BIOMENG 261

TISSUE AND BIOMOLECULAR ENGINEERING

Module I: Reaction kinetics and systems biology

Oliver Maclaren
oliver.maclaren@auckland.ac.nz

MODULE OVERVIEW

Reaction kinetics and systems biology (*Oliver Maclaren*)
[11-12 lectures/3 tutorials/2 labs]

1. *Basic principles: modelling with reaction kinetics* [5-6 lectures]
   Physical principles: conservation, directional and constitutive. Reaction modelling. Mass action. Enzyme kinetics. Enzyme regulation. Mathematical/graphical tools for analysis and fitting.
2. *Systems biology I: signalling and metabolic systems* [3 lectures]
   Overview of systems biology. Modelling signalling systems using reaction kinetics. Introduction to parameter estimation. Modelling metabolic systems using reaction kinetics. Flux balance analysis and constraint-based methods.
3. *Systems biology II: genetic systems* [3 lectures]
   Modelling genes and gene regulation using reaction kinetics. Gene regulatory networks, transcriptomics and analysis of microarray data.

LECTURE 9: FLUX BALANCE ANALYSIS CONTINUED

- Flux balance/constraint-based analysis continuued
- Null spaces and spans (linear algebra)
- Geometry of constraints
- Extra constraints
- Optimality conditions (linear programming)

WHAT IS FLUX-BASED ANALYSIS?

Orth et al. (2010) in Nature Biotechnology:

What is flux balance analysis?

Jeffrey D Orth, Ines Thiele & Bernhard Ø Palsson

Flux balance analysis is a mathematical approach for analyzing the flow of metabolites through a metabolic network. This primer covers the theoretical basis of the approach, several practical examples and a software toolbox for performing the calculations.

(see Canvas)

WHAT IS FLUX-BASED ANALYSIS?

Orth et al. (2010) in Nature Biotechnology:

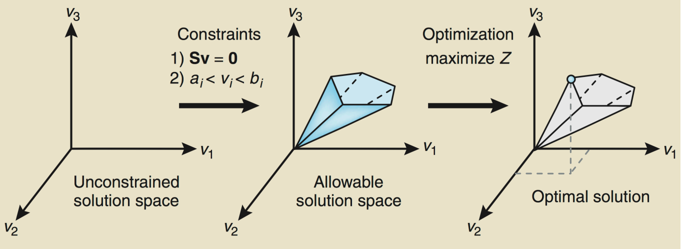


Figure 1 The conceptual basis of constraint-based modeling. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix **S** (labeled 1) and capacity constraints imposed by the lower and upper bounds (a_i and b_i) (labeled 2) are applied to a network, it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space.

RECALL: FLUX BALANCE ANALYSIS

Instead of the dynamic (ODE) problem, we aim to solve the *steady-state* equation

$$\mathbb{S}\mathbf{J} = \mathbf{0}$$

for the vector of *fluxes* $\mathbf{J}$, *here treated as unknown*.

- No consitutive equations/no rate parameters involved here.
- We don't need to know the metabolite concentrations, just solve for fluxes

RECALL: FLUX BALANCE ANALYSIS

For a given metabolic network there are *typically* (not always) more reactions than species/metabolites i.e.

More columns (unknowns) than rows (equations)

The problem is *underdetermined*, i.e. there are typically *multiple solutions*.

There is a non-trivial *null space*.

NULL SPACE?

For a matrix $\mathbb{A}$ the *null space* is just the set of solutions to the zero problem

$$\mathbb{A}\mathbf{x} = \mathbf{0}$$

i.e. here

$$N(\mathbb{S}) = \{\mathbf{x} \mid \mathbb{S}\mathbf{x} = \mathbf{0}\}$$

Example.

SPAN?

The *span* of a set of vectors is just the set of all linear combinations of these, i.e. the *hyperplane* these define.

Here we have

$$N(\mathbb{S}) = \text{span}\{\text{ indep. solutions of } \mathbb{S}\mathbf{x} = \mathbf{0}\}$$

Example.

UNIQUNESS? CONSTRAINT-BASED ANALYSIS

Clearly, there are multiple compatible solutions. To explore these further we can

- Add *bounds* (capacity constraints) on fluxes
- Add *directional* constraints (from thermodynamics)
- Look for special *'optimal'* solutions (e.g. maximum ATP production)

We say we are carrying out a *constraint-based analysis*...for obvious reasons! (FBA is a particular type of constraint-based analysis).

EXAMPLE

See handout.

GENERAL OPTIMISATION FRAMEWORK

We can formulate our problem as

$$\min z = \mathbf{c}^T \mathbf{J}$$

subject to

$$\mathbb{S}\mathbf{J} = \mathbf{0}$$
$$l_i \leq J_i \leq u_i$$

for $i = 1, \ldots N$ and a vector $\mathbf{c}$ of scalar 'costs' (weights), i.e. a

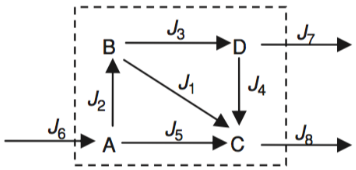
linear programming optimisation problem
(see EngSci OpsRes courses!)

WHAT DO I NEED TO BE ABLE TO DO?

- Given a network, find $\mathbb{S}$
- Find the nullspace for a simple $\mathbb{S}$ (see handout)
- Describe/list some constraints or conditions that we might add to explore our null space and find special solutions
- Write down an optimisation problem given a problem description
- Solve a simple optimisation problem

EXTRA: BOUNDARIES, INTERNAL FLUXES AND INEQUALITY VS EQUALITY CONSTRAINTS

We often want to 'draw boundaries' around a 'system' of interest. We can either *include* these boundary fluxes as usual or treat them like *'slack'* variables for *inequality* constraints



EXTRA: BOUNDARIES, INTERNAL FLUXES AND INEQUALITY VS EQUALITY CONSTRAINTS

- *Equality* constraints $\mathbb{S}\mathbf{J} = \mathbf{0}$ define *hyperplanes* in the space of *all fluxes* (including boundary etc fluxes).
- *Inequality* constraints $\mathbb{S}\mathbf{J} \geq \mathbf{0}$ define *polyhedra* in the *reduced* set of fluxes (e.g. internal only).

Equivalent, given proper care, but just be aware of which.

EXTRA: BOUNDARIES, INTERNAL FLUXES AND INEQUALITY VS EQUALITY CONSTRAINTS

Implicit inequality constraints give the polyhedra seen in:

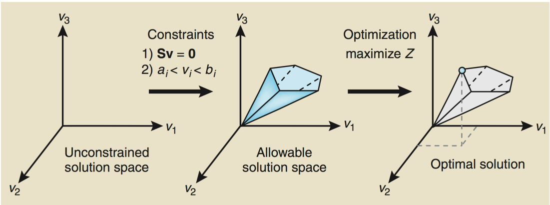


Figure 1 The conceptual basis of constraint-based modeling. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix **S** (labeled 1) and capacity constraints imposed by the lower and upper bounds (a_i and b_i) (labeled 2) are applied to a network, it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space.

Biomeng 261 Lecture 09

- Flux Balance Analysis (FBA)
  /Constraint-based analysis
  cont'd.

Some background math

- Null spaces & spans
- Geometry of constraints

} Linear algebra

- Optimality conditions & optimisation problems

} Linear programming

What is FBA?

→ another paper!
Orth et al (2010) in Nature Biotechnology

What is flux balance analysis?

Jeffrey D Orth, Ines Thiele & Bernhard Ø Palsson

Flux balance analysis is a mathematical approach for analyzing the flow of metabolites through a metabolic network. This primer covers the theoretical basis of the approach, several practical examples and a software toolbox for performing the calculations.

see Canvas.

Idea:
- Take a system of 'fluxes' $J_1, J_2, \ldots$ etc
- Could write as $\frac{d\bar{C}}{dt} = \underline{S}\,\bar{J}$

FBA
- Instead, just construct $\underline{S}$
- Solve $\underline{S}\,\bar{J} = \bar{0}$ for fluxes $\bar{J}$ as unknown.

Example: Determine $\underline{S}$ for the system

$$? \xrightarrow[\text{ATP} \to \text{ADP}]{J_1} \text{Gluc-6-P} \xrightarrow[\text{ATP} \to \text{ADP}]{J_2} ?$$

$$\text{Gluc-6-P} \xrightarrow{J_3} \text{Fruc-6-P}$$

$$\text{Fruc-6-P} \xrightarrow[\text{ATP} \to \text{ADP}]{J_4} \text{Fruc-1,6-P}_2 \xrightarrow{J_5} ?$$

$$\text{ADP} \xrightarrow{J_6} \text{ATP}$$

$$\text{ATP} \xrightarrow{J_7} \text{ADP}$$

$$\text{ATP} + \text{AMP} \xrightarrow{J_8} 2\,\text{ADP}$$

Answer:

$$\underline{S} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cccccccc} J_1 & J_2 & J_3 & J_4 & J_5 & J_6 & J_7 & J_8 \end{array}$$

$$\underline{S} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & -1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{array}{l} \leftarrow \text{Gluc-6-P} \\ \leftarrow \text{Fruc-6P} \\ \leftarrow \text{Fruc-1,6-P}_2 \\ \leftarrow \text{ATP} \\ \leftarrow \text{ADP} \\ \leftarrow \text{AMP} \end{array}$$

What's the catch? $\underline{S}\,\bar{J} = \bar{0}$ soln?

Here:

6 rows $\leftarrow$ equations/constraints

8 columns $\leftarrow$ unknowns (J)

$\boxed{\text{unknowns} > \text{eq}^{n}\text{s}}$

In general

- more reactions/fluxes (unknowns) than metabolites/concentrations (equations)
- often don't know all metabolites involved (lose a rows)

Shape:

$$\begin{array}{c} \\ c_1 \\ \vdots \\ c_M \end{array} \begin{array}{c} J_1 \;\cdots\cdots\; J_N \\ \left[\begin{array}{c} \\ \xleftrightarrow{\quad \text{wider} \quad} \\ \\ \end{array} \right] \end{array} \qquad \underline{N > M}$$

$\rightarrow$ $\underline{S}\,\bar{J} = \bar{0}$ is usually underdetermined

$\rightarrow$ ie multiple solutions

$\rightarrow$ makes sense since only using conservation of mass.

To understand this mathematically / geometrically, we need some <u>linear algebra</u>

Null spaces (of $\underline{S}$ say)

- The <u>nullspace</u> of a matrix $\underline{A}$ is the <u>set</u> of <u>all solutions</u> to $\underline{A}\,\bar{x} = \bar{0}$

- zero vector is <u>always</u> in nullspace
$$\underline{A}\,\bar{0} = \bar{0}$$

- A non-trivial null-space is when we have non-zero solutions in the null space

---

<u>Mathematically</u>: vectors — they satisfy condition.

$$\boxed{N(\underline{A}) = \{\ \bar{x}\ \mid\ \underline{A}\,\bar{x} = \bar{0}\ \}}$$

null space of $\underline{A}$ — set of — such that

<u>Example</u>

Consider:
$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ie
$$\left.\begin{aligned} x_1 - x_2 &= 0 \quad (1) \Rightarrow x_1 = x_2 \\ -x_1 + x_2 &= 0 \quad (2) \Rightarrow x_1 = x_2 \end{aligned}\right\} \text{same eq}^n!$$

$$N(\underline{A}) = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;\middle|\; x_1 = x_2 \right\}$$

- <u>Two</u> unknowns, one <u>independent</u> constraint
$2 - 1 = 1$ 'free var'

→ <u>free choice</u> of eg x_2 (say)
↳ <u>then</u> x_1 is determined.

→ a <u>particular</u> solution is eg
$$\bar{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

→ what about <u>all</u> solutions or the <u>'general'</u> solution?

<u>Note</u>
Any vector $a \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ a \end{pmatrix}$ <u>also</u> satisfies above.

Leads to idea of

Span : $\boxed{\text{span}\{\bar{v}_1, \bar{v}_2, \dots \bar{v}_n}$

$a\bar{v}_1 + b\bar{v}_2 + \dots$

- set of all linear combinations of a set of vectors

- Every vector in the null space can be written as a linear combination of independent solutions to $\underline{S}\bar{\bar{J}} = \bar{0}$

ie

$$\boxed{N(\underline{S}) = \text{span}\left\{ \begin{array}{c}\text{independent vectors} \\ \text{solving } \underline{S}\bar{\bar{J}} = \bar{0}\end{array} \right\}}$$

Steps - reduce to minimal set of eq^{n}s

Two free vars $\Rightarrow$ Two indep. vectors etc.

- choose free vars (eg m−n of them)

- find implied set of independent vectors

- write $N(\underline{S}) = \text{span}\{ \dots \}$

Eg previous example just have

$$N(\underline{S}) = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Example

$n = 6$

$$\begin{array}{cc} m=4 & \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 2 & -1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 3 & -3 & 3 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix} \end{array}$$

Expect $n - m = 6 - 4 = 2$ free vars & hence 2 independent vectors

(if all eq^{n}s independent → I'll usually make this true!)

Approach: use elementary row ops to make upper triangular...

OR

just expand out & solve for in terms of free vars by being sensible

---

Then ↳ setting each free var non-zero & rest as zero in turn generates independent vector solutions.

Example cont'd

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 2 & -1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 3 & -3 & 3 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \\ J_5 \\ J_6 \end{bmatrix} = \bar{0}$$

$$\Rightarrow \quad -J_1 + J_2 = 0 \qquad (1)$$

$$2J_2 - J_2 - 2J_3 = 0 \qquad (2)$$

$$\cancel{3}J_3 - \cancel{3}J_4 + \cancel{3}J_5 = 0 \qquad (3)$$

$$J_4 - J_5 - J_6 = 0 \qquad (4)$$

4 equations, 6 vars.

choose 2 & make sure rest are determined

eg choose J_1 $\xrightarrow{(1)}$ gives J_2

$\xrightarrow{(2)}$ gives J_3

Leaves eg J_4 or J_5 ← choose

Choosing J_1 & J_5 as free gives

$$J_2 = J_1$$

$$J_3 = \tfrac{1}{2} J_2$$

$$J_4 = J_3 + J_5 = \tfrac{1}{2} J_2 + J_5$$

$$J_6 = J_4 - J_5 = \tfrac{1}{2} J_2$$

Make J_1 & J_5 non-zero in turn to get independent sol^{n}s eg

$J_1 = 2,\ J_5 = 0$	$J_1 = 0,\ J_5 = 1$
$\bar{J}^{(1)} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$	$\bar{J}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$

Finally, write

$$\boxed{N(\underline{S}) = \text{span}\left\{ \bar{J}^{(1)}, \bar{J}^{(2)} \right\}}$$

where $\bar{J}$s are as above

(phew!)

'Special' solutions: uniqueness?

- Regardless of specific form used, we have $\boxed{\text{linear constraints}}$ on $\boxed{\text{fluxes}}$ $\bar{J}$

(Note: would typically be non-linear in concentrations if using mass action)

The constraints can be thought of as

- Defining a non-trivial null space ($\underline{S}\,\bar{J} = \bar{0}$ form)

or

- A polyhedral feasible region ($\underline{S}\,\bar{J} \geq \bar{0}$ form for internal)

$\Rightarrow$ either way we might want to look at ways to 'pick out' particular solutions from these sets of solutions.

Constraints & Optimality conditions

We can add many types of extra constraints or conditions to narrow down possible sol'ns:

- bounds/signs of fluxes
- thermodynamic feasibility (directional constraints)
- optimality or extreme cases (eg max. energy prod.)

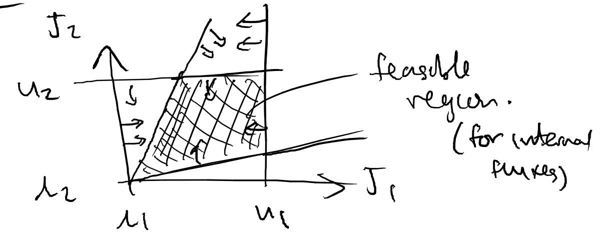
The most natural to include first in general are lower & upper flux bounds, ie

$\boxed{\text{capacity constraints}}$:

$$\boxed{l_i \leq J_i \leq u_i}$$

(note: without these only know ratios)

Gives



For irreversible, use $\boxed{J_i \geq 0}$ (why?)

'Optimal' solutions

A useful way to pick out 'special' solutions is to use $\boxed{\text{optimality}}$ (max/min) conditions

- these tell us about limits on what is possible

eg 'maximal rate of ATP production is...'

Note:

- a real system may or may not reach these limits!
  - many competing goals so may not be optimal for any one
  - still a good way to predict/understand
  - many sensible constraints can be rewritten as max/min conditions.

Example: cont'd

Recall our example where we wrote our 2D null space in terms of J_1 & J_5

$$\underbrace{\boxed{\begin{aligned} J_2 &= J_1 \\ J_3 &= J_1/2 \\ J_4 &= J_1/2 + J_5 \\ J_6 &= J_1/2 \end{aligned}}}$$

can write as $\underset{\sim}{S}\,\bar{J} = \bar{0}$

ie

$J_2, J_3, J_4, J_6 = f(J_1, J_5)$

6 vars, 2 free

4 det.

To find a particular solⁿ.

- suppose $\underbrace{2 \leq J_1, J_5 \leq 10}$ bounds apply to both
- $J_i \geq 0$ for all i
- J_4 is ATP production & J_6 is lactate prod. } for example

Goals Case 1. Max ATP prod.

Case 2. Max ATP prod. while minimising lactate prod.

Key: rewrite objective in terms of free as well!

Case I.

$$\boxed{\begin{array}{l} \max \; J_4 \\ \text{subject to} \\ \bullet \quad \underline{S}\bar{J} = \bar{0} \\ \& \\ \bullet \quad 2 \le J_1 \le 10 \\ \bullet \quad 2 \le J_5 \le 10 \end{array}}$$

$\underline{S}\bar{J} = \bar{0} \iff$ we re-wrote as
$$\begin{cases} J_2 = J_1 \\ J_3 = J_1/2 \\ J_4 = J_1/2 + J_5 \\ J_6 = J_1/2 \end{cases}$$

free

Sol^n: want $J_4 = \frac{J_1}{2} + J_5$ max.

$\Rightarrow$ set $\boxed{J_1 = J_5 = 10}$ (can vary indep.) ('free')

$\Rightarrow \boxed{J_4 = 15}$

Case 2

$$\boxed{\begin{array}{l} \underline{\max}\; aJ_4 - bJ_6 \\ a, b > 0 \text{ weights} \\ (\text{'values' or 'costs'}) \\ \text{s.t. same constraints} \end{array}}$$

$\boxed{\max -J \equiv \min J}$

$\underline{\max}\;\; aJ_4 - bJ_6 = \frac{(a-b)}{2}J_1 + aJ_5$

- if $a > b$: same: $J_1 = J_5 = 10$
- if $a < b$: $J_1 = 2,\; J_5 = 10$
- if $a = b$: $J_1 = ?,\; J_5 = 10$ — not enough info for unique

General Optimisation Framework

$$\boxed{\begin{array}{ll} \min & z = \bar{c}^T \bar{J} \\ \text{st.} & \underline{S}\bar{J} = 0 \\ & l_i \le \bar{J}_i \le u_i \end{array}}$$

$\min z = \bar{c}^T\bar{J}$ } objective function

$\underline{S}\bar{J} = 0$, $l_i \le \bar{J}_i \le u_i$ } constraints

z: scalar (number)

$\bar{c}$: vector of weights (here, 'costs')

eg

$$\bar{c}^T\bar{J} = \begin{pmatrix} 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix} = J_1 + 2J_2 - J_3 = z$$

Linear objective function + Linear constraints } "Linear Programming"

L Engsci 255 etc.

Um?

What should I be able to do?

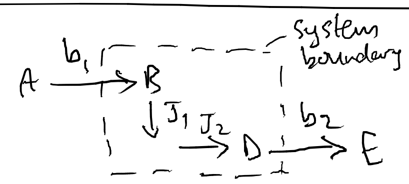
- Given network find $\underline{S}$
- Find nullspace for simple $\underline{S}$
  - ↳ reduce to indep. eq's
  - ↳ choose free fluxes
  - ↳ find indep. vectors
  - ↳ write as span{ }
- Describe typical/useful additional constraints
- Write down optimisation problem for given description
- Solve very simple optimisation problems
  - ↳ write in terms of free vars!

Appendix

Boundary vs Internal Fluxes

Sometimes we want to draw 'system boundaries' & call some fluxes 'boundary' fluxes & some 'internal' fluxes.

eg



b: boundary
J: internal
(fluxes)

We can either include these as usual

$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{cccc} J_1 & J_2 & b_1 & b_2 \end{array} \\
\begin{bmatrix} 0 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

(include b in $\bar{J}$)

OR

focus on internal

$$
\begin{array}{c} \\ B \\ C \\ D \end{array}
\begin{array}{c}
\begin{array}{cc} J_1 & J_2 \end{array} \\
\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}
\end{array}
$$

In this later case we can write $\boxed{\underline{S}\bar{J} \geq 0}$, if we choose signs carefully.

$(\underline{S}\bar{J} = \bar{b}_{net} \geq 0)$

Appendix

Geometry: Nullspace vs Feasible region

$\underbrace{\text{Nullspace}}_{=}$ vs $\underbrace{\text{Feasible region}}_{\geq}$

The two forms

$\underbrace{\underline{S}\bar{J}=\bar{0}}_{\text{all fluxes}}$ or $\underbrace{\underline{S}\bar{J}\geq 0}_{\text{subset of fluxes eg internal only.}}$

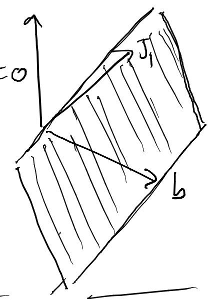
are equivalent but have slightly different geometric pictures

$\underline{S}\bar{J}=\bar{0}$

eg

$J_2 - J_1 + b = 0$

2D plane in 3D.



3 vars

1 indep equality constr.

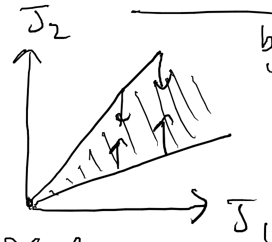
3−1 = 2 dimensional (hyper-) plane

[nullspace]

$\underline{S}\bar{J}\geq\bar{0}$ eg $2J_1 - J_2 \geq 0$

$J_2 - \frac{1}{2}J_1 \geq 0$

b becomes 'slack' variable



2D cone

2 vars

2 indep. inequality constraints

2D cone/polyhedra

[feasible region]

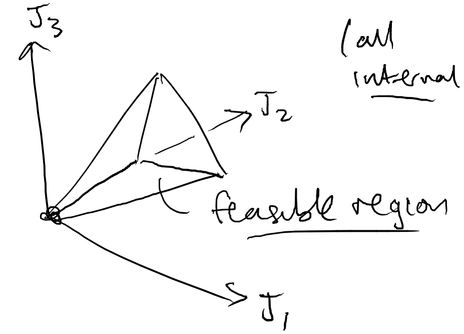
Appendix

What? Why?

In general we will put in 'standard form'

$\boxed{\underline{S}\bar{J}=\bar{0}}$ & use [nullspace] (include boundary fluxes)

But you may see pictures like:



(all internal)

which come from inequality constraints version.

→ Just be aware

(General principle/trade-off: simple in higher dim or complex in lower d)