j \leq v_j^{\max}$$

Constraints #3

What is constraint based modelling?

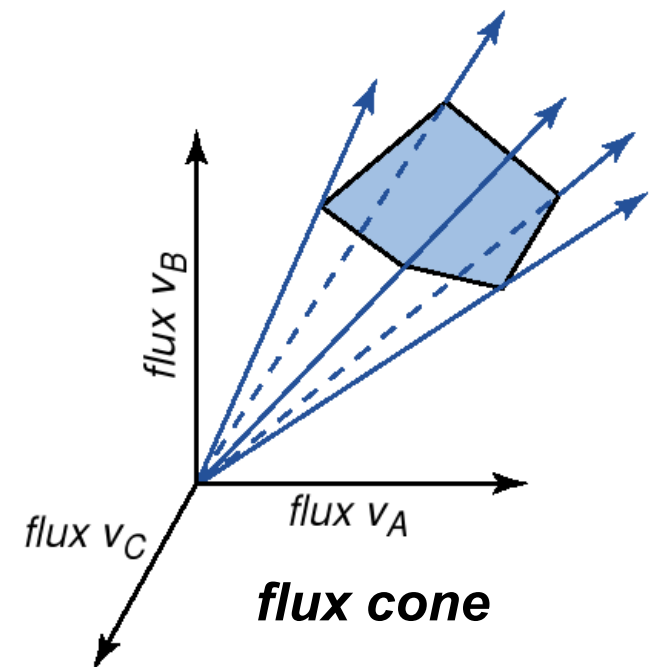
Fluxes in metabolic networks are subject to *constraints*

- Thermodynamic (directionality) $v_i \geq 0$
- Enzyme concentrations $v_i \leq v_{i,max}$

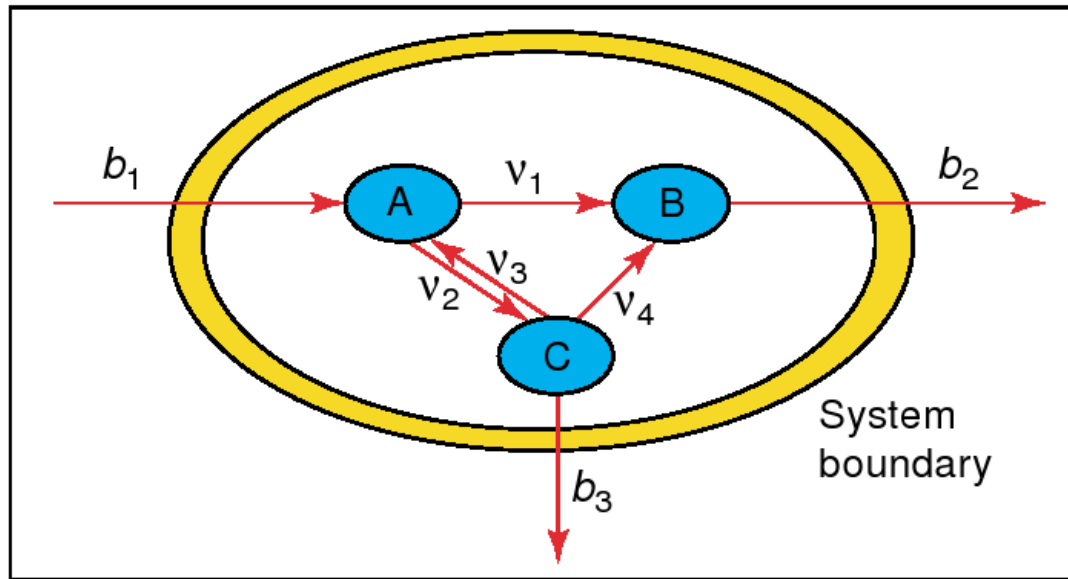
Constraint based models analyse steady state solutions which fulfill the given constraints.

Find a solution vector $v = (v_1, \dots, v_r)^T$ such that

$$N \cdot v = 0 \quad \text{and} \quad a_i \leq v_i \leq b_i$$



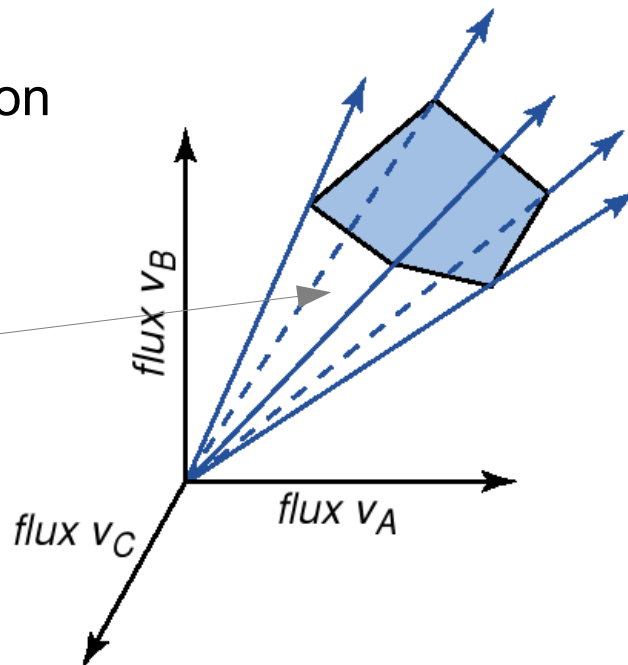
Example: constraint based model



(Kauffman et. al, 2003)

In general, the solution is a convex cone:

flux cone



constraints

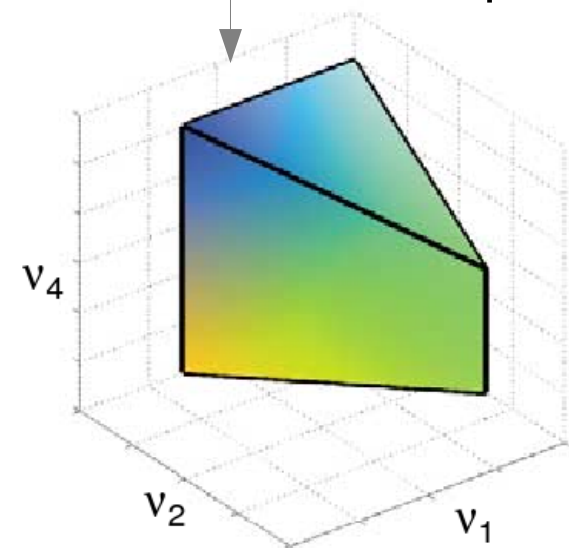
$v_3 = 0$ (thermodynamic)

$0 \leq b_1 \leq 1 \Rightarrow 0 \leq v_1 + v_2 \leq 1$

$0 \leq b_2 \leq 2 \Rightarrow 0 \leq v_1 + v_4 \leq 2$

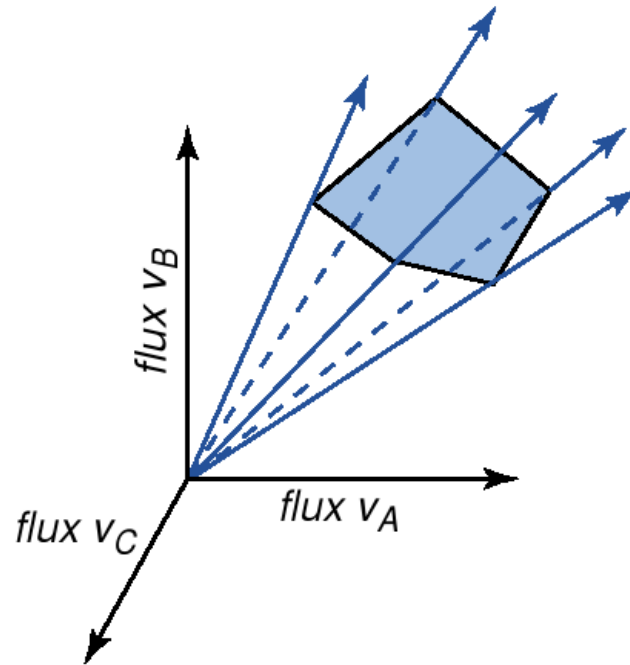
$0 \leq b_3 \Rightarrow 0 \leq v_2 - v_4$

Solution space



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

Which solution?



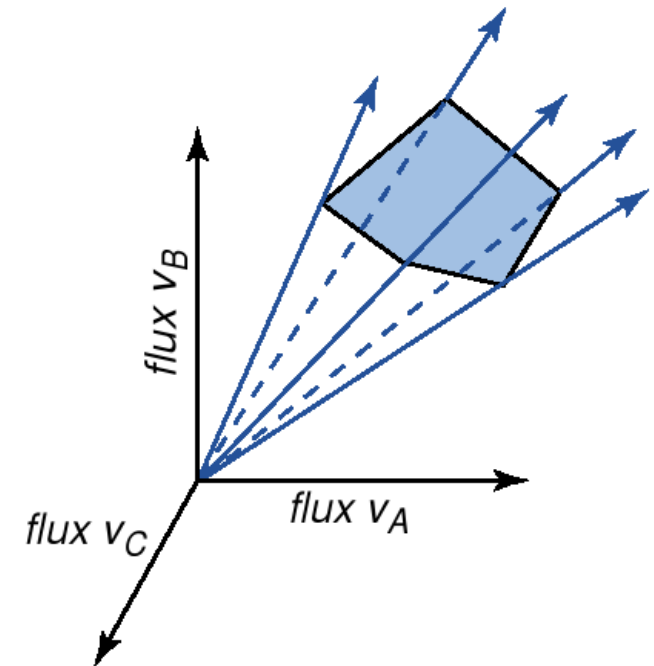
Application to metabolic networks

Variables: *FLUXES* $v_i, \quad i=1 \dots R$

Constraints: *Stationarity, maximal rates* $N \cdot v = 0, \quad 0 \leq v_i \leq v_i^{\max}$

Objective: *?*

The whole purpose of linear programming is to find one flux distribution from the solution cone which is “optimal”



What is optimal?

No general answer!

Plausible assumptions:

- maximal growth / biomass production
- most 'economic' solution (minimal enzyme usage)

Even if the objective is not 'correct', the computation is useful:

We can investigate the question “what if...?”

A typical LP problem maximising biomass

- assemble $r \times n$ stoichiometry matrix N (r reactions, n metabolites)
- identify irreversible reactions $R \subset \{1 \dots r\}$
- define boundary fluxes $B \subset \{1 \dots r\}$
- define “biomass reaction” $v_{\text{biomass}} : \sum_i \alpha_i \cdot S_i \rightarrow \text{biomass}$

Example from *E.coli* model (Feist et al, 2007)

(54.613) cpd00001 + (59.98) cpd00002 + (0.001787) cpd00003 + (0.000045) cpd00004 + (0.000335) cpd00005 + (0.000112) cpd00006 + (0.000168) cpd00010 + (0.01128) cpd00013 + (0.000223) cpd00015 + (0.000223) cpd00016 + (0.000223) cpd00017 + (0.000279) cpd00022 + (0.2557) cpd00023 + (0.000223) cpd00028 + (0.003008) cpd00030 + (0.5953) cpd00033 + (0.003008) cpd00034 + (0.4991) cpd00035 + (0.2091) cpd00038 + (0.3334) cpd00039 + (0.2342) cpd00041 + (0.000223) cpd00042 + (0.00376) cpd00048 + (0.2874) cpd00051 + (0.1298) cpd00052 + (0.2557) cpd00053 + (0.2097) cpd00054 + (0.000223) cpd00056 + (0.003008) cpd00058 + (0.1493) cpd00060 + (0.1401) cpd00062 + (0.004512) cpd00063 + (0.05523) cpd00065 + (0.18) cpd00066 + (0.134) cpd00069 + (0.000031) cpd00070 + (0.000098) cpd00078 + (0.08899) cpd00084 + (0.000223) cpd00087 + (0.004512) cpd00099 + (0.4378) cpd00107 + (0.02481) cpd00115 + (0.03327) cpd00118 + (0.0921) cpd00119 + (0.000223) cpd00125 + (0.2148) cpd00129 + (0.2342) cpd00132 + (0.003008) cpd00149 + (0.1542) cpd00155 + (0.4119) cpd00156 + (0.2465) cpd00161 + (0.000223) cpd00166 + (0.000223) cpd00201 + (0.1692) cpd00205 + (0.000223) cpd00216 + (0.000223) cpd00220 + (0.02561) cpd00241 + (0.007519) cpd00254 + (0.006744) cpd00264 + (0.2823) cpd00322 + (0.000223) cpd00345 + (0.02561) cpd00356 + (0.02481) cpd00357 + (0.000223) cpd00557 + (0.000055) cpd02229 + (0.000223) cpd03453 + (0.006767) cpd10515 + (0.006767) cpd10516 + (0.000223) cpd11313 + (0.003008) cpd11574 + (0.000223) cpd15353 + (0.002944) cpd15428[p] + (0.00229) cpd15429[p] + (0.00118) cpd15431[p] + (0.008151) cpd15432[e] + (0.000223) cpd15499 + (0.001345) cpd15501[p] + (0.000605) cpd15503[p] + (0.005381) cpd15505[p] + (0.005448) cpd15506[p] + (0.000673) cpd15508[p] + (0.0318) cpd15531[p] + (0.02473) cpd15532[p] + (0.01275) cpd15534[p] + (0.004897) cpd15538[p] + (0.003809) cpd15539[p] + (0.001963) cpd15541[p] + (0.000223) cpd15561 => (59.81) cpd00008 + (58.8062) cpd00009 + (0.7498) cpd00012 + (59.81) cpd00067 + cpd11416

- define upper bounds for uptake rates (boundary fluxes): $v_i \leq v_i^{\max}$ for $i \in B$

The LP-problem:

maximise
under the constraints

v_{biomass}

$N \cdot v = 0$

$v_i \leq v_i^{\max}$ for $i \in B$

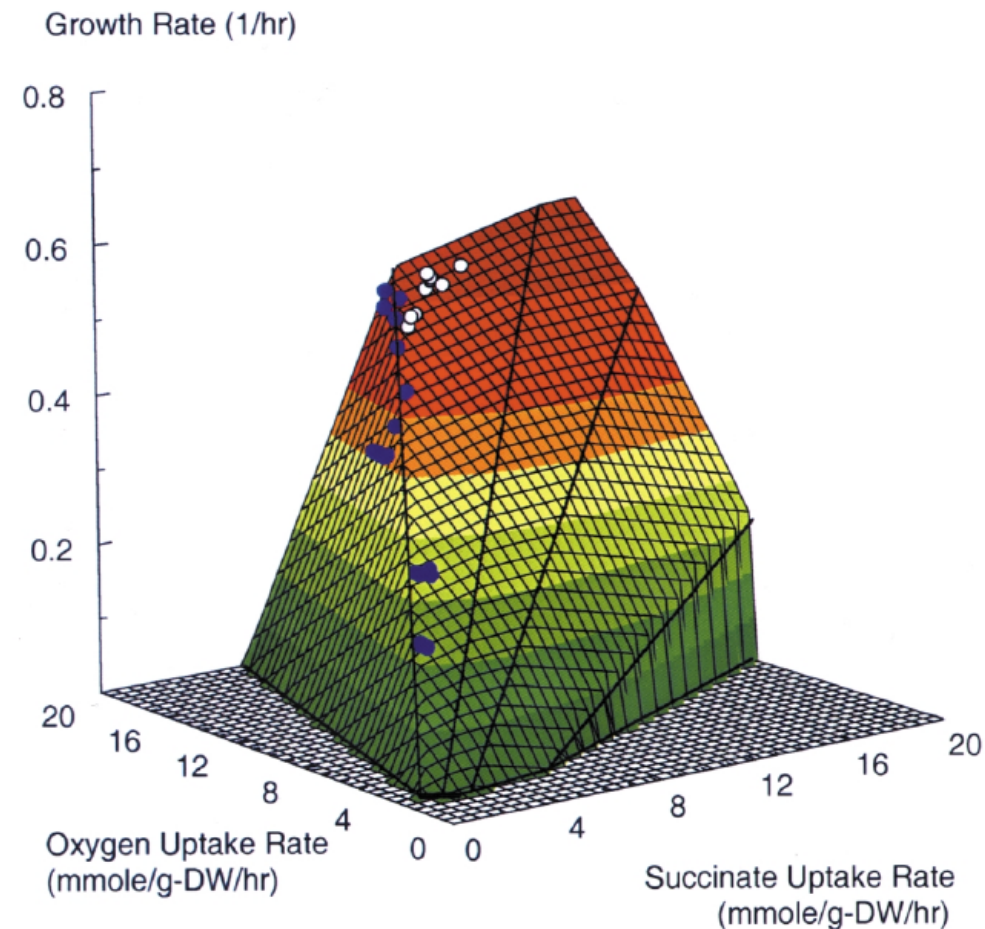
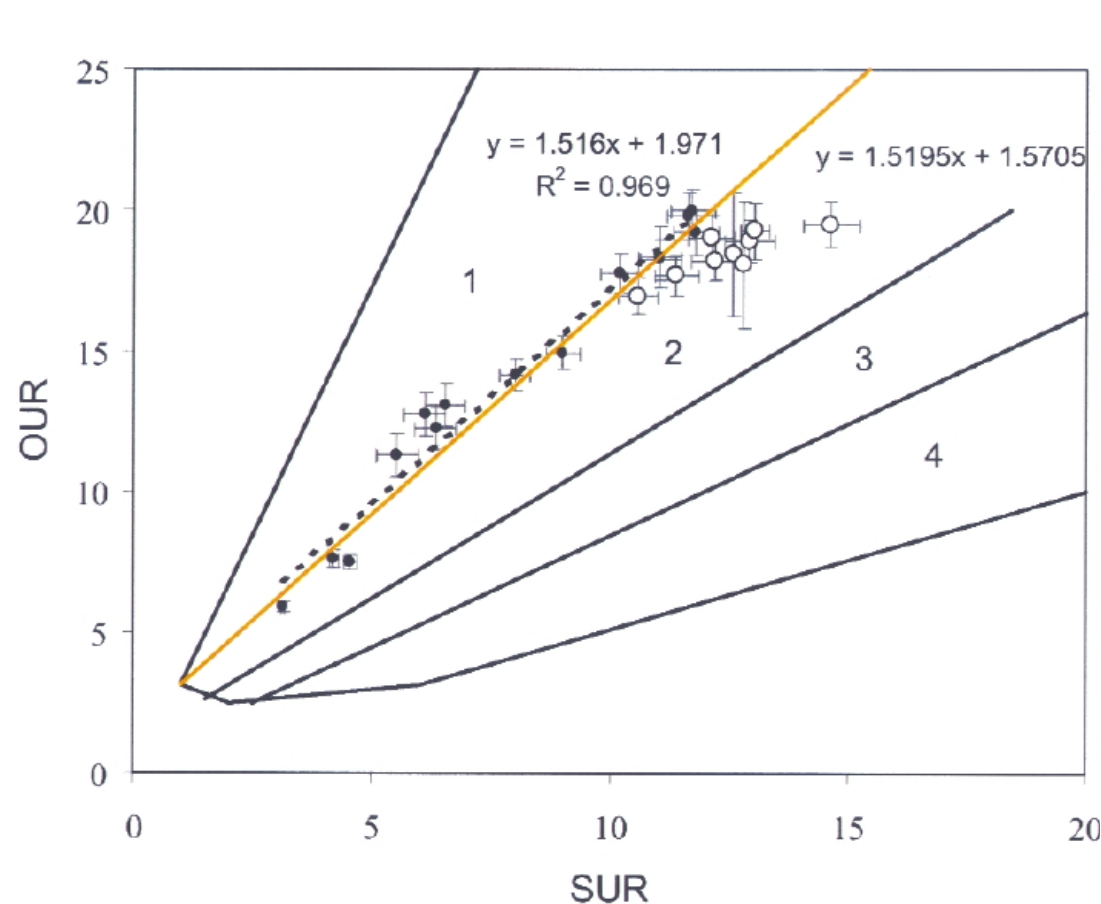
$v_j \geq 0$ for $i \in R$

Result:

Flux distribution v

Optimality studies in *E. coli*

- *E. coli* was grown on succinate
- Optimal growth rates were predicted as extreme fluxes
- Oxygen and succinate uptake rates were measured



(Edwards and Palsson, 2000)

A typical LP problem minimising costs

- assemble $r \times n$ stoichiometry matrix N (r reactions, n metabolites)
- identify irreversible reactions $R \subset \{1 \dots r\}$
- define boundary fluxes $B \subset \{1 \dots r\}$
- define “biomass reaction” $v_{\text{biomass}} : \sum_i \alpha_i \cdot S_i \rightarrow \text{biomass}$
- Fix biomass (e.g. from experiments) $v_{\text{biomass}} = v_{\text{biomass}}^{\text{exp}}$

The LP-problem:

minimise	$\sum_i^r w_i \cdot v_i$
under the constraints	$N \cdot v = 0$ $v_{\text{biomass}} = v_{\text{biomass}}^{\text{exp}}$ $v_j \geq 0 \quad \text{for } i \in R$

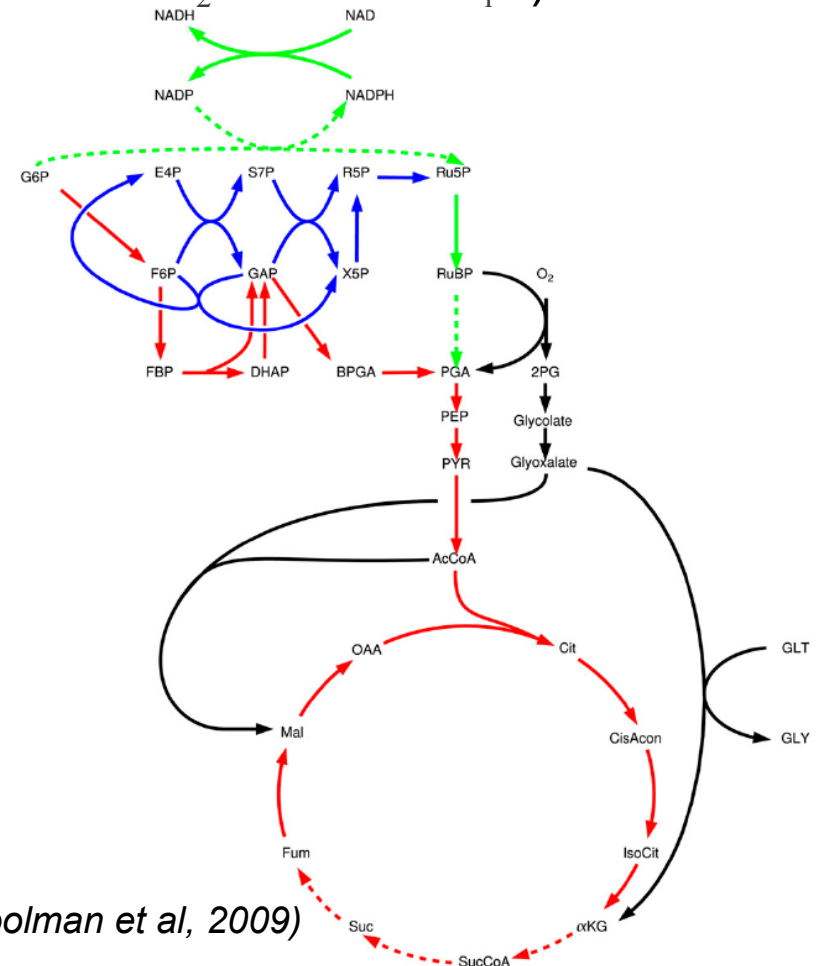
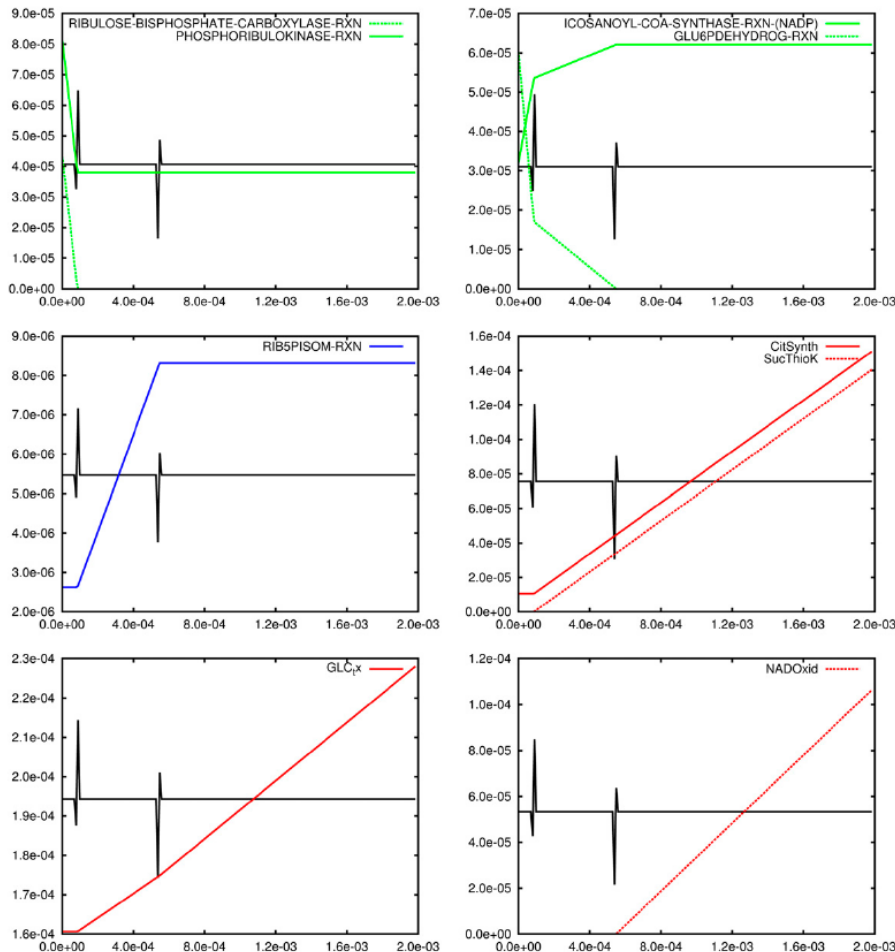
Variation of constraints to query the model

Objective: study how optimal fluxes change upon perturbation of external conditions

Example: impose additional ATP demand (reflecting e.g. external stress conditions)

⇒ Additional constraint $v_{\text{ATPdemand}} = \gamma$ ← tunable parameter

(Additional ATP consuming process: $\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_i$)



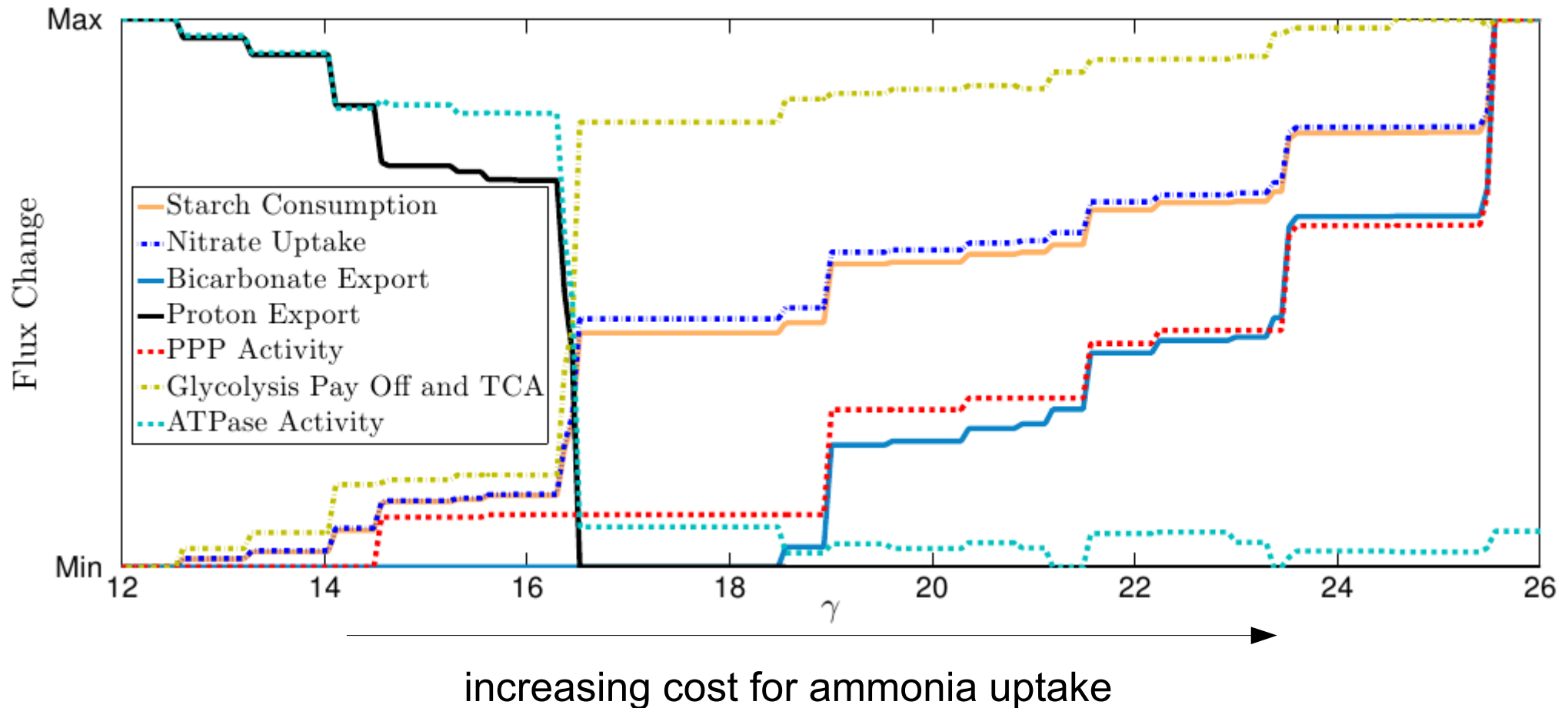
(Poolman et al, 2009)

Simulating availability of nitrogen sources

The LP-problem:

$$\begin{array}{ll}\text{minimise} & \gamma \cdot v_{\text{NO3_uptake}} + \sum_i^r v_i \\ \text{under the constraints} & N \cdot v = 0 \\ & v_{\text{biomass}} = v_{\text{biomass}}^{\text{exp}} \\ & v_j \geq 0 \text{ for } i \in R\end{array}$$

Simulating availability of nitrogen sources



Results for a network of *Medicago truncatula*

“What if” questions

Assume, we want to know what is the 'cheapest' metabolic route to produce a certain compound X

Add consuming reaction $v_X: X \rightarrow \emptyset$

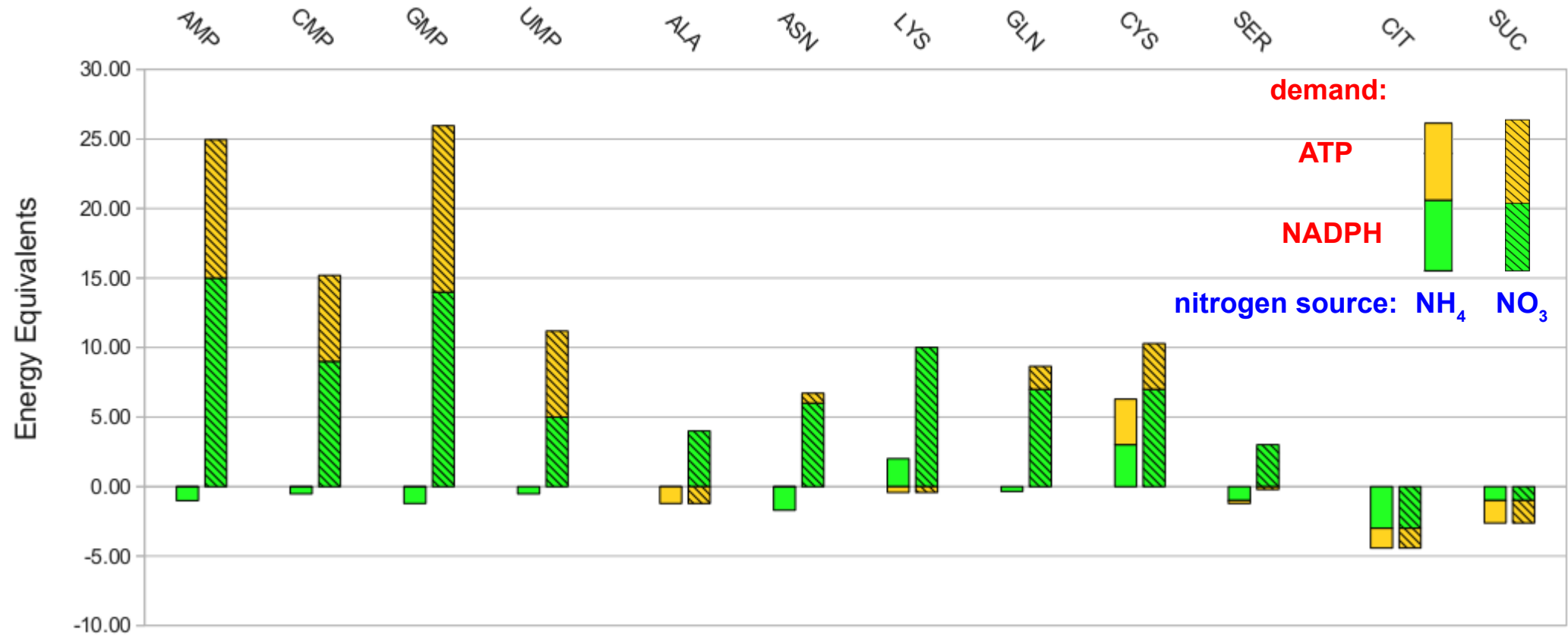
Define 'cheap'

- minimal energy requirement (ATP)
- minimal redox requirement (NADPH)

The LP-problem:

minimise	$w_1 v_{\text{ATPproduction}} + w_2 v_{\text{NADPHproduction}}$
under the constraints	$N \cdot v = 0$
	$v_X = 1$
	$v_j \geq 0 \text{ for } i \in R$

Energy Requirement of Metabolites in Terms of Reductant and ATP



Results for a network of *Medicago truncatula*