■

Flux estimation using

incomplete isotopomer information

■

Juho Rousu, Ari Rantanen and Esko Ukkonen

■

■

■

Flux estimation using incomplete isotopomer information

Juho Rousu, Ari Rantanen and Esko Ukkonen

Department of Computer Science
P.O. Box 26, FIN-00014 University of Helsinki, Finland
{Juho.Rousu,Ari.Rantanen,Esko.Ukkonen}@cs.Helsinki.FI
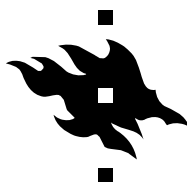
Abstract

Metabolic flux analysis – finding out the rates of reactions in metabolic pathways – is an important problem areas in the study of metabolism. The most accurate technique for this task today is the use of isotopic tracer experiments, where a mixture of differently isotope-labeled substrates is fed to a cell and the propagation of the labels is observed from the products and intermediate metabolites, where possible. We present a generic methodology for solving the fluxes of a metabolic network. The method differs from most previous approaches by not making prior assumptions about the topology of the metabolic network. Also, only very mild assumptions are made about the available measurement data, for example, positional enrichment and mass isotopomer data can be used side by side.

Computing Reviews (1998) Categories and Subject Descriptors:
G.1.3   Numerical Linear Algebra
J.3   Life and Medical Sciences

General Terms:
Algorithms, Experimentation, Theory

Additional Key Words and Phrases:
Systems biology, Metabolic flux estimation

Contents

Chapter 1

Introduction

Metabolic flux estimation – the problem of determining intracellular reaction velocities – is an important problem in metabolic engineering [17], where the goal is to optimize the production of some metabolite in microbial cells. Knowledge of the steady-state reaction velocities – the fluxes – along different pathways is a prerequisite for pathway optimization [12]. Flux estimation has potential to be an important tool in a wider context of systems biology, for example, characterizing the physiology of the organism [6] and metabolic reconstruction, where one aims at reverse engineering the metabolic network of an unfamiliar organism [10, 3, 4, 1, 15].

Currently the most appealing framework for flux estimation is the use of isotopic tracer experiments, where one feeds the cell culture with a predefined mixture of natural and ^{13}C-labeled nutrients. The fate of the ^{13}C atoms can be observed by measuring the *isotopomer distributions* of metabolic products and intermediates [16] with NMR [18, 8] or mass spectrometer [2, 20, 5, 11].

The mathematical tools for relating the isotopomer distributions of metabolites to the fluxes have developed over the last decade. On the other hand, the existing linear programming framework for flux analysis, relying on elementary balances of chemical compounds, has been extended to isotopomer analysis [9, 13, 17, 7, 19] to analyze key pathways in metabolism. This approach attempts to solve the fluxes explicitly. However, a rigorous theory of these extensions, enabling automatic flux estimation given an arbitrary metabolic network, has been lacking.

On the other hand, an iterative approach to flux estimation have been studied by Wiechert, Möllney, Isermann, Wurzel and de Graaf [19]. This framework, to the authors' knowledge, is the only existing method that is not tied to a particular metabolic topology. The method for computing the isotopomer distributions of metabolites based on the guessed flux distribution is particularly intriguing. However, although the iterative method can propose flux values in many cases where the direct approach fails, it is not trivial to assess the quality of the proposed solutions.

In this paper, we propose a generic methodology for flux estimation in a direct manner, so that the (possibly partial) solution is given in terms of linear constraints to the flux values, coupled with estimates of the error. The proposed

framework is completely free of topological assumptions; the user can input any network topology to base the analysis upon.

In addition, we do not make any *prior* assumptions of the kind and quality of isotopomer and metabolite data, although the quality of the solution is affected by the data. In particular, we allow any set of linear constraints to the isotopomer distribution as measurements. Both NMR and mass spectrometric data can be used side by side in this framework.

The paper is organized as follows. Chapter 2 presents notation and concepts to be used throughout the article. Also, we review the linear modelling approach to flux estimation, assuming complete isotopomer information. Chapter 3 presents the mathematical techniques that are used to compute estimates of isotopomer distributions of products of a reaction from the reactants and vice versa. The novelty in these techniques is that the isotopomer distributions of the metabolites need not to be fully determined. In Chapter 4 we present algorithms to propagate isotopomer information towards metabolic junctions, using the techniques described in the previous section. In Chapter 5 we discuss the potential and the limitations of the methodology and point out routes for improving the framework.

Chapter 2

Preliminaries

2.1   Metabolic networks

A *metabolic network* is a four-tuple $G = (\mathcal{M}, \mathcal{M}_I, \mathcal{L}, \mathcal{R})$, where $\mathcal{M} = \{M_1, \ldots, M_m\}$ is a set *metabolites*, $\mathcal{M}_I \subset \mathcal{M}$ is the set of *internal metabolites* – metabolites that are not uptaken nor excreted by the cell, $\mathcal{L} = \{1, \ldots, L\}$ is a set of (carbon) *locations*, $\mathcal{R} = \{\rho_1, \ldots, \rho_n\}$ is a set of *reactions*.

In a slight abuse of notation, we associate a metabolite M with the set of its carbon locations, that is, $M = \{M(1), \ldots, M(|M|)\} \subset \mathcal{L}$. Each location $l \in \mathcal{L}$ exists in exactly one metabolite $M \in \mathcal{M}$.

A reaction $\rho_j = (\vec{\alpha}_j, \lambda_j) \in \mathcal{R}$ consists of a vector $\vec{\alpha}_j \in \mathbb{Z}^m$ of stoichiometric coefficients $\alpha_{ji} \in \mathbb{Z}$, denoting the number of molecules of metabolite M_i consumed or produced in one reaction event, and a *carbon mapping* $\lambda_j : \mathcal{L}_{j,r} \mapsto \mathcal{L}_{j,p}$ from the set of *reactant locations* $\mathcal{L}_{j,r} = \cup_{\alpha_{ji}<0} M_i$ to the set of *product locations* $\mathcal{L}_{j,p} = \cup_{\alpha_{ji}>0} M_i$. Metabolites M_i with $\alpha_{ji} < 0$ and $\alpha_{ji} > 0$ are called *reactants* and *products* of ρ_j, respectively. The mapping $\lambda_j(l) = l'$ denotes that in the reaction a carbon in location l of the reactant $M \ni l$ is transferred to the location l' of the product $M' \ni l' \mathcal{M} : l' \in M$. The *reverse reaction* of $\rho = (\vec{\alpha}, \lambda)$ is the pair $\rho^{-1} = (-\vec{\alpha}, \lambda^{-1})$, where λ^{-1} is the inverse[1] of λ. We will assume that reactions keep their reactant fragments intact, changes are confined to the fragment borders. Note that one can always represent a reaction that does not fulfill the assumption by a sequence of reactions that do.

For reaction ρ_j, the product locations $F' \subset \mathcal{L}_{j,p}$ corresponding to a subset of reactant locations $F \subset \mathcal{L}_{j,r}$ are given by the *image* $F' = \lambda_j(F) = \{\lambda_j(l) | l \in F\}$ of F. Correspondingly, the reactant locations F corresponding to a subset of product locations F are given by the *pre-image* $F = \lambda_j^{-1}(F') = \{\lambda_j^{-1}(l') | l' \in F'\}$ of F'.

A *reactant fragment* is a set of locations $F \subset M_i$ that in ρ_j is destined to a single product $M_{i'}$, defined by $F = F_j^{-1}(i, i') = M_i \cap \lambda_j^{-1}(M_{i'})$. Correspondingly, a *product fragment* is a set of locations $F' \subset M_{i'}$ that originate from a single reactant M_i, defined by $F' = F_j(i, i') = M_{i'} \cap \lambda_j(M_i)$.

---

[1]The assumption that λ is invertible, or even a function, is not always valid, e.g. with symmetric molecules and reactions consuming or producing more than one molecule of the same metabolite. We treat these cases separately later on.
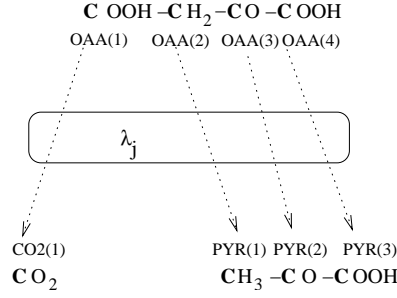
Figure 2.1: Carbon mapping for a reaction converting oxaloacetate into puryvate and carbondioxide.

Example 1 *In Figure 2.1 the carbon mapping for a reaction ρ_j converting oxaloacetate into pyruvate and carbondioxide is shown. There are four reactant locations*

$$\mathcal{L}_{j,r} = \{OAA(1), OAA(2), OAA(3), OAA(4)\} = OAA$$

all belonging to oxaloacetate, and four product locations

$$\mathcal{L}_{j,p} = \{CO_2(1), PYR(1), PYR(2), PYR(3)\} = CO_2 \cup PYR$$

belonging to carbondioxide and pyruvate.

We have $\lambda(OAA) = \mathcal{L}_{j,p}$ since OOA is the sole reactant. The product fragments are $F_j(OAA, CO_2) = \lambda_j(OAA) \cap CO_2 = CO_2$ and $F_j(OAA, PYR) = \lambda_j(OAA) \cap PYR = PYR$. The preimages of product locations are $\lambda_j^{-1}(CO_2) = \{OAA(1)\}$ and $\lambda^{-1}(PYR) = \{OAA(2), OAA(3), OAA(4)\}$. These are reactant fragments as well since both sets contain indices from exactly one reactant.

The *metabolic state* of the cell population is described by the triple $\sigma = (G, \vec{v}, \mathtt{C})$, where G is a metabolic network, vector $\vec{v} = [v_1, \ldots, v_n] \in \mathbb{R}^n$ is a set of *reaction velocities* in G and $\mathtt{C} = \{\mathtt{C}_1, \ldots, \mathtt{C}_L\}$ is the collection of subsets $\mathtt{C}_l \subset \Omega$ of carbons occupying location $l \in \mathcal{L}$. A *molecule* of metabolite M is then a sequence of carbons $\mu = (\mathtt{c}_1, \ldots, \mathtt{c}_{|M|})$, $\mathtt{c}_i \in \mathtt{C}_{M(i)}$. The set of molecules composes a *metabolite pool* $\mathtt{C}_M \subset \mathtt{C}_{M(1)} \times \cdots \times \mathtt{C}_{M(|M|)}$. Each carbon in location $M(i)$ is tied to some molecule $\mu \in \mathtt{C}_M$.

2.2   Steady-state modelling of metabolic networks

The methods presented in this article are subject to the following assumptions about the underlying biological system.

(A-1) The reactions draw their reactants independently, uniformly randomly from the respective pools.

(A-2) The reaction velocities are the same for the whole population and do not change over time.

The first assumption implies that the metabolic system is completely mixed so that the *spatial dimensions* can be overlooked (i.e. no concentration gradients exist). In essence, the cell population is viewed as homogenous, the differences between metabolic states of different cells are averaged out. This assumption is consistent with the measurement data; that also represents an average of the cell population.

The second assumption is the so called *steady-state assumption.* It is a crucial one for the methods described in this article. The steady state assumption relieves us from considering the *temporal dimension* and does away with the difficult dynamics. The downside of course is that only steady states can be handled, for example data from batch cultivations is difficult to analyze as such.

Consider now a sequence of metabolic states $\sigma(t_0), \sigma(t_1), \ldots, \sigma(t_k)$, $\sigma(t_i) = (G, \vec{v}(t_i), \mathtt{C}(t_i))$ corresponding to time points $t_0 < \cdots < t_k$. By assumption (A-2) we have $\vec{v}(t_0) = \vec{v}(t_1) = \cdots = \vec{v}(t_k)$. Furthermore, the assumption indirectly implies that $|\mathtt{C}_M(t_0)| = |\mathtt{C}_M(t_1)| = |\mathtt{C}_M(t_k)|, M \in \mathcal{M}$, that is, the sizes of the metabolite pools remain constant over time. This is because otherwise some metabolite pools could exhaust as time elapses, which would force the system to alter some of the reaction rates. The intermediates cannot accumulate inside the cell either because the obvious physical limitations.

Let us write down this assumption more formally. In a metabolic steady-state, the following balance is assumed to hold true for all intermediate metabolites M_i

$$\sum_{j=1}^{n} \alpha_{ji} v_j = 0, \tag{2.1}$$

where v_j is the reaction rate of λ_j and α_{ji} is the stoichiometric coefficient of metabolite M_i in the reaction ρ_j.

Balance equations of the above kind can be constructed for every intermediate metabolite. For external metabolites, the production or consumption rate can usually be measured. In that case we get equations like

$$\sum_{j=1}^{n} \alpha_{ji} v_j = \beta_i, \tag{2.2}$$

where $\beta_i \in \mathbb{R}$ is the measured net rate of consumption or production of metabolite M_i. Balance equations can be collected into a matrix equation

$$A\vec{v} = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,j} & \cdots & \alpha_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha_{i,1} & \cdots & \alpha_{i,j} & \cdots & \alpha_{i,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha_{m,1} & \cdots & \alpha_{m,j} & \cdots & \alpha_{m,n} \end{bmatrix} \times \begin{bmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_m \end{bmatrix} = \vec{\beta} \tag{2.3}$$

where there is a column for each reaction and a row for each metabolite. This equation system is usually under-determined, that is, there usually is more than one flux vector $\vec{v}$ that satisfies (2.3). This happens when there exists alternative routes between a pair of metabolites, that is, one metabolic route can 'mimic' the other with no net difference in the flows in or out of the metabolic system.

2.3   Balance equations and perfect isotopomer data

The problem of under-determination of the matrix (2.3) is tackled by using isotopic tracing experiments. Let us first define some vocabulary by which the principles of isotope tracing can then be explained.

We assume that the basic set of carbons Ω is divided into two disjoint, non-empty subsets, the *isotopes*[2] ^{12}C and ^{13}C.

We associate with each location a random variable *label* $\ell_l : \mathtt{c}_l \mapsto \{0, 1\}$, given by

$$\ell_l(\mathtt{c}) = \begin{cases} 0, & \text{if } \mathtt{c} \in {}^{12}\text{C}, \\ 1, & \text{if } \mathtt{c} \in {}^{13}\text{C}. \end{cases}$$

The *labeling* of a sequence of carbons $(\mathtt{c}_1, \ldots, \mathtt{c}_k)$ - This can be, for example, a metabolite or its fragment – occupying locations $F = \{F(1), \ldots, F(k)\}$ is the sequence $\ell_{F(1)}(\mathtt{c}_1) \cdots \ell_{F(k)}(\mathtt{c}_k) \in \{0, 1\}^k$. To denote different labelings of a set of carbons we use the shorthands $^\mathbf{b}F = {}^jF$, where $\mathbf{b} = \mathbf{b_1} \cdots \mathbf{b_k}$ is the k bits long the binary representation of number j.

Molecules $\mu = (\mathtt{c}_1, \ldots, \mathtt{c}_{|M|})$ and $\mu' = (\mathtt{c}'_1, \ldots, \mathtt{c}'_{|M|})$ of metabolite M that have different labelings (i.e. for some location $M(i) \in M$, $\ell_{M(i)}(\mathtt{c}_i) \neq \ell_{M(i)}(\mathtt{c}'_i)$) are called *isotopomers*. If for some $F = \{M(f_1), \ldots, M(f_{|F|})\} \subset M$, the labelings of the corresponding fragments $\mu_F = (\mathtt{c}_{f_1}, \ldots, \mathtt{c}_{f_{|F|}})$, and $\mu'_F = (\mathtt{c}'_{f_1}, \ldots, \mathtt{c}'_{f_{|F|}})$, the molecules are called *cumulative* (F)-*isotopomers* or (F-)*cumomers* of M.

The probability of randomly drawing from the pool C_M a molecule with labeling $\mathbf{b}$, that is, the event

$$\{\ell_{F(1)}(\mathtt{c}_1) = \mathbf{b_1}\} \wedge \cdots \wedge \{\ell_{F(k)}(\mathtt{c}_k) = \mathbf{b_k}\},$$

is given by

$$\mathbb{P}\{^\mathbf{b}M\} = \frac{|\{(\mathtt{c}_1, \ldots, \mathtt{c}_{|M|}) \in \mathtt{C}_M | \ell_{M(i)}(\mathtt{c}_i) = \mathbf{b_i}, \forall M(i) \in M\}|}{|\mathtt{C}_M|}.$$

The *isotopomer distribution* of the metabolite is then the vector

$$\mathbb{I}_M = [\mathbb{P}\{^0M\}, \mathbb{P}\{^1M\} \ldots, \mathbb{P}\{^{2^{|M|}-1}M\}]^T \in [0, 1]^{2^{|M|}}.$$

The isotopomer distribution of a metabolite *fragment* $F = \{F(1), \ldots F(|F|)\} \subset M$, also called *cumulative isotopomer distribution* or *cumomer distribution* [19], is defined in analogous manner by

$$\mathbb{P}\{^\mathbf{b}F\} = \frac{|\{(\mathtt{c}_1, \ldots, \mathtt{c}_{|M|}) \in \mathtt{C}_M | \ell_{F(i)}(\mathtt{c}_j)) = \mathbf{b_i}, \forall i, j : F(i) = M(j)\}|}{|\mathtt{C}_M|}$$

and $\mathbb{I}_F = [\mathbb{P}\{^0F\}, \mathbb{P}\{^1F\} \ldots, \mathbb{P}\{^{2^{|F|}-1}F\}]^T \in [0, 1]^{2^{|F|}}$. In accordance to [19], we refer to all F-cumomer distributions, where $|F| = k$, as k-cumomer distributions.

---

[2]We ignore the radioactive ^{14}C.

The term cumulative isotopomer is justified as follows. For a F-cumomer $\mathbf{b} = \mathbf{b_1} \cdots \mathbf{b_{|F|}}$ of metabolite M, we denote by

$$\mathcal{U}(\mathbf{b}) = \{\mathbf{a} \in \{0,1\}^{|M|} \,|\, \mathbf{a_i} = \mathbf{b_j} \forall i,j : M(i) = F(j)\}$$

the set of labelings of M that contain the labeling pattern $\mathbf{b}$ within the locations F. It is easy to see that $|\mathcal{U}(\mathbf{b})| = 2^{|M|-|F|}$ for any $\mathbf{b}$. We can write

$$\mathbb{P}\{{}^{\mathbf{b}}F\} = \sum_{\mathbf{a} \in \mathcal{U}(\mathbf{b})} \mathbb{P}\{{}^{a}M\} \tag{2.4}$$

In other words, the cumomer frequency is a sum of isotopomer frequencies. The equation is a re-statement of the basic probability theoretic result between a joint distribution of random variables and the marginal distributions related to its subsets. Equation (2.4) can be expressed in matrix form as $\mathbb{I}_F = U_F^T \mathbb{I}_M$, where

$$[U_F]_{ij=}[U_F]_{\mathbf{ab}} = \begin{cases} 1 & \text{if } \mathbf{a} \in \mathcal{U}(\mathbf{b}) \\ 0 & \text{otherwise.} \end{cases}$$

Example 2 *The isotopomer distribution of alanine ($C3NH7O2$) is the vector*

$$\mathbb{I}_{Ala} = [\mathbb{P}\{{}^{000}Ala\}, \mathbb{P}\{{}^{001}Ala\}, \ldots, \mathbb{P}\{{}^{110}Ala\}, \mathbb{P}\{{}^{111}Ala\}]^T.$$

The 1-cumomer distributions of alanine are the following: The $Ala(1)$-cumomer distribution is

$$\mathbb{I}_{Ala(1)} = \begin{bmatrix} \mathbb{P}\{{}^{0}Ala(1)\} \\ \mathbb{P}\{{}^{0}Ala(1)\} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \mathbb{I}_{Ala} =$$

$$= \begin{bmatrix} \mathbb{P}\{{}^{000}Ala\} + \mathbb{P}\{{}^{001}Ala\} + \mathbb{P}\{{}^{010}Ala\} + \mathbb{P}\{{}^{011}Ala\} \\ \mathbb{P}\{{}^{100}Ala\} + \mathbb{P}\{{}^{101}Ala\} + \mathbb{P}\{{}^{110}Ala\} + \mathbb{P}\{{}^{111}Ala\} \end{bmatrix},$$

the $Ala(2)$-cumomer distribution is

$$\mathbb{I}_{Ala(2)} = \begin{bmatrix} \mathbb{P}\{{}^{0}Ala(2)\} \\ \mathbb{P}\{{}^{1}Ala(2)\} \end{bmatrix} = \begin{bmatrix} \mathbb{P}\{{}^{000}Ala\} + \mathbb{P}\{{}^{001}Ala\} + \mathbb{P}\{{}^{100}Ala\} + \mathbb{P}\{{}^{101}Ala\} \\ \mathbb{P}\{{}^{010}Ala\} + \mathbb{P}\{{}^{011}Ala\} + \mathbb{P}\{{}^{110}Ala\} + \mathbb{P}\{{}^{111}Ala\} \end{bmatrix},$$

and the $Ala(3)$-cumomer distribution is

$$\mathbb{I}_{Ala(3)} = \begin{bmatrix} \mathbb{P}\{{}^{0}Ala(3)\} \\ \mathbb{P}\{{}^{1}Ala(3)\} \end{bmatrix} = \begin{bmatrix} \mathbb{P}\{{}^{000}Ala\} + \mathbb{P}\{{}^{010}Ala\} + \mathbb{P}\{{}^{100}Ala\} + \mathbb{P}\{{}^{110}Ala\} \\ \mathbb{P}\{{}^{001}Ala\} + \mathbb{P}\{{}^{011}Ala\} + \mathbb{P}\{{}^{101}Ala\} + \mathbb{P}\{{}^{111}Ala\} \end{bmatrix}.$$

The 2-cumomer distributions of alanine are the following

$$\mathbb{I}_{Ala(1,2))} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \cdot \mathbb{I}_{Ala} = \begin{bmatrix} \mathbb{P}\{{}^{000}Ala\} + \mathbb{P}\{{}^{001}Ala\} \\ \mathbb{P}\{{}^{010}Ala\} + \mathbb{P}\{{}^{011}Ala\} \\ \mathbb{P}\{{}^{100}Ala\} + \mathbb{P}\{{}^{101}Ala\} \\ \mathbb{P}\{{}^{110}Ala\} + \mathbb{P}\{{}^{111}Ala\} \end{bmatrix},$$

$$\mathbb{I}_{Ala(1,3)} = \begin{bmatrix} \mathbb{P}\{{}^{00}Ala(1,3)\} \\ \mathbb{P}\{{}^{01}Ala(1,3)\} \\ \mathbb{P}\{{}^{10}Ala(1,3)\} \\ \mathbb{P}\{{}^{11}Ala(1,3)\} \end{bmatrix} = \begin{bmatrix} \mathbb{P}\{{}^{000}Ala\} + \mathbb{P}\{{}^{010}Ala\} \\ \mathbb{P}\{{}^{001}Ala\} + \mathbb{P}\{{}^{011}Ala\} \\ \mathbb{P}\{{}^{100}Ala\} + \mathbb{P}\{{}^{110}Ala\} \\ \mathbb{P}\{{}^{101}Ala\} + \mathbb{P}\{{}^{111}Ala\} \end{bmatrix},$$

$$\mathbb{I}_{Ala(2,3)} = \begin{bmatrix} \mathbb{P}\{^{00}Ala(2,3)\} \\ \mathbb{P}\{^{01}Ala(2,3)\} \\ \mathbb{P}\{^{10}Ala(2,3)\} \\ \mathbb{P}\{^{11}Ala(2,3)\} \end{bmatrix} = \begin{bmatrix} \mathbb{P}\{^{000}Ala\} + \mathbb{P}\{^{100}Ala\} \\ \mathbb{P}\{^{001}Ala\} + \mathbb{P}\{^{101}Ala\} \\ \mathbb{P}\{^{010}Ala\} + \mathbb{P}\{^{110}Ala\} \\ \mathbb{P}\{^{011}Ala\} + \mathbb{P}\{^{111}Ala\} \end{bmatrix}.$$

The Ala(3, 3)-cumomer distribution is naturally the isotopomer distribution of alanine, $\mathbb{I}_{Ala}$.

In order to analyze the data originating from isotope tracing experiments in the linear programming framework, we make a some further assumptions, namely

(A-5) The system in in an isotopomeric steady-state, that is isotopomer distributions of the metabolites stay approximately constaint over time.

(A-6) The reaction rates are independent of the labeling of the carbons. That is, we assume that the same reaction rate holds for all isotopomers of the reactants.

(A-7) The isotopomer distribution of each metabolite is fully determined.

(A-8) The reactions in the network are bijective.

We relax the assumption A-7 in chapter 3; it is the main contribution area of this report. The assumption A-8 is relaxed in the section 3.4, which makes it possible to tackle symmetrical metabolites and reactions that consume of product more than one molecule of the same kind in one reaction event.

Given assumptions A-5 and A-6, for an internal metabolite M_i, we can write

$$\sum_{j=1}^{n} v_j \alpha_{ji} \mathbb{P}\{^{\mathbf{b}}M_{ji}\} = 0, \tag{2.5}$$

for any isotopomer $\mathbf{b} = \mathbf{b_1} \cdots \mathbf{b_{|M_i|}}$, where M_{ji} denotes the metabolites M_i entering ($\alpha_{ji} > 0$) or exiting ($\alpha_{ji} < 0$) the metabolite pool C_{M_i} via reaction ρ_j; if $\alpha_{ji} = 0$, we take the corresponding isotopomer probabilities as zero as well. Here, by assumption A-8, $\alpha_{i,j} \in \{-1, 0, 1\}$. The above equation states that the production and consumption rates need to be equal for each isotopomer individually, not just for the metabolite as a whole. The assumption requires care from the experimental point of view: samples from the metabolic system should be taken only after the isotopomer distributions have been stabilized. This may take a considerable time after a metabolic steady-state has been reached.

The balance equations (2.5) for individual isotopomers can be collected into a system of equations

$$\sum_{j=1}^{n} v_j \alpha_{ji} \mathbb{I}_{M_{ji}} = 0, \tag{2.6}$$

Assuming that all distributions $\mathbb{I}_{M_{ji}}$ are known, (2.6) is a linear equation where the only unknowns are fluxes v_j. Since the isotopomer distributions $\mathbb{I}_{M_{ji}}$ depend on how the reaction – and the pathway upto that point – manipulates the carbon chain, some of the equations within the system are typically linearly

independent. Therefore, the system constrains the fluxes around the metabolite M_i more so than the single metabolic balance equation (2.1).

Assuming that $\mathbb{I}_{M_i}$ is measured, and given that $\mathbb{I}_{M_{ji}} = \mathbb{I}_{M_i}$ for all consumers ρ_j of metabolite M_i – this holds because all consumers draw their reactants from the same metabolite pool – we only need to provide distribution estimates $\mathbb{I}_{M_{ji}}$ for the flows entering the junction at M_i; these flows have $\alpha_{ji} < 0$ in (2.6). Note that there is no way of measuring those distributions directly: all flows are inherently mixed. However, as we will see, given the isotopomer distributions of the reactants we can easily compute the isotopomer distribution of any product of that reaction

A bijective reaction $\rho = (\vec{\alpha}, \lambda)$ defines a one-to-one mapping, the *cumomer mapping*

$$\pi_{\lambda,F} : \{0,1\}^{|F|} \mapsto \{0,1\}^{|\lambda(F)|} \tag{2.7}$$

between the isotopomers of a reactant fragment F and product fragment $F' = \lambda(F)$, defined by $\pi_{\lambda,F}(\mathbf{b}) = \mathbf{b}'$ where $\mathbf{b}' = \mathbf{b}'_1 \cdots \mathbf{b}'_{|\mathbf{F}|}$ satisfies: $\mathbf{b}'_\mathbf{h} = b_g$ if $F'(h) = F(g)$. The mapping can be convieniently represented in $2^{|F|} \times 2^{|F'|}$ *permutation* matrix $\Pi_{\lambda,F}$, defined by

$$[\Pi_{\lambda,F}]_{\mathbf{b}',\mathbf{b}} = \begin{cases} 1, & \text{if } \pi_{\lambda,F}(\mathbf{b}) = \mathbf{b}' \\ 0, & \text{otherwise.} \end{cases}$$

Consider now a reactant fragment $F \subset M$ of a reaction ρ. Its isotopomer distribution is $\mathbb{I}_F$, computed via (2.4) from $\mathbb{I}_M$. Since the reactions are assumed to defined so that carbons within a reactant fragment F remain in physical contact, the frequencies of the F-labeling patterns of M are preserved in the reaction. In other words, let $F' = \lambda(F) \in M'$ be the corresponding product fragment. For any labeling pattern $\mathbf{b} \in \{0,1\}^{F_i}$ it holds that

$$\mathbb{P}\{^{\pi_{\lambda,F}(\mathbf{b})}F'\} = \mathbb{P}\{^{\mathbf{b}}F\}. \tag{2.8}$$

Thus, by applying (2.4) and (2.8) to the reactant fragments of a reaction we obtain isotopomer distributions for all product fragments. This computation is easily done using the cumomer mapping matrix:

$$U_{F'}^T \mathbb{I}_{M'} = \mathbb{I}_{F'} = \Pi_{\lambda,F} \mathbb{I}_F = \Pi_{\lambda,F} U_F^T \mathbb{I}_M \tag{2.9}$$

The matrix product $(\Pi_{\lambda,F} U_F)$ bears some similarity to the *isotopomer mapping matrix* used by Schmidt et al [13], the difference being that their matrix does not directly produce *isotopomer distributions* of the fragments. The property (2.9) will prove useful in Chapter 3 when handling incomplete isotopomer information.

Example 3 *Let us examine the metabolic reaction below, converting two 2-carbon reactants to a 4-carbon product.*
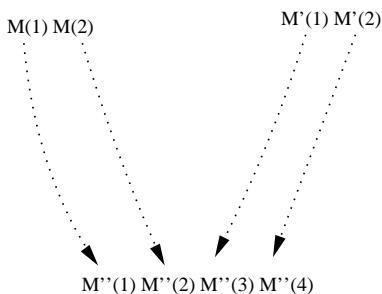
The mappings of the carbons are shown by the dashed arrows. Let the isotopomer distributions of the reactants be $\mathbb{I}_M = [0.25, 0.25, 0.25, 0.25]^T$ and $\mathbb{I}_{M'} = [0.5, 0.1, 0.2, 0.2]^T$.

The two cumomer mapping matrices of the two reactant fragments M_2 and M_3 are

$$\Pi_{\lambda,M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \Pi_{\lambda,M'} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The first one is the identity matrix since the ordering of the carbons does not change, the second flips the isotopomers 01 *and* 10 *around, in accordance to the carbon mapping. The $\lambda(M)$- and $\lambda(M')$- cumomer distributions are then $\mathbb{I}_{\lambda(M)} = \Pi_{\lambda,M} \mathbb{I}_M = [0.25, 0.25, 0.25, 0.25]^T$ and $\mathbb{I}_{\lambda(M')} = \Pi_{\lambda,M'} \mathbb{I}_{M'} = [0.5, 0.2, 0.1, 0.2]^T$*

It remains to compute the isotopomer distribution of a product given the isotopomer distributions of the product fragments. This is easily accomplished by noticing that due to the assumption (A-1), the isotopomer distributions of the fragments are independent as well.

Thus, the probability of observing a labeling $\mathbf{b}$ for $F = F' \cup F''$ is given by the product

$$\mathbb{P}\{^\mathbf{b}F\} = \mathbb{P}\{^{\mathbf{b}'}F'\} \cdot \mathbb{P}\{^{\mathbf{b}''}F''\}$$

where the labelings $\mathbf{b}'$ and $\mathbf{b}''$ match $\mathbf{b}$:

$$\mathbf{b_i} = \begin{cases} \mathbf{b'_k}, & \text{if } F(i) = F'(k) \text{ for some } k, \text{ and,} \\ \mathbf{b''_k}, & \text{if } F(i) = F''(k) \text{ for some } k. \end{cases}$$

Let, then, M_i be a product of a reaction λ_j, and let $F_1, \ldots, F_k, \sum_r F_r = M_i$, be the set of product fragments composing M_i. For any labeling $\mathbf{b_1} \cdots \mathbf{b_{|M|}} \in \{0, 1\}^{|M|}$ it holds that

$$\mathbb{I}_{M_i}(b_1 \cdots b_{|M_i|}) = \prod_{r=1}^{k} \mathbb{I}_{F_r}(b_{f_{r,1}} \cdots b_{f_{r,|F_r|}}), \tag{2.10}$$

where $F_r = \{F_r(1), \ldots, F_r(|F_r|)\}$ satisfies $M(f_{r,j_r}) = F_r(j_r)$ for all $1 \leq r \leq k$ and $1 \leq j_r \leq |F_r|$

In other words, the resulting isotopomer distribution for M_i is a product distribution of the constituent fragments. Given a set of reactions producing M_i, by applying (2.4), (2.8) and (2.10) we can compute an estimate of each $\mathbb{I}ji$ in (2.5).

Example 4 *Continuing the previous example, let us compute the isotopomer distribution of M'', using the isotopomer distributions of the product fragments $\lambda(M)$ and $\lambda(M')$, namely $\mathbb{I}_{\lambda(M)} = \Pi_{\lambda,M}\mathbb{I}_M = [0.25, 0.25, 0.25, 0.25]^T$ and $\mathbb{I}_{\lambda(M')} = \Pi_{\lambda,M'}\mathbb{I}_{M'} = [0.5, 0.2, 0.1, 0.2]^T$.*

We apply the formula $\mathbb{P}\{^{b_1 b_2 b_3 b_4}M''\} = \mathbb{P}\{^{b_1 b_2}M\} \cdot \mathbb{P}\{^{b_4 b_3}M'\}$ for each labeling $\mathbf{b} \in \{0, 1\}^4$. This gives us the isotopomer frequencies

$$\mathbb{P}\{^{0000}M''\} = \mathbb{P}\{^{00}M\} \cdot \mathbb{P}\{^{00}M'\} = 0.25 \cdot 0.5 = 0.125,$$
$$\mathbb{P}\{^{0001}M''\} = \mathbb{P}\{^{00}M\} \cdot \mathbb{P}\{^{01}M'\} = 0.25 \cdot 0.2 = 0.05, \ldots,$$
$$\mathbb{P}\{^{1110}M''\} = \mathbb{P}\{^{11}M\}\mathbb{P}\{^{10}M'\} = 0.25 \cdot 0.1 = 0.025, \ and$$
$$\mathbb{P}\{^{1111}M''\} = \mathbb{P}\{^{11}M\}\mathbb{P}\{^{11}M'\} = 0.25 \cdot 0.2 = 0.05.$$

In order to solve the fluxes in the metabolic network, an isotopomeric counterpart of (2.3) is constructed. It has the form

$$D\vec{v} = \begin{bmatrix} D_1 \\ \vdots \\ D_m \end{bmatrix} \times \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \vec{z_1} \\ \vdots \\ \vec{z_m} \end{bmatrix} = Z, \qquad (2.11)$$

where for each metabolite M_i, there is a block

$$D_i = \begin{bmatrix} d_{i,1,1} & \cdots & d_{i,j,1} & \cdots & d_{i,n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i,1,h} & \cdots & d_{i,j,h} & \cdots & d_{i,n,h} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i,1,2^{|M_i|}} & \cdots & d_{i,j,2^{|M_i|}} & \cdots & d_{i,n,2^{|M_i|}} \end{bmatrix} \qquad (2.12)$$

where $d_{i,j} = [d_{i,j,1} \cdots d_{i,j,2^{|M|}}]^T = a_{i,j}\mathbb{I}_{\lambda_j,M_i}$. The blocks $z_i = [z_{i,1} \cdots z_{i,2^{|M|}}]^T$ of the vector Z are defined by

$$z_i = \begin{cases} \vec{0}, & \text{if } M_i \text{ is an intermediate metabolite, and} \\ \vec{\beta_i}\mathbb{I}_{M_i}, & \text{if } M_i \text{ is an external metabolite.} \end{cases}$$

where $\vec{\beta_i}$ is taken from (2.3). For metabolites that are produced by a single reaction the blocks D_i and Z_i consist of one row, exactly corresponding to the metabolic balance equation (2.1). Solving (2.11) gives us the flux distribution of the metabolic network, with the exception of pathological cases where some alternative routes between two metabolites modify the carbon chain exactly the same way, or the choice of isotope labeling has been particularly unlucky.

Chapter 3

Flux estimation with incomplete data

The above presentation assumed that the isotopomer distributions of metabolic intermediates have been completely measured (assumption A-7). Such an assumption may not be well justified, given the capabilities of current techniques to chemical analysis of metabolites. In practise, one needs to be able to handle situations where

- isotopomer data is not available at all for some metabolites, and

- the isotopomer distribution is only partially resolved for some metabolites.

In the following, we develop methdos that – unlike the naive model presented earlier – takes into account these complications.

3.1 A generalized model of isotopomer measurement

To generalize the approach, we associate with each labeling $\mathbf{b}$ of a metabolite M a vector $\vec{e}_{\mathbf{b}} \in \{0,1\}^{2^{|M|}}$ that contains 0's as all other components except the $\mathbf{b}$'th location. The set of vectors $\vec{e}_0, \vec{e}_1, \ldots, \vec{e}_{2^{|M|}-1}$ form the standard basis of the *isotopomer space* $\mathcal{I}_M = \mathbb{R}^{2^{|M|}}$ of metabolite M.

Assume now that we have partial knowledge about the isotopomer distributions of metabolites in the form of a system of linear equations

$$S^T \mathbb{I}_M = \vec{d}, \tag{3.1}$$

where the columns of $S = [\vec{s}_1, \ldots, \vec{s}_r], \vec{s}_j \in \mathbb{R}^{2^{|M|}}$ are linearly independent and $\vec{d} = [d_1, d_2, \ldots, d_r] \in \mathbb{R}^r$ is an arbitrary vector. The columns of S form a basis for some r-dimensional vector subspace of $\mathcal{I}_M$; each vector $\vec{s}_j$ can be thought of as a normal of the hyperplane and d_j the offset of the hyperplane. As a whole, (3.1) represents an intersection of r such hyperplanes. The projection of all isotopomer distributions $\mathbb{I}_M$ that satisfy (3.1) to the space $\mathcal{S}$ is $\vec{d}$, represented in the coordinate system S. Furthermore, we can represent each isotopomer distribution $\mathbb{I}_M$ that is consistent with (3.1) decomposed as $\mathbb{I}_M = S\vec{d} + S_\perp \vec{d}_\perp$

where the columns of $S_\perp$ span the *orthogonal complement* $\mathcal{S}_\perp$ of the column space of S and $\vec{d}_\perp \in \mathcal{S}_\perp$ is an arbitrary vector.

The vector space representation unifies many concepts previously used in analyzing isotopic tracer experiments:

- An isotopomer distribution is a point $[x_1, \ldots, x_{|M|}]^T \in \mathcal{I}_M$ satisfying $x_j \geq 0, \sum_j x_j = 1$.

- A cumomer distribution related to a fragment $F \subset M$ lies in a subspace $\mathcal{U}_F \subset \mathcal{I}_M$ spanned by the vectors $\vec{u}_{F,0}, \ldots, \vec{u}_{F,2^{|F|}-1}$, where

$$\vec{u}_{F,\mathbf{a}} = \sum_{\{\mathbf{a_j}=\mathbf{b_i}, \text{ if } F(j)=M(i)\}} \vec{e}_{\mathbf{b}}.$$

- A *mass isotopomer distribution* – the relative frequency distribution of equal mass labelings – lies in the subspace $\mathcal{W}_M \subset \mathcal{I}_M$ spanned by the vectors $\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_{|M|}$, where $\vec{w}_i = \sum_{weight(\mathbf{b})=i} \vec{e}_{\mathbf{b}}$, and $weight(\mathbf{b_1} \cdots \mathbf{b_{|M|}}) = \sum_j \mathbf{b_j}$. Tandem mass spectrometers [2, 5, 11] produce data that, in addition to the above, also contains mass isotopomer distributions of some of the fragments, so we get extra basis vectors $\vec{w}_{i,F} = \sum_{weight(\mathbf{b})=i} \vec{u}_{F,\mathbf{b}}$. where $\mathbf{b} = \mathbf{b_1} \cdots \mathbf{b_{|F|}}$ is a labeling of the fragment F.

- *Positional labeling enrichment* data lie in the subspace $\mathcal{P}_M \subset \mathcal{I}_M$, spanned by the vectors $\vec{p}_1, \ldots, \vec{p}_{|M|}$, where each vector $\vec{p}_j = \sum_{\mathbf{b_j}=1} \vec{e}_{\mathbf{b}}$ denotes the sum of labeling frequencies where location j contains a ^{13}C isotope.

So, in general, instead of isotopomer distributions $\mathbb{I}_M$ we handle linear *isotopomer constraints* of the form $S^T \mathbb{I}_M = \vec{d}$ (i.e., different linear combinations of isotopomers as given by S), where matrix $S = [\vec{s}_1 \cdots \vec{s}_r]$ contains basis vectors spanning some subspace $\mathcal{S} \subset \mathcal{I}_M$.

Example 5 *For alanine, the above-mentioned subspaces are the following*

- *If $S = I_8$, the 8×8- identity matrix, (3.1) corresponds to measuring the isotopomer distribution in its entirety.*

- *The trivial isotopomer 'measurement', stating the fact that the relative isotopomer frequencies sum up to unity, is obtained by selecting the coefficient matrix S as*
$$S^T = [1, \ldots, 1].$$

- *The positional enrichment data, indicating the labeling degrees in different locations is modelled by setting*

$$S^T = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

- *Mass isotopomer distribution (obtained from basic mass spectrometric analysis of labeled metabolites) is represented by constructing*

$$S^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Consider now the system of isotopomer balance equations (2.5)

$$\sum_j v_j \alpha_{i,j} \mathbb{I}_{\lambda_j, M_i} = 0$$

of metabolite M_i. By multiplying the above by S^T and by denoting $\mathbb{I}_{\lambda_j, \mathcal{S}} = S^T \mathbb{I}_{\lambda_j, M_i}$, we immediately get

$$\sum_j v_j \alpha_{i,j} \mathbb{I}_{\lambda_j, \mathcal{S}} = 0, \tag{3.2}$$

illuminating the fact that an isotopomer balance equation that holds true in the space $\mathbb{I}_M$ also holds true in an arbitrary subspace $\mathcal{S} \in \mathbb{I}_M$. Thus, if we have estimates of $\mathbb{I}_{\lambda_j, \mathcal{S}}$ we can use (3.2) in place of the isotopomeric balance (2.5). The central theme in this section is, for each metabolite in a metabolic junction, to define (3.2) using as high-dimensional S as possible. That way, we aim at maximizing the number of linearly independent equations constraining the fluxes around the junction.

In the following, we describe methods for obtaining these estimates. First, in section 3.2 we show how a set of constraints can be inferred to a product of a reaction given generalized isotopomer measurements (3.1) of the reactants. Then, in section 3.3, we show how a similar process can be perfomed in backward direction, inducing sets of constraints to the reactants given measurements of the products.

3.2   Forward propagation of isotopomer information

Given a generalized isotopomer measurements of a reactants M of a reaction $\rho = (\vec{\alpha}, \lambda)$, we wish to compute as tight as possible constraints to the isotopomer distribution of a product M'. For this end, we need to generalize the equations (2.4) and (2.10), whereas (2.8) can, in essence, be applied in its original form.

Let us first construct a generalization of (2.4). Consider a reactant fragment $F \subset M$, for which $\lambda(F) = F' \subset M'$. Assume that we have a measurement (3.1) of the metabolite M. Our intent is to first infer linear constraints of the form

$$S_F^T \mathbb{I}_M = \vec{d}_F \tag{3.3}$$

to the isotopomer distribution of F. We set the following criteria to the resulting system of equations:

(i) First, it must be expressible as a linear combinations of the basis vectors of $\mathcal{U}_F$, that is, all columns $\vec{s}$ of S_F must satisfy $\vec{s} = \sum_j a_j \vec{u}_{F,j}$ where $a_j \in \mathbb{R}$. The rationale is that no other carbons than those in F contribute to the isotopomer distribution – as induced by the reaction ρ – of the product fragment F' and the product M'.

(ii) Second, the feasible set of vectors of the resulting system should contain all isotopomer distributions that are in the feasible set of system (3.1), that is if $\vec{x}$ satisfies $S^T\vec{x} = \vec{d}$, then we should also have $S_F^T\vec{x} = \vec{d}_F$. This is because otherwise we are excluding isotopomer distributions that could have generated our data.

(iii) Third, the number of linearly independent equations in the system should maximal.

We claim that the intersection space $\mathcal{W} = \mathcal{S} \cap \mathcal{U}_F$ has the largest dimension, among subspaces of $\mathcal{I}_M$ that satisfy above criteria. In detail, the best possible system of equations has the form $W_\perp^T \mathbb{I}_M = \vec{d}_F$, where $\vec{w}_1, \ldots, \vec{w}_r$ are orthonormal basis vectors of $\mathcal{W}$ and $\vec{d}_F = W_\perp^T \vec{x}_*$ is a projection into the space $\mathcal{W}$ of the least-squares solution $\vec{x}_{LS} = (S^T)^+ \vec{d}$ to the equation $S^T\vec{x} = \vec{d}$. (Figure 3.1). By the properties of Moore-Penrose pseudo-inverse $(S^T)^+$ of matrix S^T, $\vec{x}_{LS}$ is in fact the orthogonal projection into $\mathcal{S}$ of all least-squares solutions $\hat{\vec{x}}$ to $S^T\vec{x} = \vec{d}$. The computation of an intersection of vector spaces is described in Appendix A.1.

That (i) is satisfied, follows from the fact that $\mathcal{W} \subset \mathcal{U}_F$. The satisfcation of requirement (ii) is seen from the following (see also Figure 3.1). We can express any $\vec{x}$ consistent with $S^T\vec{x} = \vec{d}$ as a sum of three orthogonal vectors:

$$\vec{x} = \vec{x}_W + \vec{x}_S + \vec{x}_{S_\top}, \qquad (3.4)$$

where $\vec{x}_W \in \mathcal{W}$, $\vec{x}_S = S\vec{d} - \vec{x}_W \in \mathcal{S} \cap \mathcal{W}_\top$ and $\vec{x}_{S_\top} = \vec{x} - S\vec{d} \in \mathcal{S}_\top$, where $\mathcal{W}_\top$ and $\mathcal{S}_\top$ are orthogonal complements of $\mathcal{W}$ and $\mathcal{S}$, respectively.

The projection of $\vec{x}$ to W is then given by

$$W_\perp W_\perp^T \vec{x} = W_\perp W_\perp^T \vec{x}_W + W_\perp W_\perp^T \vec{x}_S + W_\perp W_\perp^T \vec{x}_{S_\top}, \qquad (3.5)$$

However, since the vectors $\vec{x}_S$ and $\vec{x}_{S_\top}$ both lie in $\mathcal{W}_\top$, they are orthogonal to every vector in $\mathcal{W}$, especially the basis vectors $\vec{w}_1, \ldots, \vec{w}_r$. Thus the two last terms zero out. Moreover, as $\vec{x}_W$ already lies in $\mathcal{W}$, we have

$$W_\perp W_\perp^T \vec{x} = \vec{x}_W$$

for any $\vec{x}$ consistent with (3.1). The constraint (3.3) is then satisfied by setting $\vec{d}_F = W_\perp \vec{x}_W$.

The requirement (iii) follows from the fact that the columns of W form a basis to $\mathcal{W}$, so one cannot construct a larger set of linearly independent equations.

The constraint (3.3) can be computed by projecting the least-squares solution to (3.1), namely the vector $(S^T)^+ \vec{d}$, to W:

$$\vec{x}_W = W_\perp W_\perp^T (S^T)^+ \vec{d} \qquad (3.6)$$
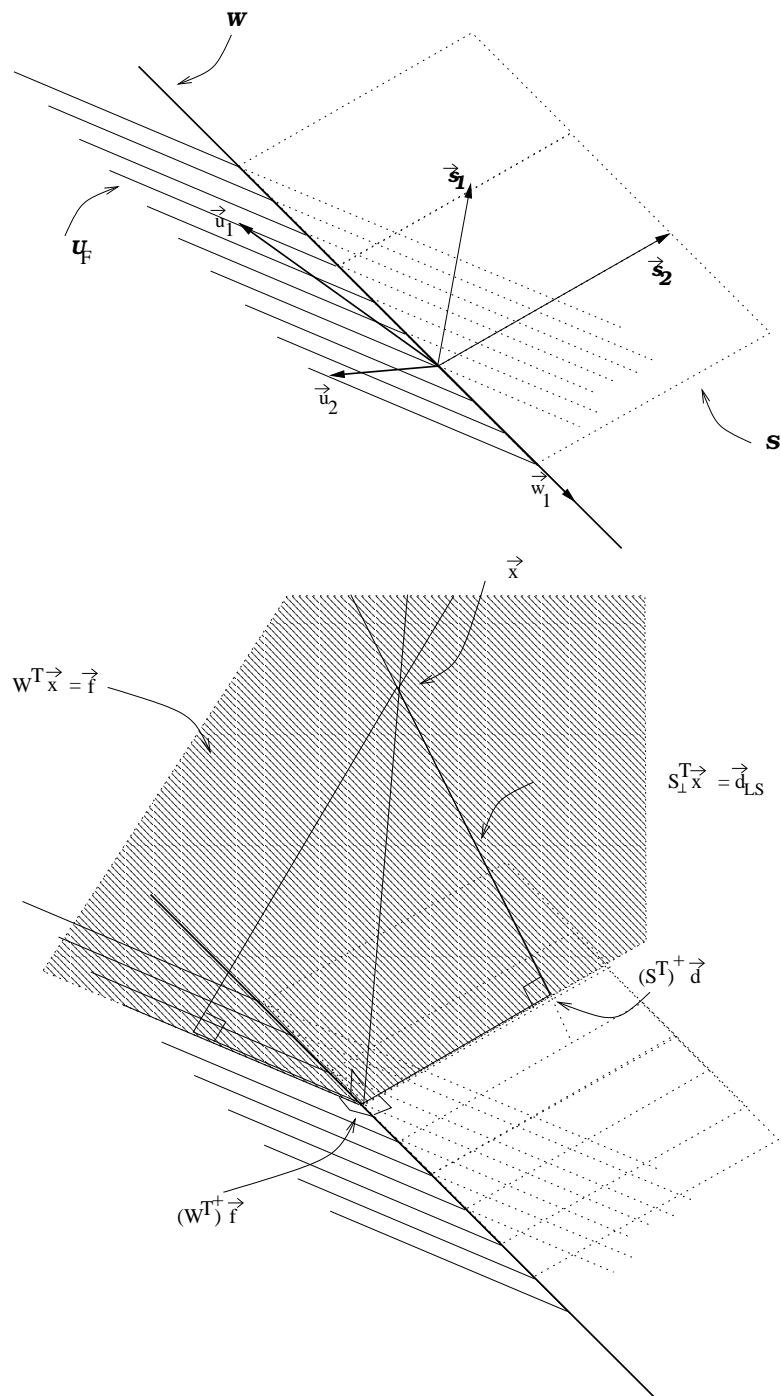
Figure 3.1: At top, the column spaces $\mathcal{S}$ and $\mathcal{U}_F$ of matrices $S = [\vec{s}_1, \vec{s}_2]$ and $U_F = [\vec{u}_1, \vec{u}_2]$ together with their intersection $W = [\vec{w}_1]$ are depicted. At bottom, the otrhogonal projection of isotopomer distribution $\vec{x}$ into $\mathcal{W}$ is depicted. The same projection can be computed directly or via either of the subspaces $\mathcal{S}$ and $\mathcal{U}_F$. The set of solutions consistent with $W^T \vec{x} = \vec{f}$ is shown in grey.

For an orthonormal S, the same projection is obtained by computing

$$\vec{x}_W = W_\perp W_\perp^T S \vec{d}.$$

On the other hand, the same projection can be computed incrementally, by first projecting $\vec{x}$ orthogonally to $\mathcal{U}_F$ to obtain $\vec{x}_W + \vec{x}_U$ where, again, $\vec{x}_U \in \mathcal{U}_F - \mathcal{W}$ is orthogonal to $\vec{x}_W$, and then projecting the result to $\mathcal{W}$. The first projection is obtained by

$$\vec{x}_W + \vec{x}_U = U_{F,\perp} U_{F,\perp}^T \vec{x},$$

where the columns of $U_{F,\perp} = U_F \cdot diag(\| \vec{u}_{F,\mathbf{b}} \|)^{-1}$ form an orthonormal basis for $\mathcal{U}_F$. The second projection is given by

$$\vec{x}_W = W_\perp W_\perp^T (\vec{x}_W + \vec{x}_U) = W_\perp W_\perp^T U_{F,\perp} U_{F,\perp}^T \vec{x}$$

Thus we have the equality

$$W_\perp W_\perp^T U_{F,\perp} U_{F,\perp}^T \vec{x} = W_\perp W_\perp^T (S^T)^+ \vec{d} \qquad (3.7)$$

for any isotopomer distribution $\vec{x}$ consistent with (3.1). For orthonormal S the equality can also be written as

$$W_\perp W_\perp^T U_{F,\perp} U_{F,\perp}^T \vec{x} = W_\perp W_\perp^T S S^T \vec{x} \qquad (3.8)$$

We can apply this to the cumomer mapping (2.9) to get

$$W_\perp W_\perp^T (S^T)^+ \vec{d} = W_\perp W_\perp^T U_{F,\perp} U_{F,\perp}^T \mathbb{I}_M =$$
$$W_\perp W_\perp^T U_{F,\perp} diag(\| \vec{u}_{F,\mathbf{b}} \|)^{-1} U_F^T \mathbb{I}_M =$$
$$W_\perp W_\perp^T U_{F,\perp} diag(\| \vec{u}_{F,\mathbf{b}} \|)^{-1} \Pi_{\lambda,F}^T U_{F'}^T \mathbb{I}_{M'} = S'^T \mathbb{I}_{F'}, \quad (3.9)$$

above denoting $S'^T = W_\perp W_\perp^T U_{F,\perp} diag(\| \vec{u}_{F,\mathbf{b}} \|)^{-1} \Pi_{\lambda,F}^T$. The left-hand side can be computed from (3.1) and the right-hand side is an isotopomer constraint for F'.

By the above method, we can compute a constraint to the isotopomer distribution of each product fragment of metabolite M'. To determine a constraint to the isotopomer distribution of $\mathbb{I}_{M'}$ as a whole, we still need to combine the fragment constraints. Let, therefore, $F', F'' \subset M'$ be two arbitrary product fragments with constraints $S'^T \mathbb{I}_{F'}$ and $S''^T \mathbb{I}_{F''}$, respectively. Let us compute a constraint for their union $F = F' \cup F''$. This is done by, in a sense, generalizing the equation (2.10).

Consider the constraints $\vec{s}'^T \mathbb{I}_{F'} = y$ and $\vec{s}''^T \mathbb{I}_{F''} = z$ corresponding to an arbitrary pair of rows in matrices S' and S''. The product

$$yz = (\vec{s}'^T \mathbb{I}_{F'})(\vec{s}''^T \mathbb{I}_{F''}) = (\sum_{\mathbf{b}'} s'_{\mathbf{b}'} \mathbb{P}\{{}^{\mathbf{b}'}F'\}) \cdot (\sum_{\mathbf{b}''} s''_{\mathbf{b}''} \mathbb{P}\{{}^{\mathbf{b}''}F''\}),$$

by above reasoning, simplifies into

$$w = yz = \sum_{\mathbf{b}',\mathbf{b}''} s'_{\mathbf{b}'} s''_{\mathbf{b}''} \mathbb{P}\{{}^{\mathbf{b}'}F'\} \mathbb{P}\{{}^{\mathbf{b}''}F''\} = \sum_{\mathbf{b}} s_{\mathbf{b}} \mathbb{P}\{{}^{\mathbf{b}}F\} = \vec{s}^T \mathbb{I}_F,$$

choosing here $\mathbf{b}$ according to (2.3) and denoting $s_{\mathbf{b}} = s'_{\mathbf{b}'} \cdot s''_{\mathbf{b}''}$ and $w = yz$.

Similar constraints can be computed for each pair of rows and collected to a matrix $S^T \mathbb{I}_F = \vec{w}$. The whole process can be applied iteratively to a union of product fragments $\cup_{t<k} F_t$ and the next unprocessed fragment F_k, $2 \leq k \leq l$, finally resulting in a constraint $S'^T \mathbb{I}_{M'}$ to the product metabolite $M' = \cup_{k=1}^l F_k$.

However, the columns of matrix S are not necessarily of unit length. Also, whether they are linearly independent, we do not know. Hence, to guarantee that the system conforms to the form of (3.1), a transformation should be made to ensure that (see Appendix A.2).

The procedure described above can be used to compute a constraint to the isotopomer distribution of a product M of any reaction λ, given any linear constraints to the distributions of the reactants. In particular, the absense of a measurement can be tolerated, this will be treated as a trivial isotopomer measurement. Moreover, the approach can be applied iteratively to a unbranched pathway $\Lambda = \{\lambda_1, \ldots, \lambda_r\}$ of reactions (see section ??).

3.3 Backward propagation of isotopomer information

In the complete information case (section 2.3), to analyze the fluxes around a metabolite M, we only needed to provide estimates of the flows *producing* the metabolite. The isotopomer distribution of metabolite M itself, and thus the distributions of the flows consuming the metabolite, were assumed to be measured. In the general case, however, it may be that the isotopomer distribution M is not measured, or a very weak measurement $S^T \mathbb{I}_M = \vec{d}$ is given (i.e. the rank of S is low), we may want to obtain a better estimate using the measurements of the products.

Assume now that M is consumed by a reaction ρ. Let $F \subset M$ be a reactant fragment induced by ρ, and let $F' = \lambda(F) \subset M'$ be the corresponding fragment in the product M'. Let us further assume that we have at our disposal an isotopomer measurement

$$ S'^T I_{M'} = \vec{d'} $$

of M'. Exactly like in forward propagation, using (3.6), we can compute the projection of the measurement to the intersection space $\mathcal{W}' = \mathcal{S}' \cap \mathcal{U}_{F'}$ as

$$ \vec{x}_{W'} = W'_\perp W'^T_\perp (S^T)^+ \vec{d'}, $$

and also as

$$ \vec{x}_{W'} = W_\perp W^T_\perp U_{F,\perp} U^T_{F,\perp} \vec{x} $$

where $\vec{x}$ is any isotopomer distribution consistent with $S'^T \vec{x} = \vec{d'}$. Applying this to the cumomer mapping (2.9) gives

$$ W'_\perp W'^T_\perp (S'^T)^+ \vec{d'} = W'_\perp W'^T_\perp U_{F',\perp} U^T_{F',\perp} \mathbb{I}_{M'} = $$
$$ W'_\perp W'^T_\perp U_{F',\perp} D_{F'}^{-1} U^T_{F'} \mathbb{I}_{M'} = $$
$$ W'_\perp W'^T_\perp U_{F',\perp} D_{F'}^{-1} \Pi_{\lambda,F} U^T_F \mathbb{I}_M $$

$$ = S^T \mathbb{I}_F, \quad (3.10) $$

denoting above $S = (W'_\perp W'^{T}_\perp U_{F',\perp} diag(\parallel \vec{u}_{F',\mathbf{b}} \parallel))^{-1}\Pi_{\lambda,F})^T$ and $D_{F'} = diag(\parallel \vec{u}_{F',0} \parallel, \ldots, \parallel \vec{u}_{F',2^{F'}} \parallel)$.

The procedure in (3.10) is in essence the inverse of (3.9), starting now from products instead of reactants and applying the inverse of the cumomer mapping. However, the approach used in the previous section of multiplying the fragment constraints to get their product distribution does not work. We illustrate this with a simple example.

Example 6 *Let $M_1 = \{M_1(1), M_1(2)\}$ be two carbon metabolite and let $\rho = (\vec{\alpha}, \lambda)$ be a reaction taking M_1 as reactant and producing two products, $M_2 = \{M_2(1)\}$ and $M_3 = \{M_3(1)\}$. The mapping λ induces two reactant fragments $M_1(1)$ and $M_1(2)$. The corresponding product fragments are $\lambda(M_1(1)) = M_2(1)$ and $\lambda(M_1(2)) = M_3(1)$.*

Assume that the isotopomer distributions of M_2 and M_3 are uniform, that is,

$$\mathbb{I}_{M_2} = \mathbb{I}_{M_3} = [0.5, 0.5]^T. \qquad (3.11)$$

Consider the isotopomer distributions of M_1 that could result in such distributions in the products.

In forward propagation we assumed fragments combining independently. Using that reasoning we would get $\mathbb{P}\{^{\mathbf{b_1 b_2}}M_1\} = \mathbb{P}\{^{b_1}M_1(1)\} \cdot \mathbb{P}\{^{b_2}M_1(2)\}$, for each labeling $\mathbf{b_1 b_2} \in \{0,1\}^2$. The resulting distribution in this case would then be $\mathbb{I}_{M_1} = [0.25, 0.25, 0.25, 0.25]^T$, that obviously satsifies the marginals (3.11).

However, the distributions $\mathbb{I}_{M_1} = [0.5, 0, 0, 0.5]$ and $\mathbb{I}_{M_1} = [0, 0.5, 0, 0.5]$ also have marginals that satisfy (3.11).

Hence, only when there is some quarantee that $\mathbb{I}_M$ is a product distribution – for example, that all metabolic paths leading to M have detached the carbons in different fragments in some point, or the carbons in the fragments originate from different substrate metabolites – one can use the analogous computation as in the forward propagation. We leave exploration of this thread as further work.

In the general case, in fact, no more can be said about the distribution $\mathbb{I}_M$, than the fact that the the reactant fragments have to satisfy the very same constraints that hold in the product fragments. Let $F_1, \ldots, F_k$ compose the set of reactant fragments of a reaction ρ consuming M. By the above backward propagation procedure, we obtain a constrain $S_i^T \mathbb{I}_M = \vec{d_i}$ corresponding to each F_i. These constraints can be grouped into a common matrix equation:

$$S_{\rho,M}^T \mathbb{I}_M = [S_1, S_2, \ldots, S_k]^T \mathbb{I}_M = [\vec{d}_1^T, \ldots, \vec{d}_k^T]^T = \vec{d}_{\rho,M}^T$$

Moreover, in case there are more than one consumer of M, the constraints obtained for each consumer $\rho_1, \ldots, \rho_l$ can be grouped as well: we have

$$S_M^T \mathbb{I}_M = \vec{d}_M,$$

where we denote $S_M = [S_{\rho_1,M}, \ldots, S_{\rho_l,M}]$ and $\vec{d}_M^T = [\vec{d}_{\rho_1,M}^T, \ldots, \vec{d}_{\rho_l,M}^T]^T$. Note that the columns of S_M are not generally linearly independent, so the rank of the system is typically not as high as the number of constraints would suggest.

3.4 Handling non-bijective reactions

The above presentation assumed that the atom mappings of the reactions are bijective. Not all reactions are of this kind, however. In the following we deal with to kinds of non-bijective reactions, namely,

- reactions involving symmetric reactants or products, and

- reactions producing or consuming more than one copy of some metabolite. (De)polymerization reactions are a typical example of reactions of this kind.

3.4.1 Symmetrical metabolites

From the perspective of isotopomer analysis of metabolites, symmetry in molecules manifests in symmetry of the isotopomer distribution: if the molecule has 2 symmetrical orientations, there are labeling pairs that always have equal frequency. For example, the isotopomer distribution of ethane ($CH3 - CH3$) always satisfies $\mathbb{I}_{Ethane}(01) = \mathbb{I}_{Ethane}(10)$. In general, if there are k symmetrical orientations, such *equivalence* sets of equal frequency have maximum size k. For some labelings, e.g. $\mathbb{I}_{Ethane}(00)$ and $\mathbb{I}_{Ethane}(11)$, the equivalence set may be a singleton. We denote the symmetry equivalence set of labeling $\mathbf{b}$ as $E_M(\mathbf{b})$. Note that for any $\mathbf{b}' \in E_M(\mathbf{b})$ $E_M(\mathbf{b}) = E_M(\mathbf{b}')$. For ethane the equivalence sets are $E_{Ethane}(00) = \{00\}, E_{Ethane}(01) = E_{Ethane}(10) = \{01, 10\}, E_{Ethane}(11) = \{11\}$.

To handle the symmetry of k-symmetrical metabolite M, we introduce a symmetry matrix $\Pi^{Symm,M} \in \mathbb{R}^{2^{|M|} \times 2^{|M|}}$ that contains a row and a column for each labeling. The entries of the matrix are defined by

$$\Pi_{i,j}^{Symm,M} = \begin{cases} 1/|E_M(\mathbf{b}_M(i))|, & \text{if } \mathbf{b}_M(j) \in E_M(\mathbf{b}_M(i)), \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Pre-multiplication of a non-symmetrical vector $\vec{d} \in \mathcal{I}_M$ by the symmetry matrix $\Pi^{Symm,M}$, produces a vector $\vec{d'} = \Pi^{Symm,M}\vec{d}$ where each equivalence set has uniform distribution. Intuitively, one can think of this multiplication as measuring the isotopomer distribution of the molecule in every possible symmetrical orientation, summing up the measurements and normalizing to a frequency distribution. The reader can verify that if the distribution already is symmetrical (with respect to equivalence sets $E_M(\mathbf{b})$, multiplication.

The approach taken by us is to ensure before backward or forward propagation the symmetry of the constraint that is to be propagated. Let M be a symmetrical metabolite and let $S^T \mathbb{I}_M = \vec{d}$ be the constraint propagated to M, either forward or backward. The multiplication $S^T \Pi^{Symm,M} \mathbb{I}_M = \vec{d}$ makes the constraint symmetrical. So the constraint that is to be propagated forward or backward from M, is $\tilde{S}^T \mathbb{I}_M = \vec{d}$, where $\tilde{S} = (\Pi^{Symm,M})^T S$.

Non–symmetrical   Symmetrical   Symmetrical

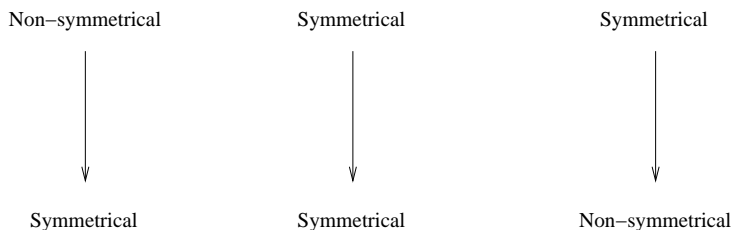Symmetrical   Symmetrical   Non–symmetrical

Figure 3.2: Symmetry transitions in reactions.

3.4.2   Reactions with non-disjoint reactant or product locations

Reactions that consume or produce more than one copy of the same metabolite are another class of non-bijective reactions. In such reactions, there are one or more carbon locations that are images or pre-images of more than one location. We call such locations non-disjoint, since, the carbon mapping cannot distinguish these locations from each other.

We start by converting a such reaction λ to a bijective one λ' by introducing special *shadow* metabolites for every metabolite that takes part in the reaction with a coefficient different from 1. The shadow metabolites related to a single metabolite together with the associated coefficients are called a *shadow set*. The shadow set of a disjoint reactant is a singleton.

Example 7 *Let λ be a non-bijective reaction consuming two (non-symmetrical) metabolites M_1 and M_2 and producing two molecules of the same metabolite M_3. The bijective reaction λ' is then a reaction consuming M_1 and M_2 and producing a pair of shadow molecules M_3' and M_3''.*
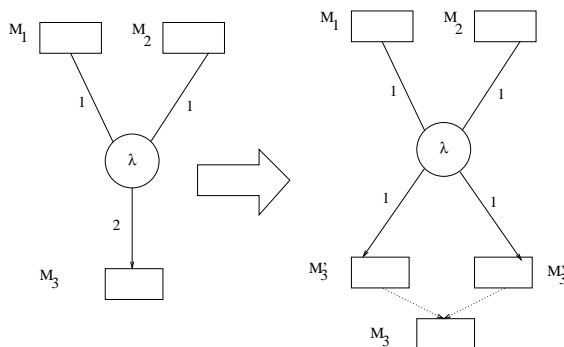
The propagation of isotopomer information is performed with the bijective reaction followed by post-processing taking into account the underlying non-disjointness of the reactant or product locations.

Forward propagation

The forward propagation procedure is the following:

1. Convert the reactants and products of λ to a disjoint set by introduction of shadow metabolites. Denote the bijective atom mapping as λ'.

2. Copy the isotopomer constraint $S^T \mathbb{I}_M = \vec{d}$ of each non-disjoint reactant to the members of its shadow set $\{M_{s,1}, \ldots, M_{s,k}\}$.

3. Perform normal forward propagation over the reaction λ' to obtain a constraint $S_{j'}^T \mathbb{I}_{M'} = \vec{d}_{j'}$ to each shadow metabolite of each product M'.

4. For each product M', process the constraints of the shadow set as follows:

   (a) Compute the intersection of column spaces of the coefficient matrices $S_{1'}, \ldots, S_{k'}$. Let the columns of $S_{i'}$ compose an orthonormal basis of the intersection space.

   (b) Project the vectors $\vec{d}_{1'}, \ldots, \vec{d}_{k'}$ to the intersection space to obtain a constraint $S_{M'}^T \mathbb{I}_{M'} = \vec{f}_{j'}$ for each member of the shadow set.

   (c) Compute an average of the constraints:

   $$S_{M'}^T \mathbb{I}_{M'} = \frac{1}{k'} \sum_j S_{M'}^T \mathbb{I}_{M'} = \frac{1}{k'} \sum_{j'} \vec{f}_{j'} = \vec{d}_{M'}$$

   to obtain a constraint for M'.

To understand step (4), note that each shadow metabolite represents a quotient of the metabolite pool. The relationship between isotopomer distribution of the metabolite pool as a whole and the isotopomer distributions of the shadow pools is given by the equation

$$\mathbb{I}_{M'} = \frac{1}{k'} \sum_{h=1}^{k'} \mathbb{I}_{M'_{s,h}}.$$

For any vector $\vec{s} \in \mathcal{I}_{M'}$, the following holds:

$$\vec{s}^T \mathbb{I}_{M'} = \sum_{h=1}^{k} \vec{s}^T \mathbb{I}_{M'_{s,h}}$$

Thus, if $s^T \mathbb{I}_{M_{s,h}} = d_h$ is known for each h, the constraints can be averaged. This calls for computation of a set of vectors $\vec{s}$ for which d_h's can be determined. It should be clear that any basis of the intersection space $\mathcal{S} = \mathcal{S}_1 \cap \cdots \cap \mathcal{S}_k$ contains such vectors. Moreover, it is a maximal-sized linearly independent set of such vectors.

Backward propagation

Backward propagation is somewhat more complicated. To see the problem, consider reaction of example (7)

Example 8 *Let the product fragments be the following* $F_\lambda(M_1, M_3') = \{M_{3'}(1), M_{3'}(2)\}$, $F_\lambda(M_2, M_3') = \{M_{3'}(3)\}$, $F_\lambda(M_1, M_3'') = \{M_{3''}(1)\}$, $F_\lambda(M_2, M_3'') = \{M_{3''}(2), M_{3''}(3)\}$.

Consider now the location $M_3(2)$. Half of the carbons in the location originate from metabolite M_1 and the other half from M_2. The $(3,1)$-cumomer distribution of M_3 satisfies

$$2\mathbb{I}_{3,2} = \mathbb{I}_{(3',2)} + \mathbb{I}_{(3'',2)} = \mathbb{I}_{(1,2)} + \mathbb{I}_{(2,2)}.$$

Thus, from $\mathbb{I}_{3,2}$ we do not obtain constraints for $\mathbb{I}_{M_1}$ and $\mathbb{I}_{M_2}$ separately.

On the other hand the carbons on location $(3,1)$ always originate from metabolite M_1, either from the location $(1,1)$ or $(1,3)$. Thus we have the following relationship

$$2\mathbb{I}_{3,1} = \mathbb{I}_{(3',1)} + \mathbb{I}_{(3'',1)} = \mathbb{I}_{(1,1)} + \mathbb{I}_{(1,3)},$$

which gives us the constraint $1/2\mathbb{I}_{(1,1)} + 1/2\mathbb{I}_{(1,3)} = \mathbb{I}_{(3,1)}$ for the isotopomer distribution of M_1. The $(3,3)$-cumomers induce a similar constraint to $\mathbb{I}_{M_2}$.

The outline of the backward propagation procedure is the following. Assume a constraint $S_{M'}^T \mathbb{I}_{M'} = \vec{d}_{M'}$, where for simplicity S is assumed orthonormal and the eqaution system consistent (see Appendix A.2 how to achieve that), as the starting point:

1. Convert the reactants and products of λ to a disjoint set by introduction of shadow metabolites. Denote the bijective atom mapping as λ'.

2. For each disjoint product, perform normal backward propagation over λ'.

3. For a non-disjoint product M', find out the set of locations $F' \subset M'$ that in each shadow product $M'_{s,1}, \ldots, M'_{s,k'}$ originate from the same shadow reactant $M_{s,l}$: $F'_l = \{M'(j) | j \in J_l\}$, where $J_l = \{\lambda'^{-1}(M'_{s,h}(j)) \in M_{s,l} | 1 \le h \le k'\}$. Denote the corresponding shadow fragments by $F_{s,1}, \ldots, F_{s,k'}$.

4. Compute the vector space intersection $\mathcal{Y}' = \mathcal{S}' \cap \mathcal{U}_{F'}$

5. Utilizing the relationship $\mathbb{I}_{M'} = \frac{1}{k'} \sum_h \mathbb{I}_{M'_{s,h}}$, write down the constraint in the intersection space as

$$Y'^T S_{M'} S_{M'}^T \mathbb{I}_{M'} = \frac{1}{k'} \sum_h Y'^T S_{M'} S_{M'}^T \mathbb{I}_{M_{s,h}} = Y'^T S_{M'} \vec{d}_{M'}$$

6. Apply the projection equality (3.8) to each term of the sum to obtain

$$\frac{1}{k'} \sum_{h=1}^{k'} Y'^T U_{F'_{s,h},\perp} U_{F'_{s,h},\perp}^T \mathbb{I}_{M_{s,h}} = Y'^T S_{M'} \vec{d}_{M'}$$

7. Apply the cumomer mapping (2.9) to each term in the sum to obtain the equation

$$S_{s,l}^T \mathbb{I}_{M_{s,l}} = \frac{1}{k} \sum_{h=1}^{k} Y'^T U_{F'_{s,h}} D_{F'}^{-1} \Pi_{\lambda^{-1}(F'_{s,h})} U_{F_{s,h}}^T \mathbb{I}_{M_{s,l}} = Y'^T S_{i'} \vec{d}_{i'} = \vec{d}_{M_{s,l}},$$

8. For each reactant, process its shadow set by collecting the constraints $S_{s,l}^T \mathbb{I}_{M_{s,l}} = \vec{d}_{M_{s,l}}$ into a common system

$$S_M^T \mathbb{I}_M = [S_{s,1}, \ldots, S_{s,k}]^T \mathbb{I}_M = [\vec{d}_{s,1}^{\,T}, \ldots, \vec{s}_{s,k}^{\,T}]^T = \vec{d}_M.$$

Note that due to the computation of the intersection of the fragments, we may lose information: locations where carbons originate from more than one source metabolite are not incluluded in the fragments. Note however, that in many cases there is only one source metabolite. In that case, no carbon locations are excluded; the only weakness over a bijective reaction is that the isotopomer distribution is less spiky due to the summation over the shadow sets.

Chapter 4

An algorithm for flux estimation

To maximize our possibilities in estimating the fluxes, we would like to propagate as many as possible independent linear constraints to the isotopomer distributions of the flows meeting in each metabolic junction.

The techniques described in previous sections enable us to compute linear constraints to the isotopomer distirbution of products of a reaction given similar constraints to the reactants, and vice versa.

We can extend these operations to pathways in straight-forward way, as the following exmaple suggests.

Example 9 *Let us consider the pathway of two reactions below*
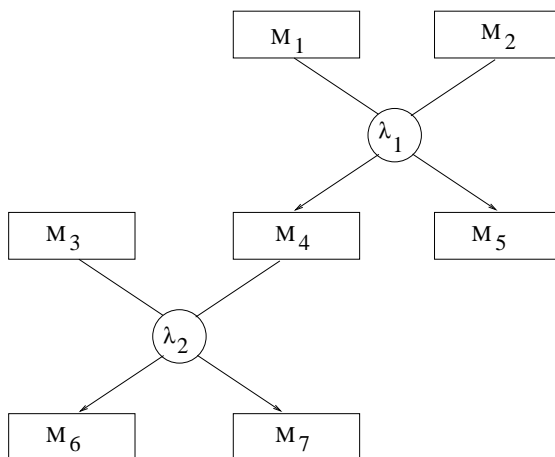
In addition let the atom mappings be such that for each reactant -product pair (M, M') of reaction $\lambda \in \{\lambda_1, \lambda_2\}$, $F_\lambda(M, M') \neq \emptyset$.

First, let us consider computing constraints to the isotopomer distribution $\mathbb{I}_{M_7}$, assuming that we possess constraints to the isotopomer distributions $\mathbb{I}_{M_1}, \mathbb{I}_{M_2}$ and $\mathbb{I}_{M_3}$.

The distribution $\mathbb{I}_{M_7}$ depends on the distributions $\mathbb{I}_{M_3}$ and $\mathbb{I}_{M_4}$. The distribution $\mathbb{I}_{M_4}$, on the other hand, depends on the distributions $\mathbb{I}_{M_1}$ and $\mathbb{I}_{M_2}$. Thus, we need to first compute constraints to $\mathbb{I}_{M_4}$ by propagating forward from $\mathbb{I}_{M_1}$ and $\mathbb{I}_{M_2}$ and then, to propagate forward the information from $\mathbb{I}_{M_3}$ and $\mathbb{I}_{M_4}$, using the above propagation result.

Note that the above procedure can be used even if we already had some mea-surement constraining $\mathbb{I}_{M_4}$: we just combine the propagation result to the ex-isting measurement to obtain combined – hopefully tighter – constraints to the distribution.

Backward propagation generalizes to simple pathways exactly as easily:

Example 10 *Let us consider the pathway of the previous example. Assume that our goal is to compute constraints to the isotopomer distribution $\mathbb{I}_{M_2}$. It is obviously dependent on $\mathbb{I}_{M_4}$ and $\mathbb{I}_{M_5}$. The distribution $\mathbb{I}_{M_4}$, on the other hand, depends on $\mathbb{I}_{M_6}$ and $\mathbb{I}_{M_7}$. So, to compute constraints to $\mathbb{I}_{M_1}$, we first propagate backward information form $\mathbb{I}_{M_6}$ and $\mathbb{I}_{M_7}$ to constrain $\mathbb{I}_{M_4}$. Then we propagate backward to $\mathbb{I}_{M_1}$ from $\mathbb{I}_{M_5}$ and $\mathbb{I}_{M_4}$, first combining with the measured constraints to $\mathbb{I}_{M_4}$, if any.*

4.1   Utilizable sub-pathways in forward and backward propagation

In order to constrain the isotopomer distributions of metabolites as well as possible, in forward (backward) propagation we would like to utilize as many upstream (downstream) metabolites as possible. Which metabolites then can be used? Let the reactions $\Lambda = \{\lambda_1, \ldots, \lambda_k\}$ and metabolites $\{M_0, \ldots, M_k\}$ compose a pathway so that M_j is the product of λ_j and a reactant of λ_{j+1}. Let us denote by $F_\Lambda(M_0, M_k)$ the locations within M_k that are mapped from reactant M_0 by Λ

Clearly, if for each M_j there are no producers other than λ_j, the $F_\Lambda(M_0, M_k)$-cumomer distribution can be computed iterative forward propagation starting from M_0. However, if there are two or more producers of some $M_j, j > 0$, the distribution $\mathbb{I}_{M_2}$ only constributes to a fraction of the metabolite pool C_k. The size of the fraction is decided by the fluxes of the different producers of M_j.

Hence, the maximal sized-pathway that can be utilized is defined as $\{\lambda_{h+1}, \ldots, \lambda_k\}$ where $h = \max_{j=0}^{k}\{j | M_j$ has two or more producers$\}$ Note that, if M_k has two or more producers, the above pathway is empty. In other words, the are no upstream metabolites that can be used to constrain $\mathbb{I}_{M_k}$.

Considering backward propagation to M_0, then, we see that if there are no backward junctions among the metabolites $M_0, \ldots, M_k$, iterative backward propagation starting from M_k can be used to obtain a constraint for $\mathbb{I}_{M_0}$. How-ever, if $M_j, j > 0$ has two or more producers, its isotopomer distribution and that of the downstream metabolites $M_{j+1}, \ldots, M_k$ only explain a fraction of the distribution $\mathbb{I}_{M_0}$, the size of this fraction being dependent on the fluzes again. So, in an analogy to forward propagation, we define the utilizable pathway as $\{\lambda_1, \ldots, \lambda_{h-1}\}$ where $h = \min_{j=1}^{k}\{j | M_j$ has two or more producers$\}$ Note that if M_1 has two producers, then the utilizable pathway is empty.

4.2 The backward propagation algorithm

Let $\mathcal{M}_B$ denote the set of metabolites that have to or more producers (the backward junctions), $\mathcal{M}_R$ denote external reactants and $\mathcal{M}_P$ denote the external products. The pseudocode of the bachward propagation algorithm is the following.

begin
$\Lambda_{work} = \emptyset$;
$\mathcal{M}_{work} = \mathcal{M}_P$;
for each $M \in \mathcal{M}_B$ **do**
    for each product edge $e = (\lambda, M)$ of M **do**
        Set trivial isotopomer constraint for e and mark e ready;
        if all product edges of λ are ready **then**
          Put λ into Λ_{work};
    od
od
while $\mathcal{M}_{work} \neq \emptyset$ **or** $\Lambda_{work} \neq \emptyset$ **do**
    for each $M \in \mathcal{M}_{work}$ **do**
        Compute constraint for $\mathbb{I}_M$ from its reactant edges
        (M, λ) and the measurement for $\mathbb{I}_M$, if any;
        If $M \notin \mathcal{M}_B \cup \mathcal{M}_R$ **then**
          Propagate backward the isotopomer constraint from
          M to the (only) product edge $e = (\lambda, M)$ and mark e ready;
          if all product edges of λ are ready **then**
            Put λ into Λ_{work};
    od
    Remove M from $\mathcal{M}_{work}$;
    for each $\lambda \in \Lambda_{work}$ **do**
        Propagate the constraints from the product edges $e = (\lambda, M)$
        to the reactant edges $e' = (M', \lambda)$ of λ;
        Mark the reactant edges e' ready;
        for each reactant edge $e = (M, \lambda)$ **do**
        if all reactant edges of M are ready **then**
          Put M into $\mathcal{M}_{work}$;
    od
    Remove λ from Λ_{work};
    od
od
end

4.3   The forward propagation algorithm

Forward propagation starts from the expternal reactants and backward junctions of the network and push isotopomer constraints towards the product edges leading to backward branched junctions and external metabolites. Note that applying forward propagation after backward propagation is crucial.

begin
$\Lambda_{work} = \emptyset$;
$\mathcal{M}_{work} = \mathcal{M}_R \cup \mathcal{M}_B$;
while $\mathcal{M}_{work} \neq \emptyset$ **or**$\Lambda_{work} \neq \emptyset$ **do**
       for each $M \in \mathcal{M}_{work}$ **do**
              Obtain a constraint for $\mathbb{I}_M$ from the measurement, if any, for it, and,
              if $M \notin \mathcal{M}_B$ the product edge (λ, M), if any.
              for each reactant edge $e = (M, \lambda)$ **do**
                     Propagate forward the isotopomer constraint from
                     M to e and mark e ready;
                     if all reactant edges $e = (M, \lambda)$ of λ are ready **then**
                         Put λ into Λ_{work};
              od
              Remove M from $\mathcal{M}_{work}$;
       od
       for each $\lambda \in \Lambda_{work}$ **do**
              Propagate the constraints from the reactant edges $e = (M, \lambda)$
              to the product edges $e' = (\lambda, M')$of λ;
              Mark the product edges e' ready;
              for each product edge $e' = (\lambda, M')$ **do**
                  if $M' \notin \mathcal{M}_B \cup \mathcal{M}_P$ **then**
                     Put M' into $\mathcal{M}_{work}$;
               od
               Remove λ from Λ_{work};
       od
od
end

Chapter 5

Discussion

The presented framework for flux estimation fullfills several important criteria of a tool that is intended for biochemical engineers and biologists:

- The system is designed to tackle any network topology that is given to it for the basis of analysis. So the mathematics does not need to be changed whenever a new organism with a topology different from before is encountered.

- The system uses the data that is available, with no prior assumption of completeness. For example, if metabolic intermediates are not measured, the system still estimates the fluxes as far as possible. Moreover, both NMR and mass spectroscopic data can be used seamlessly side by side.

- In situations where there is no unique point-solution to the problem or the system could not find one, the system outputs the feasible (in its view) solution set as a whole, instead of one solution from that set. More-over, outputs the error, 'lack-of-fit' between the experimental data and the fluxes, which can be used as a basis for, for exmaple, revising the topological assumptions.

The framework is by no means perfect in its presented form. Several areas of improvement can be distinguished:

- Compartments. The current system works in the 'bag-of-enzymes' model, which ignores any membranes separating metabolic pools. Taking the compartments into account is certainly possible. However, we expect flux estimation to be more involved in that case.

- Bidirectional reactions. In the current version, bi-directional reactions conceptually induce metabolic junctions in the network, hence requiring the measurement of all reactants and products of the reaction in order to compute the exchange flux. We believe that this can be done more cleverly so that the measurement requirements are not as great.

- White-areas of the network, places where no isotopomer information is available for propagation towards a junction, present a problem for anal-

ysis. This is mostly due to the linear programming framework which dictates that balance equations need to be constructed junction-per-junction basis. Using higher-order balance equations could enable us to analyze larger subnetworks than one-junction systems.

In any case, we believe that the presented framework shows that using direct flux estimation in place of the current iterative methods, has potential of which only a part has been explored.

Bibliography

[1] Arita, M. Metabolic reconstruction using shortest paths. Simulation Practise and Theory **8** (2000), 109–125

[2] Christensen, B., Nielsen, J. Isotopomer Analysis Using GC-MS. Metabolic Engineering **1** (1999), (E8)–(E16)

[3] Gaasterland, T., Selkov, E. Reconstruction of metabolic networks using incomplete information. Proc. 3rd International Conference on Intelligent Systems for Molecular Biology, ISMB-95, 1995, 127–135

[4] Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K., Kanehisa, M. Organizing and computing metabolic pathway data in terms of binary relations. Pacific Symposium on Biocomputing, 1997, 175–186

[5] Jeffrey, M., Roach, S., Storey, C., Sherry, D., Malloy, C. ^{13}C Isotopomer Analysis of Glutamate by Tandem Mass Spectrometry. Analytical Biochemistry **300** (2002), 192–205

[6] Kelleher, J., Flux estimation Using Isotopic Tracers: Common Ground for Metabolic Physiology and Metabolic Engineering. Metabolic Engineering **3** (2001), 100–110

[7] Klapa, M., Park, S., Sinskey, A., Stephanopoulos, G. Metabolite and Isotopomer Balancing in the Analysis of Metabolic Cycles: I. Theory. Biotechnology and Bioengineering **62** (1999), 375–391

[8] Maaheimo, H., Fiaux, J., Cakar, P., Bailey, J., Sauer, U., Szyperski, T., Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional 13C labeling of common amino acids. European Journal of Biochemistry **268** (2001), 2464–2479

[9] Marx, A., de Graaf, A., Wiechert, W., Eggeling, L., Sahm, H. Determination of the fluxes in the central metabolism of Corynebacterium glutamicum by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnolgy and Bioengineering **49** (1996), 111–129.

[10] Mavrovouniotis, M., Stephanopoulos, G., Stephanopoulos, G. Synthesis of Biochemical Production Routes. Computers and Chemical Engineering **16**, 6 (1992), 605–619

[11] Rantanen, A., Rousu, J., Kokkonen, J., Tarkiainen, V., Ketola, R., Computing Positional Isotopomer Distributions from Tandem Mass Spectrometric Data. Metabolic Engineering, to appear

[12] Sanford, K., Soucaille, P., Whited, G., Chotani, G. Genomics to fluxomics to physiomics – pathway engineering. Current Opinion in Microbiology **5** (2002), 318–322

[13] Schmidt, K., Carlsen, M., Nielsen, J., Viladsen, J. Modeling Isotopomer Distributions in Biochemical Networks Using Isotopomer Mapping Matrices. Biotechnology and Bioengineering **55** (1997), 831–840

[14] Schott, J. Matrix analysis for statistics. Wiley, 1997

[15] Širava, M., Schäfer, T., Eiglsperger, M., Kaufmann, M., Kohlbacher, O., Bornberg-Bauer, E. and H.-P. Lenhof. Biominer— Modeling, analysing, and visualizing buichemical pathways and networks. Proc. Europan Conference on Computational Biology 2002. Bioinformatics **18** suppl. 2 (2002), S219–S230

[16] Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M. and Nishioka, T. Simultaneous Determination of Anionic Intermediates for Bacillus subtilis Metabolic Pathways by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry. Analytical Chemistry **74**, 10 (2002), 2233–2239

[17] Stephanopoulos, G., Aristidou, A., Nielsen, J. (1998) Metabolic engineering: Principles and Methodologies. Academic Press.

[18] Szyperski, T., Glaser, R., Hochuli, M., Fiaux, J., Sauer, U., Bailey, J., Wütrich, K. Bioreaction Network Topology and Metabolic Flux Ratio Analysis by Biosynthetic Fractional ^{13}C Labeling and Two-Dimensional NMR Spectrometry. Metabolic Engineering **1** (1999), 189–197

[19] Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., de Graaf, A. Bidirectional Reaction Steps in Metabolic Networks: III. Explicit Solution and Analysis of Isotopomer Systems. Biotechnology and Bioengineering **66** (1999), 69–85

[20] Wittmann, C., Heinzle, E., Mass Spectrometry for Metabolic Flux Analysis. Biotechnology and Bioenginering **62** (1999), 739–750

Appendix A

Matrix algebra

A.1 Computing a basis for the intersection of vector spaces

Let $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{S}$ be two vector subspaces with orthonormal bases $\{\vec{u}_{11}, \ldots, \vec{u}_{1r}\}$ and $\{\vec{u}_{21}, \ldots, \vec{u}_{2r'}\}$, respectively. We need to find a set of vectors $\vec{w}_1, \ldots, \vec{w}_{r''}$ satisfying

- $\vec{w}_j \in S_1$ and $\vec{w}_j \in S_2$ for all j,

- the set of vectors is linearly indepenent, and

- the set has maximal size

A feasible vector satisfies $\vec{w} = \sum_{i=1}^{r} \alpha_i \vec{u}_{1i}$ for some $\alpha_1, \ldots, \alpha_r$ and $\vec{w} = \sum_{i=1}^{r'} \beta_i \vec{u}_{2i}$ for some $\beta_1, \ldots, \beta_{r'}$ meaning that

$$\sum_{i=1}^{r} \alpha_i \vec{u}_{1i} - \sum_{i=1}^{r'} \beta_i \vec{u}_{1i} = 0 \tag{A.1}$$

which can be stated in matrix form as

$$[U_1 U_2]\vec{c} = [\vec{u}_{11}, \ldots, \vec{u}_{1r}, \vec{u}_{21}, \ldots, \vec{u}_{2r'}] \begin{bmatrix} \vec{a} \\ -\vec{b} \end{bmatrix} = 0$$

The vectors $\vec{c}$ consistent with the above system make up the null space of U. Let thus the columns of the matrix

$$C = [\vec{c}_1, \ldots, \vec{c}_k] = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_k \\ \vec{b}_1 & \cdots & \vec{b}_k \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix}$$

define a basis for the null space of U. For any j, the vector $\vec{v}_j = U_1 \vec{a}_j = -U_2 \vec{b}_j$ satisfies (A.1) and thus lies in the intersection space. Moreover, any vector $\vec{w}$ in the intersection space must satisfy $w = U_1 A \vec{d} = U_2 B \vec{d}$, for some $\vec{d} \in \mathbb{R}^k$. It follows that the columns of the matrix $S' = U_1 A = -U_2 B$ span the intersection space $\mathcal{S}_1 \cap \mathcal{S}_2$. A basis for the intersection space is finally obtained by selecting a set of linearly independent columns from S' spanning the same space.

A.2   Computing a linearly independent equation system consistent with the least-squares solution

Let us consider the system of linear equations

$$A^T \vec{x} = \vec{y} \tag{A.2}$$

where the columns of A may be linearly dependent and hence the system (A.2) may be inconsistent. Our goal is to transform the system into another system

$$S^T \vec{x} = \vec{d} \tag{A.3}$$

where S has linearly independent columns – thus the system is consistent – and the solution to (A.3) is a least-squares solution to (A.2).

Let the columns of U form an orthonormal basis for the column space of A and let the columns of $U_\perp$ form an orthonormal basis for the orthogonal complement of A's column space. Now all least-squares solutions to (A.2) can be expressed as

$$\vec{x}_{LS} = [U U_\perp] \begin{bmatrix} \vec{d} \\ \vec{d}_\perp \end{bmatrix}.$$

where $\vec{d}$ is a fixed vector and $\vec{d}_\perp$ is arbitrary. Let $\vec{x}^*_{LS}$ be the vector that is closest to origin, that is, $\vec{d}^*_\perp = 0$. Now we have

$$[U U_\perp] x^*_{LS} = \begin{bmatrix} U^T x^*_{LS} \\ U_\perp^T x^*_{LS} \end{bmatrix} = \begin{bmatrix} \vec{d} \\ 0 \end{bmatrix},$$

which gives us $\vec{d} = U^T \vec{x}^*_{LS}$. Since the projection of all vectors $\vec{x}_{LS}$ to A's column space needs to be $\vec{x}_{LS}$, the solution we want is obtained by substituting $S = U$ and $\vec{d} = U^T \vec{x}^*_{LS}$.

$\vec{x}^*_{LS}$ can be computed by the pseudo-inverse function *pinv* of Matlab.