university of groningen

faculty of science and engineering

mathematics and applied mathematics

Elementary Interaction Modes provide a molecular description for communities of microorganisms

Bachelor's Project Mathematics

December 2020

Student: T.M. Papousek

First supervisor VU Amsterdam: D.H. de Groot Msc.

Second supervisor VU Amsterdam: A. Tjalma Msc.

First supervisor University of Groningen: Prof. dr. A.J. van der Schaft

Second assessor University of Groningen: dr. ir. B. Besselink

Abstract:

This research was done in cooperation with the Systems Biology Lab of the VU Amsterdam. In this paper a generalization of Elementary Flux Modes (EFMs) is made towards communities of micro-organisms. To do so a community of microorganisms is assumed to grow at a steady-state with a fixed growth rate across all species in the community. The resulting "Elementary Interaction Modes" (EIMs) provide a mathematical basis for analyzing such communities of microorganisms that grow in a chemostat (a type of bioreactor). The model used to compute the EIMs was constructed after analyzing different earlier efforts made to generalize "Flux Balance Analysis" towards communities of microorganisms. With this knowledge the model was tested on a toy model of 2 organisms and altered until it functioned properly, biologically speaking. Afterwards, the resulting model was generalized towards any community of 2 species and subsequently towards any community of n species. Based on this developed notion of EIMs for communities of n species, several theorems were proven which together put an upper bound on the complexity (in the number of EFMs used) in a community of n species that grows at its maximal growth rate.

Contents

1 Introduction

1.1 Research problem

Micro-organisms often form large communities that by themselves play a part in an even larger system. For example, some communities of micro-organisms help digest food inside the human intestines. In fields such as medicine or industrial engineering a better understanding of these communities of micro-organisms could benefit for example treatments of diseases or improve production processes. A sound general understanding of such communities and their flexibility as a whole may also become a basis for further academic research.

Micro-organisms often use metabolic [1] interactions with other micro-organisms (e.g. cross-feeding or parasitism) to grow faster than they would have done on their own. Due to the development in the last few decades of genome-scale sequencing it is becoming more and more possible to reconstruct all the possible reactions that can occur inside a (simple) organism by reading out and analyzing their DNA. Based upon this information one may now be able to describe the complete metabolic functioning of a single organism. But although the metabolic functioning of each individual micro-organism might be known, due to the complexity in which the species interact with each other in a community, a qualitative approach would be far too time consuming to learn to understand how the community functions. However, due to evolution, it is very likely that the microorganisms each grow in an optimal way, that is, with maximized steady-state growth. This is because organisms with an optimal approach have an edge over rivaling organisms with a non-optimal approach, as they grow faster and therefore, in the long run will be more abundant, which ensures their survival. Thus, over time, it is very likely that only the most optimal organisms survive [1]. Therefore, it might be possible to compute the way in which microorganisms grow by using mathematical optimization techniques.

For single organism metabolism, mathematical models have been used for quite some time now in this field. But here, two applications are especially relevant:

- In order to better understand the combinations of reactions that lead to steady state growth that can be used by an organism, the concept of Elementary (Flux) Modes (EFMs) [2] [3] was developed. Although computationally intensive, this gives a set of elementary metabolic pathways of reactions which form a basis for all the possible metabolic combinations of reactions that lead to steady state growth of the organism.
- The second group of noteworthy applications are formed by optimization techniques such as the renowned Flux Balance Analysis (FBA)[4] which use apart from the stoichiometry [2] also the biochemical constraints of the environment. In this way, by maximizing a chosen objective function

[1] The whole range of chemical processes that occur within a living organism.
[2] The ratio of reactants and products in chemical reactions.

(e.g. reproduction) the technique computes the optimal flow through the metabolic network.

Previous efforts have already been made to extend the optimization techniques as used for metabolic networks of single organisms to microbial communities, by for example community Flux Balance Analysis (cFBA) [5] and SteadyCom [6]. For the elementary modes however, there has not yet been made a generalization towards communities of microorganisms. In this paper an attempt is made to define such modes.

Note that due to the fact that there are different species active in such a community the ratios between the species may vary. Therefore, the single species approaches cannot strait away be generalized, as they assume a very rigid constant ratio between species that can only be changed by extensively altering the network.

1.2 Research question:

Can we, for any given set of metabolic interactions, mathematically define a minimal set of modes that can be used to describe all forms of metabolic interactions amongst microorganisms that together lead the community as a whole to achieve steady-state growth?

Besides this main research question we define the following sub-research questions:

Sub-research question:

- Given a community of 2 species, can we define a set of elementary modes that can be used to describe all its possible metabolic behaviour? Can we generalize this set of modes to any community of 2 species? And to any community of n species?
- Can we relate a single one of these modes to one or more EFMs? Is there an upper bound on the number of EFMs that are related to one mode?
- Given a community that grows at an optimal rate, how many different elementary modes are used by that community? Is there a non-trivial upper bound?

1.3 Sketch of approach to research question:

First, as this research is applicable for chemostat[3]-growth we will give a brief introduction of the chemostat environment. Afterwards, we will follow with a mathematical basis of the methods used in the rest of the paper. Once we have provided the basic framework, we will follow with a section on two earlier

[3] A certain type of bioreactor

published optimization methods for analyzing communities of microorganisms. Then, by using these methods as inspiration, we will design our own method to find a set of elementary modes for communities. After this method starts to take form, we will test it on a toy model, to see if the output corresponds with what would be expected biologically. If the method indeed functions properly, we will start generalizing it. First for communities of 2 species, and then, for communities of n species. After we have derived at this general method for communities of n species, we will start analyzing the general elementary modes that it has as output and prove several theorems which together will put an upper bound on the expected complexity of communities of microorganisms in terms of how many EFMs are used.

2 Research

2.1 Application to bioreactor

As mentioned before, this paper is written towards an application in a chemostat. A chemostat is a bioreactor in which medium constantly enters and leaves the reactor at a fixed rate which is set by the experimenter [7]. In order to keep the medium inside the chemostat well mixed, it is continuously stirred.

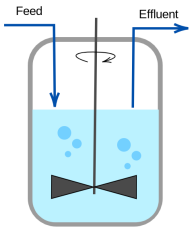

Figure 1: A schematic picture of a chemostat.

In this way a community can be grown at steady state under constant environmental conditions. Since the inflow of nutrients can be controlled, the system can be set up such that the availability of the nutrients is the only limiting factor for the growth of the microorganisms. By controlling the dilution rate the excess material can flow out (the medium contains waste products as well as nutrients and the microorganisms itself). In that way the micro-organisms can grow continuously, without depriving nutrients, overpopulating the chemostat nor creating too much waste products that limit the growth. While the global dilution rate is set, the dilution rate of a of a certain metabolite[4] is dependent on its metabolite concentration in the medium. As the community of microorganisms grows, the concentration of each organism in the medium rises. So at a constant dilution rate and constant inflow the concentration of the metabolites that are consumed drop, while the concentration of metabolites that are excreted increase. Therefore, in order for the system to remain in steady state, it is crucial that the inflow and dilution rate scale with the consumption of the metabolites. i.e. the inflow and dilution rate need to scale with a factor of the growth rate of the micro-organisms.

To illustrate this, let us describe the environment mathematically: We start with a chemostat of volume V filled with medium. The inflow and outflow of the chemostat together give rise to a certain amount of flow (e.g. in liters/hour) through the chemostat. Let us call this volumetric flow rate Q. As both V and Q are based on volumes, and we are in fact only interested in the ratio between them, let us introduce $q = \frac{Q}{V}$ as the dilution rate which has unit ratio per unit

[4] A biochemical compound

of time.

Now let us look at the medium in a chemostat. In such a medium firstly there will be micro-organisms whose concentrations will vary throughout time, and secondly metabolites concentrations which will also vary throughout time. In biology it is standard practice to measure the concentration of micro-organisms in grams of dry weight and metabolite concentrations in grams per liter. If there are n species active in the medium, we can measure for each of the species its concentration in the medium. Let $X_i(t)$ be the mass of micro-organism i at time t in the medium. Furthermore, if there are m metabolites in the medium, we can measure the concentration of each of the m metabolites in the medium. let $c_j(t)$ be the concentration of metabolite j. Additionally, let us describe the growth rate of the community as a whole by μ. In this paper, we assume that growth is enzyme mediated, which means that growth depends on reactions catalyzed by enzymes. These reactions in turn depend on the concentrations of the metabolites in the medium. Hence, we have that μ is a function of the $c_i(t)$, which together form the vector $c(t)$. Now that we have a basic framework for the components in the chemostat, we may describe the fluctuations of both the amount of microorganisms and the concentrations of metabolites in the chemostat by a set of differential equations derived from mass balance equations and the laws of mass action:

For each species i we have that the amount of micro-organisms in the medium grows with the growth rate $\mu(c(t))$. At the same time a part of the medium, containing micro-organisms flows out with a dilution rate q. This together gives us:

$$\frac{dX_i}{dt}(t) = \mu(c(t))X_i(t) - qX_i(t) \tag{1}$$

which states that the change in biomass $X_i(t)$ over time is equal to the difference between the biomass that was created through growth $(\mu(c(t))X_i(t))$ minus the amount of biomass that is diluted out of the system$(qX_i(t))$. For every metabolite i we have that it can flow into the medium through the inflow, it can flow out of the medium through the overflow or it can be consumed or produced by one of the micro-organisms. Here both the inflow and outflow are coupled to the dilution rate, and therefore the metabolite concentrations also depend on the dilution rate. The production and consumption of the metabolite by the micro-organism of course depend on the amount of the micro-organism in the medium. Together this gives us for the ith metabolite in the medium the following equation:

$$\frac{dc_i}{dt}(t) = qc_{i0} - qc_i(t) - \sum_{j=1}^{n} c_{ij}\mu(c_i(t))X_j(t) \tag{2}$$

where c_{i_0} signifies the concentration of the metabolite in the inflowing medium. c_{ij} is a constant which indicates how much of the metabolite is consumed or pro-

duced by organism j. The equation states that the change in concentration of a metabolite over time equals the inflow (qc_{i0}) minus the outflow ($qc_i(t)$) minus the consumption (plus the production) by the microorganisms of the metabolite ($\sum_{j=1}^{n} c_{ij}\mu(c_i(t))X_j(t)$).

In order to operate a chemostat continuously the system needs to be at steady state, i.e. the micro-organisms need to be able to grow continuously, without depriving nutrients, overpopulating the chemostat nor creating too much waste products that limit the growth. Hence we are interested in the case where the populations of the micro-organisms are stable over time in the chemostat while the concentrations of the metabolites remain constant (this is done by controlling the in- and outflow). This implies that we need to fullfil the following equations:

$$\frac{dX_i}{dt}(t) = 0 \tag{3}$$

$$\frac{dc_i}{dt}(t) = 0 \tag{4}$$

Note that together with equation (1), equation (3) also implies that $q = \mu(c(t))$. Which shows that indeed the dilution rate needs to be equal to the growth rate, just as we argued before.

2.2 Mathematical groundwork

Here we will look at the two single organism orientated applications mentioned in the introduction: the concepts of "Elementary Flux Modes (EFMs)" and "Flux Balance Analysis (FBA)". Both techniques were created for the analysis of metabolic networks. To do so, the metabolic network of an microorganism is reconstructed by analyzing its DNA. This can be done as in the DNA all the enzymes of a micro-organism are encoded. Since each enzyme typically catalyzes a specific reaction, this information can be used to reconstruct the entire metabolic network. Subsequently, this information can be transferred into a matrix:

Definition 1: Stoichiometric Matrix
The stoichiometric matrix (S) is a $n \times m$ matrix, containing the information of all the metabolic reactions of a micro-organism. Every reaction has a column which denotes exactly how much of which metabolite(s) is used and how much of which metabolite(s) is produced by that particular reaction. In each row, one can read for every metabolite exactly which reactions use and and which produce it. Together this contains all the information on the metabolic network. [8]

From this metabolic network can be read what metabolites are taken from the environment, how they react inside the cell and what metabolites are exported back out of the cell. Each reaction is typically catalyzed by an enzyme and can be either reversible or irreversible catalyzed (most reactions are reversible but not all are within the environment of the cell). Every reaction has a certain flux (v_i) associated to it, this flux is the amount of the reaction that occurs. The vector with as its elements the fluxes, is called a flux vector (v). For the micro-organism to be at steady state (so that no metabolite is accumulated nor depleted) the rate at which any metabolite is consumed needs to equal the rate at which it is produced. This implies that the flux vector needs to meet the relation: $Sv = 0$. Now, for mathematical convenience, we will without loss of generality assume that all fluxes are non-negative (i.e. that all reactions are irreversible). This can be done by splitting any flux (v_i) of a reversible reaction into two fluxes: a forward flux (v_i^+) and a backward flux (v_i^-) so that $v = v_i^+ - v_i^-$ where $v_i^+ \geq 0$ and $v_i^- \geq 0$. Note that this slightly alters matrix S as well. Thus for any micro-organism at steady state we have :

$$Sv = 0, \quad v \geq 0.$$

Example 1: Stoichiometric Matrix
Let us consider the metabolic sample network in figure 2 to illustrate how such a stoichiometric matrix is constructed:

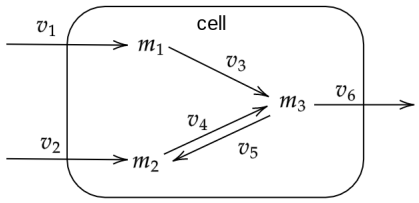

Figure 2: A metabolic sample network where the m_i's are the metabolites and the v_i's are the fluxes.

This metabolic network can be translated into the following stoichiometric matrix:

$$S = \begin{array}{c} \\ m_1 \\ m_1 \\ m_1 \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{bmatrix} \end{array}, \quad v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}$$

The j-th column describes which metabolites reaction j transfers into what products. Here the i-th row describes which reactions make or use metabolite i.

If we now assume steady state for this sample network (i.e. $Sv = 0$ where $v \geq 0$), we see that certain relations between the v_i's must hold.

In general, the solution space for a non-negative flux vector v can be described geometrically by a cone (a "flux cone"). From now on this cone will be referred to as FC.

2.2.1 Elementary Flux Modes:

Let us, for a flux vector v define $supp(v) = \{i|v_i \neq 0\}$, i.e. the set of nonzero fluxes in a flux vector. As we assumed without loss of generality that all fluxes are non-negative, in this paper we have $supp(v) = \{i|v_i > 0\}$. We may now use this notion for the following definition:

Definition 2: Elementary Flux Mode (EFM)
The Elementary Flux Modes are the extreme rays of the flux cone FC. That is: $e \in FC$ is an EFM if there does not exist a ray $r \in FC$ such that

$$supp(r) \subset supp(e)$$

As we already assumed all elements of v to be non-negative we know that the EFMs span the entire solution space (the entire solution space is non-negative and therefore the EFMs, which are also non-negative, can describe the entire solution space). As is known from linear programming, The solution space FC, forms a pointed cone [9]. Therefore any steady-state metabolic behaviour of a micro-organism can be described by a convex combination of its EFMs.

Example 2: Let us illustrate the calculation of the EFMs with the example network in figure 3, where we assume the model to be in steady-state:

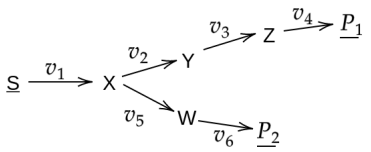

Figure 3: Practice Network: The capital letters represent metabolites, where the underlined metabolites are external, and therefore are set by the environment. As is standard practice S denotes a substrate, and P the products. The v_i arrows represent the fluxes per reaction.

Based on this network we can construct the stoichiometric matrix. Because the figure already features an S we will call the stoichiometric matrix here N:

$$N = \begin{array}{c} \\ X \\ Y \\ Z \\ W \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \end{array}$$

where the rows represent the metabolites and the columns the reactions.
Since we assumed the model to be in steady-state, we have:

$$Nv = 0 \text{ and } v_i \geq 0 \text{ where } v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}$$

We note that we have 6 variables with 4 equality constraints and 6 inequality constraints. Since we know that an EFM will be located on an extreme ray which is by definition 1-dimensional, and as the space is 6-dimensional, the EFM will be located at the intersection of 6-1=5 (independent) constraints. Therefore, aside from the 4 metabolic constraints at least one flux will be set to zero to saturate the inequality constraint. Furthermore, we note that if we for example set $v_5 = 0$ it forces v_6 to zero as well. The W constraint however would become a trivial constraint (a row with only zeros, so not an independent constraint anymore). So in that case we would again fulfill exactly 5 nontrivial constraints (2 saturated inequalities and 3 metabolite constraints). Since for any solution of v we have $Nv = 0$ we note that setting $v_5 = 0$ forces $v_6 = 0$ and vice versa. In the same way setting one of v_2, v_3 or v_4 to zero forces the other two to zero as well. We note that in all of these cases there are 5 nontrivial constraints active. Setting $v_1 = 0$ forces all other fluxes to zero. Hence there are only two viable nontrivial EFMs:

•EFM$_1$: $v_1 = 1$, $v_2 = 1$, $v_3 = 1$, $v_4 = 1$, $v_5 = 0$ and $v_6 = 0$.

•EFM$_2$: $v_1 = 1$, $v_2 = 0$, $v_3 = 0$, $v_4 = 0$, $v_5 = 1$ and $v_6 = 1$.

An FC is typically, without environmental conditions, an infinite solution space. Because in practice there often are boundaries to the solution space, it makes sense to also define finite versions of the EFMs. We do so as follows: For a given EFM e with $supp(e) = \{i|e_i > 0\}$ we can take a (finite) vector v with $supp(v) = supp(e)$. In this paper, we will call such a vector v an Elementary Flux State (EFS). The difference here between e and v is that e is a pointed infinite extreme ray where v is a finite vector in the direction of e.

Definition 3: Elementary Flux State (EFS)
An Elementary Flux State is a finite vector on an EFM (which in this case coincides with the extreme rays) of the flux cone FC.

This definition will be valuable later on in the paper. But let us for now leave it as it is, and dive into the other fundamental technique in this paper, called Flux Balance Analysis.

2.2.2 Flux Balance Analysis:

The other application that we will look at in this paper is running a linear program over the solution space FC to find the optimal solution and most often reproduction is chosen as the objective function. Clearly, running a linear program is only possible if the flux cone is bounded (in the direction of the objective function) and in order to make this happen, additional physiological constraints are added to bind the FC into a "Flux Polyhedron", from now on referred to as a FP. After these additional constraints are added such that the solution space becomes a bounded polyhedron we can use the main theorem from linear programming which says that the optimal solution will lie on one of the vertices of the bounded polyhedron.

Flux Balance Analysis maximizes the objective function over this solution space. Typically it assumes the network to be in steady state and uses constant upper- and lower bounds for each flux, based on experimental data. To keep notation further on in the paper relatively neat, we will here use a bit quirky notation for the upper and lower bounds on the fluxes. As an objective function the production of the desired substance is used, often the production of biomass is chosen. If we put all of this together we obtain a linear program of the form:

$$Maximize: \ c^T v$$
$$Subject\ to: \ Sv = 0$$
$$and: \ LB \leq v \leq UB$$

Where $c^T v$ is the objective function, $Sv = 0$ is the steady-state assumption and LB and UB are the upper bound and lower bound vectors of v.

2.2.3 Outlook

Both the concepts of EFMs and of FBA are designed for metabolic networks of one single organism, hence one might question:

Can we just as easily apply the concepts of Elementary Flux Modes and Flux Balance Analysis to communities of micro-organisms? i.e. Are these methods scalable to function on communities of micro-organisms instead of on single organisms?

From a mathematical point of view this would be a logical next step. We have the metabolic networks, so let us paste them together and reapply these techniques. The devil here however is in the ratios between species, and their variability. Theoretically if we take a community of 2 species with an exact fixed ratio of 1:1, you could add the two networks together and look at them as a single organism. Or if it was an exact fixed ratio of 2:1 instead of 1:1, you could add the whole network of the first species again to the model so that the community again could be seen as a single organism. In practice however, that gives these methods a very limited reach in the application on communities of micro-organisms. In any community of micro-organisms the ratios might vary throughout time or they could only be expressed in integers using really large numbers which would skyrocket the computing time. These issues can as of now not be fixed with an easy trick. Therefore other approaches are needed to analyze such communities of micro-organisms. For Flux Balance Analysis, as we will see in the next section, such techniques already exist. In this paper we will try to somehow mimic those techniques so that we can extend the use of Elementary Flux Modes to communities of micro-organisms as well.

To fully understand how the ratios between species mess with the single organism techniques mathematically, let us go back to the constraints these methods function upon. Up to now we had for an single micro-organism:

$$Sv = 0$$

here v is the flux vector with all the specific fluxes of the micro-organism in $\frac{mmol}{hour*gdw}$ and S is the stochiometric matrix of the micro-organism. gwd here, stands for the amount of the micro-organism in grams of dry weight. So it is

the amount of mol going through as flux per hour, per amount of the micro-organism. So if the amount of a certain micro-organism changes, it is not reflected in its specific flux v (as the amount of (absolute) $mmol$ rises with the total gwd and it is therefore divided out). This means that the model, if applied to a community of micro-organisms, is not altered if the ratio between species changes and therefore the model will become a false representation of the community.

To mediate this we will rewrite v to $v = \frac{V}{X}$ where V is the absolute flux in $\frac{mmol}{hour}$ and X is the biomass of the species in gwd. Hence our equation becomes:

$$S\frac{V}{X} = 0$$

Because for any non-trivial community $X \neq 0$ we can now also rewrite this to:

$$SV = 0.$$

In the case of a single organism, this equation is interchangeable with $Sv = 0$ because all specific fluxes v are multiplied by the same value of X to obtain the absolute fluxes V. But if we have more than one species there will be different values of X for different specific fluxes, which means that the absolute fluxes V will change along with the different abundances X of the different species, which solves the problem the equation $Sv = 0$ had. If we now apply this to a community of n species we will have for every species i, that we multiply all the specific fluxes v_i with different amounts of X_i to get all the different absolute fluxes V_i. Note that here every species also has its own stoichiometric matrix S_i so that in the case of n species we would get:

$$\sum_{i=1}^{n} S_i V_i = 0.$$

Which exactly states that the whole community of micro-organisms as a whole is at steady state. As this solves the issue that arose by changing ratios amongst species, this will now become the basis for our approach of communities of micro-organisms.

An additional practical reason that this change is necessary is that we cannot write V as $V = vX$ (with v as our variables because the specific abundances(X_i) are all unknown and only the dry weight of the entire community ($\sum_{i=1}^{n} X_i$) can be measured, which inhibits us from using the specific fluxes as variables.

A final issue we should name before we dive any deeper, is caused by the inevitable change in objective function in the case of the FBA. As we will now consider a community of different species of microorganism there will not be a clear objective anymore. Each species is in principle selected for its own reproduction and not for the reproduction of the whole community. On top of

that, due to the often occurring inter-dependencies of the metabolic objectives of the different organisms in the community, each organism's objective function is often directly or indirectly dependent on all the the other objective functions.

This can, generally speaking, best be solved by using a multi-objective optimization technique. That however runs rapidly into computational issues. To bypass that situation, it is common to assume a steady state community growth rate i.e. a (constant across species) community growth rate. Although at any specific moment in time this might not be the case, due to the steady state assumption it should be the case on average throughout time. Therefore it is likely that this average community growth rate will be a good estimator for the growth rate of each specific species growth rate. All the methods from here onward considered in this paper, including our own story, adhere to this assumption.

2.3 Previous efforts to generalize FBA to communities

Later on in this paper our goal is to generalize Elementary Flux Modes (EFMs) towards communities of micro-organisms. In this section we will look at two techniques that do exactly that, but for Flux Balance Analysis (FBA). By analyzing these two techniques we hope to find inspiration on how to do this for Elementary Flux Modes (EFMs) as well. The two techniques we will consider in this section are named cFBA and SteadCom. Both methods originate from a normal (single organism) FBA, which is extended to form a "joint" FBA which is tweaked until it is functioning properly. Here we should note that extensive papers were written on both techniques and that the goal in this paper is to find inspiration in them on how to go from a single organism to a community, not to go thoroughly through their details. Therefore in this section sometimes we will not go into detail in the exact reasoning of every step and be more concerned with the crucial parts of them.

2.3.1 cFBA:

cFBA [2013] (community Flux Balance Analysis) is a technique that shows a way around an earlier published method called OptCom [10], which uses an multi-objective function as objective function for the entire community to adjust for the presence of several different species. OptCom however, is a tool that is very heavy in computing and practically unfeasible for applications in larger communities. By instead assuming balanced growth of the entire community (as we discussed earlier), cFBA can use just a single objective function. And by fixing one of the variables in the nonlinear model that they came up with, the problem becomes linear, and in that way also solvable.

Balanced growth is here assumed to be: internal metabolism at steady-state for all the organisms in the community, while the entire community grows at a fixed (community) growth rate. This also implies that all metabolite levels

in the community are at steady-state. Hence the mass balance equations of all metabolites are set to zero.

As a start in the original paper, it is noted that three types of reactions influence the metabolite levels in the medium of the community: reactions that produce or consume a metabolite, the production of biomass which uses or produces a metabolite, and the flow in or out of the environment of a metabolite.

Based on this observation, the researchers of [5] come up with the following equations for all metabolites m_i:

$$\frac{d}{dt}m_i = \sum_{j=1}^{n_X}\sum_{k=1}^{n_R} n_{ik}q_{kj}X_j + \sum_{j=1}^{n_X} g_{ij}\mu_j X_j + \sum_{l=1}^{n_E} n_{il}J_l = 0$$

where n_{ik}, g_{ij} and n_{il} are the dimensionless stoichiometric coefficients. q_{kj} is the reaction rate of reaction k in organism j, μ_j is the biomass production rate of organism j and J_l is the total rate at which metabolite m_i flows in or out of the environment. X_j is the total biomass of species j. Furthermore, n_X is the number of species, n_R is the number of metabolic reactions and n_E is the number of reactions that exchange substances in or out of the environment.

Since $\frac{d}{dt}m_i = 0$ and the q_{kj}'s and μ_j's are constant, the researchers argue, there are two possible cases:

• Either the biomasses X_j remain fixed $\forall j$, while J_l remains fixed $\forall l$, so that the two balance out against each other. Where a constant biomass also implies (biologically) that for all j $\mu_j = 0$ i.e. that there is no growth, implying that the entire second term is equal to zero.

• Or $\mu_i = \mu_j = \mu_C \neq 0 \ \forall \ i, j$ i.e. all species grow at the same rate and J_l also increases with this same factor μ_C. Here μ_C is the community growth rate. In this case the biomasses X_j scale with the in/out flow J_l so that the total equation remains equal to zero.

The first case could be biologically possible in certain specific occasions, but the focus in the cFBA paper is on the second scenario.

The next step used in the cFBA method is to normalize the balanced growth constraints by dividing both sides by the total sum of biomass $\sum_{j=1}^{n_X} X_j(t)$. Resulting in:

$$\sum_{j=1}^{n_X} \phi_j(\sum_{k=1}^{n_R} n_{ik}q_{kj} + g_{ij}\mu_C) + \sum_{l=1}^{n_E} n_{il}q_{il} = 0$$

Where:

$$\phi_j = \frac{X_j(t)}{\sum_{j=1}^{n_X} X_j(t)} \text{ and } q_{il} = \frac{J_{il}(t)}{\sum_{j=1}^{n_X} X_j(t)}$$

In this equation the ϕ's (the mass-percentage of the total community that consists out of species j) and the q's are the only variables that can change in order to maximize μ_C. This is however still a nonlinear problem as it has a product of ϕ's and q's. Additionally, there are more variables than equations. To solve the latter, first several constraints are added based on thermodynamic relations that have to be measured/estimated for each flux experimentally (just like in single species FBA) such that: $q_{ij,min} \leq q_{ij} \leq q_{ij,max}$. Now it is noted that for the optimization of μ_C, μ_C is dependent on one or more $q_{ij,max}$ that is/are attained. And thus μ_C is optimized subject to some q.

To make this problem linearly solvable the ϕ's are set to a fixed value in this method. By then running the optimization over all values in the estimated range of the ϕ's the global optimum of μ_C can be found. This has been used successively for communities of up to 5 species.[5]

By extensive rewriting, this model can be written as a linear program and solved accordingly as follows:

$$\text{maximize: } \mu_C$$
$$\text{subject to:}$$
$$C\ \Phi\ q = 0$$
$$\text{And:}$$
$$q_{ij,min} \leq q_{ij} \leq q_{ij,max}$$

Where C is a matrix filled with constants, Φ is a matrix containing the ϕ_j's, and thus the biomass's (X_j) whose variables get new values in each iteration and q is a vector consisting of the reaction variables (q) and the community growth rate(μ_C).

To wrap it up, we have seen that cFBA has found a way to extrapolate FBA to (small) communities of micro-organisms as well. Although it is not very scalable, it is neatly written down and quite intuitive. A big problem with it is however that it is required that the total biomass of all the species as well as their ratios between species are known. Especially in larger communities this is often not the case. The assumption of a community growth rate is definitely something we will take from this paper but on the further execution it might be worth to explore other similar methods. Let us therefore look at another technique in the next subsection.

2.3.2 SteadyCom:

SteadyCom [2017] is the second technique we will analyze, it could be seen as a reformulation of the cFBA but here we will derive SteadyCom from a combination of single organism FBA models to keep it as intuitive as possible.

Such a "joint" FBA is created by extending a single organism FBA model to a multi-organism FBA model, which is basically achieved by pasting several FBA models together.

For a single organism k we have the following FBA model:

$$\begin{gathered}\text{max: } v^k_{biomass} \\ \text{subject to:}\end{gathered}$$

$$\sum_{j \in J_k} S^k_{ij} v^k_j = 0 \quad \in I^k$$

$$LB^k_j \leq v^k_j \leq UB^k_j \;\; \forall j \in J^k$$

Where: $v^k_{biomass}$ is the flux of biomass for species k(i.e. the reproduction flux), which is assumed to be one of the fluxes of v^k. v^k_j is the flux of reaction j of species k, S^k_{ij} is the stoichiometric coefficient for metabolite i in reaction j of species k. And LB^k_j and UB^k_j are the upper and lower bounds for fluxes v^k_j. I^k and J^k are respectively the set of metabolites and the set of reactions for organism k.

By combining all these models we cover all the activity inside each of the cells. However, next to the organisms themselves the space in between the cells ("The community space" or medium) has also to be taken into account. So in order to obtain a steady state for all metabolites the researchers of SteadyCom argue we also need to have:

$$\frac{d}{dt} m_i = u^c_i - e^c_i + \sum_{k \in K} v^k_{ex(i)} = 0 \;\; \forall i \in I^{com} \tag{5}$$

where: u^c_i is the community uptake (inflow) of metabolite i, e^j_i is the community export (outflow) of metabolite i and $v^k_{ex(i)}$ are the exchange reactions of metabolite i between the community space and the individual organisms. And I^{com} is the set of metabolites that are active in the community space. To stay away from multiple objective functions the objective of the joint FBA becomes:

maximize: $\sum_{k \in K} \alpha^k v^k_{biomass}$

where α^k is the objective coefficient for the biomass flux of organism k ($v^k_{biomass}$). This however still does not guarantee a stable steady state growth. Therefore, it is also assumed that all organisms grow at the same rate: $v^k_{biomass} = \mu_C \; \forall \; k \in K$ Where μ_C is the (time averaged) community growth rate (just as in cFBA). Equation (5) however, still has a problem with it. It implies that the exchange reactions are a direct sum of the fluxes. This indirectly assumes that all species have identical biomasses. Therefore, in a community application equation (5) is incorrect. This is the same issue we ran into in the previous section. And this is also where the earlier discussed approach stems from.

To improve the issue SteadyCom scales the optimization problem to the full population by multiplying the specific fluxv^k_j by the biomass X^k (as we showed before) so that we have: $V^k_j = X^k v^k_j$. Where V^k_j is the flux of reaction j for the full population of species k. Now by replacing v^k_j by V^k_j in all the constraints and by adding a constraint that the total biomass is nonzero, the optimization problem becomes:

$$max \;\; \mu_C$$

$$\text{subject to: } \begin{bmatrix} \sum_{j \in J_k} S^k_{ij} V^k_j = 0 \; \forall i \in I^k \\ LB^k_j X^k \leq V^k_j \leq UB^k_j X^k \; \forall j \in J^k \\ V^k_{biomass} = X^k \mu_C \\ X^k \geq 0 \end{bmatrix} \forall k \in K$$

$$u^c_i - e^c_i + \sum_{k \in K} V^k_{ex(i)} = 0 \; \forall i \in I^{com}$$

$$\sum_{k \in K} X^k = X_0$$
$$e^c_i \geq 0 \;\; \forall i \in I^{com}$$

where X_0 is the total community biomass. These new constraints force all the constraints to hold for the full population of all the species. By the third constraint the community composition remains stable. And by the second constraint a species can only generate a nonzero flux if both the flux of the cell as well as the species biomass are nonzero. By fixing μ_C this problem becomes linear and by iterating over values of μ_C and checking its feasibility $\mu_{C,max}$ can be found. This generally takes less than 10 iterations and only depends on the accuracy needed and the closeness of the initial guess to the actual value of $\mu_{C,max}$. [6]

2.4 Comparison

Now, before we continue to create our own approach to analyze the "EFMs" of communities, let's look at what we can take from these two models. Both cFBA and SteadyCom assume a community growth rate so that the community stays in steady state. This assumption, as introduced by cFBA, is also quite practical as it bypasses the need for several objective functions. This is the first thing we will take from these FBA approaches. The second thing we can take from both methods is that the community growth rate μ_C needs to be fixed in order for the problem to become linearly solvable as neither of the methods found a way to bypass this. The newer SteadyCom does have several advantages over cFBA, most noteworthy for this paper is that it does not assume given ratios of species in the community. A disadvantage of SteadyCom however is that it is also less intuitive, as the population wide (absolute) flux is used instead of the flux per organism. Both techniques have previously been shown to give very similar results [6] so the approach we choose is mostly based on practicality in this perspective. Because in this paper we rather not make the quite strong assumption of knowing the ratios between species, we will in this paper use the approach of SteadyCom as a starting point for our technique.

2.5 EFM generalization for communities

Instead of the optimization of the growth rate as in the last section, we now want to generalize the notion of Elementary Flux Modes (EFMs) to the application in a community. To do so we will use the SteadyCom approach, as it is more widely applicable. Therefore, by completely following the SteadyCom approach, but then in an "elementary modes" application we would have the following model:

$$\begin{bmatrix} \sum_{j \in J_k} S_{ij}^k V_j^k = 0 \; \forall i \in I^k \\ LB_j^k X^k \leq V_j^k \leq UB_j^k X^k \; \forall j \in J^k \\ V_{biomass}^k = X^k \mu_C \\ X^k \geq 0 \end{bmatrix} \forall k \in K$$

$$u_i^c - e_i^c + \sum_{k \in K} V_{ex(i)}^k = 0 \; \forall i \in I^{com}$$

$$\begin{gathered} \textstyle\sum_{k \in K} X^k = X_0 \\ \mathrm{e}_i^c \geq 0 \;\; \forall i \in I^{com} \end{gathered}$$

where it is noteworthy that for an elementary modes application we are no longer looking to optimize the community growth rate but we are rather look-

ing for the extreme rays of the solution space. To obtain a model that is more biology based we will remove the upper and lower bounds on the fluxes. but we do need to replace them with a constraint that keeps them from going to infinity as this is biologically impossible. To do so we assume that all reactions are enzyme mediated and therefore we may introduce a constraint that states that there is an enzyme constraint for each organism. Which leads to an upper bound on the total flux for an organism. We write this as follows: $a_i V^k_{biomass} + \sum_{j \in J_k} c_{ij} V^k_j \leq UB^k X^k \ \forall \ k \in K$, where UB^k is the upper bound on the energy an organism k can spend on its reactions (which can be obtained experimentally). And a_i and c_{ij} are the corresponding weights of the reactions. Since these last constraints are inequalities, slack variables (s^k) are added to form a buffer. Furthermore, in order for the chemostat to remain in steady state the difference between the community uptake (u^c_i) and community export (e^c_i) needs to grow with the community growth rate. Hence, we may replace $u^c_i - e^c_i$ with $\mu_C C_i$ where C_i is a constant computed by taking the difference between the concentration of metabolite i in the feed and the concentration of metabolite i in the medium. Hence, we obtain the following model:

$$\left[\begin{array}{c} \sum_{j \in J_k} S^k_{ij} V^k_j = 0 \ \forall i \in I^k \\ V^k_{biomass} = X^k \mu_C \\ a_i V^k_{biomass} + \sum_{j \in J_k} c_{ij} V^k_j + s^k = UB^k X^k \\ V^k_j \geq 0 \ \forall j \in J^k \\ X^k \geq 0 \\ s^k \geq 0 \end{array}\right] \forall k \in K$$

$$\sum_{k \in K} V^k_{ex(i)} - \mu_C C_i = 0 \ \forall i \in I^{com}$$

2.6 Functionality check via toy model

To learn to better understand the constraints of our new model, let us consider the toy model in figure 4. The model has two organisms (A and B) that both have two internal metabolites (m^A_1, m^A_2, m^B_1 and m^B_2) and with two external metabolites (m^E_1 and m^E_2) in the community space. The community is placed in a chemostat such that the uptake and export of the external metabolites only depends on the concentration of the external metabolites itself and some constant which is set by the experimenter. In that way, the steady-state community

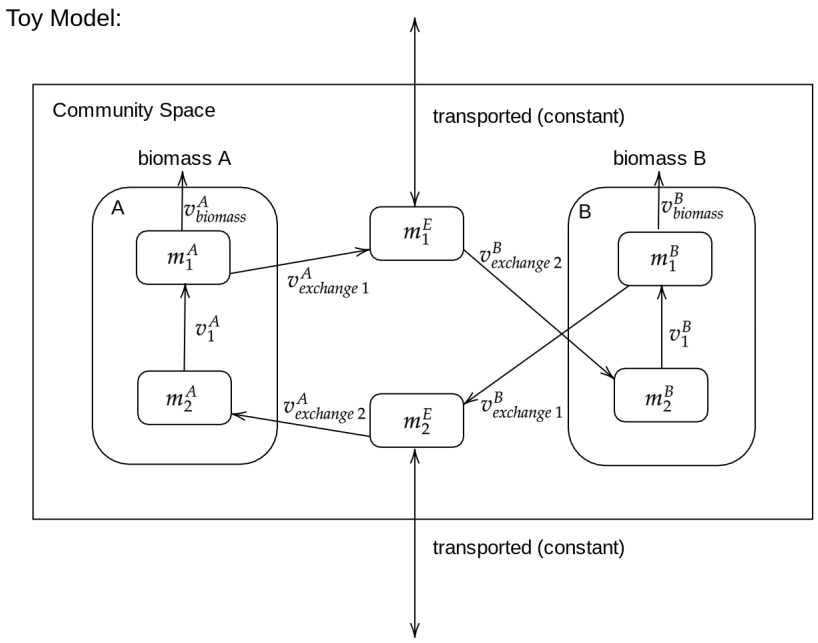

Figure 4: A toy model with 2 species in a community.

growth can be assured by not providing enough external metabolites for either of the organisms to overgrow the other.

If we now apply our model from the previous section to this situation we see that from the steady-state assumption we already have eight constraints for the metabolites: a constraint for each of the four inner metabolites, a constraint for the two biomasses and a constraint for the two external metabolites. Additionally, we have a constraint on the total reactions of each of the organisms. If we now put all of this into a matrix format we get the following (in-homogeneous) system of equations:

$$
\begin{array}{c}
 \\
m^I_{A1} \\
m^I_{A2} \\
m^I_{B1} \\
m^I_{B2} \\
biomassA \\
biomassB \\
m^E_1 \\
m^E_2 \\
species\ A \\
species\ B
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccc}
V^A_1 & V^B_1 & V^A_{mass} & V^B_{mass} & V^A_{ex1} & V^A_{ex2} & V^B_{ex1} & V^B_{ex2} & X^A & X^B & slack1 & slack2
\end{array} \\
\begin{bmatrix}
1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\mu_c & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\mu_c & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & -UB_A & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & -UB_B & 0 & 1
\end{bmatrix}
\end{array}
\begin{bmatrix}
V^A_1 \\ V^B_1 \\ V^A_{mass} \\ V^B_{mass} \\ V^A_{ex1} \\ V^A_{ex2} \\ V^B_{ex1} \\ V^B_{ex2} \\ X^A \\ X^B \\ slack1 \\ slack2
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ C^{tot}_{trans\ 1} \\ C^{tot}_{trans\ 2} \\ 0 \\ 0
\end{bmatrix}
$$

Let us define this system of equations as:

$$Pq = r \tag{6}$$

where P is the matrix and q and r are the vectors.

Now we would like to analyze the "EFMs" of this model, to see what an "Elementary Interaction Mode" of this community might look like. In order to do so we want to look for support minimal solutions with one degree of freedom (so that it forms an extreme ray). As this is quite a tedious task we use a MATLAB package "efmtool" [11]. The calculation of elementary modes with this tool however is only possible for homogeneous systems of equations while our model is in-homogeneous. Hence we will now apply the transformation as suggested by [12] and further worked out by [13] to homogenize our model.

let $\lambda \geq 0 \in \mathbb{R}$, Now let us set:

$$N = \begin{pmatrix} P & r \end{pmatrix} \ and \ v = \begin{pmatrix} q' \\ \lambda \end{pmatrix} \tag{7}$$

So that we have:

$$Nv = \begin{pmatrix} P & r \end{pmatrix} \begin{pmatrix} q' \\ \lambda \end{pmatrix} = \ 0 \tag{8}$$

If we now obtain an EIM solution q' for v, we may rescale it such that $q = \frac{q'}{\lambda}$ if $\lambda > 0$ and $q = q'$ if $\lambda = 0$. In this way we make a distinction between homogeneous solutions (where $\lambda = 0$) and in-homogeneous solutions (where $\lambda > 0$) while all the sizes are scaled back to the actual fluxes. Applied to our

model this looks like:

$$
\begin{array}{c}
 \\
m^I_{A1} \\
m^I_{A2} \\
m^I_{B1} \\
m^I_{B2} \\
biomassA \\
biomassB \\
m^E_1 \\
m^E_2 \\
species\ A \\
species\ B
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccccc}
V_1^{A\prime} & V_1^{B\prime} & V_{mass}^{A\ \prime} & V_{mass}^{B\ \prime} & V_{ex1}^{A\ \prime} & V_{ex2}^{A\ \prime} & V_{ex1}^{B\ \prime} & V_{ex2}^{B\ \prime} & X^{A\prime} & X^{B\prime} & slack\ 1' & slack\ 2' & \lambda
\end{array} \\
\left[\begin{array}{cccccccccccccc}
1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\mu_c & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\mu_c & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & C^{tot}_{trans\ 1} \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & C^{tot}_{trans\ 2} \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & -UB_A & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & -UB_B & 0 & 1 & 0
\end{array}\right]
\end{array}
\begin{array}{c}
\left[\begin{array}{c}
V_1^{A\prime} \\ V_1^{B\prime} \\ V_{mass}^{A\ \prime} \\ V_{mass}^{B\ \prime} \\ V_{ex1}^{A\ \prime} \\ V_{ex2}^{A\ \prime} \\ V_{ex1}^{B\ \prime} \\ V_{ex2}^{B\ \prime} \\ X^{A\prime} \\ X^{B\prime} \\ slack\ 1' \\ slack\ 2' \\ \lambda
\end{array}\right]
\end{array}
=
\left[\begin{array}{c}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{array}\right]
$$

where $\lambda \geq 0$. Now, let:

$$q' = (V_1^{A\prime}, V_1^{B\prime}, V_{mass}^{A\ \prime}, V_{mass}^{B\ \prime}, V_{ex1}^{A\ \prime}, V_{ex2}^{A\ \prime}, V_{ex1}^{B\ \prime}, V_{ex2}^{B\ \prime}, X^{A\prime}, X^{B\prime}, s^{A\prime}, s^{B\prime})^T$$

Then for each EIM q:

$$q = (q', \lambda)^T \tag{9}$$

where q corresponds to a bounded EIM of the system of equations. To this model we can now apply the MATLAB efmtool [11] that can calculate all the EIM for the system, as can be seen in the figure below. For the code used, see Appendix A.

	efv1	efv2	efv3	efv4
VA1	0.16	1.3333	0.16	0.3008
VB1	0.16	1.3333	0.3008	0.16
VAmass	0.16	0.16	0.0192	0.3008
VBmass	0.16	0.16	0.3008	0.0192
VAex1	0	1.1733	0.1408	0
VAex2	0.16	1.3333	0.16	0.3008
VBex1	0	1.1733	0	0.1408
VBex2	0.16	1.3333	0.3008	0.16
XA	0.2	0.2	0.024	0.376
XB	0.2	0.2	0.376	0.024
slack 1	3.52	0	0	6.6176
slack 2	3.52	0	6.6176	0
lambda	1	1	1	1

Figure 5: The four elementary flux modes (vectors) of the toy model in the columns. On the row the size of the fluxes for each of the "elementary flux modes".

As can be seen from the figure there are four elementary flux modes for the toy model. In efv1 V^A_{ex1} and V^B_{ex1} are set to zero, therefore, this mode signifies the case in which the species do not cooperate and both work separately on their own biomass production. In efv2 all fluxes are positive, so this mode signifies the case in which both work on their own biomass production but where additional energy is put into producing metabolites for the other species as well (in exact equilibrium). In efv3 only V^B_{ex1} is set to zero, and thus efv3 corresponds to the mode in which the species together maximize the production of biomass of species B. In efv4 only V^A_{ex1} is set to zero and this mode thus corresponds to the mode where the species work together to maximize production of biomass A.

As these modes correspond to the modes that a biologist would expect for such a model, we may now start to generalize our model towards modelling any community of 2 species.

3 Model analysis:

3.1 Generalization for communities of 2 species

Now that we have a feeling for what such elementary modes look like, let us rewrite the problem as sketched in the toy model, into a generally applicable one of the form $NV = 0$, given a community with two species living in it, let us set the following: if species i has p internal reactions and k exchange reactions:

$$V_{in}^i = \begin{bmatrix} V_1^i \\ V_2^i \\ \vdots \\ V_p^i \end{bmatrix} \qquad V_{ex}^i = \begin{bmatrix} V_{ex1}^i \\ V_{ex2}^i \\ \vdots \\ V_{exk}^i \end{bmatrix}$$

Now that those are defined let us define V given two species (A and B) as follows:

$$V = \begin{bmatrix} V_{in}^A \\ V_{ex}^A \\ V_m^A \\ V_{in}^B \\ V_{ex}^B \\ V_m^B \\ X^A \\ X^B \\ s^A \\ s^B \\ \lambda \end{bmatrix}$$

where for species i: V_m^i is the total biomass flux of species i, X^i is the total biomass of species i, s^i is the slack variable for species i and λ is the additional variable that was added to homogenize the problem.

Now that V is defined, we can define the matrix N in the system of equations $NV = 0$ with $N = \begin{bmatrix} S \\ C \end{bmatrix}$ where:

$$
S = \begin{array}{c} \\ (m_1^A \ldots m_{k_1}^A)^T \\ (m_1^B \ldots m_{k_2}^B)^T \\ (m_1^E \ldots m_{k_3}^E)^T \\ biomass\ A \\ biomass\ B \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} V_1^A \ldots V_{j_1}^A & V_{ex\ 1}^A \ldots V_{j_2}^A & V_m^A & V_1^B \ldots V_{j_3}^B & V_{ex\ 1}^B \ldots V_{ex\ j_4}^B & V_m^B & X^A & X^B & s^A & s^B & \lambda \end{array} \\
\begin{bmatrix}
S_{in}^A & S_{ex\ 1}^A & S_m^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & S_{in}^B & S_{ex\ 1}^B & S_m^B & 0 & 0 & 0 & 0 & 0 \\
0 & S_{ex\ 2}^A & 0 & 0 & S_{ex\ 2}^B & 0 & 0 & 0 & 0 & 0 & S_{trans}^{com} \\
0 & 0 & 1 & 0 & 0 & 0 & -\mu & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -\mu & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

.

Here the first three columns represent all the reactions of species A (j_1 internal, j_2 exchange and 1 biomass production reactions), the second three represent all the reactions corresponding to species B (j_3 internal, j_4 exchange and 1 biomass production reactions). The next two columns represent the produced biomass of the two species, and the last three columns represent the slack variables for the two species and the λ variable which is used to homogenize the model. The first two set of rows of S represent all the steady-state constraints for the internal metabolites of the two species (k_1 for species A and k_2 for species B). The third set of rows represents all the steady-state constraints for the k_3 metabolites in the medium. The last two rows of S represent the constraint that both the communities grow with the community growth rate. S_{in}^A represents the stoichiometric submatrix of all the internal reactions of species A, and similarly for the other S sub-matrices. Note here that S_m^A, S_m^B and S_{trans}^{com} are vectors instead of matrices. So that the dimensions of S are:

$$
\begin{aligned}
\#\ of\ columns\ of\ S\ &= \sum_{i=1}^{4} j_i\ +\ 7 \\
\#\ of\ rows\ of\ S\ &= \sum_{i=1}^{3} k_i\ +\ 2.
\end{aligned}
$$

Furthermore:

$$
C = \begin{array}{c} \\ species\ A \\ species\ B \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} V_1^A \ldots V_{j_1}^A & V_{ex\ 1}^A \ldots V_{j_2}^A & V_m^A & V_1^B \ldots V_{j_3}^B & V_{ex\ 1}^B \ldots V_{ex\ j_4}^B & V_m^B & X^A & X^B & s^A & s^B & \lambda \end{array} \\
\begin{bmatrix}
c_1 & c_2 & c_3 & 0 & 0 & 0 & -UB_A & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & c_4 & c_5 & c_6 & 0 & -UB_B & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

where, the columns coincide with S and the two rows represent the constraints on the two species' energy usage. The c_i's represent the row vectors

with the weights (in energy) of each of the reactions. These two constraints put the upper bound on the energy usage of the organisms. Here we note that the sizes of C are:

$$\#\ of\ columns\ of\ C\ = \sum_{i=1}^{4} j_i\ +\ 7$$
$$\#\ of\ rows\ of\ C\ =\ 2$$

If we now put S and C together to form the matrix N we obtain the following system of equations:

$$NV\ =\ \begin{array}{c} \\ \left(m_1^A\ \ldots m_{k_1}^A\right)^T \\ \left(m_1^B\ \ldots m_{k_2}^B\right)^T \\ \left(m_1^E\ \ldots m_{k_3}^E\right)^T \\ biomass\ A \\ biomass\ B \\ species\ A \\ species\ B \end{array} \begin{array}{c} \begin{array}{ccccccccccc} V_1^A\ \cdots V_{j_1}^A & V_{ex\ 1}^A\ \cdots V_{j_2}^A & V_m^A & V_1^B\ \cdots V_{j_3}^B & V_{ex\ 1}^B\ \cdots V_{ex\ j_4}^B & V_m^B & X^A & X^B & s_1 & s_2 & \lambda \end{array} \\ \left[\begin{array}{ccccccccccc} S_{in}^A & S_{ex\ 1}^A & S_m^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{in}^B & S_{ex\ 1}^B & S_m^B & 0 & 0 & 0 & 0 & 0 \\ 0 & S_{ex\ 2}^A & 0 & 0 & S_{ex\ 2}^B & 0 & 0 & 0 & 0 & 0 & S_{trans}^{com} \\ 0 & 0 & 1 & 0 & 0 & 0 & -\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\mu & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & 0 & 0 & 0 & -UB_A & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 & c_6 & 0 & -UB_B & 0 & 1 & 0 \end{array}\right] \end{array} \begin{bmatrix} V_{in}^A \\ V_{ex}^A \\ V_m^A \\ V_{in}^B \\ V_{ex}^B \\ V_m^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix} = 0$$

Here we note that the sizes of N are:

$$\#\ of\ columns\ of\ N\ = \sum_{i=1}^{4} j_i\ +\ 7$$
$$\#\ of\ rows\ of\ N\ =\ \sum_{i=1}^{3} k_i\ +\ 4$$

We also note that the matrix N only depends on the community growth rate μ (both direct as well as through S_{trans}^{com}) and on the community metabolite concentrations through S_{trans}^{com}.

3.2 Moving towards the space of external metabolites

Since analysis on the variables presented in the previous paragraph will be rather chaotic, we will in this section apply a variable transformation on the model. This variable transformation only applies for the flux variables and therefore we will split the matrix up in the flux part of N (left) and the non-flux part of N (right). The non-flux variables will be multiplied by an identity matrix and in that way remain unchanged. To properly explain the implications of this variable transformation, let us, for now, also reduce our model back to a community of just one species. So that we have:

$$NV\ =\ \begin{array}{c} \\ \left(m_1\ \ldots m_{k_4}\right)^T \\ \left(m_1^E\ \ldots m_{k_5}^E\right)^T \\ biomass \\ species \\ =0 \end{array} \begin{array}{c} \begin{array}{cccccc} V_1\ \cdots V_{j_5} & V_{ex\ 1}\ \cdots V_{j_6} & V_m & X & s_1 & \lambda \end{array} \\ \left[\begin{array}{cccccc} S_{in} & S_{ex\ 1} & S_m & 0 & 0 & 0 \\ 0 & S_{ex\ 2} & 0 & 0 & 0 & S_{trans}^{com} \\ 0 & 0 & 1 & -\mu & 0 & \lambda \\ c_7 & c_8 & c_9 & 0 & -UB & 0 \end{array}\right] \end{array} \begin{bmatrix} V_{in} \\ V_{ex} \\ V_m \\ X \\ s_1 \\ \lambda \end{bmatrix} \tag{10}$$

where there are k_4 internal metabolites, k_5 external metabolites, j_5 internal fluxes, j_6 exchange

fluxes, X is the biomass of the organism, s_1 is the slack variable and λ is the homogenezation variable.
Let us first look just at the metabolite constraints on this community of one species, so that we have:

$$N_{1L}V = \begin{matrix}\begin{pmatrix} m_1 & \dots m_{k_4}\end{pmatrix}^T \\ \begin{pmatrix} m_1^E & \dots m_{k_5}^E\end{pmatrix}^T\end{matrix} \begin{matrix}\begin{bmatrix} V_1 & \cdots V_{j_5} \end{bmatrix} & \begin{bmatrix} V_{ex\ 1} & \cdots V_{j_6} \end{bmatrix} & \begin{bmatrix} V_m \end{bmatrix} \\ \begin{bmatrix} S_{in} \end{bmatrix} & \begin{bmatrix} S_{ex\ 1} \end{bmatrix} & \begin{bmatrix} S_m \end{bmatrix} \\ \begin{bmatrix} 0 \end{bmatrix} & \begin{bmatrix} S_{ex\ 2} \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix}\end{matrix} \begin{bmatrix} V_{in} \\ V_{ex} \\ V_m \end{bmatrix} \tag{11}$$

where N_{1L} stands for the first left part of the constraint matrix N. We will now rewrite the vector V into the product of the EFM matrix E times a scalar vector β as follows:

$$\begin{bmatrix} V_{in} \\ V_{ex} \\ V_m \end{bmatrix} = \begin{bmatrix} EFM_1 & \cdots EFM_d \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{12}$$

where the d EFMs of the organisms are the column vectors of matrix E and where for all elements of vector β, $\beta_i \geq 0$. If we now plug equation 12 into equation 11, we obtain the following equation:

$$N_{1L}V = N_{1L}E\beta = \begin{bmatrix} S_{in} & S_{ex\ 1} & S_m \\ 0 & S_{ex\ 2} & 0 \end{bmatrix} \begin{bmatrix} EFM_1 & \cdots EFM_d \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{13}$$

we note that in the matrix N, the first set of rows represent the constraints on the internal metabolites and the second set of rows represent the constraints on the external metabolites. We will now multiply matrix N with matrix E to obtain a new matrix which we will call matrix G. To understand what matrix G will look like, let us first look at the product of the first set of rows of matrix N and matrix E, and then afterwards we will look at the second set of rows.

As we assumed internal steady state for all organisms in a community, also this organism has an assumed internal steady-state. This implies that no internal metabolite will be accumulated nor depleted. As the EFMs of the organism are the extreme rays of the solution space of the internal metabolism, and therefore part of the solution space, none of them will have an accumulation or depletion of any of the metabolites. Therefore, the product of the stoichiometry of any internal metabolite times any EFM is equal to zero. And thus the top rows of matrix G will all consist of only zeros.

If we now look at the second set of rows of N, we note that these represent the constraints of the external metabolites. So if we take one of these rows, we can read in it, from the perspective of the medium, in which reactions that specific metabolite is used and in which it is produced. As this only considers the fluxes that flow from the medium into the cell and the fluxes that flow out of the cell into the medium, typically the product with E of these rows is not equal to zero. Instead, the columns of these products will contain only parts of the EFMs. Specifically, the only entries of these column vectors are the metabolites that enter or leave the organism per EFM. We call these column vectors Elementary Conversion Modes.

Definition 4: Elementary Conversion Mode (ECM)
an ECM of an organism is the metabolic conversion of substrates into products in the medium by the organism via the corresponding EFM (i.e. the conversion made by the EFM, as it is seen from the medium).

Although often several EFMs correspond to the same ECM and the set of ECMs therefore is often smaller than the set of EFMs of an organism, there could be at most "the number of

EFMs" of ECMs.

The product of the matrices in equation 13 therefore becomes:

$$N_{1L}E\beta = \begin{bmatrix} S_{in} & S_{ex\ 1} & S_m \\ 0 & S_{ex\ 2} & 0 \end{bmatrix} [EFMs] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} = G_{1L}\beta = \begin{bmatrix} 0 \\ ECMs \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{14}$$

where EFMs and ECMs are the (sub)matrices with in its columns the Elementary Flux / Conversion Modes of the organism.

Let us now continue with the second part of the variable transformation. In the beginning of this section we only considered the constraints on the metabolites. Let us now consider the two additional constraints: the one on the growth rate of the organism and the one on the enzyme-usage upper bound of the organism. In matrix format this looks as follows:

$$N_{2L}V = \begin{matrix} \\ biomass \\ species \end{matrix} \begin{matrix} V_1 \cdots V_{j_5} & V_{ex\ 1} \cdots V_{j_6} & V_m \end{matrix} \begin{bmatrix} 0 & 0 & 1 \\ c_7 & c_8 & c_9 \end{bmatrix} \begin{bmatrix} V_{in} \\ V_{ex} \\ V_m \end{bmatrix} \tag{15}$$

where N_{2L} is the lower left part of matrix N, of the one species model. Here we again apply the following variable transformation:

$$\begin{bmatrix} V_{in} \\ V_{ex} \\ V_m \end{bmatrix} = \begin{bmatrix} EFM_1 & \cdots EFM_d \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{16}$$

So that we obtain the following product of matrices:

$$N_{2L}V = N_{2L}E\beta = \begin{matrix} \\ biomass \\ species \end{matrix} \begin{matrix} V_1 \cdots V_{j_5} & V_{ex\ 1} \cdots V_{j_6} & V_m \end{matrix} \begin{bmatrix} 0 & 0 & 1 \\ c_7 & c_8 & c_9 \end{bmatrix} \begin{bmatrix} EFM_1 & \cdots EFM_d \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{17}$$

.

Here the product of N_{2L} and E will form the lower left part of the new matrix G. The product of the first row of N_{2L} with matrix E now becomes a row vector of non-negative elements per EFM which indicate whether and how much biomass that specific EFM produces. We will call this row vector the biomass indicator of the organism denoted by e.

The second row of N_{2L} contains the weights on the enzyme usage of every reaction and thus, multiplied with the EFMs these become the weights of each of the EFMs (and so also of each of the ECMs) we will denote this row vector by w. Note that here, it does make sense to keep double ECMs coming from different EFMs in the model, because although the ECMs might coincide with each other (in reactions), they come from different EFMs and so the enzyme usage might differ. Thus, the product of these matrices will become:

$$\begin{aligned} N_{2L}E\beta = \begin{matrix} \\ biomass \\ species \end{matrix} \begin{matrix} V_1 \cdots V_{j_5} & V_{ex\ 1} \cdots V_{j_6} & V_m \end{matrix} \begin{bmatrix} 0 & 0 & 1 \\ c_7 & c_8 & c_9 \end{bmatrix} [EFMs] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} = \\ = G_{2L}\beta = \begin{bmatrix} e \\ w \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \end{aligned} \tag{18}$$

.

If we now put this together with equation 13 we obtain the complete left side of matrix G for

this single species community:

$$G_L\beta = \begin{bmatrix} 0 \\ ECMs \\ e \\ w \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \end{bmatrix} \tag{19}$$

.
Together with the unchanged right hand side this gives us:

$$G\beta = \begin{matrix} \\ (m_1 \ ... m_{k_4})^T \\ (m_1^E \ ... m_{k_3}^E)^T \\ biomass \\ species \end{matrix} \begin{matrix} \begin{matrix} \beta_1 \ ... \ \beta_{l_0} & X & s_1 & \lambda \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ ECMs & 0 & 0 & S_{trans}^{com} \\ e & -\mu & 0 & 0 \\ w & -UB & 1 & 0 \end{bmatrix} \end{matrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{l_0} \\ X \\ s_1 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

where the upper set of rows of course are trivial constraints. In the next subsection we will apply this variable transformation to the 2 species model of the previous section.

3.2.1 Variable Transformation on 2 Species Model

Let us remind ourselves of the model of a community with 2 species of microorganisms we had before:

$$NV = \begin{matrix} \\ (m_1^A \ ... m_{k_1}^A)^T \\ (m_1^B \ ... m_{k_2}^B)^T \\ (m_1^E \ ... m_{k_3}^E)^T \\ biomass \ A \\ biomass \ B \\ species \ A \\ species \ B \end{matrix} \begin{matrix} \begin{matrix} V_1^A \ ... V_{j_1}^A & V_{ex\ 1}^A \ ... V_{j_2}^A & V_m^A & V_1^B \ ... V_{j_3}^B & V_{ex\ 1}^B \ ... V_{ex\ j_4}^B & V_m^B & X^A & X^B & s_1 & s_2 & \lambda \end{matrix} \\ \begin{bmatrix} S_{in}^A & S_{ex\ 1}^A & S_m^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{in}^B & S_{ex\ 1}^B & S_m^B & 0 & 0 & 0 & 0 & 0 \\ 0 & S_{ex\ 2}^A & 0 & 0 & S_{ex\ 2}^B & 0 & 0 & 0 & 0 & 0 & S_{trans}^{com} \\ 0 & 0 & 1 & 0 & 0 & 0 & -\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\mu & 0 & 0 & 0 \\ c_1 & c_2 & c_3 & 0 & 0 & 0 & -UB_A & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 & c_6 & 0 & -UB_B & 0 & 1 & 0 \end{bmatrix} \end{matrix} \begin{bmatrix} V_{in}^A \\ V_{ex}^A \\ V_m^A \\ V_{in}^B \\ V_{ex}^B \\ V_m^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix} = 0$$

.

Here we will apply the variable transformation, as written out in the beginning of this section, on the vector V. As there are now 2 species in the community we will apply this transformation to both species. Also, instead of the 3 non-flux variables we had for a one species model we now have 5 non-flux variables (an additional biomass and an additional slack variable). Therefore we will apply the following transformation:

$$\begin{bmatrix} V_{in}^A \\ V_{ex}^A \\ V_m^A \\ V_{in}^B \\ V_{ex}^B \\ V_m^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix} = \begin{bmatrix} [EFMs^A] & & \\ & [EFMs^B] & \\ & & [I_{5\times 5}] \end{bmatrix} \begin{bmatrix} \beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix}$$

where the 5 non-flux variables are multiplied with the identity matrix so that they remain unchanged. Furthermore, l_1 are the number of EFMs of organism A and l_2 are the number of EFMs of organism B. If we now plug this transformation into the 2 species model we had before, we obtain:

$$N \cdot E \cdot \beta = \begin{bmatrix} S_{in}^A & S_{ex\ 1}^A & S_m^A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{in}^B & S_{ex\ 1}^B & S_m^B & 0 & 0 & 0 & 0 & 0 \\ 0 & S_{ex\ 2}^A & 0 & 0 & S_{ex\ 2}^B & 0 & 0 & 0 & 0 & 0 & S_{trans}^{com} \\ 0 & 0 & 1 & 0 & 0 & 0 & -\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\mu & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -UB_A & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & -UB_B & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} [EFMs^A] & & [EFMs^B] & & [I_{5\times 5}] \end{bmatrix} \begin{bmatrix} \beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix}$$

.

Here we should note that this equation is just twice equation 11 on the first 6 sets of columns (column 3 and 6 are just regular columns) and a multiplication with the identity matrix for the last 5 columns. Therefore, after multiplication we obtain:

$$N \cdot E \cdot \beta = G \cdot \beta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ ECMs^A & ECMs^B & 0 & 0 & 0 & 0 & S_{trans}^{com} \\ e_A & 0 & -\mu & 0 & 0 & 0 & 0 \\ 0 & e_B & 0 & -\mu & 0 & 0 & 0 \\ w_A & 0 & -UB_A & 0 & 1 & 0 & 0 \\ 0 & w_B & 0 & -UB_B & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix}$$

where the first two sets of columns are just twice the column as displayed in equation 18 and the other 5 columns are unchanged with respect to matrix N. As the upper two sets of rows have now become trivial constraints we may reduce our model to the following system of equations:

$$G \cdot \beta = \begin{array}{c} \left(m_1^E \ \ldots m_{k_3}^E\right)^T \\ biomass\ A \\ biomass\ B \\ species\ A \\ species\ B \end{array} \begin{array}{c} \begin{matrix} \beta_1^A \ \ldots\ \beta_{l_1}^A & \beta_1^B \ \ldots\ \beta_{l_2}^B & X^A & X^B & s_1^A & s_1^B & \lambda \end{matrix} \\ \begin{bmatrix} ECMs^A & ECMs^B & 0 & 0 & 0 & 0 & S_{trans}^{com} \\ e_A & 0 & -\mu & 0 & 0 & 0 & 0 \\ 0 & e_B & 0 & -\mu & 0 & 0 & 0 \\ w_A & 0 & -UB_A & 0 & 1 & 0 & 0 \\ 0 & w_B & 0 & -UB_B & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{bmatrix} \beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where k_3 is the number of metabolites in the medium, l_1 is the number of EFMs for organism

A and l_2 is the number of EFMs for organism B. Here we additionally note that:

$$\#\ of\ columns\ of\ G = \sum_{j=1}^{2}(l_j + 2) + 1 = l_1 + l_2 + 5$$
$$\#\ of\ rows\ of\ G = k_3\ + 4.$$

We will now look at the interdependence between the constraints in this model. First, we note that aside from the $k_3 + 4$ constraints in the matrix format, all variables in the vector β have an additional non-negativity constraint. If we now further analyze the model we see that:

- Setting the biomass of organism A (X^A) to zero forces all the scalars of the ECMs of organism A (β^A's) and the slack variable of organism A (s_1) to zero because the enzyme-usage upper bound then becomes zero. The same reasoning holds for organism B.
- Setting either one of the slack variables to zero implies that the upper bound inequality of the corresponding species becomes an equality constraint, which forces all its ECMs to perfectly balance with its biomass. Which makes many combinations of ECMs infeasible.
- Setting λ to zero excludes all inhomogeneous solutions. Forcing the ECMs of both organisms to perfectly balance out each metabolite in the community.

Therefore, it should be clear that the constraints of this model are, without further assumptions, generally not independent.

3.3 Generalizations for communities of n species

In this section we will prove a few theorems on communities of n species. To do so, let us start by generalizing the matrix G towards a community of n species. As this already was a rather chaotic task with two species, we will here for n species only write out the changes it implies for the dimensions of G. As we saw before, every added organism adds a biomass variable and a slack variable to the system of equations and of course its own ECM scalars, it also ads two constraints (the growth rate constraint and the enzyme-usage constraint). Therefore, the new dimensions of G for an community with n species become:

$$\#\ of\ columns\ of\ G = \sum_{j=1}^{n}(l_j + 2) + 1 = l_{tot} + 2n + 1$$
$$\#\ of\ rows\ of\ G = k_3\ + 2n$$

where: l_j is the number of ECMs (equal to the number of EFMs) of species i and $l_{tot} = \sum_{j=1}^{n} l_j$. k_3 is the number of active metabolites in the community space.

We also note that in a community, which grows at a certain community growth rate μ and which has a given vector C_0 (which depend on the external metabolite concentrations), the solution space for a non-negative flux vector v can again be described geometrically by a pointed cone (a "flux cone") From now on this solution space will be referred to as FC_{com}.

Definition 5: Elementary Interaction Modes (EIMs)
We define the set of Elementary Interaction Modes as the extreme rays of the community flux cone FC_{com} which depends on the community growth rate μ and the external metabolite concentrations through C_i.

Note that the EIMs can be either an infinite extreme ray, in the case where $\lambda = 0$ or a finite extreme ray for the case in which $\lambda = 1$ as the set of EIMs span the entire community flux cone FC_{com}. Any steady-state metabolic behaviour of a community of micro-organism can be described by a convex combination of its EIMs.

3.3.1 Upper Bound on the number of ECMs used in an EIM

Theorem 1: Every EIM of a community in which n species are present, which has a positive biomass (X) for all species and a positive particularity variable (λ) uses at most k_3+s_0 ECMs. Where k_3 is the number of active metabolites in the community space and s_0 is the number of slack variables equal to zero.

Proof: If we have an EIM β, by definition such an extreme ray has one degree of freedom. Hence, we may rewrite its constraint matrix into a matrix with one column more than the number of independent rows. We may assume the constraint matrix to be full row rank (since if it were not full row rank it could be reduced to a full row rank matrix without changing the solution space of vectors satisfying its constraints). Hence, if there are p independent rows in the constraint matrix of β, it must now have exactly $p+1$ nonzero columns. In order for the constraint matrix to fulfill this definition, all variables in vector β will be set to zero except for a subset of $p+1$ variables. This reduces the number of columns of the constraint matrix to the required amount. Note: if such a subset of variables does not exist, the corresponding EIM also does not exist. Therefore, β will have at most $k_3 + 2n + 1$ nonzero variables as there are at most $k_3 + 2n$ independent rows ($p \leq k_3 + 2n$). We note that all variables also have non-negativity constraints so that all nonzero variables have to be positive. Since we assumed all the X variables and λ to be strictly positive (together $n + 1$ variables) we have at most $k_3 + n$ additional variables that can be strictly positive. Let us now set s_0 to the number of slack variables that are equal to zero. Since there are n slack variables we have $s_0 \in \{0, 1, 2, \ldots, n\}$. Since we have a total of at most $k_3 + n$ additional positive variables we have at most $k_3 + n - (n - s_0) = k_3 + s_0$ positive variables for the ECMs. Hence at most $k_3 + s_0$ ECMs (and thus at most $k_3 + s_0$ EFMs) will be used in such an EIM.

3.3.2 Community uses at maximal μ at most n EIMs

We note that the model is dependent only on the variables μ and S^{com}_{trans}. Where S^{com}_{trans} in itself is dependent on μ and C, where C is the vector with constant C_i for every metabolite i (the difference between the concentration of metabolite i in the feed, and the concentration of metabolite i in the medium). As we know, Theorem 1 holds for any constant community growth rate μ. Therefore, we know it also holds for the maximal μ. This is interesting as it is likely that a community will grow at a maximal growth rate as the organisms are selected to do so by evolution. In this subsection we will analyze the complexity of the model when it operates at a maximal community growth rate μ. Let us start doing so by making our model clearer by rewriting $S^{com}_{trans} = \mu C$ where C is the vector with constants C_i for every metabolite i. This rewriting yields the following model:

$$
\begin{array}{c}
\\
\left(m_1^E \ \ldots m_{k_3}^E\right)^T \\
\textit{biomass A} \\
\textit{biomass B} \\
\textit{species A} \\
\textit{species B}
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
\beta_1^A & \ldots & \beta_{l_1}^A & \beta_1^B & \ldots & \beta_{l_2}^B & X^A & X^B & s_1^A \quad s_1^B \quad \lambda
\end{array} \\
\left[\begin{array}{ccccccccc}
ECMs^A & & & ECMs^B & & & 0 & 0 & 0 \quad 0 \quad \mu C \\
e_A & & & 0 & & & -\mu & 0 & 0 \quad 0 \quad 0 \\
0 & & & e_B & & & 0 & -\mu & 0 \quad 0 \quad 0 \\
w_A & & & 0 & & & -UB_A & 0 & 1 \quad 0 \quad 0 \\
0 & & & w_B & & & 0 & -UB_B & 0 \quad 1 \quad 0
\end{array}\right]
\end{array}
\begin{bmatrix}
\beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B \\ \lambda
\end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

note here that the upper bounds are intrinsic to the organisms and therefore generally are fixed. If we now also de-homogenize the model back to its original form, we obtain:

$$
\begin{array}{c}
\\
\left(m_1^E \ \ldots m_{k_3}^E\right)^T \\
\textit{biomass A} \\
\textit{biomass B} \\
\textit{species A} \\
\textit{species B}
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
\beta_1^A & \ldots & \beta_{l_1}^A & \beta_1^B & \ldots & \beta_{l_2}^B & X^A & X^B & s_1^A \quad s_1^B
\end{array} \\
\left[\begin{array}{ccccccccc}
ECMs^A & & & ECMs^B & & & 0 & 0 & 0 \quad 0 \\
e_A & & & 0 & & & -\mu & 0 & 0 \quad 0 \\
0 & & & e_B & & & 0 & -\mu & 0 \quad 0 \\
w_A & & & 0 & & & -UB_A & 0 & 1 \quad 0 \\
0 & & & w_B & & & 0 & -UB_B & 0 \quad 1
\end{array}\right]
\end{array}
\begin{bmatrix}
\beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B
\end{bmatrix}
= \mu \begin{bmatrix} C \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

which from now on we will refer to as the system of equations $G(\mu)\beta = \mu b(C)$, where $\beta \geq \mathbf{0}$.

Let us also, in order to be able to better analyze the implications for the model, write out the system of equations $G(\mu)\beta = \mu b(C)$, with for clarity $n = 2$, we obtain:

$$\forall m_i : \quad \sum_{j=1}^{l_{tot}} g_{ij}\beta_j = \mu C_i$$

$$biomass\ A : \quad \sum_{j=1}^{l_1} \mathbb{1}_j \beta_j^A \ - \mu X^A \ = 0$$

$$biomass\ B : \quad \sum_{j=1}^{l_2} \mathbb{1}_j \beta_j^B \ - \mu X^B \ = 0$$

$$species\ A : \quad \sum_{j=1}^{l_1} w_j^A \beta_j^A \ + s_1 \ - UB^A X^A = 0$$

$$species\ B : \quad \sum_{j=1}^{l_2} w_j^B \beta_j^B \ + s_2 \ - UB^B X^B = 0$$

$$non-negativity : \quad \beta_j^i \geq 0, \quad X^i \geq 0, \quad s_i \geq 0$$

where m_i are the metabolites in the community space, g_{ij} are the values of $G(\mu)$ in the left upper block corresponding to the ECM coefficients, $\mathbb{1}_j$ are the indicator variables which indicate whether the corresponding ECM produces biomass. w_j are the enzyme-usage weights of the ECMs. We note from the metabolite constraints that if we now set $\mu = (1+\alpha)\mu_0$, we obtain $\beta_j = (1+\alpha)\beta_{j_0}$ for all β_j. If we now increase μ we see that all β_j increase linearly with it. Therefore, to maintain all equations (specifically the combination of both the metabolite equations and the the mass equations) we obtain that both X^A and X^B are independent of μ. Hence with increasing μ and thus increasing β's, the last two equations indicate that since the UB^A and UB^B are also independent of μ these equations can only be maintained in case both s^A and s^B decrease as well. Therefore, we obtain: $s^A = s_0^A - (1+\alpha)\sum_{j=1}^{l_1} w_j^A \beta_{j_0}^A$ and $s^B = s_0^B - (1+\alpha)\sum_{j=1}^{l_2} w_j^B \beta_{j_0}^B$. Since we have $s^A \geq 0$ and $s^B \geq 0$, μ cannot be increased infinitely. Hence, If we start with a set of EIMs for a system of equations given a low μ, by increasing μ, the lower bounds of the slack variables will be violated at some point. Hence for both the slack variables there will be some EIM that is the last EIM to violate the lower bound (and since the point at which this happens is dependent on the specific weights of the specific ECMs it is practically impossible that more than one EIMs remain upto the same μ). The EIS that exactly meets this lower bound will have an optimal $\mu = \mu_{max_i}$. Hence, the overall maximal μ will be either μ_{max_A}, μ_{max_B} or a convex combination of two (possibly other) growth rates Since in the application of a chemostat we can indeed aim the evolutionary power towards the maximal growth rate, we expect that after a sufficient amount of time this optimal growth rate will be obtained and that therefore at most 2 (number of species) EIMs will be used in this setting. This in turn, by theorem 1, gives us an upper bound on the number of ECMs and thus on the number of EFMs used by the community. Let us now generalize this upper bound in the following theorem.

Theorem 2: In a community with n species that grows at the optimal community growth rate, the community will use a convex combination of at most n EIMs.

Proof:
Let us take a community with n species that grows at the optimal community growth rate. Then for each species there will be an upper bound constraint that binds the reactions of the corresponding organism, hence for each of these organisms there will be an EIM that is the last to remain feasible while increasing the community growth rate. Therefore either one of these n EIMs or a convex combination of a subset of these n EIMs will attain the optimal community growth rate. Hence a convex combination of at most n EIMs will be used in the community.

As mentioned before, by combining Theorem 1 and Theorem 2, we can give an upper bound on the number of EFMs that is used in a community. To show this let us consider the following example:

Example 3: In a community with 2 active species and 3 metabolites in the community space, that grows at the optimal community growth rate, we know by Theorem 2 that at most 2 EIMs are used. For each of those EIMs we know by Theorem 1 that at most 3(external metabolites)+2(at most two slack variables will be equal to zero) =5 ECMs (thus 5 EFMs) are used. Hence in the community at most 10 EFMs will be used.

Especially in small controlled environments such as research facilities or industrial reactors such an upper bound can give a good indication of the expected complexity of the community. In Appendix B, an additional theorem is featured which would prove that EIMs are continuous with respect to the community transfer vector C and the community growth rate μ. Due to its complexity and lack of time however, it has not been possible to include it in this paper. Together with the two theorems of this paper this continuity theorem would make the framework of the EIMs a more practical approach to determine the expected complexity of communities of micro-organisms.

4 Discussion

As is well known by now, micro-organisms form a big part of life on earth. Until quite recently however, the understanding of those micro-organisms were relatively low, but due to the technological development in the research area, the level of understanding has massively been improved. Especially for single micro-organisms many methods have already been designed to analyze their behaviour. The two examples of such techniques in this paper are Flux Balance Analysis (FBA) and the concept of Elementary Flux Modes (EFMs). By using these kind of techniques a lot of the behaviour of these micro-organisms can be quantified and in that way be predicted. With this information, it would now also be very interesting to be able to analyze how micro-organisms interact with one another in a larger environment. To do so, efforts are made to extend these techniques for single micro-organisms to also function in communities of micro-organisms. As was shown in the first part of the paper, several approaches have already been published to do just that for FBA. In this paper an attempt is made to design such an extension towards communities for EFMs too. The resulting Elementary Interaction Modes (EIMs) form mathematically speaking the basic building blocks for communities of microorganisms that grow together at a steady state in a controlled environment. To construct this method, the earlier published methods on the "FBA extension" were taken as inspiration. This new generalized method was subsequently mathematically analyzed in the end of this paper. Resulting in two theorems that together provide an upper bound on the expected metabolic complexity of communities of micro-organisms. This can have many applications in varying fields such medicine or industrial engineering. Furthermore, it can also streamline further research on communities of microorganisms.

As this paper gives a good starting point for computationally analyzing communities of microorganisms it also creates several interesting opportunities that were not covered in this paper. Topics for further research might be:

- the proposal in Appendix B, the model now assumes constant community growth rate μ and C, the vector with constant C_i for every metabolite i (which is the difference between the concentration of metabolite i in the feed, and the concentration of metabolite i in the medium). But biologically speaking these variables might vary and it would therefore be interesting to see if EIMs are continuous with respect to μ and C.
- We already looked at the EIMs at a maximal community growth rate but if the EIMs are indeed continuous with respect to μ, it would be interesting to see how the EIMs react to slight changes in μ.
- If the EIMs are indeed also continuous with respect to C, it would also be interesting to see how the EIMs react to changes in C. As it is biologically speaking not very logical that C, the communities exchange with the environment, would remain constant at all times.

A final word:

This bachelor project covers some new ground with some actual new results, and this was very exiting for me. The idea was that although the subject was potentially a bit challenging for a bachelors project, with on site supervision and my interest in the subject it would be doable. However, due to the global COVID-19 pandemic I was forced to do my research from home instead of full-time on site at the Systems Biology department of the VU in Amsterdam. Both content wise as well as socially this negatively impacted my progress and because of that it has been more difficult than anticipated to conduct my research. In particular I would have liked to also finalize theorem 3 (Appendix B) and to include it in the body text of my thesis. Nevertheless, I trust that my project can still be helpful to me or other researchers in the future.

References

[1] Charles Darwin. *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.

[2] Stefan Schuster, Claus Hlgetag, and Fachbereich Biologic. On elementary flux modes in biochemical reaction systems at steady state. 1994.

[3] Julien Gagneur and Steffen Klamt. Computation of elementary modes: A unifying framework and the new binary approach. *BMC bioinformatics*, 5:175, 02 2004.

[4] D A Fell and J R Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal*, 238(3):781–786, 09 1986.

[5] Ruchir A. Khandelwal, Brett G. Olivier, Wilfred F. M. Röling, Bas Teusink, and Frank J. Bruggeman. Community flux balance analysis for microbial consortia at balanced growth. *PLOS ONE*, 8(5):1–10, 05 2013.

[6] Siu Hung Joshua Chan, Margaret N. Simons, and Costas D. Maranas. Steadycom: Predicting microbial abundances while ensuring community stability. *PLOS Computational Biology*, 13(5):1–25, 05 2017.

[7] H.T. Banks and Marie Davidian. Introduction to the chemostat. `https://www4.stat.ncsu.edu/~davidian/stma810c/lectures/chemostat.pdf`.

[8] Osbaldo Resendis-Antonio. *Stoichiometric Matrix*, pages 2014–2014. Springer New York, New York, NY, 2013.

[9] Annika Röhl and Alexander Bockmayr. Finding memo: Minimum sets of elementary flux modes. *bioRxiv*, 2019.

[10] Ali R. Zomorrodi and Costas D. Maranas. Optcom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLOS Computational Biology*, 8(2):1–13, 02 2012.

[11] ETH Zurich Marco Terzer. *Matlab efmtool*. Universitätstrasse 19, 8092 Zürich, Switzerland, 2008-2009.

[12] R. Urbanczik. Enumerating constrained elementary flux vectors of metabolic networks. *IET Systems Biology*, 1(5):274–279, 2007.

[13] Steffen Klamt, Georg Regensburger, Matthias P. Gerstl, Christian Jungreuthmayer, Stefan Schuster, Radhakrishnan Mahadevan, Jürgen Zanghellini, and Stefan Müller. From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLOS Computational Biology*, 13(4):1–22, 04 2017.

[14] Daan H. de Groot, Josephus Hulshof, Bas Teusink, Frank J. Bruggeman, and Robert Planqué. Elementary growth modes provide a molecular description of cellular self-fabrication. *PLOS Computational Biology*, 16(1):1–33, 01 2020.

A Matlab Code

```
clear all

cgr = 0.8;                          %community growth rate
C1 = 0.5;                           %uptake constant metabolite mE1
C2 = 0.6;                           %uptake constant metabolite mE2
concmE1 = 0.3;                      %concentration external metabolite 1
concmE2 = 0.4;                      %concentration external metabolite 2
comm_tr_mE1 = cgr*(C1-concmE1);     %total community transport external metabolite 1
comm_tr_mE2 = cgr*(C2-concmE2);     %total community transport external metabolite 2
UB_A = 20;                          %upper bound for all reactions in organism A
UB_B = 20;                          %upper bound for all reactions in organism B


% Stoichiometric matrix:
S = [1  0 -1  0 -1  0  0  0   0    0  0 0      0;
    -1  0  0  0  0  1  0  0   0    0  0 0      0;
     0  1  0 -1  0  0 -1  0   0    0  0 0      0;
     0 -1  0  0  0  0  0  1   0    0  0 0      0;
     0  0  1  0  0  0  0  0 -cgr   0  0 0      0;
     0  0  0  1  0  0  0  0   0  -cgr 0 0      0;
     0  0  0  0  1  0  0 -1   0    0  0 0 comm_tr_mE1;
     0  0  0  0  0 -1  1  0   0    0  0 0 comm_tr_mE2];


% Sum of fluxes is maximally UB_I * X_I
CF=[1  0  1  0  1  1  0  0 -UB_A  0   1 0 0;
    0  1  0  1  0  0  1  1   0  -UB_B 0 1 0];

S=[S;CF];
%metabs = [mA1,mA2,mB1,mB2,massA,massB,mex1,mex2]
%r =[VA1; VB1; VAmass; VBmass; VAex1; VAex2; VBex1; VBex2; XA; XB; slack1; slack2;
     lambda]
r= zeros(length(S(1,:)),1); %all values have to be positive --> all "irreversible"

mnet = CalculateFluxModes(S,r);

efvs= mnet.efms;
[numRows,numCols] = size(efvs);

 for i = 1:numCols %normalize EFVs to lambda=1 so that all values are actual fluxes.
     if efvs(numRows,i) == 0
         efvs(:,i) = efvs(:,i);
     else
         efvs(:,i)= efvs(:,i)/efvs(numRows,i);
     end
 end

rowNames = {'VA1', 'VB1', 'VAmass', 'VBmass', 'VAex1', 'VAex2', 'VBex1', 'VBex2', '
     XA', 'XB', 'slack 1', 'slack 2', 'lambda'};
efv = efvs;
EFV_table = array2table(efv,'RowNames',rowNames)
```

B Are EIMs continuous w.r.t. μ and C ?

This section is still a draft version. The idea here was to slightly alter theorem 3 of [14] which relies on the Implicit Function Theorem, so that it would fit this paper. During this bachelor project I have not been able to work this out properly but it would have been an interesting feature. If it was finished it would have been fitted in after paragraph 3.3.1. The setup here in the appendix will be as follows: First we will discuss the additional mathematics needed to understand the use of the theorem. Then, we will consider the original theorem in [14] and how this needs to be altered to fit this paper. Afterwards we will do the same for the proof.

Let us start with the basics: μ and C were up to now assumed to be constant but biologically speaking this is a bit awkward as they often may vary. Therefore, we would like to understand how the EIMs react to changes in μ an C. To do so we will analyze how a (unique) fixed point on an EIM changes after a change in the variables. But in order to achieve that we will first have to introduce a bit more mathematical background for our model. Let us go back to our model:

$$
\begin{array}{c}
\left(m_1^E \ \ldots m_{k_3}^E\right)^T \\
\textit{biomass A} \\
\textit{biomass B} \\
\textit{species A} \\
\textit{species B}
\end{array}
\begin{array}{c}
\begin{array}{cccccc}
\beta_1^A \ \ldots \ \beta_{l_1}^A & \beta_1^B \ \ldots \ \beta_{l_2}^B & X^A & X^B & s_1^A & s_1^B
\end{array} \\
\begin{bmatrix}
ECMs^A & ECMs^B & 0 & 0 & 0 & 0 \\
e_A & 0 & -\mu & 0 & 0 & 0 \\
0 & e_B & 0 & -\mu & 0 & 0 \\
w_A & 0 & -UB_A & 0 & 1 & 0 \\
0 & w_B & 0 & -UB_B & 0 & 1
\end{bmatrix}
\end{array}
\begin{bmatrix}
\beta_1^A \\ \vdots \\ \beta_{l_1}^A \\ \beta_1^B \\ \vdots \\ \beta_{l_2}^B \\ X^A \\ X^B \\ s_1^A \\ s_1^B
\end{bmatrix}
= \mu
\begin{bmatrix}
C \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
$$

which we will refer to as the system of equations $G(\mu)\beta = \mu b(C)$, where $\beta \geq \mathbf{0}$.

G is here an $(k_3 \ + 2n) \times (l_{tot} + 2n + 1)$ matrix. Let us assume $l_{tot} \geq k_3$ so that G has more columns than rows. Let us now set $rank(G) = m$. And let us reduce the matrix G to a matrix with m rows such that there are no row-dependencies in G. Now because G has more columns than rows there are degrees of freedom in the matrix. Therefore, in general, the solution of β is not unique (This is also why we can find EIMs with one degree of freedom). But at this moment we are looking for unique solutions of β. In order to be able to find an unique solution of β, G needs to be square and invertible. To make G square we may choose a set D of columns of G, so that G_D (G restricted to the columns in D) is a square matrix. If G_D is also invertible we will call D a basis for G. For every basis D we have a unique solution for β namely, $\beta_D = \mu(G_D)^{-1}b(C)$, where all elements of β that are not in D are set to zero. If all the elements of β_D also adhere to the constraint that $\beta \geq \mathbf{0}$, then we call basis D a feasible basis and the corresponding β_D a basic feasible solution. It can however be, that one of the elements of β_D is zero. Therefore we have in general that:

$$supp(v) \subseteq D.$$

Let us now define these unique solutions as follows:

Definition 6: Elementary Interaction State (EIS)
An EIS is a fixed unique solution β_D on an EIM where $\beta_D = \mu(G_D)^{-1}b(C)$.

With this definition we now have a framework of EISs and EIMs which works analogously to the framework of EGSs (Elementary Growth States) and EGMs (Elementary Growth

Modes) as was constructed in [14]. The theory behind that framework is not very relavant in this paper but if you are interested in it, it can be found in the original paper. Theorem 3 of that paper states the following:

"Theorem 3. For a given growth rate μ_0 and set of metabolite concentrations x_0, there exists an open neighbourhood U such that for all $(x, \mu) \in U$, each EGS with support equal to its feasible basis, $\alpha(x_0, \mu_0)$, can be continuously extended to a vector $\alpha(x, \mu)$ that solves the balanced growth equations and belongs to the same EGM."[14]

Here the "balanced growth equations" refer to the system of equations that are used in the paper which are similar to our system of equations $G(\mu)\beta = \mu b(C)$. If we now alter this theorem to fit this paper, we replace EGS with EIS, EGM by EIM and instead of the set of metabolite concentrations x we have the "community transfer vector" C. Additionally our EIS will be called β instead of α as in the original paper. Together these alterations give us the following theorem:

Theorem 3: For a given growth rate μ_0 and a given community transfer vector C_0 there exists an open neighbourhood U such that for all $(\mu, C_0) \in U$, each EIS with support equal to its feasible basis $\beta(\mu_0, C_0)$ can be continuously extended to a vector $\beta(\mu, C)$ that solves the balanced growth equations and belongs to the same EIM.

Let us for the proof of this theorem first look at the proof of the original paper, here the exact interpretation of all the formulas are not of interest, but the blueprint is (especially the use of the Implicit Function Theorem).

"Proof: Choose an arbitrary EGS $\alpha_0 \in \mathbb{R}^{n+1}$. We know that it solves:

$$B(x_0, \mu_0)\alpha_0 = \mu_0 u_{m+1}, \quad \alpha_0 \geq 0.$$

The EGS has a feasible basis D, and without loss of generality we restrict $B(x_0, \mu_0)$ to the columns indexed in D. As discussed before, we can select a set of rows corresponding to independent metabolites with non-zero concentrations such that the resulting matrix is square and invertible. We may therefore choose $n = m$ and the dimension of B to be $(m+1)\times(m+1)$.

We would like to apply the Implicit Function Theorem to see that there are continuously differentiable functions $\hat{\alpha_j}$, $j = 1, \ldots, m+1$, such that in an open environment of $(x0, \mu_0)$, the Balanced Growth Equations are still met: $B(x_0, \mu_0)\alpha_0 = \mu_0 u_{m+1}$. Since the $\hat{\alpha_j}(x, \mu)$ are continuous, and since no $\alpha_j(x_0, \mu_0)$ is equal to zero by the assumption in the Theorem, we can then also choose a neighbourhood in which $\hat{\alpha}(x, \mu) \geq 0$. This $\hat{\alpha}(\mu)$ would thus indeed be a continuous extension of the EGS that belongs to the same EGM, since its support does not change.

Let s be the number of components in x. For the Implicit Function Theorem we need a function F: $\mathbb{R}^{m+1} \times \mathbb{R}^{s+1} \longrightarrow \mathbb{R}^{m+1}$ that is zero at (α_0, x_0, μ_0). For this function, we can use the Balanced Growth equations as components: the first m are given by:

$$\begin{aligned} F_k(\alpha, x, \mu) := \sum_{j=1}^{m}((x_k\alpha_j - P_{kj})\frac{f_j(x)}{\mu} + x_k b_j - M_k j)a_j g_j(x) + \\ (x_k b_{m+1} - M_{k(m+1)})\alpha_{m+1} g_{m+1}(x) = 0 \end{aligned} \tag{21}$$

where $k = 1, \ldots, m$. The last component is given by:

$$F_k(\alpha, x, \mu) := a_{m+1} g_{m+1}(x) - \mu = 0 \tag{22}$$

Let us check the conditions for the Implicit Function Theorem:

The function F: $\mathbb{R}^{m+1} \times \mathbb{R}^{s+1} \longrightarrow \mathbb{R}^{m+1}$ *with components given by (21) and (22) is continuously differentiable in* α, x, *and* μ *in a neighbourhood of* (α_0, x_0, μ_0).

By assumption, (α_0, x_0, μ_0) *is a solution of the balanced growth equations:* $F(\alpha_0, x_0, \mu_0) = 0$.

The entries of the Jacobian of F *at* (α_0, x_0, μ_0) *with respect to* α *are, for* $k = 1 \ldots, m$, *given by*

$$J_{kj} = \frac{\delta F_k}{\delta \alpha_j} = ((x_{0,k} a_j - P_{kj}) \frac{f_j(x_0)}{\mu_0} + x_{0,k} b_j - M_{kj}) g_j(x_0);$$

and the last row of the Jacobian is zero except for the last entry, which is $g_{m+1}(x)$. *Note that this Jacobian is exactly the matrix* $B(x_0, \mu_0)$, *which was invertible by construction.*

The IFT may therefore be invoked, which shows that for each EGS at (x_0, μ_0), *with support equal to its feasible basis, there is an open neighbourhood around* (x_0, μ_0) *such that the EGS can be continuously extended. By taking the intersection of these neighbourhoods we can indeed find a neighbourhood in which all of the EGSs can be extended."*[14]

If we now replace the EGS framework by the EIS framework we replace the balanced growth equations by our own balance growth equations $G(\mu)\beta = \mu b(C)$. To let the application of the Implicit Function Theorem work in this case as well we will replace the continuously differentiable function F by $F(\beta, \mu, C) = G(\mu)\beta - \mu b(C)$. This yields us the following proof:

Proof: Choose an arbitrary EIS: $\beta_0 \in \mathbb{R}^{l_{tot}+4}$, with support equal to its feasible basis. By definition it solves:

$$G(\mu_0)\beta_0 = \mu_0 b(C_0), \ \beta_0 \geq 0.$$

This EIS has a feasible basis D, and without loss of generality we may restrict $G(\mu_0)$ to the columns indexed in D. As discussed before we can select a set of rows corresponding to independent metabolites with non-zero concentrations such that the resulting matrix is square and invertible. We may therefore choose $l_{tot} = K_3$ where K_3 is the number of independent metabolites with non-zero concentrations. Hence the dimensions of G are $(K_3 + 4) \times (K_3 + 4)$.

Now, we would like to apply the Implicit Function Theorem to see that there exist $K_3 + 4$ continuously differentiable functions $\hat{\beta}_j(\mu, C)$ that together form the vector $\hat{\beta}$ such that in an open neighborhood around (μ_0, C_0) the system of equations $G(\mu)\hat{\beta} = \mu b(C)$ is still met. Since the functions $\hat{\beta}_j(\mu, C)$ are continuous, and since no $\beta_j(\mu_0, C_0) \in \beta_0$ is equal to zero by assumption in the theorem, we can then also choose a neighborhood in which $\hat{\beta} \geq 0$. This $\hat{\beta}$ would thus indeed be a continuous extension of the EIS that belongs to the same EIM, since its support does not change.

For the Implicit Function Theorem we need a continuous differentiable function F: $\mathbb{R}^{K_3+4} \times \mathbb{R}^{1+K_3} \longrightarrow \mathbb{R}^{K_3+4}$ that is zero at (μ_0, C_0). That has an invertible Jacobian matrix. For this function we can use the system of equations $G(\mu)\beta = \mu b(C)$ so that $F(\beta, \mu, C) = G(\mu)\beta - \mu b(C)$. We note:

- F: $\mathbb{R}^{K_3+4} \times \mathbb{R}^{1+K_3} \longrightarrow \mathbb{R}^{K_3+4}$ is continuously differentiable in β, μ and in C in a neighborhood of (β_0, μ_0, C_0). To see this, note that a composition of functions is continuously differentiable if all of its components are, which clearly is the case as all functions are mere linear combinations.
- By assumption (β_0, μ_0, C_0) is a solution of $F(\beta, \mu, C) = 0$.
- The entries of the Jacobian matrix of F ($\frac{\delta F_k}{\delta \beta_j}$) are exactly equal to the entries of matrix G, which was invertible by construction (both $G(\mu)$ and $\mu b(C)$ are independent of β, hence the Jacobian of $F = G(\mu)\beta - \mu b(C)$ is exactly equal to $G(\mu)$).

Hence the IFT may be invoked. This shows that for each EIS at (μ_0, C_0) with support equal to its feasible basis, there is an open neighborhood around (μ_0, C_0) such that the EIS can be continuously extended. By taking the intersection of these neighborhoods we can indeed find a neighborhood in which all of the EIS can be extended.