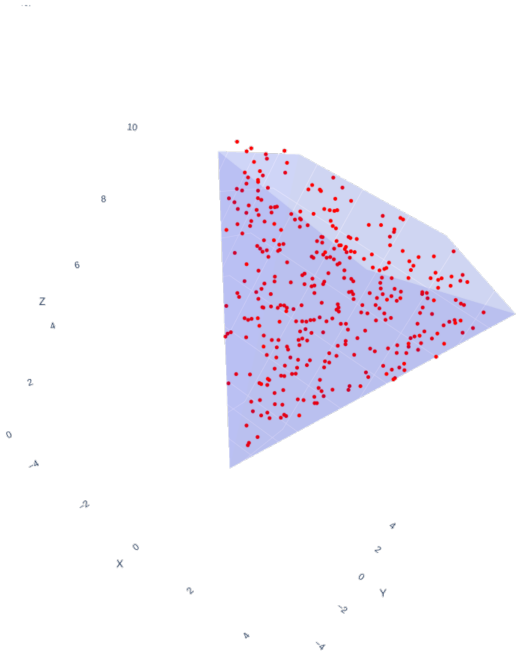
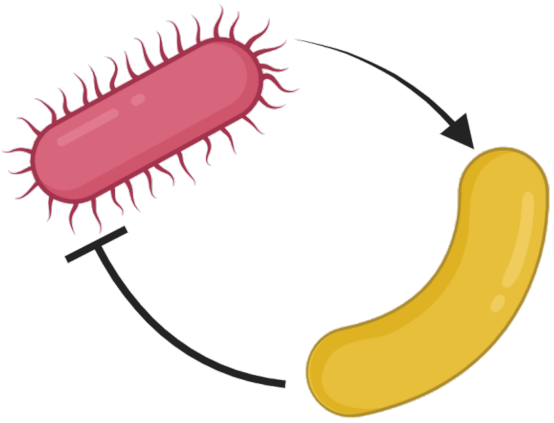
Notes on metabolic modelling analysis

Haris Zafeiropoulos

May 08, 2024

CONTENTS





STATUS IN PREPARATION

CHAPTER

ONE

INTRODUCTION

This is a *book* about microbial metabolic models, their reconstructions and analysis at the strain and the community level. It is intended to give only some insight from the user's perspective and not a thorough background on each analysis presented. Yet, the basics will do be shown but mostly *when* to use a type of analysis, *what* can we learn from it, *how* to interpret their results and what are the assumptions made.

The book contains numerous examples *as programs*, including implementations of many concepts. Each chapter is generated from a self-contained Jupyter Notebook. You can click on the "download" button at the top-right of the chapter, and then select ".ipynb" to download the notebook for that chapter, and you'll be able to execute the examples yourself. Many of the examples are generated by code that is hidden (for readability) in the chapters you'll see here. You can show this code by clicking the "Click to show" labels adjacent to these cells.

This *book* is open source, and the latest version will always be available online here. The source code is available on GitHub. If you would like to fix a typo, suggest an improvement, or report a bug, please open an issue on GitHub.

The techniques described in this book have developed out of the study of *data privacy*. For our purposes, we will define data privacy this way:

Definition 1 (M-models)

Genome-scale metabolic models (**M-models**) provide for a metabolic description of genotype–phenotype relationship without accounting explicitly for synthesis of enzymes. M-models employ Boolean logic statements relating genes, proteins, and reactions, or the Gene–Protein–Reaction associations, or Gene-Protein-Reactions (GPRs). A reaction can only carry a non-zero flux if its GPR statement evaluates to True [1].

---

integrated models of metabolism and expression (ME-Models) account explicitly for the genotype–phenotype relationship. Macromolecular expression is directly integrated with cellular metabolism [1].

CHAPTER

TWO

METABOLIC MODELS

Definition 2 (Balanced growth)

Balanced growth is the average state of a cell in a cell bacterial population growing exponentially at the specific (constant) growth rate $\$\mu \geq 0\$$, i.e. the amount of produced biomass per biomass per cell per unit of time.

Definition 3 (k_{cat})

It is the maximum rate at which an enzyme can catalyze a specific reaction when it is saturated with substrate. It indicates the number of substrate molecules converted into product per enzyme molecule per unit time under optimal conditions. In simpler terms, it reflects how fast an enzyme can convert substrate into product.

Definition 4 (reaction flux)

Reaction flux refers to the **rate** at which a biochemical reaction proceeds in a biological system. It's a measure of *how quickly* reactants are being converted into products within a specific cellular context.

In a simplified picture of balanced growth, all metabolic processes are balanced: the rate at which material flows into the cell matches the rate at which it is converted, which again matches the production rate of macromolecule precursors. In addition, we assume that these fluxes are constant, such that the whole metabolic network is in a 'steady-state'. Taken together, we thus assume that the metabolic network can take up and produce external metabolites (e.g. extracellular metabolites and macromolecular precursors), but that all internal metabolites ("inside" the metabolic network) are mass-balanced, that is, for each of these metabolites, production and consumption cancel out.

Since each enzyme has a maximal catalytic rate (the k_{cat} value), a reaction flux will require a certain (minimal) amount of enzyme, which takes up cellular space; since cellular space is limited, fluxes cannot increase infinitely since there is always an upper bound on a weighted sum of reaction fluxes. This constraint implies compromises between different reaction fluxes: one flux can only be increased at the expense of others.

The mathematical model:

- *variables* to describe: the metabolic **fluxes** in steady-state metabolism,
- *constraints* to apply: the **balance** of production and consumption of all **internal** metabolites

Importantly, the model will be able to describe compromise: for example, with a given carbon influx and assuming mass balance, the carbon atoms can either be used to generate energy **or** biomass; if one function increases, the other one goes down.

To obtain realistic predictions, we may introduce additional constraints, for example known flux directions or experimentally measured uptake rates.

All this information will not suffice to predict metabolic fluxes precisely, but it allows us to narrow down the possible flux distributions.

$$N \times v = 0 = N \times v^{+} - N \times v^{-} = [N \; -N] \begin{bmatrix} v^{+} \\ v^{-} \end{bmatrix} \tag{2.1}$$

The mass-balance constraints in the previous equation, combined with the property that v_i^+ , $v_i^- \geq 0$ can be expressed in the form

$$A \begin{bmatrix} v^{+} \\ v^{-} \end{bmatrix} \geq 0 \tag{2.2}$$

where:

$$A = \begin{bmatrix} N & -N \\ -N & N \\ I & 0 \\ 0 & I \end{bmatrix}$$

The set of constraints on (v^+ , v^-) define a **polyhedral cone** and since they are non-negative, the cone is also pointed, meaning it contains no complete line and the zero vector is the only vertex (extreme point) of the cone.

The space of solutions that satisfies is called the **flux cone**.

2.1 Kinetic models

Kinetic models are typically formulated as a set of deterministic **ordinary differential equations (ODEs)**.

```
{prf:definition} kinetic variable
:label: kinetic variable


kinetic parameters
$k_{cat}$
$\frac {k_{cat}}{ K_M}$
$K_M$
```

Assumptions used in the formulation of biological network models

Assumption	Description
Continuum assumption	Do not deal with individual molecules, but treat medium as a continuum
Finer spatial structure ignored	Medium is homogeneous
Constant-volume assumption	V is time-invariant, $\frac{dV}{dt} = 0$
Constant temperature	Isothermal systems; Kinetic properties a constant
Ignore physico-chemical factors	Electroneutrality and osmotic pressure can be important factors, but are ignored

The **stoichiometric matrix **(S) represents the reaction topology of a network. For an overview on its characteristics see [2]

```
{prf:definition} gradient matrix
:label: gradient matrix
```

(continues on next page)

(continued from previous page)

```
(from {cite}`palsson2006systems`)
Each link in a reaction map has kinetic properties with which it is associated.
The reaction rates that describe the kinetic properties are found in the rate laws,
 ↪$v(x; k)$, where the vector $k$ contains
all the kinetic constants that appear in the rate laws. Ultimately, these
properties represent time constants that tell us how quickly a link in a
network will respond to the concentrations that are involved in that link.
The reciprocal of these time constants is found in the gradient matrix G,
whose elements are

$$
g_{ij} =\frac{\partial v_i}{\partial x_j}
$$

These constants may change from one member to the next in a biopop-
ulation, given the natural sequence diversity that exists. Therefore, the
gradient matrix is a *genetically determined* matrix. Two members of the
population may have a different $G$ matrix.
```

Mathematically speaking, G has several challenging features. Unlike the stoichiometric matrix, its numerical values vary over many orders of magnitude. Some links have very fast response times, while others have long response times. The entries of G are real numbers and, therefore, are not "knowable." The values of G will always come with an error bar associated with the experimental method used to determine them. It has though, the same sparsity properties as the matrix S.

```
{prf:definition} Jacobian matrix
:label: jacobian matrix

$S$ gives us network structure and $G$ gives us kinetic parameters of the links in␣
 ↪the network.
Their product, the **Jacobian matrix** ($J$)  gives us the network dynamics.
```

CHAPTER

THREE

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Bernhard Palsson. *Systems biology*. Cambridge university press, 2015.

[2] Bernhard Ø Palsson. *Systems biology: properties of reconstructed networks*. Cambridge university press, 2006.

PROOF INDEX