Metabolic flux estimation

- So far in this course we have examined techniques that help us understanding the cell's capabilities:
  - Given genome, what kind of metabolic network (Metabolic reconstruction)
  - Given metabolic network, what kind of behaviour is possible (Flux balance analysis, elementary flux modes)
- Now we turn to a different question: how to analyze quantitatively the activity of metabolic pathways
  - Given some measurements and the stoichiometry, estimate flux vector v

The flux estimation problem

- In flux estimation the goal is to restrict the space of solutions of the steady equations

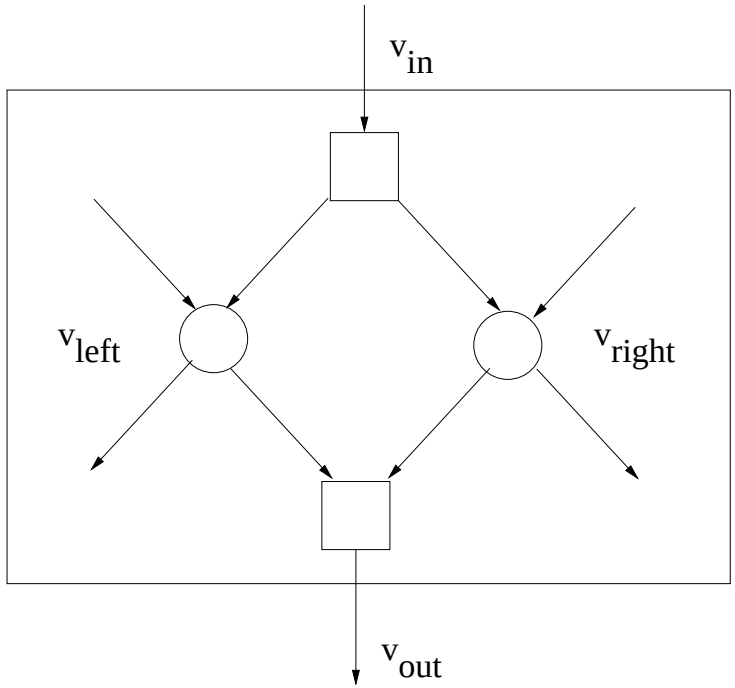
$$S\mathbf{v} = 0$$

- Ideally, a single rate vector v is left as the solution
- In practise, we will need to resort to constraining the set of solutions in the null space $\mathcal{N}(S)$.

The problem with alternative routes

- 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 just by measuring exchange fluxes, only their sum can be solved.

- In this case the null space of $S_{unknown}$ is non-empty, thus there is a choice of flux vectors that satisfy steady state



Material balance

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

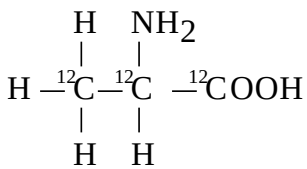
Isotope tracing experiments

- Isotope tracing experiments are the most accurate tool for estimating the fluxes of alternative pathways
- In isotope tracing experiments the cell culture is fed a mixture of natural and ^{13}C labeled substrate (e.g. 90%/10%).
- The fate of the ^{13}C labels is followed by measuring the intermediate metabolites by mass spectrometry or NMR
- From the enrichment of labels in the intermediates the fluxes are inferred

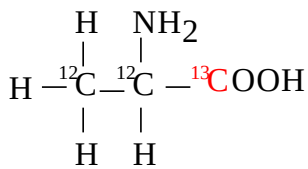
^{13}C-Isotopomers

- 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** (isotopic isomers):

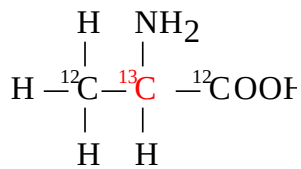
000Ala



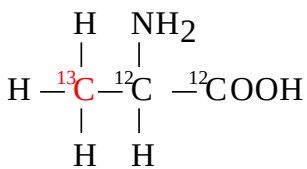
001Ala



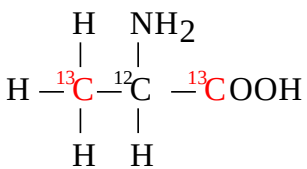
010Ala



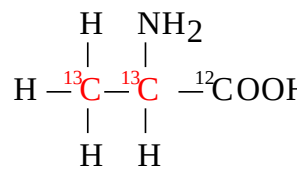
100Ala



101Ala

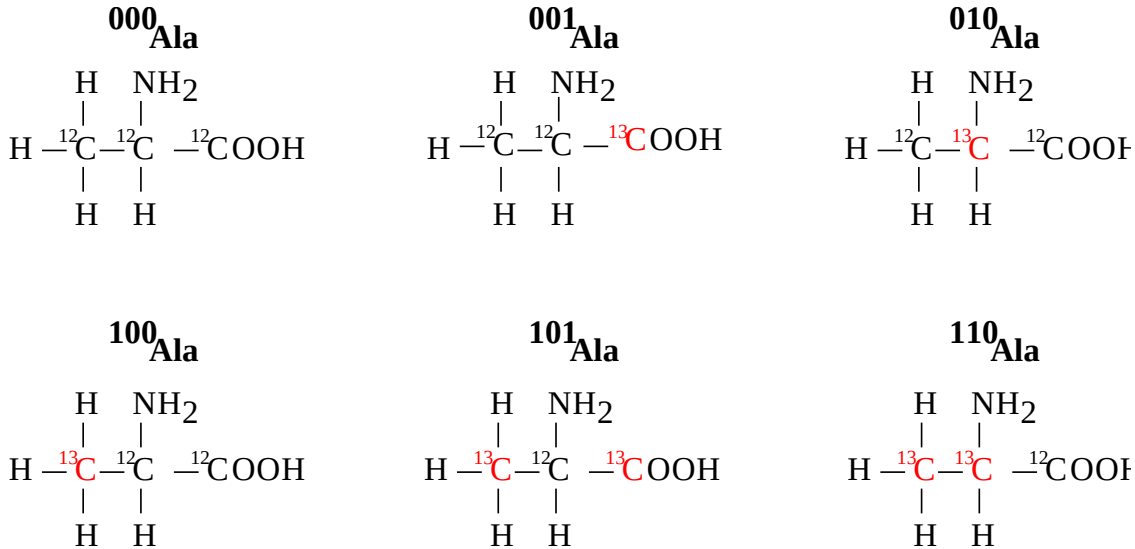


110Ala



Isotopomers and alternative pathways

- 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**
- Isotopomer distributions can give information about the fluxes of alternative pathways *if* the pathways manipulate the carbon chains of the metabolites differently

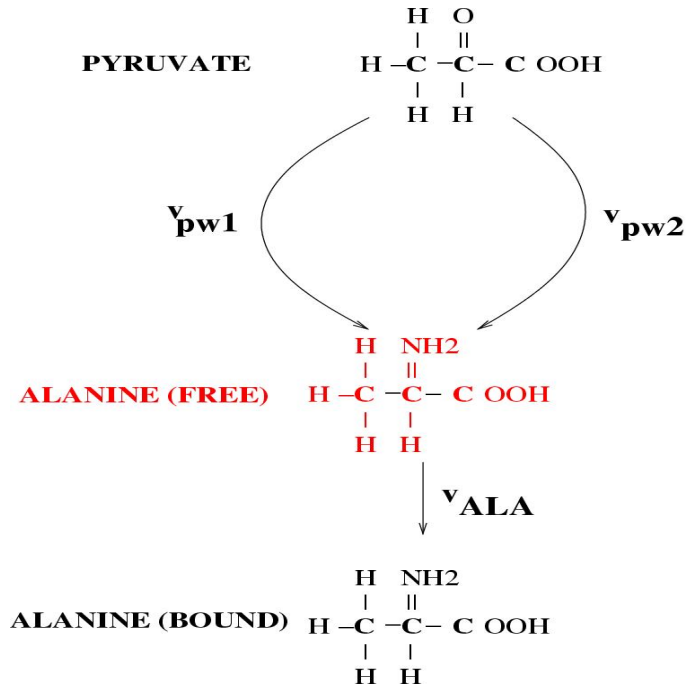


Isotopomeric balance equations

- The steady state condition for free alanine implies:

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

- The steady state assumption needs to hold for each isotopomer separately
- We can write balance equations for each isotopomer:



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

Flux estimation from incomplete isotopomer data

In practice, we are faced with incomplete isotopomer data:

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

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

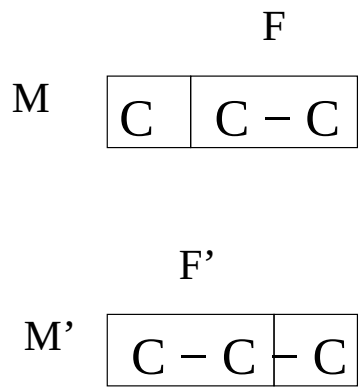
  - NMR measurements require ^{13}C in a specific position e.g. the middle carbon in alanine

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

We start by tackling the first difficulty.

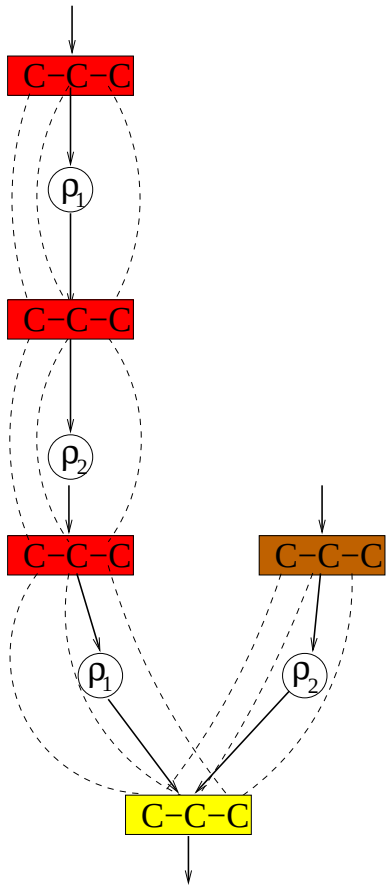
Fragment equivalence

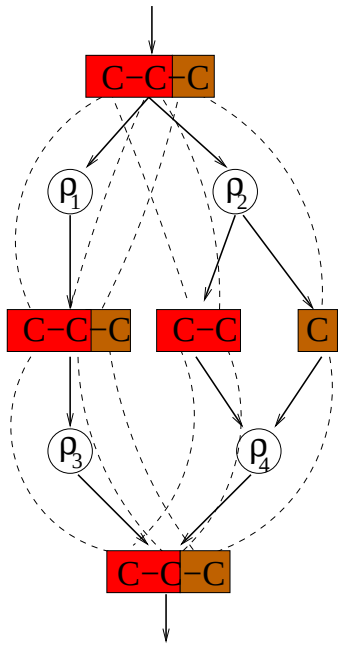
- Two fragments $F \subseteq M$ and $F' \subseteq M'$ are *equivalent* if the fragment marginal distributions of the respective isotopomer distributions of M and M' are equal, irrespectively of the fluxes of the metabolic network
- When does the fragment equivalence hold true?



Fragment equivalence in general

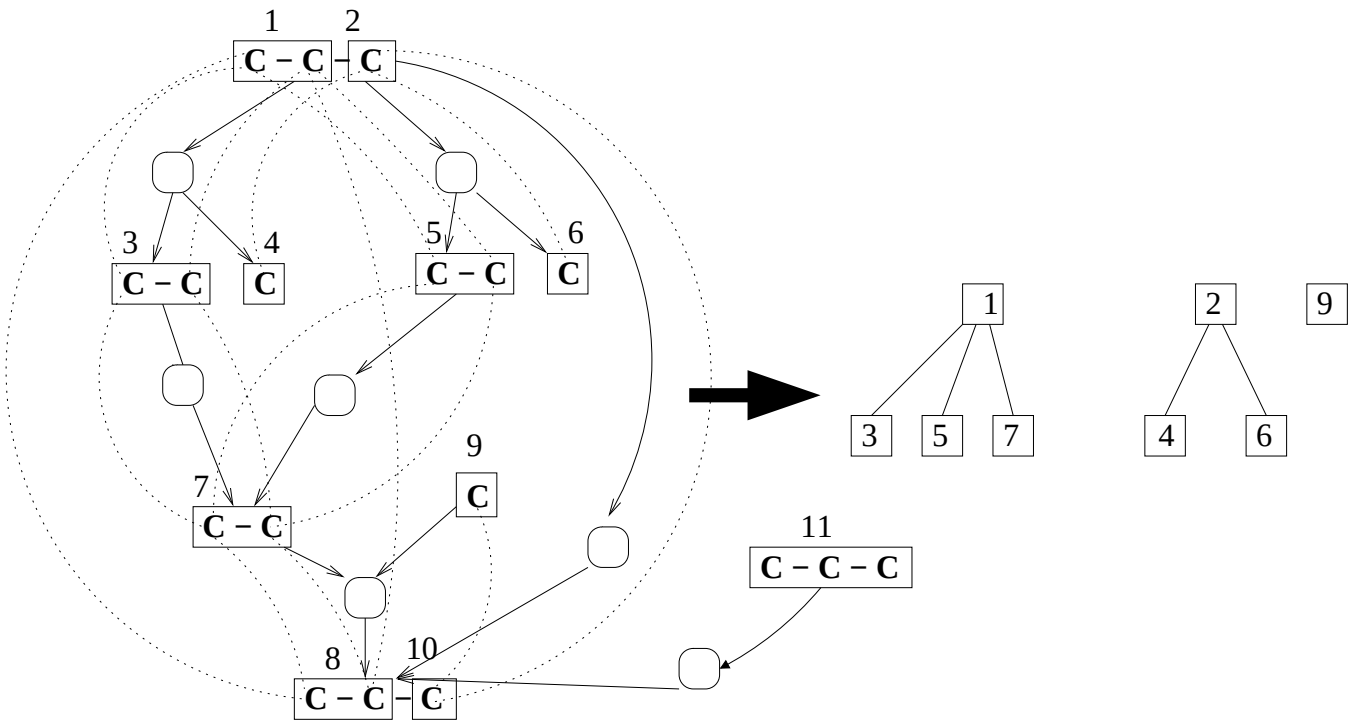
- Assume fragments produced by alternative pathways travel *intact* and *similarly oriented* (i.e. no permutation) starting from the common source fragment
- The isotopomer distribution of that fragment remain equivalent to the source along the alternative pathways





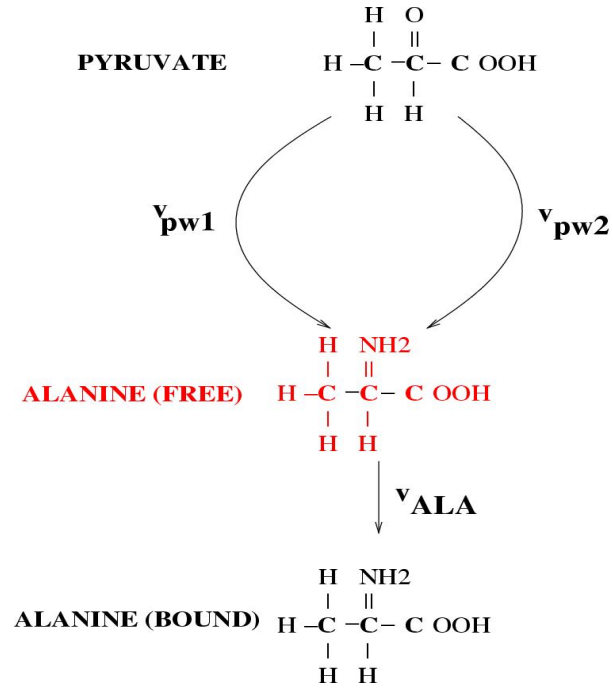
Equivalence classes

- The equivalance relation for fragments induces equivalence classes of fragments to the metabolic networks
- The isotopomer distribution is the theoretically the same for the whole equivalence class



Balance equations for fragments

- Assume we have deduced fragment marginals of ALA_{12} for both pathways
- Balance equations for the fragment ALA_{12}:



$$P(^{00}ALA_{12}|pw1) \cdot v_{pw1} + P(^{00}ALA_{12}|pw2) \cdot v_{pw2} = P(^{00}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{01}ALA_{12}|pw1) \cdot v_{pw1} + P(^{01}ALA_{12}|pw2) \cdot v_{pw2} = P(^{01}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{10}ALA_{12}|pw1) \cdot v_{pw1} + P(^{10}ALA_{12}|pw2) \cdot v_{pw2} = P(^{10}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{11}ALA_{12}|pw1) \cdot v_{pw1} + P(^{11}ALA_{12}|pw2) \cdot v_{pw2} = P(^{11}ALA_{12}) \cdot v_{ALA_{12}}$$

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

$$\mathbf{a}^T \mathbb{I}_{ALA} = \sum_{xyz} a_{xyz} \mathbb{P}\{^{xyz}ALA\} = d$$

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

$$A^T \mathbb{I}_{ALA} = \mathbf{d}$$

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 $e_{xyz}^T \mathbb{I}_{Ala} = \mathbb{P}\{^{xyz}ALA\}$
- e_{xyz} is the unit vector along the coordinate axis of isotopomer ^{xyz}Ala, i.e. $e_{000} = (1,0,0,\ldots,0)$, $e_{001} = (0,1,0,\ldots,0)$, $e_{111} = (0,\ldots,0,1)$

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.

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

  by the orthogonality of the unit vectors e_{xyz}
- The vectors have unit length $||u_{xy}|| = 1$
- Thus the matrix

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

  is orthonormal
- The matrix equation

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

  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

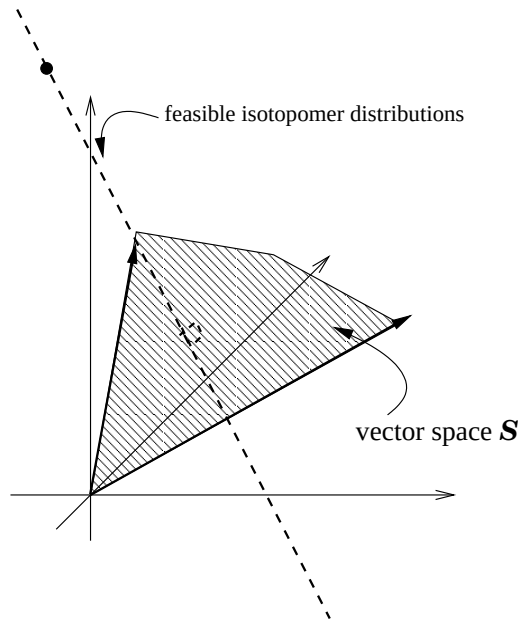
- So we can model the measurements as sets of equations

$$\mathbf{a}^T \mathbb{I}_{ALA} = \sum_{xyz} a_{xyz} \mathbb{P}\{^{xyz}ALA\} = d$$

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

$$A^T \mathbb{I}_{ALA} = \mathbf{d}$$

- If matrix A 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:

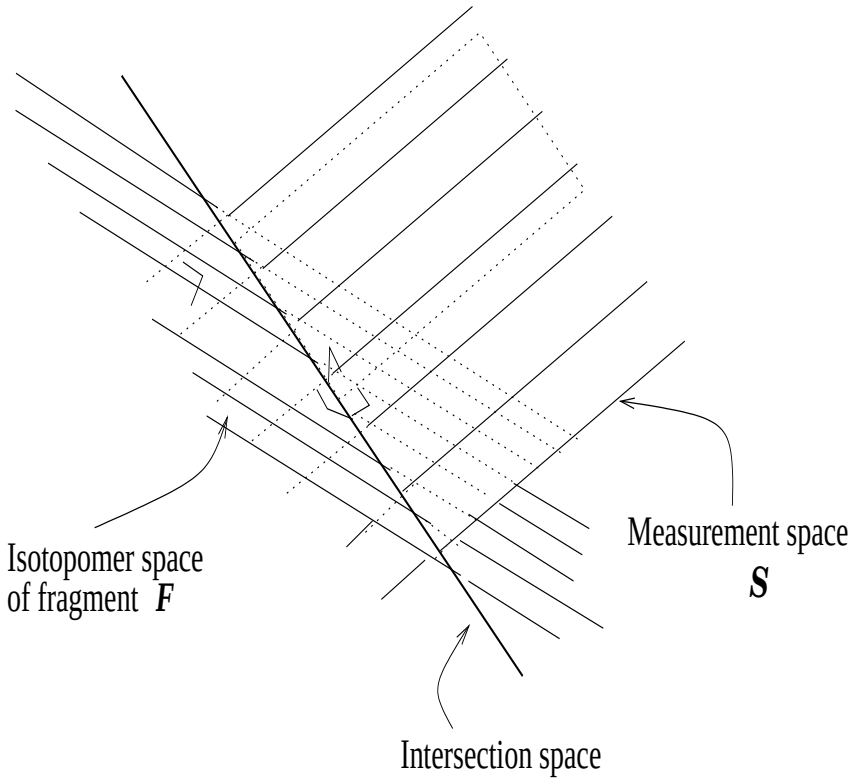
$$\frac{P(^{01}Ala_{12})}{\sum_{x1} P(^{x1}Ala_{12})} = \frac{P(^{010}Ala) + P(^{011}Ala)}{\sum_{x1y} P(^{x1y}Ala)} = d_1 + d_2,$$

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

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

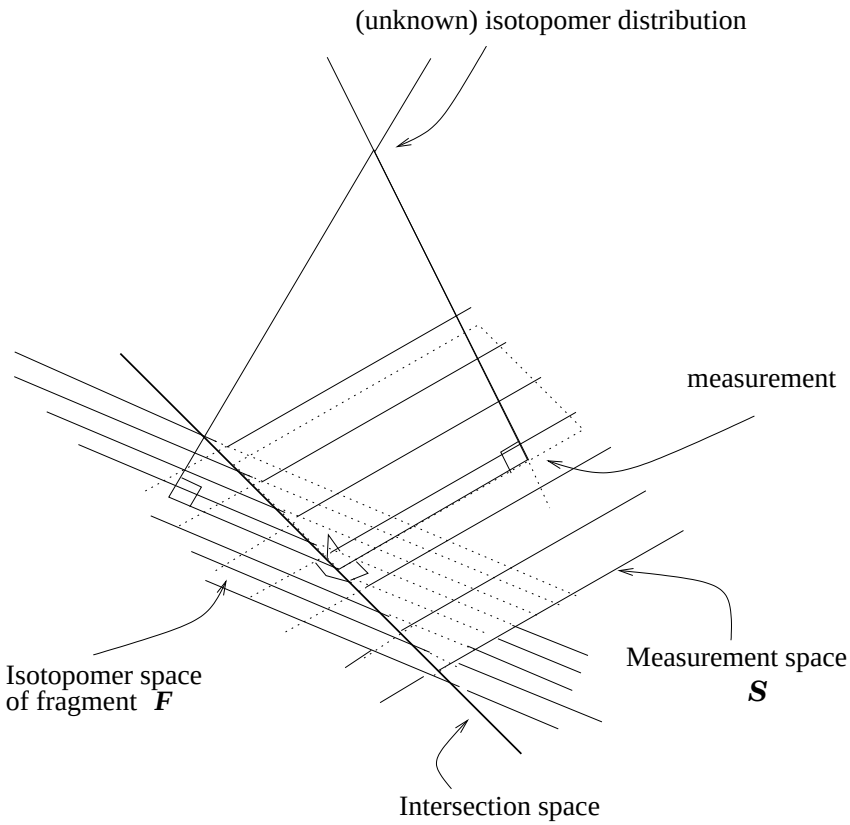
Mapping measurements to fragments

- 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



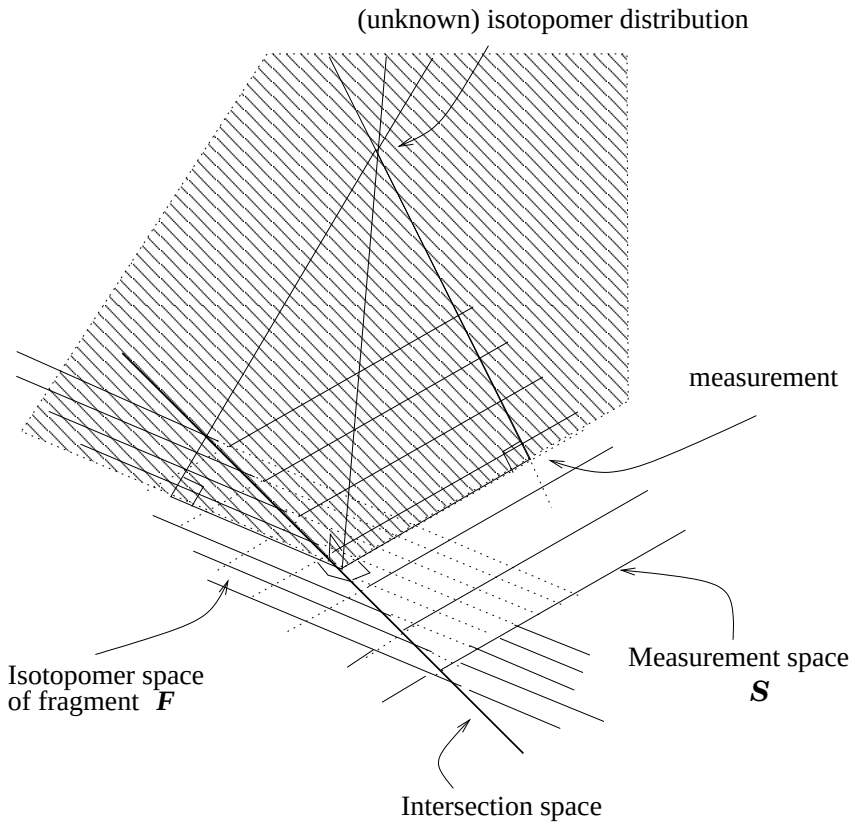
Mapping measurements to fragments

- 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



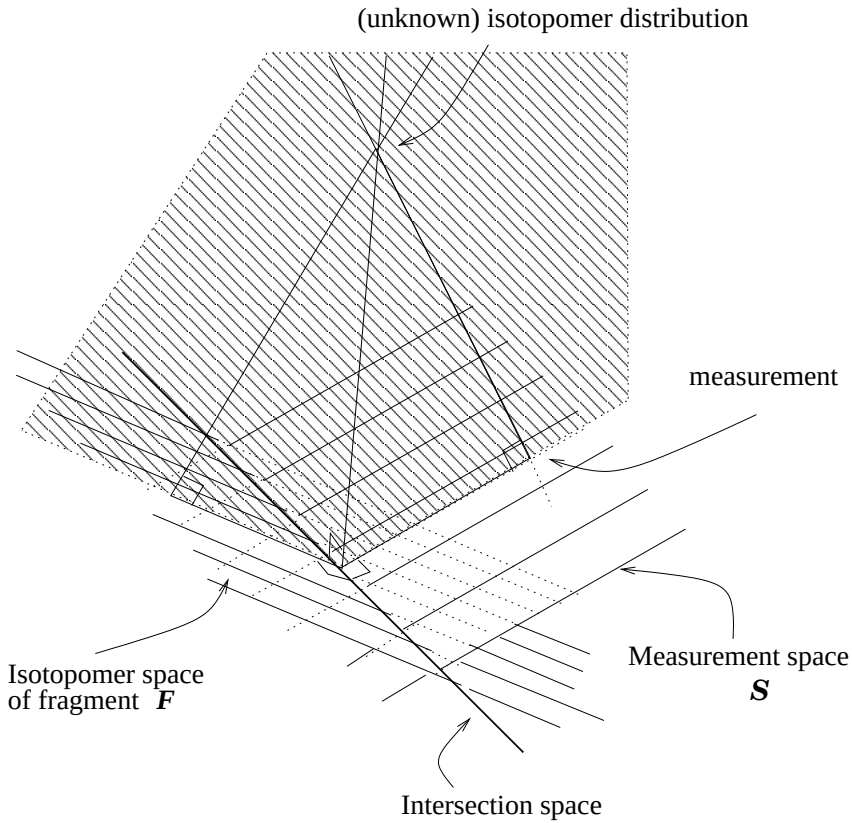
Mapping measurements to fragments

- The uncertainty about the isotopomer distribution translates to uncertainty about the fragment distributions
- The smaller the dimension of the intersection space, the less is known about the fragment



Mapping measurements to fragments

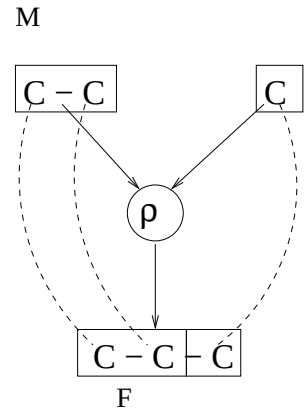
- The measurement completely determines the fragment marginal if and only if the fragment subspace is a subspace of the measurement space
- Both computing the intersection space and the projections to it can be written in terms of linear algebra



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

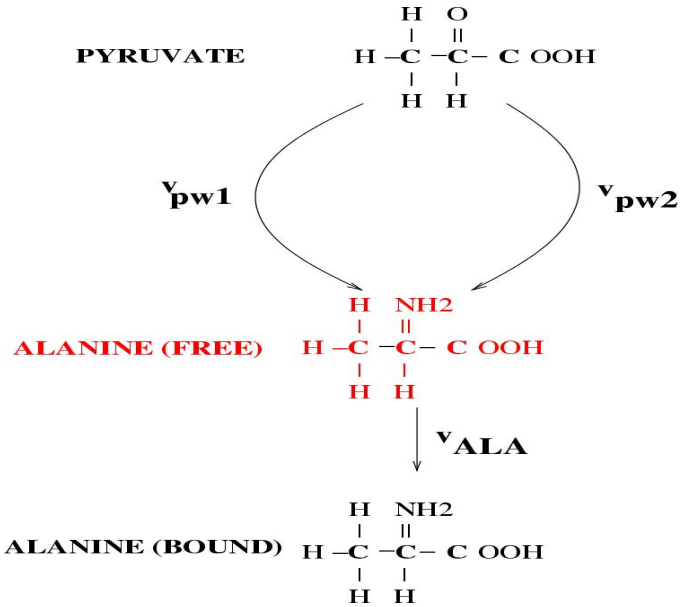
$$\sum_{xyz} c_{xyz} P(^{xyz}M_3) = \left(\sum_{xy} a_{xy} P(^{xy}M_1) \right) \cdot \left(\sum_{z} b_z P(^{z}M_2) \right),$$

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 $A^T \mathbb{I}_{ALA|pw_1} = \mathbf{d_1}$, $A^T \mathbb{I}_{ALA|pw_2} = \mathbf{d_2}$ and $A^T \mathbb{I}_{ALA} = \mathbf{d}$.
In a steady state get a generalized isotopomer balance

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

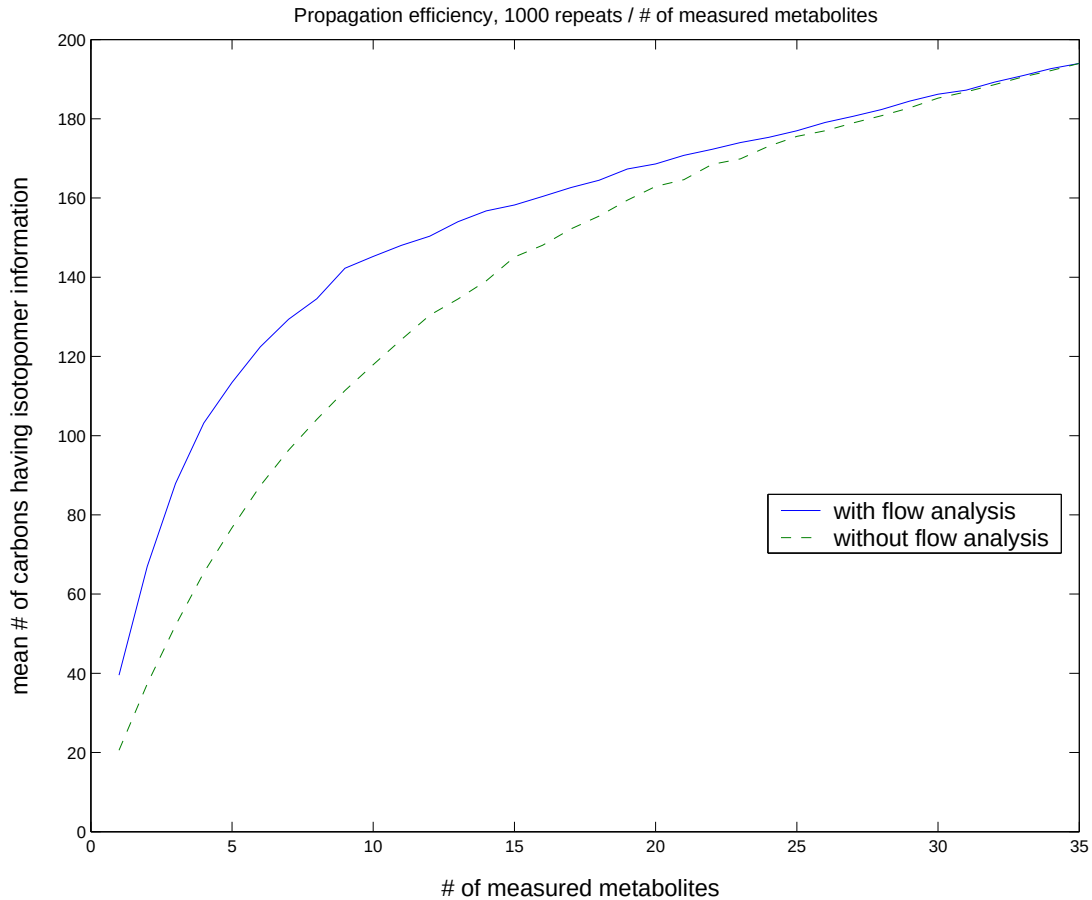


General recipe for flux estimation

- 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.



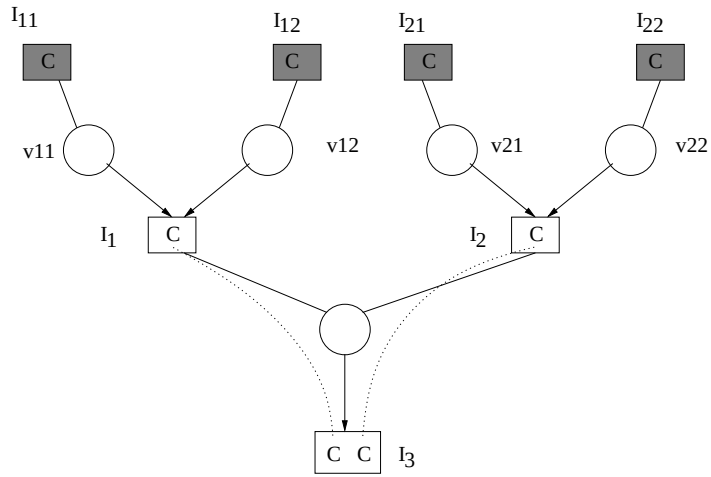
Incompleteness of the linear approach

- 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

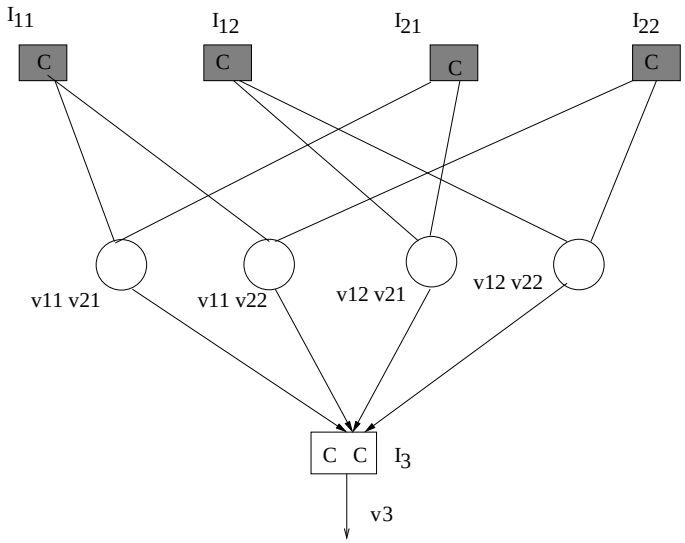
- Consider situation where we want to estimate $\mathbb{I}_3$ given $\mathbb{I}_{11}, \mathbb{I}_{12}, \mathbb{I}_{21}$ and $\mathbb{I}_{22}$ with no measurement from the intermediate metabolites
- The carbons above the junctions are not equivalent with carbons below the junctions, so it is not possible 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 balance equation that is quadratic in the fluxes 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 $\left|\left|\tilde{\mathbb{I}} - \hat{\mathbb{I}}\right|\right|$ 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