Flux estimation

In stoichiometric flux estimation the goal is to restrict the space of solutions of the balance equations

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

by fixing some of the free fluxes by measuring them (e.g. uptake or production rates of extracellular metabolites).

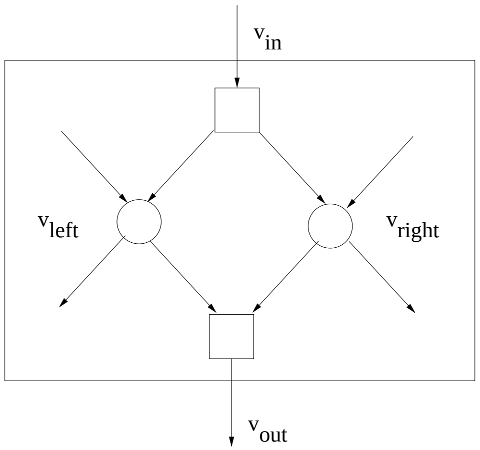
Ideally, a single point in the null space $\mathcal{N}(N)$ remains.

In practise, especially when analysing larger networks, a set of possible flux vectors will remain.

The problem with alternative routes

The biggest single reason for underdetermination of the flux system are alternative routes in the metabolism: if there are alternative routes to produce some metabolite in the metabolic network, the relative activity of the routes cannot be pinpointed.

In the example on the right, the fluxes v_{left} and v_{right} cannot be pinpointed, only their sum can be solved.



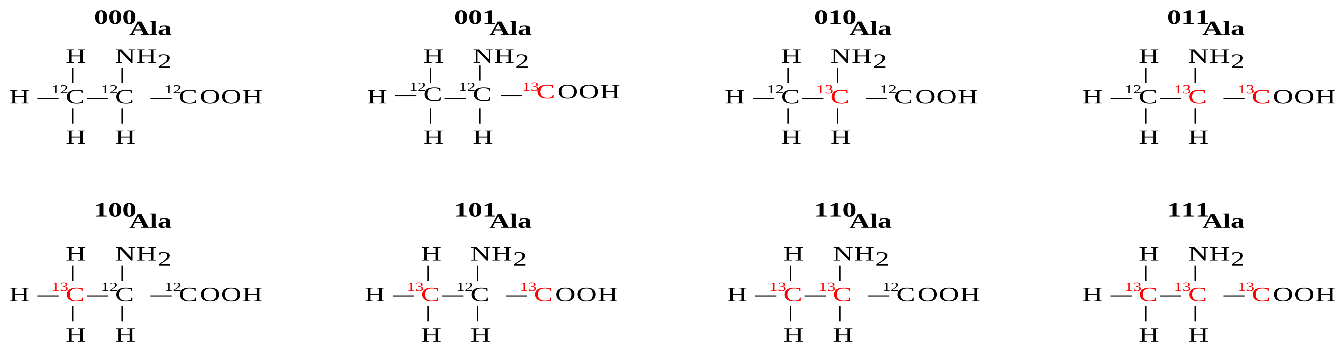
Material balance

$v_{in} = v_{left} + v_{right} = v_{out}$

Isotope tracing experiments

In isotope tracing experiments the cell culture is fed a mixture of natural and ^{13}C substrate (e.g. 90%/10%).

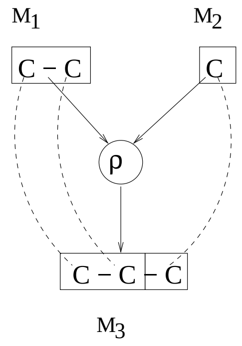
This induces different kinds of ^{13}C labeling patterns, **isotopomers**:



The vector of relative frequencies of the isotopomers $\mathbb{I}_{Ala} = [\mathbb{P}\{^{000}Ala\}, \mathbb{P}\{^{001}Ala\}, \ldots, \mathbb{P}\{^{111}Ala\}] \in [0,1]^{2^3}$, is called an **isotopomer distribution** of alanine.

Metabolite and reaction representation

- We treat metabolites as a set of unique named carbon locations.
- Metabolites M are further divided to fragments, ie. subsets of their carbons $F = M|F$.
- For each reaction a **carbon mapping** describing the transitions of carbon atoms in a reaction event is given.

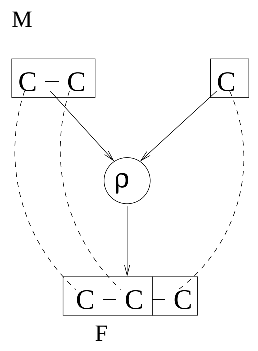


Isotopomers of a product metabolite

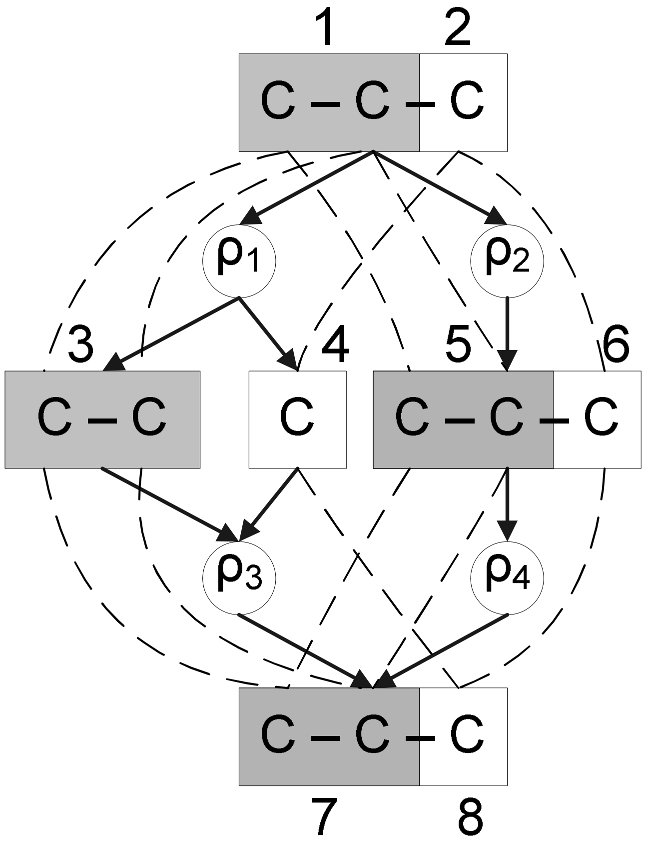
To compute the isotopomer frequencies of a reaction product, we make two assumptions

- Uniform sampling assumption: a reaction draws its reactants independently, uniformly randomly from the reactant pools
- No isotope effects assumption: the reaction does not make any difference between different isotopomer pools

These assumption ensures $P(^{xyz}M_3) = P(^{xy}M_1)P(^{z}M_2)$

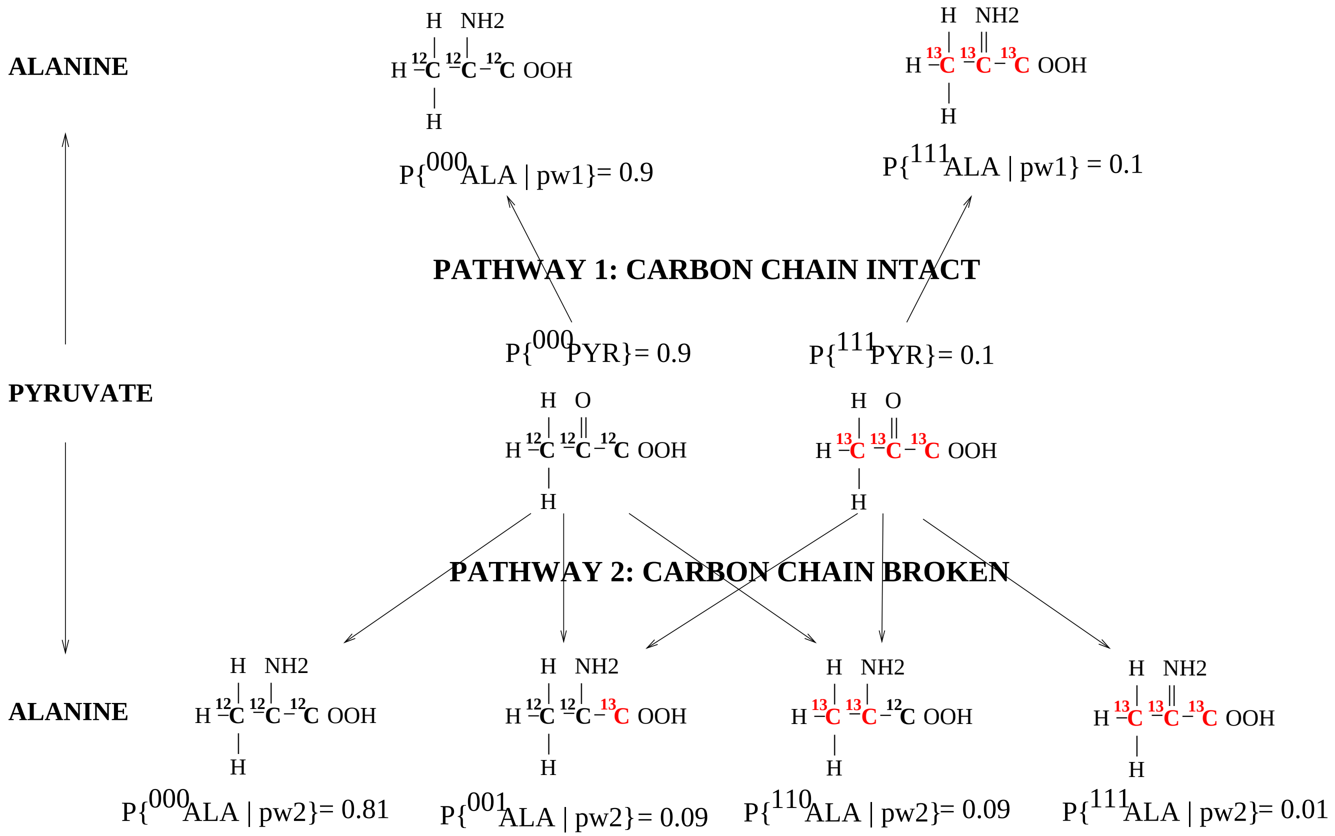


Isotopomer distributions & alternative pathways



- Metabolite 1&2 = pyruvate, metabolite 7&8 = alanine
- Pathway 1 = $\{\rho_1, \rho_3\}$, pathway 2 = $\{\rho_2, \rho_4\}$

Isotopomer distributions & alternative pathways

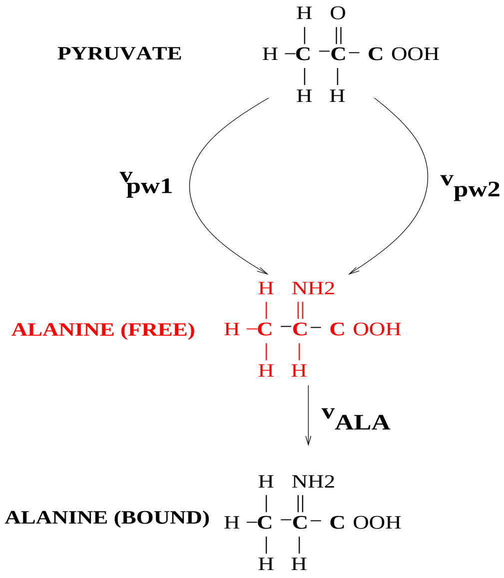


Isotopomeric balance equations

- In a steady state flow in must match the flow out of each metabolite:

$$v_{pw1} + v_{pw2} = v_{ALA}$$

- If the isotopomer frequencies are also known, additional balance equations in the fluxes v can be written (for **junction** metabolites):



$$P(^{000}ALA|pw1)\cdot v_{pw1} + P(^{000}ALA|pw2)\cdot v_{pw2} = P(^{000}ALA)\cdot v_{ALA}$$
$$P(^{001}ALA|pw1)\cdot v_{pw1} + P(^{001}ALA|pw2)\cdot v_{pw2} = P(^{001}ALA)\cdot v_{ALA}$$
$$P(^{110}ALA|pw1)\cdot v_{pw1} + P(^{110}ALA|pw2)\cdot v_{pw2} = P(^{110}ALA)\cdot v_{ALA}$$
$$P(^{111}ALA|pw1)\cdot v_{pw1} + P(^{111}ALA|pw2)\cdot v_{pw2} = P(^{111}ALA)\cdot v_{ALA}$$

Flux estimation from incomplete isotopomer data

In practice, we are faced with incomplete isotopomer data:

- Not all isotopomer distributions can be measured, due to sensitivity issues of measuring equipment.
- Complete isotopomer distributions can only rarely be measured:
  - MS data groups isotopomers of the same weight:

$$aP(^{010}ALA) + bP(^{100}ALA) = d$$

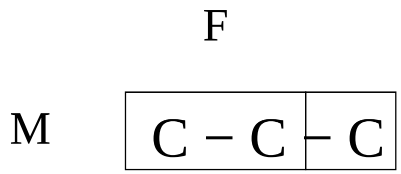
  - NMR measurements require ^{13}C in a specific position e.g. the middle carbon in alanine

$$\frac{P(^{010}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d.$$

We start by tackling the first difficulty.

Fragment marginals of isotopomer distibutions

Each fragment $F \subseteq M$ induces a *marginal distribution* to the isotopomer distribution:



$$P(^{00}F) = P(^{000}M_3) + P(^{001}M_3)$$

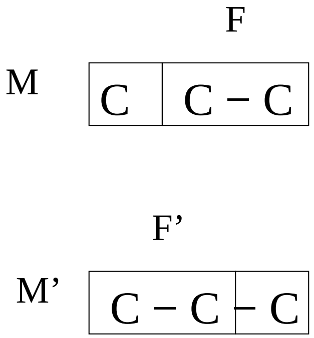
$$P(^{01}F) = P(^{010}M_3) + P(^{011}M_3)$$

$$P(^{10}F) = P(^{100}M_3) + P(^{101}M_3)$$

$$P(^{11}F) = P(^{110}M_3) + P(^{111}M_3)$$

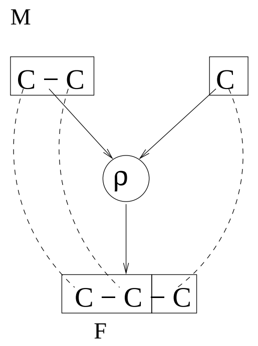
Fragment equivalence

- Two fragments $F \subseteq M$ and $F' \subseteq M'$ are *equivalent* if the fragment marginal distributions of the respective isotopomer distributions of M and M' are equal, irrespectively of the fluxes of the metabolic network
- Meaurement information can be easily propagated between equivalent fragments!
- When does the fragment equivalence hold true?



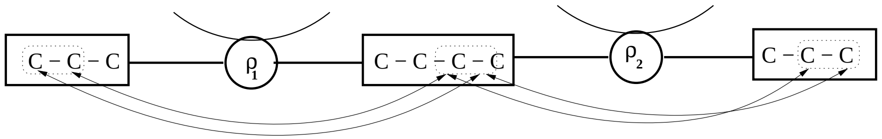
Fragment equivalence for a single reaction

In a single reaction, the reactant and product fragments are equivalent:
$P(^{xy}F) = P(^{xy}M_1), x, y = 0, 1$



Fragment equivalance for unbranched pathways

Each reaction (and transitively, an unbranched pathway) defines a one-to-one mapping between carbon atom locations

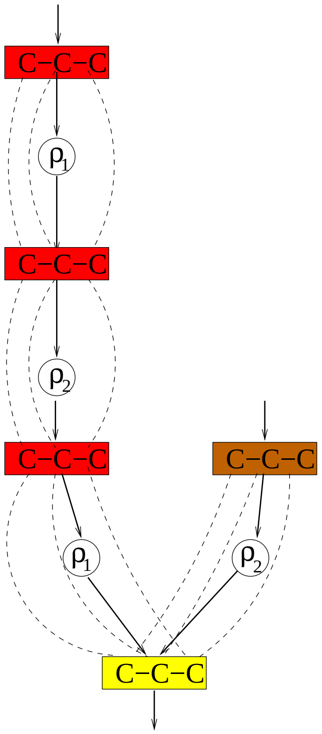


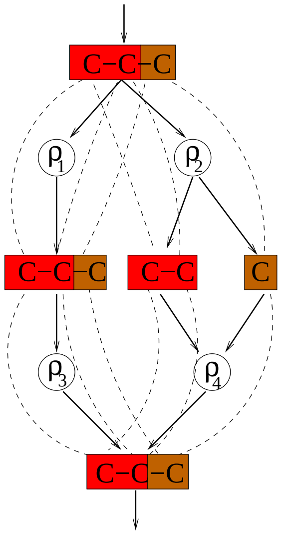
Given a reactant M and product M' of the pathway, and two fragments $F \subseteq M$ and $F' \subseteq M'$, assume that $F' = \Lambda(F)$ is the image of the fragment F under the atom mapping Λ of the pathway.

Then F and F' are equivalent

Fragment equivalence in general

If the molecules of the fragment produced by alternative pathways travel intact and similarly oriented starting from the common source fragment the isotopomer distribution of that fragment is equivalent to the source.

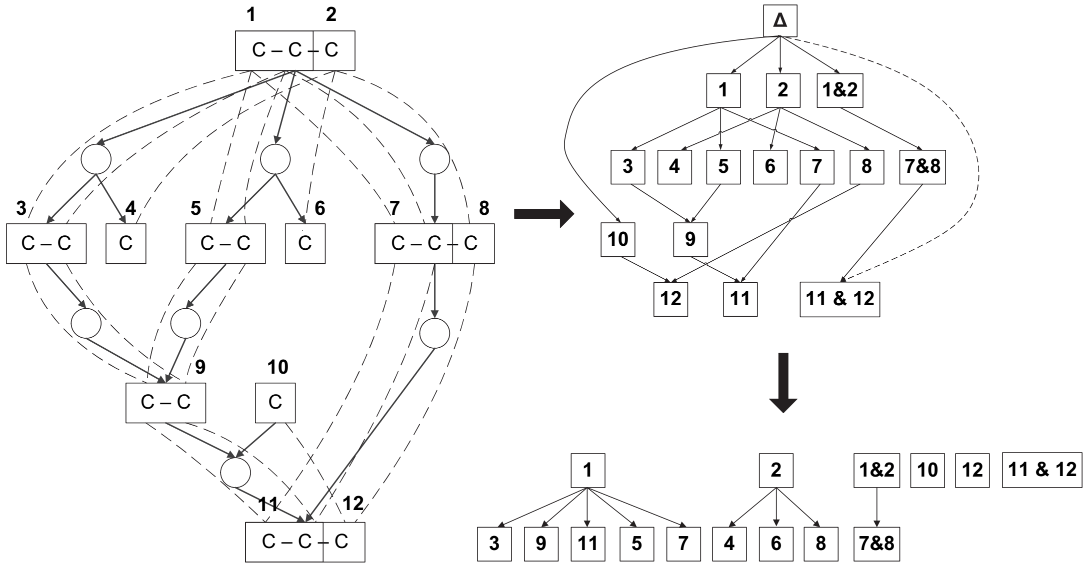




Equivalence classes

The equivalance relation for fragments induces equivalence classes of fragments to the metabolic networks

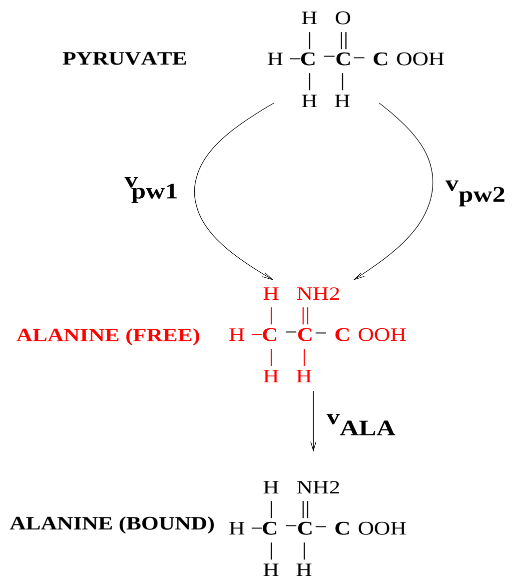
Equivalent fragments and the classes can be found in polynomial time via so called dominator tree analysis



Balance equations for fragments

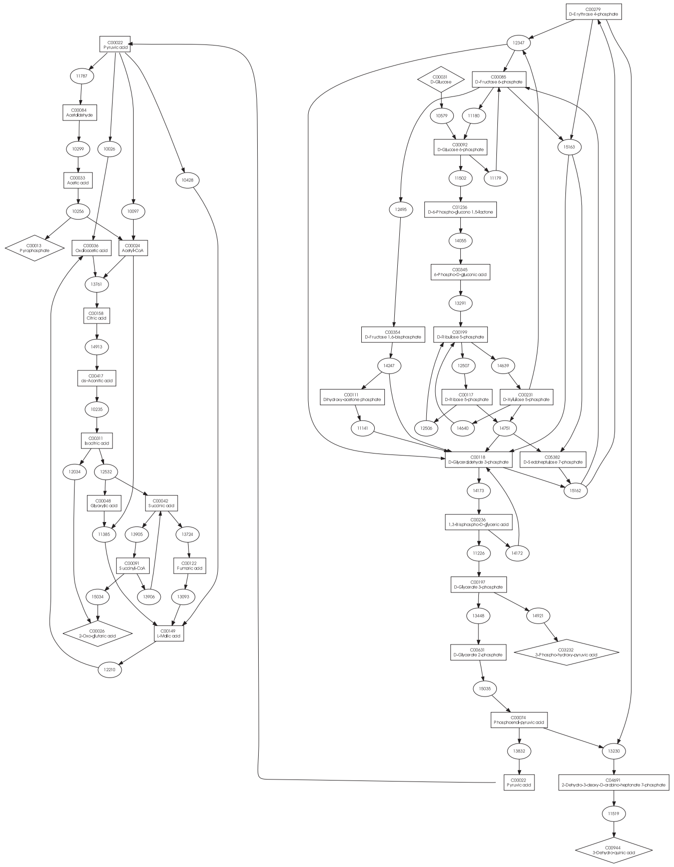
If a junction metabolite is a boundary of a equivalence sets of some fragment, the fragment potentially carries useful infomation for the flux estimation

We can write balance equations for the marginal isotopomer distribution of the fragment, some of which are linearly independent, thus constraining the fluxes:

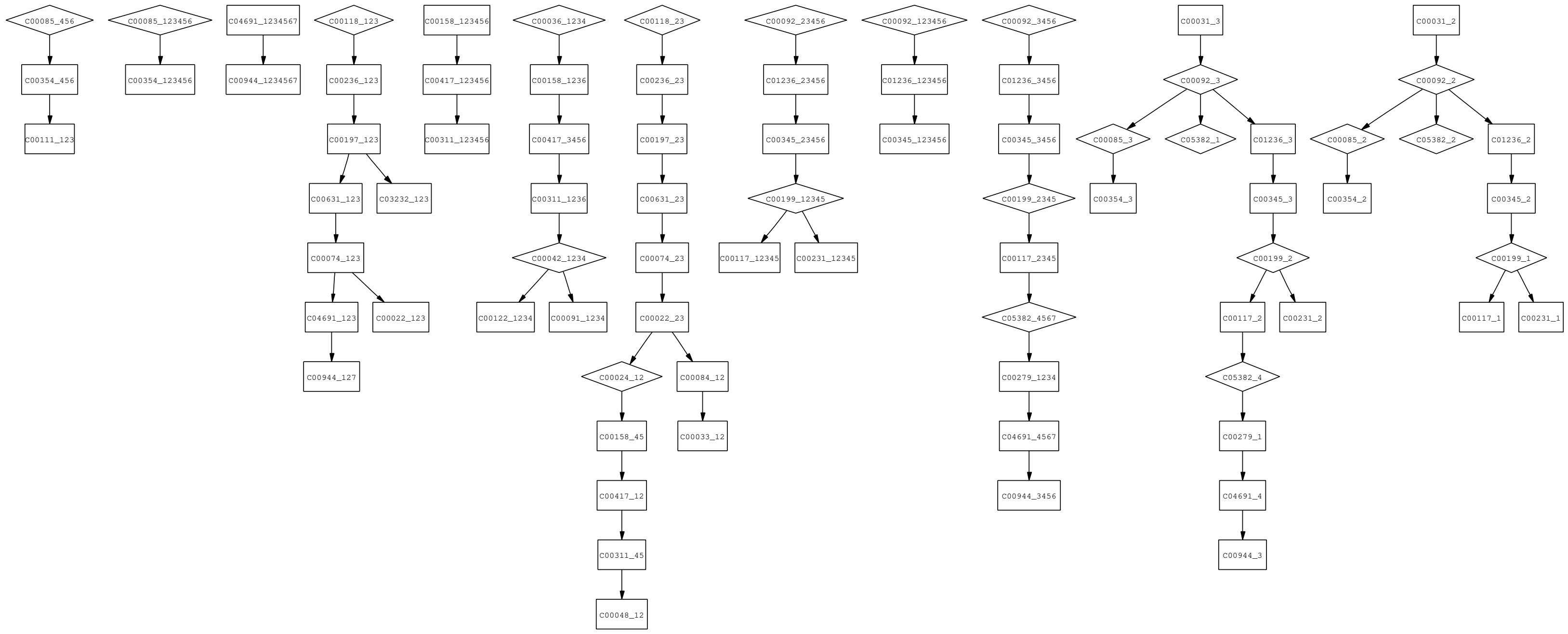


$$P(^{00}ALA_{23}|pw1) \cdot v_{pw1} + P(^{00}ALA_{23}|pw2) \cdot v_{pw2} = P(^{00}ALA_{23}) \cdot v_{ALA}$$
$$P(^{01}ALA_{23}|pw1) \cdot v_{pw1} + P(^{01}ALA_{23}|pw2) \cdot v_{pw2} = P(^{01}ALA_{23}) \cdot v_{ALA}$$
$$P(^{10}ALA_{23}|pw1) \cdot v_{pw1} + P(^{10}ALA_{23}|pw2) \cdot v_{pw2} = P(^{10}ALA_{23}) \cdot v_{ALA}$$
$$P(^{11}ALA_{23}|pw1) \cdot v_{pw1} + P(^{11}ALA_{23}|pw2) \cdot v_{pw2} = P(^{11}ALA_{23}) \cdot v_{ALA<}$$

Example network: Glycolysis of baker's yeast



Equivalence classes



Flux estimation from incomplete isotopomer data

So far we have assumed that isotopomer distributions can either be completely measured or not at all.

In practice, we are faced with incomplete isotopomer data:

- MS data: our PIDC software generally groups some isotopomers together so we get data like

$$aP({}^{010}ALA) + bP({}^{100}ALA) = d$$

- NMR measurements require ${}^{13}C$ in a specific position e.g. the middle carbon in alanine

$$\frac{P({}^{010}ALA)}{\sum_{x1y} P({}^{x1y}ALA)} = d.$$

Isotopomer measurements as linear constraints

Complete isotopomer distribution, NMR and mass spectrometric data, and the **absence** of isotopomer information all can be expressed as a set of linear constraints to the isotopomer distribution.

So we model the measurements as sets of equations

$$\vec{s}^T \mathbb{I}_{ALA} = \sum_{xyz} s_{xyz} \mathbb{P}\{^{xyz}ALA\} = d$$

to the isotopomer distribution $\mathbb{I}_{ALA}$, represented as a matrix system

$$S^T \mathbb{I}_{ALA} = \vec{d}$$

Vector space interpretation

- An n-carbon metabolite M is associated with a 2^n-dimensional **isotopomer vector space** $\mathcal{I}_M$, that has a coordinate axis for each isotopomer
- An isotopomer distribution $\mathbb{I}_{Ala} = [\mathbb{P}\{^{000}Ala\}, \mathbb{P}\{^{001}Ala\}, \ldots, \mathbb{P}\{^{111}Ala\}] \in [0,1]^{2^3}$, is a point in $\mathcal{I}_{Ala}$ and lies in the intersection of 8 hyperplanes of the form $e_{xyz}^T \mathbb{I}_{Ala} = \mathbb{P}\{^{xyz}ALA\}$ where e_{xyz} is the unit vector along the coordinate axis of isotopomer ^{xyz}Ala.

Fragment subspaces

- A n'-carbon fragment of a n-carbon metabolite defines a $2^{n'}$-dimensional subspace of the 2^n-dimensional isotopomer space
- The fragment subspace is spanned by vectors that corresponds to the fragment marginals of the isotopomer distribution, e.g. for Ala_{12} we have four basis vectors $u_{00} = [e_{000} + e_{001}]/\sqrt{2}, u_{01} = [e_{010} + e_{011}]/\sqrt{2}, u_{10} = [e_{100} + e_{101}]/\sqrt{2}$ and $u_{11} = [e_{110} + e_{111}]/\sqrt{2}$
- A fragment marginal of the isotopomer distribution is a point in the subspace spanned by the above basis vectors, given as the intersection of hyperplanes

$$u_{xy}^T \mathbb{I}_{Ala} = \mathbb{P}\{^{xy}Ala_{12}\}$$

which are collectively written as

$$U^T \mathbb{I}_{Ala} = \mathbb{I}_{Ala_{12}}$$

Fragment marginal as an orthogonal projection

The set of basis vectors $u_{00}, u_{01}, u_{10}, u_{11}$ is orthogonal: e.g.

$$\begin{aligned}2u_{00}^T u_{01} &= (e_{000} + e_{001})^T (e_{010} + e_{011}) = \\ &= e_{000}^T e_{010} + e_{000}^T e_{011} + e_{001}^T e_{010} + e_{001}^T e_{011} = 0 \quad (1)\end{aligned}$$

by the orthogonality of the unit vectors e_{xyz}

Thus the matrix

$$U = [u_{00} u_{01} u_{10} u_{11}]$$

is orthonormal and

$$U^T \mathbb{I}_{Ala} = \mathbb{I}_{Ala_{12}}$$

can be seen as the orthogonal projection of the original isotopomer distribution to the fragment subspace.

Mass spectrometric data

- In mass spectrometers, molecules with equal mass will reside in a single peak
- Thus, mass spectrum will group isotopomers with the same number of labels.
- Thus we will get data of the form

$$m_0^T \mathbb{I}_{Ala} = e_{000}^T \mathbb{I}_{Ala} = d_0$$

$$m_1^T \mathbb{I}_{Ala} = (e_{001} + e_{010} + e_{100})^T \mathbb{I}_{Ala} = d_1$$

$$m_2^T \mathbb{I}_{Ala} = (e_{011} + e_{101} + e_{110})^T \mathbb{I}_{Ala} = d_2$$

$$m_3^T \mathbb{I}_{Ala} = (e_{111} \mathbb{I}_{Ala} = d_3$$

  which is called the mass isotopomer distribution

- The vectors $m_0, \ldots, m_3$ are again an orthogonal set spanning a 'measurement' subspace of the isotopomer space
- The vector $(d_0, d_1, d_2, d_3)^T$ can be seen to be an orthogonal projection of the (unknown) isotopomer distribution to the measurement subspace

Tandem mass spectrometry

- Tandem mass spectrometers fragment the molecule to be measured.
- Thus one can measure the mass isotopomer distribution of fragments in addition to that of the whole molecule

$$m_0'^T \mathbb{I}_{Ala_{12}} = u_{00}^T \mathbb{I}_{Ala_{12}} = d_0$$

$$m_1'^T \mathbb{I}_{Ala_{12}} = (u_{01} + u_{10})^T \mathbb{I}_{Ala_{12}} = d_1$$

$$m_2'^T \mathbb{I}_{Ala_{12}} = u_{11}^T \mathbb{I}_{Ala_{12}} = d_2$$

- Typically, (at least some of) the vectors m_i' are linearly independent from the vectors m_i. Thus they constrain the isotopomer distribution more than the 'vanilla' MS spectrum would

NMR measurements

- NMR measurements require ^{13}C in a specific position e.g. the middle carbon in alanine
- The form of the data is normalized abundances of such specific isotopomers

$$\frac{P(^{010}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d.$$

- The data can be represented in the isotopomer space:

$$P(^{010}ALA) = d \sum_{x1y} P(^{x1y}ALA)$$

$$n^T \mathbb{I}_{Ala} = (d, \ldots, d-1, \ldots, d)^T \mathbb{I}_{Ala} = 0$$

- Thus, NMR data also introduced linear constraints to the isotopomer distribution

Measurements in general

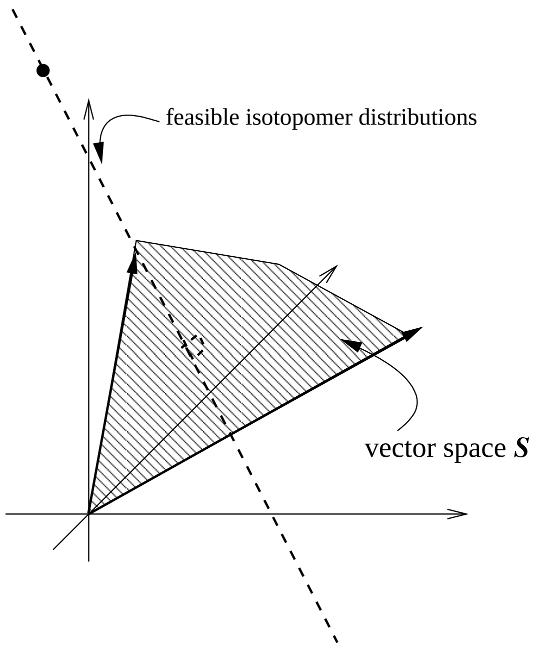
- So we can model the measurements as sets of equations

$$\vec{s}^T \mathbb{I}_{ALA} = \sum_{xyz} s_{xyz} \mathbb{P}\{^{xyz}ALA\} = d$$

  to the isotopomer distribution $\mathbb{I}_{ALA}$, represented as a matrix system

$$S^T \mathbb{I}_{ALA} = \vec{d}$$

- If matrix S is orthonormal, the interpretation is an orthogonal projection of the isotopomer distribution to the measurement subspace



Mapping measurement data to metabolite fragments

- Our goal is to propagate the measurement data within the equivalance set of the fragments
- However, in general, the measurements are made for metabolites not their fragments, so some translation is needed.
- For example, suppose we have the following measurement from NMR:

$$\frac{P(^{010}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d_1, \frac{P(^{011}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d_2,$$

$$\frac{P(^{110}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d_3, \frac{P(^{111}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d_4$$

Mapping measurement data to metabolite fragments

- For Ala_{12} we get the following isotopomer constraints:

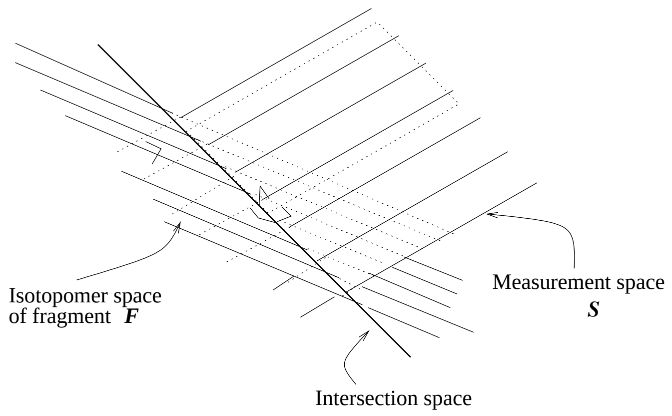
$$\frac{P(^{01}Ala_{12})}{\sum_{x1} P(^{x1}Ala_{12})} = \frac{P(^{010}Ala) + P(^{011}Ala)}{\sum_{x1y} P(^{x1y}Ala)} = d_1 + d_2,$$

$$\frac{P(^{11}Ala_{12})}{\sum_{x1} P(^{x1}Ala_{12})} = \frac{P(^{110}Ala) + P(^{111}Ala)}{\sum_{x1y} P(^{x1y}Ala)} = d_3 + d_4$$

- What is the general approach, assuming arbitrary linear equation set as the measurement?

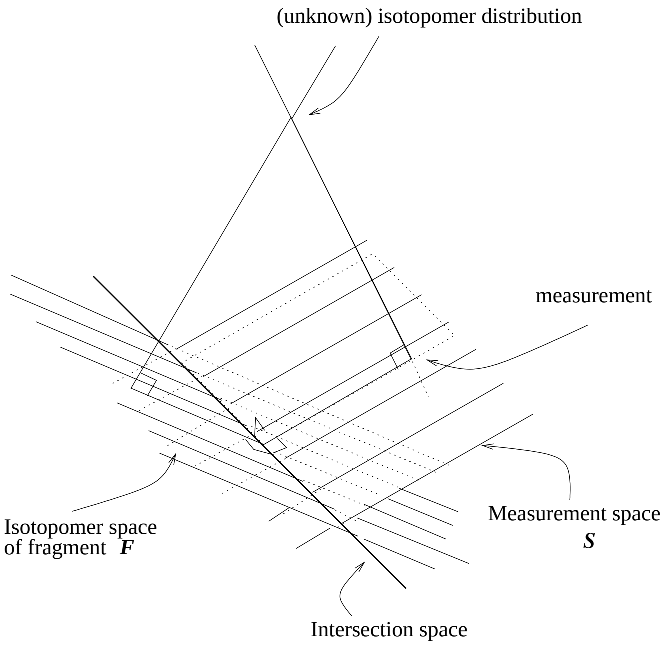
Mapping measurements to fragments

- All fragment distributions lie within a subspace of the isotopomer space of the metabolite
- Measurements lie in another subspace, the measurement space.
- What is common between the measurement and the fragment can be expressed in terms of the intersection space



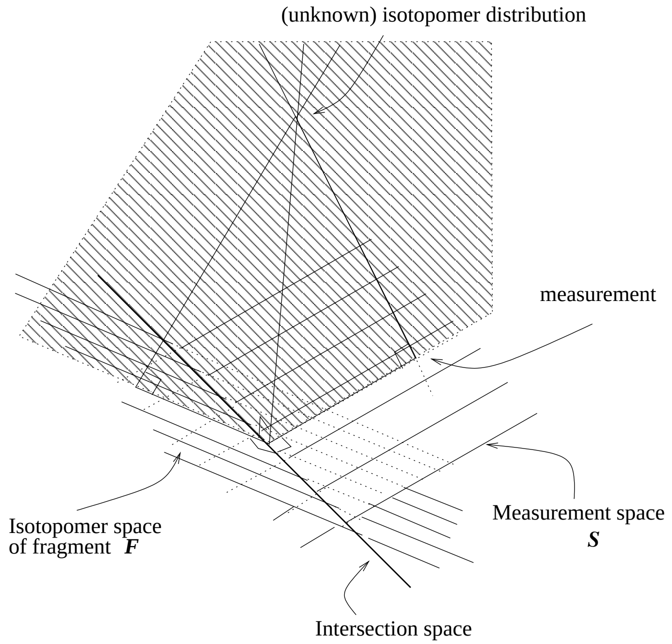
Mapping measurements to fragments

- What is known about the underlying isotopomer distribution is its projection to the measurement space
- The fragment marginal distribution is a projection to the fragment subspace
- The projection to the intersection space represented what is known about the fragment marginal based on the measurement



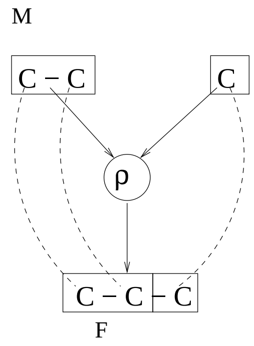
Mapping measurements to fragments

- The uncertainty about the isotopomer distribution translates to uncertainty about the fragment distributions
- The smaller the intersection space, the less is known about the fragment
- The measurement completely determines the fragment marginal if and only if the fragment subspace is a subspace of the measurement space
- Both computing the intersection space and the projections to it can be written in terms of linear algebra



Isotopomers of a product metabolite

- In the complete information case, we could compute the isotopomer distribution of a product from the distributions of the fragments via $P({}^{xyz}M_3) = P({}^{xy}M_1)P({}^{z}M_2)$
- In the case of incomplete information an analogous procedure can be used, independently for each pair of isotopomer constraints of the reactants



$$\sum_{xyz} c_{xyz} P({}^{xyz}M_3) = \left(\sum_{xy} a_{xy} P({}^{xy}M_1)\right) \cdot \left(\sum_{z} b_z P({}^{z}M_2)\right),$$

where $c_{xyz} = a_{xy} b_z$

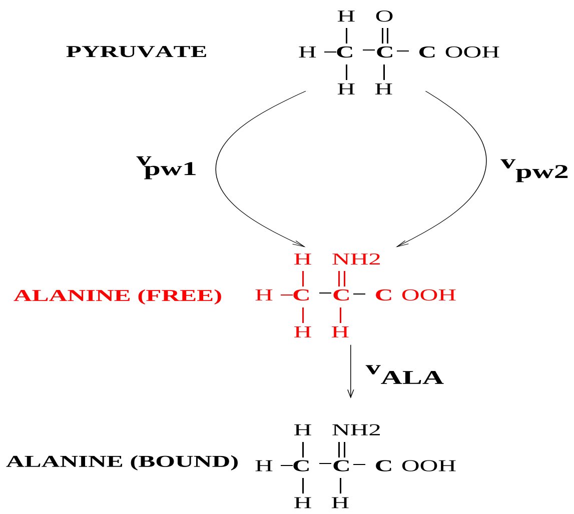
Generalized isotopomer balances

Via the above described approach we can obtain isotopomer constraints for fluxes around a junction.

Assume we have the follwing constraints $S^T \mathbb{I}_{ALA|pw_1} = \vec{d_1}$, $S^T \mathbb{I}_{ALA|pw_2} = \vec{d_2}$ and $S^T \mathbb{I}_{ALA} = \vec{d}$.

In a steady state get a generalized isotopomer balance

$$\vec{d_1} \cdot v_{pw_1} + \vec{d_2} \cdot v_{pw_2} = \vec{d} \cdot v_{ALA}$$

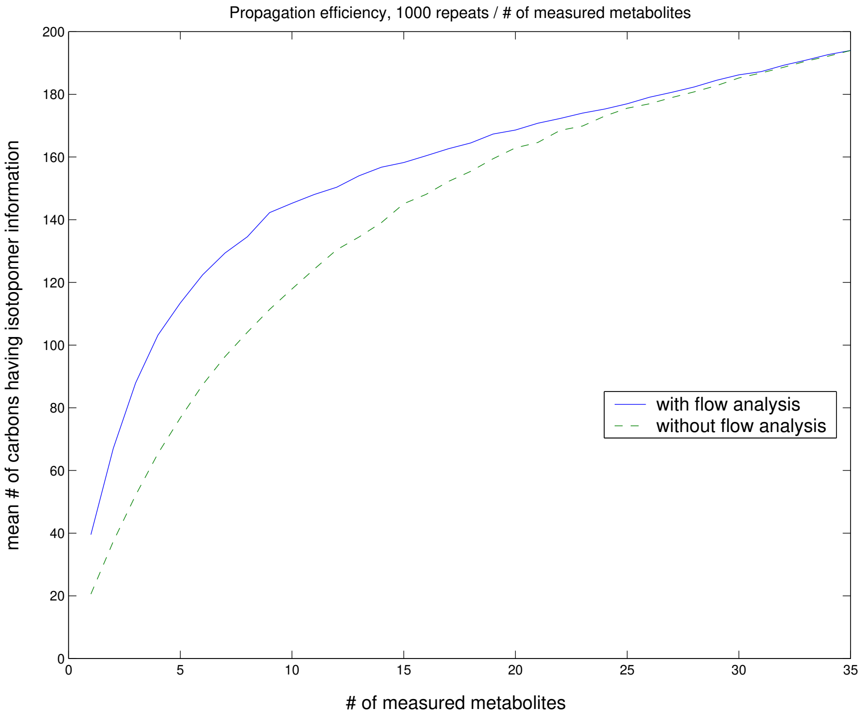


General recipe for flux estimation

- Compute equivalence sets of fragments for the metabolic network
- Project the isotopomer measurements of the metabolites to the fragments
- Propagate the obtained isotopomer constraints accross the equivalence classes
- Form balance equations for fragments at the boundaries of the equivalence classes
- Solve the resulting linear equation system

Identifiability of the fluxes

The success of the flux estimation approach depends heavily on the amount of isotopomer measurements taken from the metabolism.

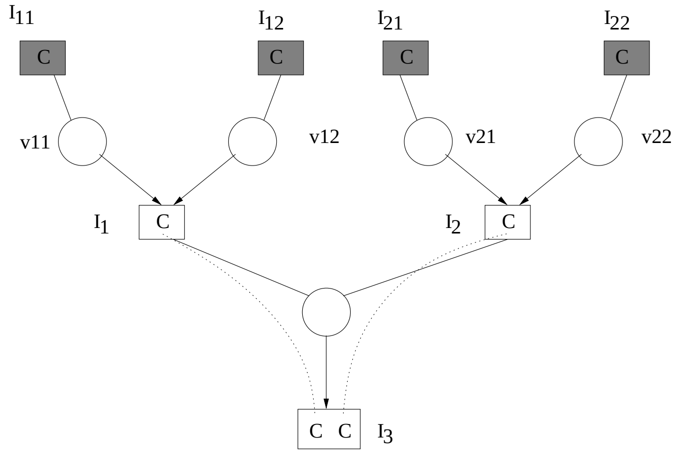


Incompleteness of the linear approach

- The preceding approach for flux estimation works in the linear framework
- We propagate isotopomer constraints accross the equivalence classes in order to compute balance equations that are linear in the fluxes
- However, given incomplete isotopomer measurements one can find situations where a non-linear balance equation can be written while a linear cannot.

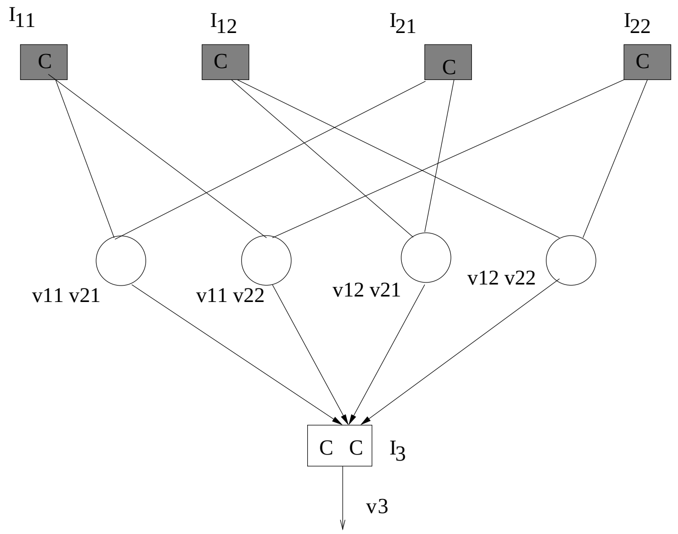
Non-linear example

- Consider situation where we want to estimate $\mathbb{I}_3$ given $\mathbb{I}_{11}, \mathbb{I}_{12}, \mathbb{I}_{21}$ and $\mathbb{I}_{22}$ with no measurement from the intermediate metabolites
- The carbons above the junctions are not equivalent with carbons below the junctions, so it is not possible to estimate $\mathbb{I}_3$ independently from the fluxes
- The example contains four elementary flux modes, corresponding to four different combinations of source carbons to the two-carbon target



Non-linear example

- We can draw and equivalent example consisting of the four elementary flux modes
- A quadratic equation can be written:
  $v_3\mathbb{I}_3 = v_{11}v_{21}\mathbb{I}_{11}\mathbb{I}_{21} + v_{11}v_{22}\mathbb{I}_{11}\mathbb{I}_{22} + v_{12}v_{21}\mathbb{I}_{12}\mathbb{I}_{21} + v_{12}v_{22}\mathbb{I}_{12}\mathbb{I}_{22}$



Isotopomer systems in general

- The above example could be tackled by switching from linear equation solver to a quadratic solver
- However, there is no principal limit to the degree of the balance equation, so we are faced with a non-linear systems of arbitrary degree, if we want to utilize isotopomer information to the fullest

Iterative approach

An alternative approach to solving fluxes is based on iteratively solving the isotopomer system as follows:

1. Start with an initial guess of the flux vector $\mathbf{v}_0$
2. Compute a predicted isotopomer distributions $\tilde{\mathbb{I}}$ that should result, were the guess correct
3. Compare the predicted isotopomer distributions to the measured ones $\hat{\mathbb{I}}$
4. If the distance $\| \tilde{\mathbb{I}} - \hat{\mathbb{I}} \|$ is small enough, stop and return the current guess flux vector
5. Otherwise, generate a new flux guess $\mathbf{v}$ and continue from step 2.

Properties of the iterative approach

- Given fixed flux guess, the predicted isotopomer distibutions can be exactly computed
- For generation of flux guesses many kinds of search methods can be used
- May be difficult to make sure what the degree of underdetermination of the flux vector is

References

- Rousu, J., Rantanen, A., Maaheimo, H., Pitkänen E., Saarela, K., Ukkonen, E.: A method for estimating metabolic fluxes from incomplete isotopomer information. Computational Methods in Systems Biology, Proc. First Intl. Workshop, CMSB 2003, Lecture notes in Computer Science 2602 (2003), 88–103.
- Rantanen, A., Rousu, J., Pitkänen, E., Maaheimo, H. and Ukkonen, E. Equivalence of metabolite fragments and flow analysis of isotopomer distributions for flux estimation. Transactions on Computational Systems Biology VI, Lecture Notes in Bioinformatics 4220, Journal subline (2006), pp. 198 - 220.