Flux Cones of Metabolic Networks

Dissertation
Zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorleget von
Frederik Laszlo Wieder

Berlin 2024

Erstgutacher: Prof. Dr. Alexander Bockmayr
Zweitgutacher: Prof. Dr. Francisco J. Planes

Tag der Disputation: 07.10.2024

Abstract

Systems biology is located at the intersection of biology, computer science, and mathematics, and is based on the translation of biological systems into mathematical models. It aims to predict the behavior of these biological systems to improve the efficiency of time- and cost-intensive research in laboratories.

In this thesis, we focus on the description and understanding of metabolic networks at steady state. These networks are mathematical models of the metabolic processes inside a cell.

The stoichiometric and thermodynamic constraints that must hold in a metabolic network at steady-state define the steady-state flux cone. An important concept to analyze flux cones in a mathematically and biologically meaningful way are elementary flux modes, which can be considered as minimal functional units of metabolic networks. In the flux cone, elementary flux modes correspond to vectors with inclusionwise minimal support.

We focus on geometric aspects of flux cones of metabolic networks and elementary flux modes. The number of elementary flux modes may be very large, even for medium-sized metabolic networks. We study the facial structure and investigate the distribution of elementary flux modes among the faces of the flux cone. We observe that they are primarily contained in faces of relatively low dimension. Due to this observation, we develop a method to enumerate subsets of elementary flux modes that are contained in a specific face of the flux cone and apply this to decompositions of flux vectors.

Empirically, we observed that elementary flux modes can always be written as a positive sum of exactly two others. Motivated by this, we investigate decompositions of elementary flux modes into others and discuss a conjecture that claims each EFM can always be decomposed into exactly 2 others or is not decomposable at all.

Our mathematical results are illustrated on real examples and the presented data can be reproduced with a Python package we developed.

Contents

Chapter 1

Introduction

Systems biology is located at the intersection of biology, computer science, and mathematics and is based on the translation of biological systems into mathematical models. Until the turn of the last century, traditional molecular biology focused on studying individual components such as specific genes or proteins in isolation. Systems biology studies interactions of multiple components within living organisms, leading to a deeper understanding of the dynamic behavior of biological systems (Kitano, 2002).

The last decades have seen a rise in high-throughput methods that allow quantifying molecular properties of biological systems with increasing accuracy. These ever growing datasets consisting of omics data (e.g., transcriptomics, genomics, proteomics, and metabolomics) escape conventional biological analysis methods, mostly due to their size and complexity. At the same time, they promise a more comprehensive understanding of biological systems than ever before (Mardis, 2008).

A metabolic network is a complex system of biochemical reactions within a cell or an organism. Metabolites are transformed by enzyme-catalyzed biochemical reactions. Chains of these reactions form metabolic pathways that enable cellular functions such as energy production, metabolite production, maintenance, and growth.

Genome-scale metabolic networks have a wide range of applications, including the study of microorganisms, metabolic engineering, drug development, prediction of enzyme functions, and understanding microbial community interactions and human disease (Fang, Lloyd and Palsson (2020), and the references therein). By studying these models, researchers can predict the behavior of microorganisms in specific environments, determine metabolic pathways for biotechnological applications, anticipate the effects of drugs on cellular metabolism, and gain insights into interactions within microbial com-

munities and their impact on human health. Metabolic networks provide a powerful tool to simulate complex metabolic processes in silico, thereby supporting time- and cost- intensive research in laboratories.

Constraint-based analysis of metabolic networks applies mathematical constraints like mass balance, thermodynamics, and capacity limitations to predict the flow of metabolites as well as functional states of metabolic networks. It has become an important research field in systems biology (Bordbar *et al.,* 2014; Fang, Lloyd and Palsson, 2020).

In contrast to dynamic modeling approaches, where changes of metabolite concentrations over time are described by differential equations, this thesis focuses on metabolic networks at steady-state. The stoichiometric and thermodynamic constraints that must hold in a metabolic network at steady-state define the (steady-state) flux cone, which comprises all (steady-state) flux distributions. A flux balance analysis (FBA) problem is a linear optimization problem to determine an optimal flux distribution in a metabolic network at steady-state to achieve a specific objective, typically maximizing biomass or metabolite production (Orth, Thiele and Palsson, 2010).

One important concept to analyze flux cones of metabolic networks in a mathematically and biologically meaningful way are elementary flux modes (EFMs) (Schuster and Hilgetag, 1994; Schuster, Hilgetag, *et al.,* 2002), which can be viewed as minimal functional units of the network. Viewed as flux vectors, they form a finite generating set of the flux cone. This means that every steady-state flux distribution in a metabolic network is a positive combination of EFMs (Gagneur and Klamt, 2004; Wagner and Urbanczik, 2005; Larhlimi and Bockmayr, 2008; Jevremović and Boley, 2013). From a mathematical point of view, we explore geometric properties of flux cones of metabolic networks with a particular focus on elementary flux modes. We apply concepts from various areas of mathematics, including linear optimization, polyhedral geometry, and matroid theory, to obtain new insights into the structure of the set of EFMs.

Structure of this thesis

In Chapter 2, we introduce the required mathematical background. First, we define polyhedral cones in general and explore their basic properties. Next, we extend the familiar definition of a graph step by step to the definition of a generalized weighted hypergraph that is used to represent a metabolic network. Metabolic networks are formally defined and the definition of the steady-state flux cone is derived from the steady state-assumption. Next, we formally introduce EFMs as vectors with inclusionwise minimal support in the flux cone and present some well-known properties with their proofs. Finally, we provide a brief overview of the most prominent methods to enumerate the set of EFMs.

In Chapter 3, we study the geometry of the flux cone and EFMs. We begin by expanding our knowledge of geometric and combinatorial properties of polyhedral cones. Next, we discuss inequality descriptions of the flux cone and the effects of adding and removing redundant constraints. Then, we investigate the relationship between faces of the flux cone and metabolic behaviors. We also generalize a result concerning the one-to-one correspondence between minimal metabolic behaviors and minimal proper faces of the flux cone. To structure the set of EFMs, we introduce the degree of a flux vector as the dimension of the inclusionwise minimal face containing it and analyze the distribution of elementary flux modes in the face lattice of the flux cone. We prove upper bounds on the degree of EFMs and show that EFMs occur in the relative interior of the flux cone only in very specific cases. We conclude the chapter with a result establishing a relationship of combinatorial properties of the flux cone and the cardinality of its minimal metabolic behaviors.

In Chapter 4, we discuss decompositions of flux vectors into EFMs, i.e., how they can be written as a positive combination of them. We show that a flux vector that is decomposed and the decomposing EFMs have to belong to the same face of the flux cone. Motivated by this observation, we develop an

algorithm to determine the face of the flux cone that is defined by a given flux vector and the subset of EFMs contained in that face. For lower-dimensional faces, the cardinality of such a subset of EFMs turns out to be significantly smaller than the total number of EFMs. We illustrate the scalability of our method by determining EFMs in the faces defined by solutions of FBA problems in a large selection of genome-scale metabolic networks. Furthermore, we introduce low-degree decompositions as an alternative to shortest decompositions. Although more EFMs are needed in a low-degree decomposition, these EFMs cannot be further decomposed into other EFMs of lower degrees.

In Chapter 5, we further investigate decompositions of EFMs into other EFMs. Empirically, we observed that EFMs can always be decomposed into two other EFMs or they are not decomposable at all. This observation leads to a conjecture claiming that this is always the case, i.e., that the length of a shortest decomposition of an EFM into other EFMs is always 2. We formalize this conjecture, define what a counterexample to this conjecture is, and prove that a counterexample needs to have at least two reversible reactions. Next, we study the relationship between matroids and metabolic networks where all reactions are reversible. Applying concepts from matroid theory allows us to prove a weaker version of our conjecture, namely that the support of every EFM in a metabolic network with only reversible reactions is contained in the support of a positive combination of two other EFMs. We derive an algorithm to generate new EFMs from a starting set by positive combinations (similar to the circuit enumeration method for matroids described by Khachiyan *et al.* (2005)) and an algorithm to determine a length-2 decomposition of an EFM. Finally, we present a computational approach to find counterexamples to the conjecture by enumerating and testing all metabolic networks of small sizes.

In Chapter 6, we present a Python package that was developed in the course of this thesis. The software is open source and available on GitHub. The package can reproduce the data presented in this thesis. We give a short overview of the structure of the software and available functions. Next, we de-

scribe different methods to enumerate the set of EFMs that are provided by the package. Finally, we discuss applications of the software and implementations of several algorithms described in Chapter 5.

Chapter 2

Background

In this chapter, we summarize common results that form the mathematical background needed for this thesis. More specific concepts will be introduced in the chapters where they are first discussed.

2.1 Notation

Throughout this thesis we will be following the notation guidelines summarized in this section. If we have k vectors in $\mathbb{R}^n$, we will denote them by $x^1, \dots, x^k$ while we will write x_i for the i-th entry of a vector $x \in \mathbb{R}^n$. For scalars in $\mathbb{R}$, we will use Greek letters $(\lambda, \gamma, \alpha, \beta, \dots)$ and if we have k scalars we will denote them $\lambda_1, \dots, \lambda_k$. We will sometimes write $\mathbb{R}_{\geq 0} := \{\lambda \in \mathbb{R} \mid \lambda \geq 0\}$ or $\mathbb{R}_{>0} := \{\lambda \in \mathbb{R} \mid \lambda > 0\}$ when we mean the subset of all non-negative or positive real numbers, respectively. In definitions of sets, colons (":") and vertical lines ("|") will be used analogously. A vector in $\mathbb{R}^n$ should always be viewed as a column vector (i.e., a $(n \times 1)$-matrix) if not specified. For better readability, when we present a set of vectors, we will often present a matrix and specify that the vectors are the rows of that matrix.

2.2 Polyhedral cones

We begin by introducing polyhedral cones in general, before we define the flux cone of a metabolic network as a specific type of polyhedral cone. For further reading, we refer to (Lauritzen, 2013; Schneider, 2013; Schrijver, 1998; Ziegler, 1995). Most of our notation, definitions and basic results are from (Schrijver, 1998).

A vector $x \in \mathbb{R}^n$ is a *linear combination* of the vectors $x^1, \dots, x^k \in \mathbb{R}^n$ if

$x = \lambda_1 x^1 + \cdots + \lambda_k x^k$, for some $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. If, in addition

$$\left\{ \begin{array}{ll} \lambda_1, \ldots, \lambda_k \geq 0, & \\ & \lambda_1 + \cdots + \lambda_k = 1, \\ \lambda_1, \ldots, \lambda_k \geq 0, & \lambda_1 + \cdots + \lambda_k = 1, \end{array} \right\} \text{ we call } x \text{ a } \left\{ \begin{array}{c} \textit{conic} \\ \textit{affine} \\ \textit{convex} \end{array} \right\}$$

combination of the vectors $x^1, \ldots, x^k$.

For a nonempty subset $X \subseteq \mathbb{R}^n$, we denote by $\text{lin}(X)$ (resp. $\text{cone}(X)$, $\text{aff}(X)$, $\text{conv}(X)$) the *linear* (resp. *conic, affine, convex*) *hull of* X, i.e., the set of all linear (resp. conic, affine, convex) combinations of finitely many vectors of X.

A nonempty set $C \subseteq \mathbb{R}^n$ is a *convex cone*, if it is closed under conic combinations, i.e., $\lambda x + \mu y \in C$, for all $x, y \in C$ and $\lambda, \mu \geq 0$. By definition, the conic hull of a subset of $\mathbb{R}^n$ is a convex cone. A convex cone C is *polyhedral* if it is the solution set of a system of finitely many homogeneous linear inequalities, i.e., if

$$C = \{x \in \mathbb{R}^n \mid Ax \geq 0\},$$

for some matrix $A \in \mathbb{R}^{m \times n}$. If a cone C is the conic hull of a finite set $X = \{x^1, ..., x^k\} \subset \mathbb{R}^n$, it is called *finitely generated* and the set X is called a *generating set* of C. By the well-known theorem of Farkas-Minkowski-Weyl (see e.g. (Schrijver, 1998)), a convex cone is polyhedral if and only if it is finitely generated. For the rest of this thesis we will only consider polyhedral cones and often simply write cone.

If $C := \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ is a polyhedral cone, an inequality $ax \geq 0$, where a denotes a row of A and ax the inner product of a and x, is called an *implicit equality in* $Ax \geq 0$, if $ax = 0$, for all $x \in C$. Following (Schrijver, 1998), we denote by $A^{=}x \geq 0$ the system of implicit equalities in $Ax \geq 0$ and by $A^{+}x \geq 0$ the remaining inequalities.

If removing an inequality $ax \geq 0$ from $Ax \geq 0$ does not change the associated cone C, the inequality is called *redundant*. If there are no redundant

inequalities, the description $Ax \geq 0$ is called *irredundant*.

The *dimension* $\dim(C)$ of a cone C is the dimension of its affine hull $\text{aff}(C) = \{x \in \mathbb{R}^n \mid A^{=}x = 0\}$ and is equal to $n - \text{rank}(A^{=})$. Note that since $0 \in C$, $\text{aff}(C)$ coincides with the linear hull $\text{lin}(C)$.

A vector $x \in C$ is in the *relative interior of* C, if there exists $\epsilon > 0$ such that $B_\epsilon(x) \cap \text{aff}(C) \subseteq C$, where $B_\epsilon(x)$ is the n-dimensional ball of radius ϵ centered at x. We will write $x \in \text{relint}(C)$, if x is in the relative interior of C. If $x \in C$ is not in the relative interior of C, it is in the *relative boundary of* C.

The *lineality space* of a cone $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ is given by $\text{lin.space}(C) := \{x \in \mathbb{R}^n \mid Ax = 0\}$, which is the inclusionwise maximal linear subspace contained in C. A cone C is called *pointed* if its lineality space is trivial, i.e., $\text{lin.space}(C) = \{0\}$. If a cone is pointed, it does not contain a line.

An inequality $ax \geq 0$ is called *valid* for C if $C \subseteq \{x \in \mathbb{R}^n \mid ax \geq 0\}$. A nonempty set $F \subseteq C$ is called a *face* of C if there exists an inequality $ax \geq 0$ valid for C such that $F = C \cap \{x \in \mathbb{R}^n \mid ax = 0\}$. The hyperplane $\{x \in \mathbb{R}^n \mid ax = 0\}$ is then called a *supporting hyperplane* of F. Alternatively, a face can be characterized as a nonempty set $F \subseteq C$ with $F = \{x \in C \mid A_{I,\star}x = 0\}$, where $A_{I,\star}$ is the submatrix of A whose rows belong to the set $I \subseteq \{1, \ldots, m\}$ (Schrijver, 1998).

A polyhedral cone C has only finitely many faces, each face F of C is itself a polyhedral cone and $F' \subseteq F$ is a face of F if and only if F' is a face of C. A k-dimensional face will also be called a *k-face*. A cone C is pointed if and only if it has a 0-face, namely the origin.

A face $F \neq C$ of C is called a *facet* if it is inclusionwise maximal, i.e., there is no other face $F' \neq C$ such that $F \subset F'$. If the description $Ax \geq 0$ of C is irredundant, there is a 1-1 correspondence between the facets of C and the inequalities in $A^{+}x \geq 0$ (Schrijver, 1998, Theor. 8.1). In particular, for every facet F there is an inequality $ax \geq 0$ from $A^{+}x \geq 0$ such that $F = \{x \in C \mid ax = 0\}$. We have $\dim(F) = \dim(C) - 1$ for every facet F of C,

and every face of C (except C itself) is the intersection of facets of C.

2.3 From graphs to hypergraphs

We refer to (Diestel, 2017) for detailed information on graphs. In this section, we describe how the definition of a graph can be adapted step by step to obtain directed hypergraphs, which we will generalize even further in the following section to represent metabolic networks.

A *finite graph* is defined as a tuple $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is a set of nodes, and E is a set of edges where each edge $e \in E$ is a subset of V. We assume that each edge is a subset of V with cardinality 2, thereby excluding graphs with loops. Such a graph is also called *undirected graph* because the edges are not oriented.

If the edges are ordered pairs, i.e. $E \subseteq (V \times V)$, we obtain a *directed graph*, where the direction from one node to another matters ($(v_i, v_j) \neq (v_j, v_i)$). The directed edges of a directed graph are often called *arcs* to distinguish them from the edges of undirected graphs. Figure 2.1 shows an example of an undirected graph G on five nodes and a directed graph D on the same set of nodes obtained by giving the edges of G an orientation and replacing the lines (edges) with arrows (arcs) to indicate the chosen orientation.

An undirected graph G on n nodes can be represented by a so-called adjacency matrix $A \in \mathbb{R}^{n \times n}$, where the entry a_{ij} (i.e., the entry in the i-th row and j-th column of A) is 1 if there exists an edge connecting the nodes v_i and v_j and 0 otherwise. For an undirected graph, the adjacency matrix is always symmetrical. To represent a directed graph D on n nodes, we only set $a_{ij} = 1$ if there exists an arc from v_i to v_j while the remaining entries are zeros. For directed graphs, the adjacency matrix need not be symmetrical. The adjacency matrices for the undirected graph (A_G) and the directed graph (A_D) in

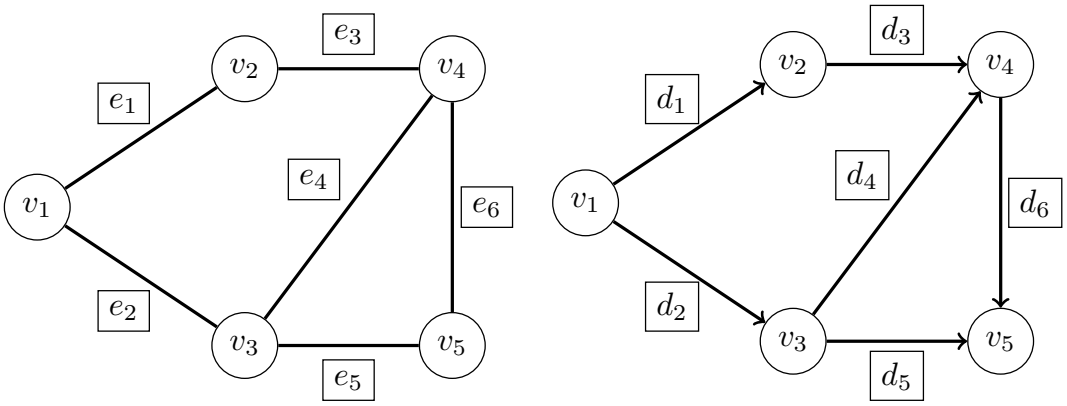

Figure 2.1: On the left: Undirected graph G on nodes $v_1, \ldots, v_5$ with edges $e_1, \ldots, e_6$. On the right: Directed graph D on the same nodes with arcs $d_1, \ldots, d_6$.

Figure 2.1 are given by

$$A_G = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad A_D = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

A directed *hypergraph* generalizes the concept of a directed graph. It is defined as a tuple $H = (V, \mathcal{H})$, where $V = \{v_1, \ldots, v_n\}$ is a set of nodes and $\mathcal{H}$ is a set of hyperarcs. Each hyperarc can connect any number of nodes, not just pairs. Hence, a hyperarc $h = (V_1 \subseteq V, V_2 \subseteq V)$ is an ordered pair consisting of two subsets of V. Note that we did not exclude the empty set for the definition of hyperarcs. Figure 2.2 shows a directed hypergraph on the same five nodes as in Figure 2.1. Observe that $h_1 = (\emptyset, \{v_1\})$ contains the empty set and therefore does not start at any other node, but ends in v_1. The arcs h_3, h_5 and h_6 only connect pairs of nodes and thus could also appear in a directed graph (cf. Figure 2.1). Arcs h_2 and h_4 are *proper hyperarcs* since

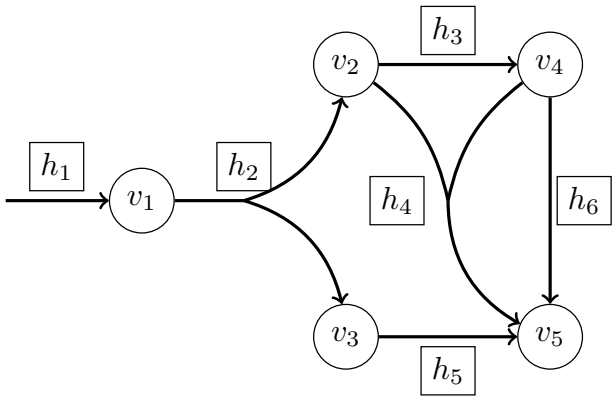

Figure 2.2: Directed hypergraph on the same five nodes as in Figure 2.1 with directed hyperarcs $h_1, \ldots, h_6$.

they connect more than just two nodes.

To represent a hypergraph H with n hyperarcs on m nodes, we introduce the incidence matrix $A_H \in \mathbb{R}^{m \times n}$. Each row of A_H represents a node and each column represents a hyperarc. For each hyperarc $h_j = (V_1 \subseteq V, V_2 \subseteq V)$ we set $a_{ij} = -1$ for all $i \in I_1$ and $a_{ij} = 1$ for all $i \in I_2$ where I_1 is the set of indices of nodes in V_1 and I_2 the set of indices of nodes in V_2. The incidence matrix for the directed hypergraph H in Figure 2.2 is given by

$$A_H = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Of course an incidence matrix can also be used to represent graphs without hyperarcs, but a proper hyperarc cannot be represented by an adjacency matrix.

2.4 Metabolic networks

A *metabolic network* $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Irr})$ is given by a set $\mathcal{M}$ of (internal) *metabolites*, a set $\mathcal{R}$ of *reactions*, a *stoichiometric matrix* $S \in \mathbb{R}^{m \times n}$, where $m = |\mathcal{M}|$ and $n = |\mathcal{R}|$, and a subset $\mathrm{Irr} \subseteq \mathcal{R}$ of *irreversible reactions*. The reactions in $\mathcal{R} \setminus \mathrm{Irr}$ are called *reversible reactions* and are denoted by Rev. Depending on the context, we will also equivalently define a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$ by its set of reversible reactions Rev instead of the set of irreversible reactions Irr. The equivalence follows from $\mathrm{Irr} = \mathcal{R} \setminus \mathrm{Rev}$ and $\mathrm{Rev} = \mathcal{R} \setminus \mathrm{Irr}$.

For $J \subseteq \mathcal{R}$, we denote by $S_{\star,J}$ the submatrix of S whose columns belong to J. Analogously for $I \subseteq \mathcal{M}$, we denote by $S_{I,\star}$ the submatrix of S whose rows belong to I.

A metabolic network can be seen as a generalized weighted hypergraph with the generalized incidence matrix S, where the metabolites are represented by nodes and the reactions by hyperarcs. To distinguish reversible and irreversible reactions in visualizations of metabolic networks, we use arrows in both directions for reversible reactions (cf. Figure 2.4). So a metabolic network is a hypergraph as in Figure 2.2 where in addition some arcs (namely the ones representing reversible reactions) have arrow tips on both sides (e.g. $\boxed{5}$ in Figure 2.4). The weights come into play when considering that chemical reactions typically transform more than single molecules. A positive entry $S_{i,j} > 0$ in the stoichiometric matrix S indicates that reaction j produces $S_{i,j}$ units of metabolite i. If $S_{i,j} < 0$, $S_{i,j}$ units of metabolite i are consumed in reaction j.

An example of a chemical reaction r and its mathematical representation as a metabolic network is the combustion of hydrogen:

$$2H_2 + O_2 \to 2H_2O$$

If we consider this reaction r to be irreversible, the corresponding metabolic network is visualized in Figure 2.3.

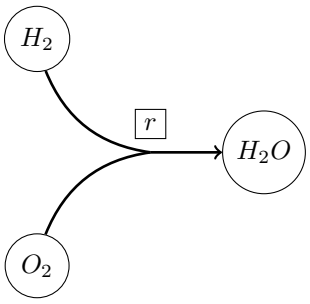

Figure 2.3: Combustion of hydrogen, visualized as a metabolic network.

The stoichiometric coefficients in this reaction are represented in the stoichiometric matrix S_{Hc}, which has 3 rows and 1 column because we have 3 metabolites and 1 reaction:

$$S_{Hc} = \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}.$$

The first entry is -2 because 2 H_2 are consumed in this reaction, the second entry is -1 because 1 O_2 is consumed in this reaction, and the last entry is 2 because 2 H_2O are produced. The full description of the combustion of hydrogen as a metabolic network $\mathcal{N}_{Hc}$ is:

$$\mathcal{N}_{Hc} = (\mathcal{M} = \{H_2, O_2, H_2O\}, \mathcal{R} = \{r\}, S_{Hc}, \text{Irr} = \{r\}).$$

The metabolic network in Figure 2.4 and variations of it will be used throughout this thesis for visualizations. Sometimes, a system boundary is added to clarify which reactions are *exchange reactions* that cross the system boundary. The corresponding generalized hyperarcs have the empty set ($\emptyset$) as one of their subsets of nodes.

For this example metabolic network we assume that all stoichiometric co-

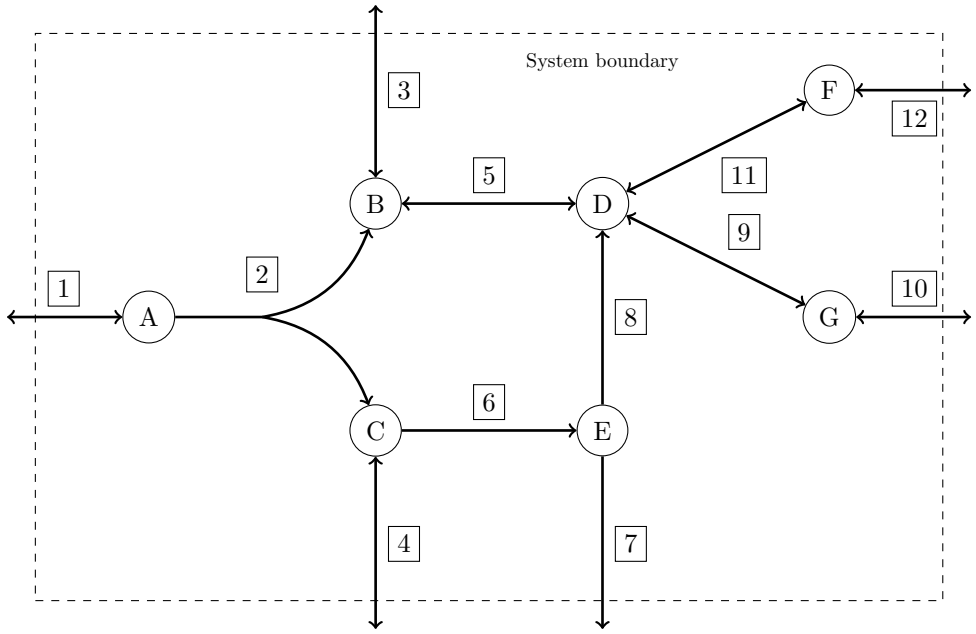

Figure 2.4: Example of a metabolic network.

efficients are -1,0 or 1. We get the stoichiometric matrix

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix}.$$

Together with $\mathrm{Irr} = \{2, 6, 7, 8\}$ our metabolic network is defined.

Throughout this thesis, we assume that the metabolic networks are at *steady-state*, i.e., for each internal metabolite, the rate of production is equal to the rate of consumption. Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$ be a metabolic network.

In matrix notation, the steady-state constraints can be written as $Sv = 0$, where $v \in \mathbb{R}^n$ denotes a *flux vector*. By adding the *thermodynamic irreversibil-*

ity constraints $v_j \geq 0$, for all $j \in \mathrm{Irr}$, and setting

$$A = \begin{pmatrix} S \\ -S \\ I_{\mathrm{Irr},\star} \end{pmatrix} \tag{2.1}$$

we obtain the polyhedral cone

$$C = \{x \in \mathbb{R}^n \mid Ax \geq 0\} = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}, \tag{2.2}$$

which is called the (steady-state) *flux cone* of $\mathcal{N}$. Here, v_{Irr} is the subvector of v, whose components belong to Irr, and $I_{\mathrm{Irr},\star}$ is the submatrix of the $(n \times n)$ identity matrix I_n, whose rows correspond to the irreversible reactions.

2.5 Elementary flux modes

Let C be the flux cone of a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$. A vector $e \in C \setminus \{0\}$ is called an *elementary flux mode* (EFM) (Schuster and Hilgetag, 1994) if it has inclusionwise minimal support, i.e., if

$$\forall v \in C \setminus \{0\} : \mathrm{supp}(v) \subseteq \mathrm{supp}(e) \implies \mathrm{supp}(v) = \mathrm{supp}(e), \tag{2.3}$$

where the *support* of $v \in \mathbb{R}^n$ is defined by $\mathrm{supp}(v) = \{i \in \mathcal{R} \mid v_i \neq 0\}$. We say that a reaction $i \in \mathcal{R}$ is *active* in $v \in C$, if $i \in \mathrm{supp}(v)$. By $\mathrm{irr.supp}(v) := \mathrm{supp}(v) \cap \mathrm{Irr}$ we denote the *irreversible support* of v, i.e., the set of active irreversible reactions in v. Analogously, $\mathrm{rev.supp}(v) := \mathrm{supp}(v) \cap \mathrm{Rev}$ denotes the *reversible support* of v. We call $v \in C$ *reversible*, if $\mathrm{supp}(v) = \mathrm{rev.supp}(v)$, i.e., if all reactions that are active in v are reversible and v is called *irreversible* if $\mathrm{irr.supp}(v) \neq \emptyset$.

The following basic properties of elementary flux modes were first described by Schuster and Hilgetag (1994) and will be used throughout this thesis. The presented proofs have been adapted to fit our notation.

Proposition 2.5.1 (Schuster and Hilgetag, 1994)**.** *Let $e \in C$ be an EFM in the flux cone of a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$. If a steady-state flux vector $v \in C$ has support equal to the support of e, i.e.* $\mathrm{supp}(v) = \mathrm{supp}(e)$*, then $v = \lambda e$ for some $\lambda \neq 0 \in \mathbb{R}$.*

Proof. Let $U := \mathrm{supp}(e) = \mathrm{supp}(v) \neq \emptyset$. If $\mathrm{irr.supp}(v) = \mathrm{irr.supp}(e) = \emptyset$, choose $k \in U$ and define $\lambda = \frac{v_k}{e_k}$. Otherwise, if $\mathrm{irr.supp}(v) \neq \emptyset$, define $\lambda = \min\{\frac{v_i}{e_i} \mid i \in \mathrm{irr.supp}(v)\}$ and choose $k \in U$ such that $\lambda = \frac{v_k}{e_k}$. Define $v' = v - \lambda e$. Then $v' \in C$ is a steady-state flux vector because $Sv' = S(v - \lambda e) = Sv - S(\lambda e) = Sv - \lambda Se = 0$ and $v'_i = v_i - \lambda e_i \geq v_i - \frac{v_i}{e_i} e_i = 0$, for all $i \in \mathrm{irr.supp}(v)$. Since $v'_k = v_k - \frac{v_k}{e_k} e_k = 0$, we have $\mathrm{supp}(v') \subsetneq \mathrm{supp}(v) = \mathrm{supp}(e) = U$. Since e is an EFM, we get from equation (2.3) that $\mathrm{supp}(v') = \emptyset$ and thus $v' = 0$. We conclude $0 = v' = v - \lambda e$ and thus $v = \lambda e$. □

Because of Prop. 2.5.1, we say that EFMs are *uniquely determined by their support* (up to scaling). The support of every vector $v \in C$ corresponds to an element of the finite power set $2^{\mathcal{R}}$, which has $2^{|\mathcal{R}|}$ elements. This is a trivial upper bound on the number of EFMs with distinct support a metabolic network can have.

Corollary 2.5.2. *Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$ be a metabolic network and let $U := \mathrm{supp}(e)$ be the support of an EFM e of $\mathcal{N}$. Then for every solution $v \neq 0$ of the system of linear equations*

$$\begin{aligned} Sv &= 0, \\ v_i &= 0 \text{ for } i \in \mathcal{R} \setminus U, \end{aligned} \tag{2.4}$$

there exists $\lambda \in \mathbb{R} \setminus \{0\}$ such that $\lambda v = e$. More precisely, the set of solutions to (2.4) has dimension 1.

Proof. By construction every solution $v \in \mathbb{R}^n$ of (2.4) has support $\mathrm{supp}(v) \subseteq U$. Since $v \neq 0$, we get $\mathrm{supp}(v) = U$ by equation (2.3) and by Prop. 2.5.1 there exists $\lambda \neq 0 \in \mathbb{R}$ such that $\lambda v = e$. □

Specifically, given the support U of an EFM, it suffices to solve (2.4) to find an EFM e with $\text{supp}(e) = U$.

Next we present a test that can be applied to check whether a given flux vector v is an EFM of a metabolic network.

Corollary 2.5.3 (Jevremovic *et al.,* 2010; Urbanczik and Wagner, 2005)**.** *A flux vector $v \neq 0$ in the flux cone C of a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$ is an EFM of $\mathcal{N}$ if and only if*

$$\text{rank}(S_{\star,\text{supp}(v)}) = |\text{supp}(v)| - 1, \tag{2.5}$$

where $\text{rank}(S_{\star,\text{supp}(v)})$ *is the rank of the submatrix of S, whose columns correspond to reactions that are active in v.*

Proof. "⇒" Let L be the set of solutions of (2.4) where $U := \text{supp}(e)$ is the support of an EFM e of $\mathcal{N}$. Note that (2.4) is equivalent to $S_{\star,U} v_U = 0$ and therefore, $1 = \dim(L) = |U| - \text{rank}(S_{\star,U})$.

"⇐" To see that v is an EFM, suppose there exists $v' \in C$ with $\text{supp}(v') \subsetneq \text{supp}(v)$. But then v' is a solution of (2.4) and $\lambda v' = v$ implies $\text{supp}(v') = \text{supp}(v)$. □

We will refer to equation (2.5) as *rank test.*

Proposition 2.5.4 (Schuster and Hilgetag, 1994)**.** *A flux vector $v \in C \setminus \{0\}$ is an EFM of the metabolic network $\mathcal{N}$ with flux cone C if and only if there exist no $v^1, v^2 \in C \setminus \{0\}$ such that*

$$\text{supp}(v^1), \text{supp}(v^2) \subsetneq \text{supp}(v) \text{ and } v = v^1 + v^2.$$

Proof. "⇒": Suppose the opposite, i.e. $v \in C \setminus \{0\}$ and $v = v^1 + v^2$ for some $v^1, v^2 \in C \setminus \{0\}$ with $\text{supp}(v^1), \text{supp}(v^2) \subsetneq \text{supp}(v)$. But then equation (2.3) implies that v is not an EFM in C.
"⇐": Suppose $v \in C$ and there exist no $v^1, v^2 \in C \setminus \{0\}$ with

$$\text{supp}(v^1), \text{supp}(v^2) \subsetneq \text{supp}(v) \text{ and } v = v^1 + v^2,$$

but v is not an EFM in C. Because of equation (2.3) there exists $v' \in C \setminus \{0\}$ with inclusionwise smaller support than v, i.e., $\text{supp}(v') \subsetneq \text{supp}(v)$. If $\text{irr.supp}(v') = \emptyset$, choose any $k \in \text{supp}(v')$ and define $\lambda = \frac{v_k}{v'_k}$. Otherwise if $\text{irr.supp}(v') \neq \emptyset$, define $\lambda = \min\{\frac{v_i}{v'_i} \mid i \in \text{irr.supp}(v')\}$ and choose $k \in \text{supp}(v')$ such that $\lambda = \frac{v_k}{v'_k}$. Define $v^1 = \lambda v^1$ and $v^2 = v - v^1$. Then $v^1, v^2 \in C \setminus \{0\}$, $\text{supp}(v^1), \text{supp}(v^2) \subsetneq \text{supp}(v)$ and $v = v^1 + v^2$ contradicts our assumption. □

In Chapter 5, we will discuss how EFMs can be written as positive combinations of other EFMs that do not have inclusionwise smaller support.

Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$ be a metabolic network with flux cone C. Since EFMs of $\mathcal{N}$ are uniquely determined by their support, we can define $\mathcal{U}$ as the set containing all subsets of $\mathcal{R}$ that are the support of an EFM of $\mathcal{N}$ (i.e. for every EFM e of $\mathcal{N}$, $\text{supp}(e) \in \mathcal{U}$). We can split the set $\mathcal{U}$ into two disjoint subsets $\mathcal{U}^r, \mathcal{U}^i$ such that $\mathcal{U}^r$ contains the supports of all reversible EFMs, while $\mathcal{U}^i$ contains the supports of all irreversible EFMs and we get $\mathcal{U} = \mathcal{U}^r \dot{\cup} \mathcal{U}^i$.

For each element $U \in \mathcal{U}^i$, choose an EFM e of $\mathcal{N}$ such that $\text{supp}(e) = U$ and for each element $U' \in \mathcal{U}^r$ choose two EFMs e^+, e^- of $\mathcal{N}$ such that $e^+ = -e^-$. We define $\mathcal{E}$ as the set containing all EFMs of $\mathcal{N}$ chosen this way and call it a *representative set of the EFMs* of $\mathcal{N}$. Note that every EFM in the flux cone C of a metabolic network is a positive multiple of an EFM in $\mathcal{E}$, i.e., if e is an EFM of a metabolic network $\mathcal{N}$ with a representative set $\mathcal{E}$ of the EFMs of $\mathcal{N}$, then there exists $\lambda > 0 \in \mathbb{R}$ such that $\lambda e \in \mathcal{E}$.

Proposition 2.5.5 (Schuster and Hilgetag, 1994)**.** *Let $\mathcal{N}$ be a metabolic network where $\mathcal{E}$ is a representative set of the EFMs of $\mathcal{N}$. Any $v \in C$ can be expressed as a conic combination of EFMs in $\mathcal{E}$:*

$$v = \sum_{e \in \mathcal{E}} \lambda_e e, \text{ for some } \lambda_e \geq 0 \in \mathbb{R}.$$

Proof. If $v \in C \setminus \{0\}$ is not an EFM, by Prop. 2.5.4 it can be written as

$$v = v^1 + v^2, \text{ with } \text{supp}(v^1), \text{supp}(v^2) \subsetneq \text{supp}(v).$$

This can be repeated for v^1 and v^2 and the resulting vectors until only EFMs remain in the conic combination, since supports of the vectors are finite subsets of $\mathcal{R}$ and become inclusionwise smaller in each repetition. □

We defined a representative set $\mathcal{E}$ of the EFMs of a metabolic network to contain both orientations of reversible EFMs. With this definition, Prop. 2.5.5 is equivalent to $C = \text{cone}(\mathcal{E})$, where C is the flux cone of a metabolic network with representative set of the EFMs $\mathcal{E}$.

Note that $\mathcal{E}$ is not a minimal generating set of the flux cone (for more details, we refer to Rezola *et al.* (2011)). A set $\mathcal{W}$ of EFMs is called a *minimal set of elementary modes* (MEMo) of a metabolic network $\mathcal{N}$, if it is an *inclusionwise minimal conic generating set* of the flux cone C of $\mathcal{N}$, i.e., removing any element from $\mathcal{W}$ leads to a set that is not a conic generating set of the flux cone (Röhl and Bockmayr, 2019). By our definition, a set of EFMs is a MEMo of a metabolic network $\mathcal{N}$ with flux cone C if and only if

$$\text{cone}(\mathcal{W}) = C \text{ and } \forall e \in \mathcal{W} : \text{cone}(\mathcal{W} \setminus \{e\}) \neq C.$$

The main advantage of MEMos is the fact that they can be computed faster than a representative set of the EFMs. We refer to (Röhl and Bockmayr, 2019) for an in-depth explanation of how a MEMo can be obtained without having to enumerate a representative set of all EFMs. It should be noted that in (Röhl and Bockmayr, 2019) MEMos are defined as an inclusionwise minimal set $\mathcal{U} \subseteq \mathcal{E}_\mathcal{N}$ such that every vector $v \in C$ can be represented as a linear combination

$$v = \sum_{e \in \mathcal{U} \cap \mathcal{E}_\mathcal{N}^{\text{Rev}}} \lambda_e e + \sum_{f \in \mathcal{U} \cap \mathcal{E}_\mathcal{N}^{\text{Irr}}} \lambda_f f,$$

for some $\lambda_e, \lambda_f \in \mathbb{R}$, with $\lambda_f \geq 0$, for all irreversible $f \in \mathcal{U} \cap \mathcal{E}_\mathcal{N}^{\text{Irr}}$. In their notation $\mathcal{E}_\mathcal{N}$ is the finite set of EFMs (finite because EFMs with equal support are considered to be the same), $\mathcal{E}_\mathcal{N}^{\text{Rev}}$ is the subset of reversible EFMs, and $\mathcal{E}_\mathcal{N}^{\text{Irr}}$ is the subset of irreversible EFMs. Our definition of a representative set of the

EFMs of a metabolic network was chosen so that we only need to consider non-negative (i.e conic) combinations. Doing so also allowed for a more compact definition of MEMos. Our MEMos are not exactly the same as in (Röhl and Bockmayr, 2019), since they contain two oppositely oriented representatives for the support of each reversible EFM.

2.6 EFM enumeration

In this section, we briefly discuss some commonly used methods to enumerate a representative set of the EFMs of a metabolic network. Most of these are based on algorithms that enumerate extreme rays of a pointed cone. While the flux cone of a metabolic network is not pointed in many cases (examples are presented in Chapter 4), it has been shown that the extreme rays of an augmented pointed cone correspond to EFMs of the original cone (Schuster and Hilgetag, 1994). This augmented cone is obtained by replacing every reversible reaction with two oppositely oriented copies that are irreversible. This augmented cone is always pointed and methods like the double description method (Fukuda and Prodon, 1996) can be used to enumerate the extreme rays of this pointed cone.

The most prominent implementation of the double description method to enumerate EFMs, `efmtool`, was introduced by Terzer (2009). Enumeration methods based on double description have the disadvantage that they are *pass-or-fail* methods, i.e., either a full representative set of EFMs is enumerated or no EFMs at all. A more recent alternative to double description for enumeration of EFMs by enumerating extreme rays of an augmented cone uses lexicographic reverse search (Buchner and Zanghellini, 2021).

The number of EFMs in a representative set for genome-scale metabolic networks is generally too high to be enumerated by these methods. To overcome this issue, de Figueiredo *et al.* (2009) developed a method to determine shortest EFMs (with the fewest number of active reactions) by repeatedly solv-

ing *mixed-integer linear programs* (MILP). This approach has the advantage that after each iteration one EFM is returned. Although it is outperformed by the extreme ray enumeration methods if they terminate, it can be used to enumerate subsets of EFMs in genome-scale metabolic networks where the other methods fail.

Chapter 3

Geometry of the flux cone and elementary flux modes

The results in this chapter were obtained in a collaboration with Martin Henk and Alexander Bockmayr and were published in the Journal of Mathematical Biology (Wieder, Henk and Bockmayr, 2023)(Reproduced with permission from Springer Nature).

3.1 Introduction

Constraint-based analysis of metabolic networks is an important area in computational biology (Bordbar *et al.,* 2014; Fang, Lloyd and Palsson, 2020). The stoichiometric and thermodynamic constraints that must hold in a metabolic network at steady-state define the steady-state flux cone, which comprises all feasible flux distributions over the network at steady-state.

An important concept to analyze the flux cone in a mathematically and biologically meaningful way are elementary flux modes (EFMs) (Schuster and Hilgetag, 1994; Schuster, Hilgetag, *et al.,* 2002). A representative the set of the EFMs of a metabolic network provides an inner description of the flux cone by a finite set of generating vectors (Gagneur and Klamt, 2004; Wagner and Urbanczik, 2005; Larhlimi and Bockmayr, 2008; Jevremović and Boley, 2013). From a mathematical point of view, a representative set of the EFMs is not necessarily a minimal generating set, except for special cases (for example, when all reactions are irreversible). Even for small networks, the cardinality of a representative set of elementary flux modes may be very large.

Larhlimi and Bockmayr (2009) introduced metabolic behaviors and studied outer descriptions of the flux cone based on *minimal metabolic behaviors*

(MMBs), which are in a one-to-one correspondence with the minimal proper faces of the flux cone. Röhl and Bockmayr (2019) introduced the concept of a minimal set of elementary flux modes (MEMos, cf. Chapter 2) and gave an algorithm to compute such a set.

The goal of this chapter is to get a deeper understanding of the structure of EFMs by further studying their geometric properties. In Section 3.2, we take a closer look at geometric properties of polyhedral cones in general and extend our knowledge beyond the basic properties introduced in Section 2.2. In Section 3.3, we discuss the effects of redundancies in the description of a flux cone of a metabolic network. In Sections 3.5 and 3.6, we study the faces of flux cones and introduce the degree of a flux vector as the dimension of the inclusionwise minimal face containing it. Intuitively, the degree of a flux vector can be viewed as a measure of how close (combinatorially) it is to the relative interior of the flux cone. If a flux cone is pointed, its extreme rays are elementary flux modes and form a MEMo. EFMs that are not extreme rays can still occur in pointed cones and can be identified by having a degree larger than 1. EFMs in a MEMo of a non-pointed cone have degree equal to the dimension of the lineality space or equal to the dimension of a minimal proper face. Here we study the distribution of EFMs among all faces of the flux cone and observe that EFMs tend to appear primarily in the lower-dimensional faces of flux cones.

Finally, in Section 3.7, we generalize a result from (Larhlimi and Bockmayr, 2009) and show that higher-dimensional faces of the flux cone can be characterized by metabolic behaviors.

3.2 Geometry of polyhedral cones

In this section, we further investigate polyhedral cones and extend the properties mentioned in Section 2.2. We explore geometric properties of flux cones of metabolic networks that will be applied in the remaining sections of this

chapter.

The *face lattice* of a polyhedral cone C is the partially ordered set $L(C)$ of all faces of C, partially ordered by set inclusion (Henk, Richter-Gebert and Ziegler, 2017; Ziegler, 1995). Two polyhedral cones C, C' will be called *combinatorially equivalent* if there is a bijection from $L(C)$ to $L(C')$ that preserves the inclusion relation. The *combinatorial type* of a polyhedral cone is the equivalence class under combinatorial equivalence.

Proposition 3.2.1. *Let $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ be a polyhedral cone and $z \in C$. Furthermore, let $A_z^=$ be the submatrix of A whose rows correspond to the inequalities in $Ax \geq 0$ that are fulfilled with equality by z. Finally, let F be the face of C defined by $F = \{x \in C \mid A_z^= x = 0\}$. Then*

i) *F is the inclusionwise minimal face of C containing z,*

ii) *$\dim(F) = n - \operatorname{rank}(A_z^=)$, and*

iii) *$z \in \operatorname{relint}(F)$.*

Proof. For $x \in \mathbb{R}^n$, define $I(x) = \{i \in \{1, \ldots, m\} \mid A_{i,\star} x = 0\}$, where $A_{i,\star}$ is the i-th row in A. Let $F' = \{x \in C \mid A_{I,\star} x = 0\}$ be a face of C containing z. Then $A_{I,\star} z = 0$ and thus $I \subseteq I(z)$, which implies $F = \{x \in C \mid A_{I(z),\star} x = 0\} \subseteq F'$ and statement i) follows.

For $x \in F$, we have $I(z) \subseteq I(x)$. Therefore, $I(z)$ has the minimal number of elements among $I(x)$, where $x \in F$. The statements ii) and iii) now follow from Prop. 4.3 in (Lauritzen, 2013) and its proof. □

If C is a cone with $\dim(\text{lin. space}(C)) = t \geq 0$, a face of dimension $t + 1$ is called a *minimal proper face* of C. For a pointed cone C, the minimal proper faces are the 1-faces, which are called the *extreme rays* of C. Equivalently, $\operatorname{cone}(\{r\}) \subseteq C, r \neq 0$, is an extreme ray of C if and only if $r = x + y$ implies $x, y \in \operatorname{cone}(\{r\})$, for all $x, y \in C$.

The *Minkowski sum* of two sets X and Y is defined as $X+Y = \{x+y \mid x \in X, y \in Y\}$. The next result states that any polyhedral cone can be decomposed into a Minkowski sum of its lineality space and a pointed cone.

Proposition 3.2.2. *Let $C \subseteq \mathbb{R}^n$ be a polyhedral cone, $L = \text{lin. space}(C)$. Let $G^1, \ldots, G^s, s \geq 0$, be the distinct minimal proper faces of C and $g^i \in G^i \setminus L$, for $i = 1, \ldots, s$. Let $P = \text{cone}(\{g^1, \ldots, g^s\})$. Then*

i) *P is a pointed cone and its extreme rays are $\text{cone}(\{g^1\}), \ldots, \text{cone}(\{g^s\})$,*

ii) *$C = L+P = L+\text{cone}(\{g^1, \ldots, g^s\})$, $L \cap P = \{0\}$ and if $L \cap \text{lin}(P) = \{0\}$ then $\dim(C) = \dim(L) + \dim(P)$.*

Proof. i) By definition P is a finitely generated cone. Assume that P is not pointed. Then there exist $\lambda_i \geq 0$, $i = 1, \ldots, s$, not all equal to zero, such that $0 = \sum_{i=1}^{s} \lambda_i g^i$. Hence, there exists $j \in \{1, \ldots, s\}$ such that $-g^j \in P \subseteq C$ and so $g^j \in L$, contradicting our choice.

To see that $g^1, \ldots, g^s$ define extreme rays of P, assume without loss of generality that $\text{cone}(\{g^1\})$ is not an extreme ray of P. Then we can find $\mu_i \geq 0$, $2 \leq i \leq s$, not all equal to zero, such that $g^1 = \sum_{i=2}^{s} \mu_i g^i$. As G^1 is a face, there exists a supporting hyperplane $H_a := \{x \in \mathbb{R}^n \mid ax = 0\}, a \in \mathbb{R}^n \setminus \{0\}$, such that $G^1 = H_a \cap C$ and $ax > 0$ for all $x \in C \setminus G^1$. Thus, from $0 = ag^1 = \sum_{i=2}^{s} \mu_i a g^i$ we conclude that $ag^k = 0$ for some $k \in \{2, \ldots, s\}$. Therefore $g^k \in G^1$ and $G^k \subseteq G^1$, which leads to $G^k = G^1$, because G^1 is a minimal proper face. But then $G^1, \ldots, G^s$ are not distinct, which is a contradiction.

ii) By Theorem 8.5 in (Schrijver, 1998), we have $C = L + P$. Since L is a face of C, there exists $a \in \mathbb{R}^n \setminus \{0\}$ such that $ax = 0$ for all $x \in L$ and $ax > 0$ for all $x \in C \setminus L$. From $ag^i > 0, i = 1, \ldots, s$, we get $ax > 0$, for all $x \in P \setminus \{0\}$, hence $L \cap P = \{0\}$. If $L \cap \text{lin}(P) = \{0\}$, we have $\text{lin}(C) = L \oplus \text{lin}(P)$, where $L \oplus \text{lin}(P)$ is the direct sum of the vector spaces L and $\text{lin}(P)$ and thus, $\dim(C) = \dim(L) + \dim(P)$. □

The combinatorial type of the pointed cone $P = \text{cone}(\{g^1, \ldots, g^s\})$ in Prop. 3.2.2 is (generally) not uniquely determined. However, if we choose $g^1, \ldots, g^s$ such that $L \cap \text{lin}(P) = \{0\}$, i.e., all the g^i are contained in some linear subspace L' complementary to L, then the combinatorial type of P is independent of the choice of the g^i from the minimal proper faces. Observe that for any complementary space L' of L, $L' \cap G^i$ is a ray, i.e., $L' \cap G^i = \text{cone}(\{g^i\})$ for some $g^i \in L' \cap G^i$.

Proposition 3.2.3. *Let $C \subseteq \mathbb{R}^n$ be a polyhedral cone, $L = \text{lin. space}(C)$, and let P_1, P_2 be pointed cones with $L + P_1 = C = L + P_2$ and $L \cap \text{lin}(P_1) = \{0\} = L \cap \text{lin}(P_2)$. Then P_1 and P_2 are combinatorially equivalent.*

Proof. Without loss of generality, let $\dim(C) = n$ and $\dim(L) = t$. With $L'_j := \text{lin}(P_j)$, by Prop. 3.2.2, we have $\dim(L'_j) = n - t$ and $C \cap L'_j = P_j$, for $j = 1, 2$. Let $u^1, \ldots, u^{n-t}$ be a basis of L'_1. As also L'_2 is a complement of L, there exist uniquely determined $v^1, \ldots, v^{n-t} \in L'_2$, $w^1, \ldots, w^{n-t} \in L$ such that $u^i = v^i + w^i$, $1 \leq i \leq n-t$. Now, let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the invertible linear map with

$$T(u) = u, \ \ \forall u \in L, \quad T(u^i) = v^i, \ \ 1 \leq i \leq n - t.$$

We get $T(C) = C$. To see this, let $y = u + w \in C$ with $u \in L'_1$ and $w \in L$. We may write

$$y = \sum_{i=1}^{n-t} \lambda_i u^i + w, \ \text{ for some } \lambda_1, \ldots, \lambda_{n-t} \in \mathbb{R}.$$

Thus,

$$\begin{aligned} T(y) = \sum_{i=1}^{n-t} \lambda_i v^i + w &= \sum_{i=1}^{n-t} \lambda_i (u^i - w^i) + (y - \sum_{i=1}^{n-t} \lambda_i u^i) \\ &= y - \sum_{i=1}^{n-t} \lambda_i w^i \in C + L = C, \end{aligned}$$

and vice versa. We conclude

$$T(P_1) = T(L'_1 \cap C) = T(L'_1) \cap T(C) = L'_2 \cap C = P_2.$$

Hence, P_1 and P_2 are affinely and thereby combinatorially equivalent. □

We point out that the relative interior of a polyhedral cone can easily be described by looking at implicit equalities in $Ax \geq 0$:

Proposition 3.2.4. *Let* $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\} = \{x \in \mathbb{R}^n \mid A^{=}x = 0, A^{+}x \geq 0\}$ *be a polyhedral cone. Then*

$$\text{relint}(C) = \{x \in \mathbb{R}^n \mid A^{=}x = 0, A^{+}x > 0\}.$$

Proof. If $x \in C$ with $A^{+}x > 0$, then for any $y \in \text{lin}(C) = \{y \in \mathbb{R}^n \mid A^{=}y = 0\}$ there exists $\epsilon > 0$ such that $A^{+}(x + \epsilon y) > 0$. Hence, $x \in \text{relint}(C)$. Conversely, let $x \in \text{relint}(C)$ and let a be an arbitrary row of A^{+}. By definition of A^{+} there exists $z \in C$ with $az > 0$. As $x \in \text{relint}(C)$, there exists $\epsilon > 0$ such $x - \epsilon z \in C$ and so $a(x - \epsilon z) \geq 0$. Thus, $ax > 0$. □

3.3 Redundancy in flux cones

In this section we will discuss properties of flux cones in the context of metabolic networks. Namely, we will establish the relationship between implicit equalities and blocked irreversible reactions, as well as redundant irreversibility constraints.

Implicit equalities. The implicit equalities (cf. Section 2.2) in the definition of a flux cone C (cf. Section 2.4) include all steady-state constraints $Sv = 0$. If any of the irreversibility constraints $v_j \geq 0, j \in \text{Irr}$, is an implicit equality, the corresponding reaction $j \in \text{Irr}$ is *blocked*, i.e., $v_j = 0$, for all $v \in C$. For some of the results in this chapter, we will assume that there are no implicit equalities in $v_{\text{Irr}} \geq 0$, or equivalently that there are no blocked irreversible reactions. In general, blocked reactions can be determined by solving two linear optimization problems for each reaction:

$$M = \max\{v_i \mid Sv = 0, v_{\text{Irr}} \geq 0\} \text{ and}$$
$$m = \min\{v_i \mid Sv = 0, v_{\text{Irr}} \geq 0\}.$$

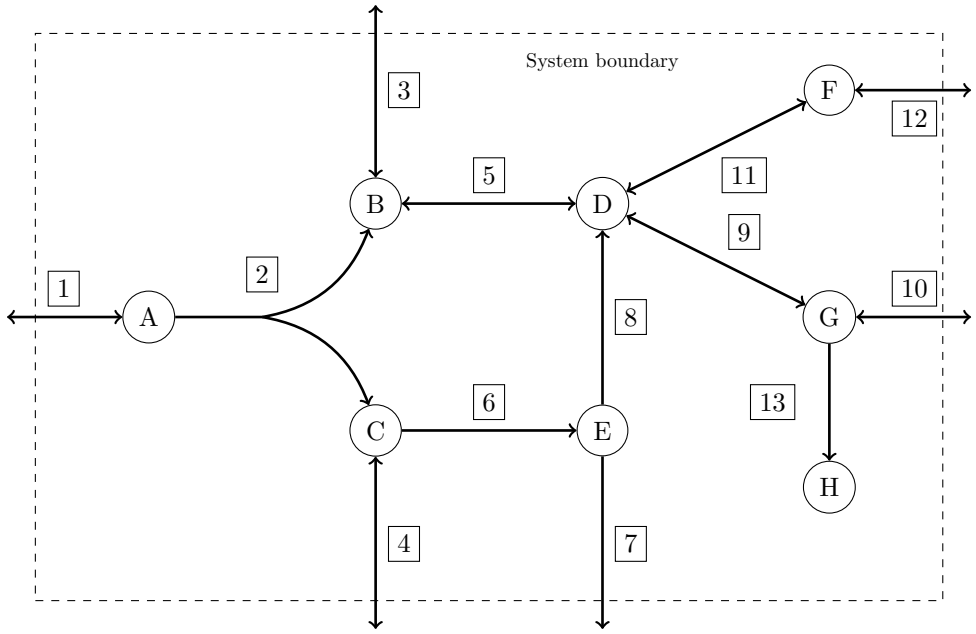

Figure 3.1: Modification of the metabolic network in Figure 2.4.

If $M = m = 0$ the flux through reaction i is always 0 and thus, the reaction is blocked. Note that for irreversible reactions it suffices to solve one linear optimization problem, namely the maximization.

Redundant inequalities. If in (2.2) one of the irreversibility constraints $v_j \geq 0, j \in \text{Irr}$, is redundant, the corresponding reaction j can be shifted from the set Irr of irreversible reactions to the set Rev of reversible reactions, without changing the flux cone C. The constraint $v_j \geq 0$ is then implied by the remaining constraints.

Example 3.3.1. *Consider a slightly adjusted version of the example metabolic network in Figure 2.4.*

The metabolic network in Figure 3.1 has the set of metabolites $\mathcal{M} = \{\text{A}, \text{B}, \ldots, \text{G}, \text{H}\}$*, the set of reversible reactions* $\text{Rev} = \{1, 3, 4, 5, 9, 10, 11, 12\}$*, and the set of irreversible reactions* $\text{Irr} = \{2, 6, 7, 8, 13\}$*. For simplicity, all stoichiometric coefficients are assumed to be 0,1, or -1. The stoichiometric*

matrix is

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where we assume that reversible reactions are oriented from left to right and from top to bottom.

In the metabolic network in Figure 3.1, the irreversible reaction 13 *is blocked, i.e.,* $v_{13} = 0$*, for all* $v \in C$*, because there is no reaction consuming metabolite* H*.*

The irreversibility constraint $v_6 \geq 0$ *is redundant. There can be no flux from metabolite* E *to metabolite* C *because there is no reaction producing* E*, other than reaction* 6*.*

The reversible reaction 1 *cannot carry flux from right to left. Thus, reaction* 1 *could be added to the set* Irr *of irreversible reactions without changing the associated flux cone* C*, but then the inequality* $v_1 \geq 0$ *would be redundant.*

If redundant inequalities are removed from the description of a flux cone, the resulting irredundant description is generally not unique, as it depends on the order in which redundant constraints are removed.

Proposition 3.3.2. *Let* $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ *be a flux cone such that none of the inequalities* $v_j \geq 0, j \in \mathrm{Irr}$*, is redundant or an implicit equality. Then* C *has exactly* $|\mathrm{Irr}|$ *facets and each facet* F *has the representation*

$$F = \{v \in C \mid v_j = 0\}, \text{ for some } j \in \mathrm{Irr}\,. \tag{3.1}$$

Proof. If there are no implicit equalities in $v_{\mathrm{Irr}} \geq 0$ and A is given by (2.1), then $A^{=} = \begin{pmatrix} S \\ -S \end{pmatrix}$ and $A^{+} = I_{\mathrm{Irr},\star}$. Since there are no redundant inequalities in $v_{\mathrm{Irr}} \geq 0$, the result follows from Theorem 8.1 in (Schrijver, 1998). □

3.4 Illustrative examples

To illustrate the theoretical results in the following sections through concrete examples, we will use the metabolic networks `Pyruvate` and `Pentose Phosphate Pathway` from the KEGG database (`https://www.genome.jp/kegg/pathway.html`, (Kanehisa and Goto, 2000)) and `Escherichia coli str. K-12 substr. MG1655 (E.coli core)` from the BiGG database (King *et al.,* 2016), where we removed the biomass reaction. The characteristics of these networks are summarized in Table 3.1. EFMs were computed with `efmtool` (`https://csb.ethz.ch/tools/software/efmtool.html`, (Terzer, 2009)).

3.5 Faces of the flux cone and metabolic behaviors

Given a metabolic network $\mathcal{N}$ with flux cone C, a *metabolic behavior* (Larhlimi and Bockmayr, 2009) of C is a nonempty set of irreversible reactions $D \subseteq \mathrm{Irr}$ with $D = \mathrm{irr.supp}(v)$, for some $v \in C$. A *minimal metabolic behavior* (MMB) is a metabolic behavior D for which there is no other metabolic behavior $D' \subsetneq D$. Larhlimi and Bockmayr (2009) have shown that minimal metabolic behaviors are in a 1-1 correspondence with the minimal proper faces of the flux cone C. In particular, if G is a minimal proper face and L the lineality space of C, then all flux vectors $v \in G \setminus L$ have the same irreversible support $D_G = \mathrm{irr.supp}(v)$, which is a minimal metabolic behavior.

	E. coli core	Pentose Phosphate	Pyruvate
$(m, n) = (\|\mathcal{M}\|, \|\mathcal{R}\|)$	(72, 94)	(34, 57)	(28, 81)
|Irr|	48	19	40
|Rev|	46	38	41
$\text{rank}(S)$	67	34	28
$n - \text{rank}(S)$	27	23	53
$\dim(C)$	23	23	53
$t = \dim(L)$	0	8	16
$\dim(P)$	23	15	37
|Facets|	39	17	37
|blocked irr|	8	0	0
|blocked rev|	2	0	0
|EFMs|	16673	5180	47854
|MMBs|	1421	19	37

Table 3.1: Characteristics of the three example networks. $|\mathcal{M}|$ and $|\mathcal{R}|$ denote the number of metabolites resp. reactions, which correspond to the number of rows resp. columns of the stoichiometric matrix. |Irr| and |Rev| denote the number of irreversible resp. reversible reactions of the network. $\text{rank}(S)$ is the rank of the stoichiometric matrix. The flux cone $C = L + P$ is the Minkowski sum of the lineality space L and a pointed cone P, with $\dim(C) = \dim(L) + \dim(P)$. |Facets| is the number of facets of the flux cone, which is equal to the number of irreversibility constraints if none of these is redundant or an implicit equality. |blocked irr| resp. |blocked rev| describe the number of blocked irreversible resp. blocked reversible reactions. |EFMs| is the number of EFMs and |MMBs| the number of minimal metabolic behaviors.

Proposition 3.5.1. *Let C be the flux cone of a metabolic network. Then each metabolic behavior is the union of MMBs.*

Proof. Let $\emptyset \neq D \subseteq \mathrm{Irr}$ be a metabolic behavior and let $v \in C$ with $D = \mathrm{irr.supp}(v)$. Let L be the lineality space and $G^1, \ldots, G^s, s \geq 0$, be the minimal proper faces of C. Since $D \neq \emptyset$ and $\mathrm{irr.supp}(l) = \emptyset$, for all $l \in L$, we have $C \neq L$. Thus C has at least one minimal proper face and $s \geq 1$. By Prop. 3.2.2, $v = l + \sum_{i=1}^{s} \lambda_i g^i$, for some $l \in L$ and $\lambda_i \geq 0, g^i \in G^i \setminus L$, for $i = 1, \ldots, s$. It follows $\mathrm{irr.supp}(v) = \bigcup_{\lambda_i > 0} \mathrm{irr.supp}(g^i) = \bigcup_{\lambda_i > 0} D^i$, where $D^i = \mathrm{irr.supp}(g^i)$ is the MMB of the minimal proper face G^i, for $i = 1, \ldots, s$. □

Next, we generalize the characterization of minimal proper faces by MMBs (Larhlimi and Bockmayr, 2009) to higher-dimensional faces.

Proposition 3.5.2. *Let C be the flux cone of a metabolic network and let F be a face of C. Then all $v \in \mathrm{relint}(F)$ have the same irreversible support or equivalently share the same metabolic behavior.*

Proof. Let $v, w \in \mathrm{relint}(F)$. Assume w.l.o.g. that there exists $j \in \mathrm{irr.supp}(w) \setminus \mathrm{irr.supp}(v)$. Then $v_j = 0$, but $w_j > 0$, and hence for any $\lambda > 1$, we have $\lambda v + (1-\lambda) w \notin C$. However, since $v, w \in \mathrm{relint}(F)$, we know $\lambda v + (1-\lambda) w \in F$, for some $\lambda > 1$. This shows $\mathrm{irr.supp}(v) = \mathrm{irr.supp}(w)$, which implies the statement. □

Example 3.5.3. *Consider the network in Figure 2.4. If we remove the redundant irreversibility constraint $v_6 \geq 0$ and assume $6 \in \mathrm{Rev}$, the MMBs of the network are $\{2\}, \{7\}, \{8\}$ (if $6 \in \mathrm{Irr}$, the MMBs are $\{2\}, \{6, 7\}, \{6, 8\}$). The face lattice, together with the supports of the EFMs contained in each face, is shown in Figure 3.2.*

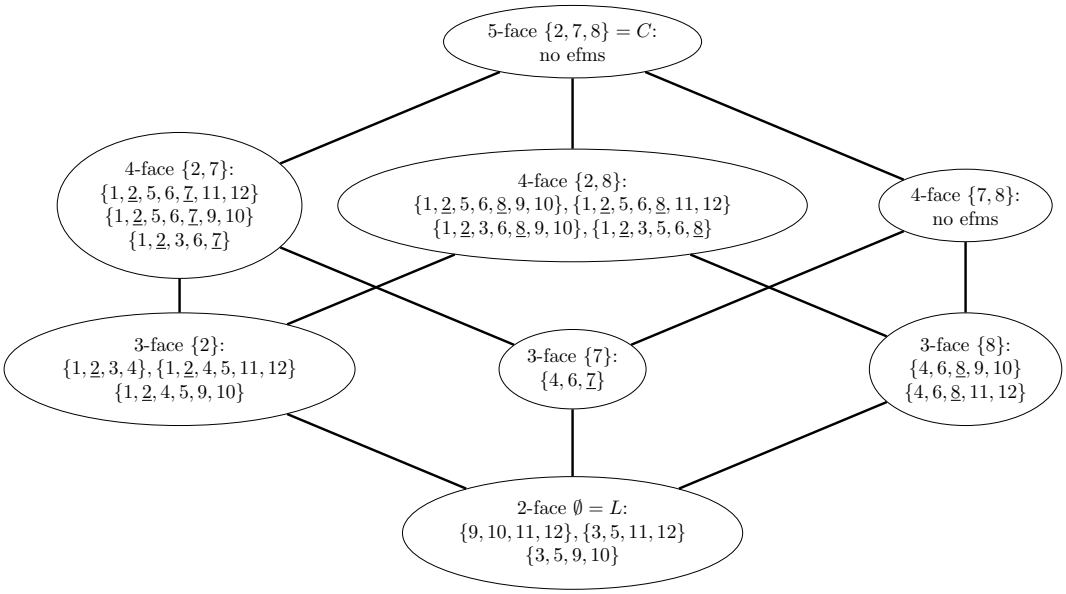

Figure 3.2: Face lattice of the network in Figure 2.4. Each node represents a face of the flux cone together with the corresponding metabolic behavior and the supports of the EFMs contained in that face. The active irreversible reactions are underlined. Edges connecting the nodes indicate the inclusion of lower-dimensional in higher-dimensional faces. The only 2-face is the lineality space L and the only 5-face is the entire flux cone C.

3.6 The degree of flux vectors

Representative sets of EFMs in a metabolic network tend to be very large and can be of exponential size in the number of reactions. Hence, there is a need to elucidate the structure of this set. In this section, we define a property of flux vectors, namely their degree and apply it to EFMs to establish such a structure. For the remainder of this section, let C be the flux cone of a metabolic network.

We define the *degree* $\deg(v)$ of a flux vector $v \in C$ as the dimension of the inclusionwise minimal face of C containing v, which for $v \neq 0$ is the unique face F of C with $v \in \operatorname{relint}(F)$. By Prop. 3.2.1, we have

$$\deg(v) = n - \operatorname{rank}(A_v^=), \quad \text{where} \quad A_v^= = \begin{pmatrix} S \\ -S \\ I_{\mathrm{Irr} \setminus \mathrm{irr.supp}(v),\star} \end{pmatrix}. \tag{3.2}$$

It follows that flux vectors in the lineality space L of C have degree $\dim(L)$ and flux vectors in minimal proper faces have degree $\dim(L)+1$. A flux vector in the relative interior of C has degree $\dim(C)$.

A widely used technique in metabolic network analysis is to split reversible reactions into two irreversible reactions (Clarke, 1980; Schilling, Letscher and Palsson, 2000; Papin *et al.,* 2004; Gagneur and Klamt, 2004; Röhl and Bockmayr, 2019). In particular, this is applied in algorithms for EFM enumeration such as `efmtool` (Terzer, 2009) and `EFMlrs` (Buchner and Zanghellini, 2021). In these algorithms, splitting all reversible reactions results in a pointed cone C' in a higher-dimensional space. The EFMs of the original flux cone C correspond to extreme rays in the reconfigured cone C' (Gagneur and Klamt, 2004), which all have degree 1 in C'. Therefore, the degree of an EFM as defined by (3.2) has to be determined in the original flux cone C and cannot be determined in the reconfigured cone C'.

The support of a vector is its set of active reactions. We can further

characterize flux vectors in the relative interior of C by their sets of active irreversible reactions, i.e. their irreversible support. The following proposition shows that in a flux vector within the relative interior of C, all irreversible reactions are active.

Proposition 3.6.1. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ be a flux cone with no implicit equalities in $v_{\mathrm{Irr}} \geq 0$. For $v \in C$ we have $\deg(v) = \dim(C)$ if and only if $v_i > 0$ for all $i \in \mathrm{Irr}$.*

Proof. Direct consequence of Prop. 3.2.4. □

For a flux vector in the relative interior of the flux cone, $\deg(v) = \dim(C)$ and by Prop. 3.6.1 all irreversible reactions are active in v. We can generalize this observation to derive an upper bound to the degree of a flux vector v which depends on the number of irreversible reactions that are active in v.

Proposition 3.6.2. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ be the flux cone of a metabolic network with lineality space L. Then for each flux vector $v \in C$*

$$\deg(v) \;\leq\; \dim(L) + |\mathrm{irr.\,supp}(v)|.$$

Proof. By definition of the lineality space L, $t := \dim(L) = n - \mathrm{rank}(A)$, with

$$A = \begin{pmatrix} S \\ -S \\ I_{\mathrm{Irr},\star} \end{pmatrix}.$$

By (3.2), we have $\deg(v) = n - \mathrm{rank}(A_v^=)$, with

$$A_v^= = \begin{pmatrix} S \\ -S \\ I_{\mathrm{Irr}\,\setminus\,\mathrm{irr.supp}(v),\star} \end{pmatrix}.$$

It follows $\mathrm{rank}(A) - \mathrm{rank}(A_v^=) = (n - t) - (n - \deg(v)) = \deg(v) - t$. Thus, at least $\deg(v) - t$ rows from A must be missing in $A_v^=$. Therefore $|\mathrm{irr.\,supp}(v)| \geq \deg(v) - t$, or equivalently $\deg(v) \leq t + |\mathrm{irr.\,supp}(v)|$. □

The following proposition establishes an upper bound on the number of active irreversible reactions in elementary flux modes. The only computation required to establish this upper bound for a given metabolic network is to determine the rank of a submatrix of the stoichiometric matrix (namely the submatrix consisting of the columns corresponding to irreversible reactions). This makes the bound easily computable for any given metabolic network.

Proposition 3.6.3. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ be the flux cone of a metabolic network. Then*

$$|\mathrm{irr.\,supp}(e)| \leq \mathrm{rank}(S_{\star,\mathrm{Irr}}) + 1,$$

for each EFM $e \in C$.

Proof. Suppose the opposite. Then

$$\begin{aligned} |\mathrm{supp}(e)| &= |\mathrm{irr.\,supp}(e)| + |\mathrm{rev.\,supp}(e)| \\ &\geq \mathrm{rank}(S_{\star,\mathrm{Irr}}) + 2 + |\mathrm{rev.\,supp}(e)| \\ &\geq \mathrm{rank}(S_{\star,\mathrm{irr.supp}(e)}) + 2 + \mathrm{rank}(S_{\star,\mathrm{rev.supp}(e)}) \\ &\geq \mathrm{rank}(S_{\star,\mathrm{supp}(e)}) + 2, \end{aligned}$$

contradicting the rank test (2.5). □

By combining Prop. 3.6.3 and Prop. 3.6.2, we get an upper bound on the degree of EFMs.

Corollary 3.6.4. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ be the flux cone of a metabolic network with lineality space L. Then for each EFM $e \in C$*

$$\deg(e) \;\leq\; \dim(L) + (\mathrm{rank}(S_{\star,\mathrm{Irr}}) + 1).$$

The following example shows that the bound from Cor. 3.6.4 is sharp.

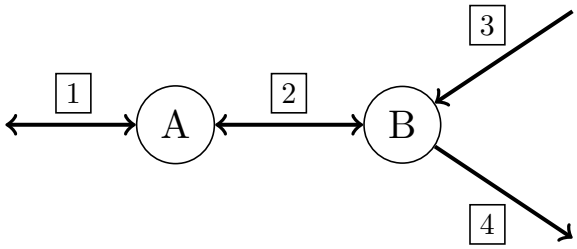

Figure 3.3: Example network showing that the bound from Cor.3.6.4 is sharp.

Example 3.6.5. *The network in Figure 3.3 contains 2 metabolites and 4 reactions, with* $\mathrm{Rev} = \{1, 2\}$ *and* $\mathrm{Irr} = \{3, 4\}$. *Given the stoichiometric matrix*

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix},$$

the network has the EFMs $e^1 = (1, 1, 0, 1)$, $e^2 = (-1, -1, 1, 0)$ *and* $e^3 = (0, 0, 1, 1)$, *with* $\deg(e^1) = \deg(e^2) = 1$ *and* $\deg(e^3) = 2$. *Note that* $C = \mathrm{cone}(\{e^1, e^2\})$ *and* $e^3 = e^1 + e^2 \in \mathrm{relint}(C)$. *Since there are no reversible flux vectors, we have* $\dim(\mathrm{lin.\,space}(C)) = 0$. *Furthermore,* $\mathrm{rank}(S_{\star,\mathrm{Irr}}) = 1$ *and thus* $\deg(e^3) = \dim(\mathrm{lin.\,space}(C)) + (\mathrm{rank}(S_{\star,\mathrm{Irr}}) + 1) = 2$.

For the example networks `E.coli core`, `Pentose Phosphate Pathway`(`PPP`), and `Pyruvate` from Section 3.4, the maximum degree of EFMs and the upper bounds from Prop. 3.6.2 and Cor. 3.6.4 are given in Table 3.2. We can see that $l = \max\{|\mathrm{irr.\,supp}(e)| : e \text{ EFM}\}$ is much smaller than the upper bound $\mathrm{rank}(S_{\star,\mathrm{Irr}}) + 1$ from Prop. 3.6.3. The actual degrees of the EFMs in these networks are summarized in Figure 3.4.

The next proposition further explains the scarcity of EFMs in the relative interior of a flux cone C, in the facets of C and in the faces of dimension $\dim(C) - 2$.

Proposition 3.6.6. *Let* $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ *be a flux cone with no redundant inequalities or implicit equalities in* $v_{\mathrm{Irr}} \geq 0$. *If* $|\mathrm{Irr}| >$

	E. coli core	PPP	Pyruvate
$(\lvert\mathcal{M}\rvert, \lvert\mathcal{R}\rvert)$	(72, 94)	(34, 57)	(28, 81)
$\lvert\mathrm{Irr}\rvert$	48	19	40
$\lvert\mathrm{Rev}\rvert$	46	38	41
$\mathrm{rank}(S)$	67	34	28
$\mathrm{rank}(S_{\star,\mathrm{Irr}})$	41	16	24
$t = \dim(\mathrm{lin.\,space}(C))$	0	8	16
$q = \max\{\lvert\mathrm{irr.\,supp}(e)\rvert : e \text{ EFM}\}$	23	9	10
$r = \mathrm{rank}(S_{\star,\mathrm{Irr}}) + 1$ (cf. Prop. 3.6.3)	42	17	25
$\max\{\deg(e) : e \text{ EFM}\}$	6	14	24
$t + q$ (cf. Prop. 3.6.2)	23	17	26
$t + r$ (cf. Cor. 3.6.4)	42	25	41
$\dim(C)$	23	23	53
$\lvert\mathrm{EFMs}\rvert$	16673	5180	47854

Table 3.2: Maximum number of active irreversible reactions and maximum degree of EFMs together with the upper bounds from Prop. 3.6.3, 3.6.2 and Cor. 3.6.4.

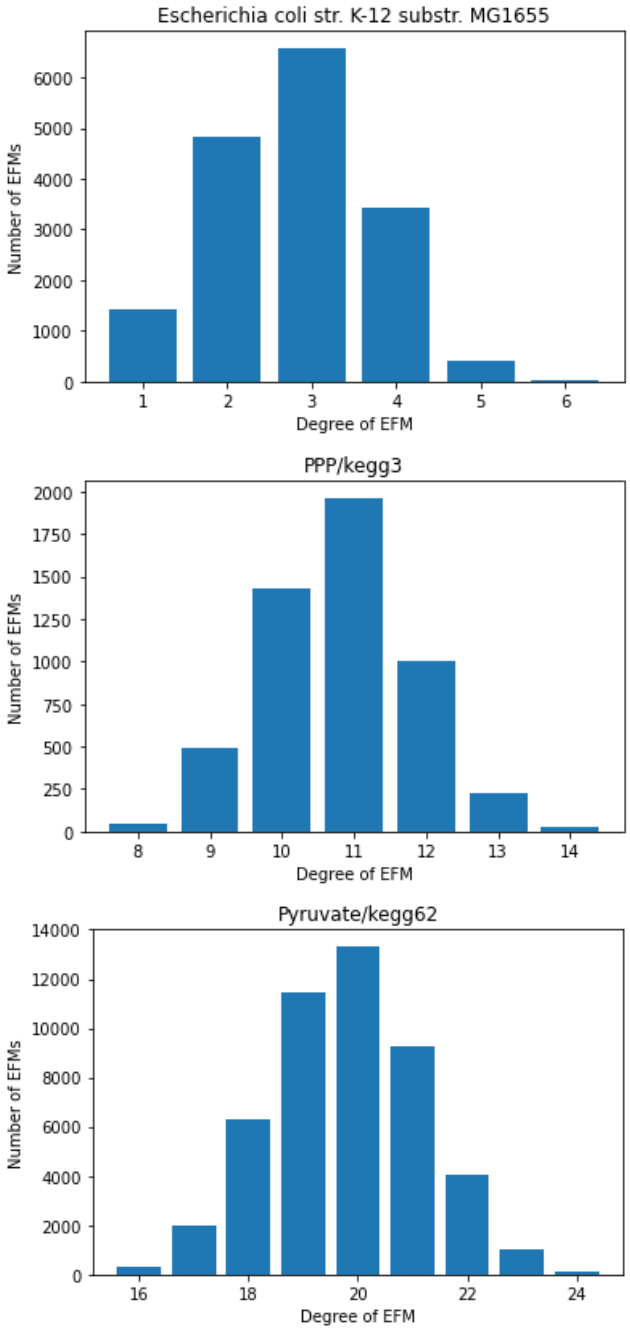

Figure 3.4: Degree distribution of EFMs in metabolic networks.

$\operatorname{rank}(S_{\star,\mathrm{Irr}}) + q$, *for some* $q \in \{1,2,3\}$, *then* $\deg(e) \leq \dim(C) - q$, *for each EFM* e *of* C.

Proof. By Prop. 3.6.3, $|\mathrm{irr.\,supp}(e)| \leq \operatorname{rank}(S_{\star,\mathrm{Irr}}) + 1$, for each EFM e of C.

Assume $\deg(e) = \dim(C) - (q-1)$, for some $q \in \{1,2,3\}$. Then, by definition, the inclusionwise minimal face F containing e has dimension $\dim(C) - (q-1) \geq \dim(C) - 2$. It follows that e resp. F is contained in exactly $q-1$ facets of C. Here we use for $q = 3$ that a $(\dim(C)-2)$-face is contained in exactly two facets, cf. (Schrijver, 1998).

By our hypothesis on the description of the flux cone, it follows $|\mathrm{irr.\,supp}(e)| = |\mathrm{Irr}| - (q-1)$. We get $|\mathrm{Irr}| - (q-1) \leq \operatorname{rank}(S_{\star,\mathrm{Irr}}) + 1$ or $|\mathrm{Irr}| \leq \operatorname{rank}(S_{\star,\mathrm{Irr}}) + q$, in contradiction to the hypothesis $|\mathrm{Irr}| > \operatorname{rank}(S_{\star,\mathrm{Irr}}) + q$. □

In the proof of Prop. 3.6.6, we used that $(\dim(C) - q)$-faces of a cone C are contained in exactly q facets of C, for $q = 0, 1, 2$. As the example of a 3-dimensional pointed cone with n facets shows, a $(\dim(C) - 3)$-face (here the origin) can be contained in an arbitrary number of facets, and thus a similar argument does not hold for such faces. To limit the number of facets a face can be contained in, we introduce the concept of *l-simple* cones and use this for a generalization of Prop. 3.6.6.

A cone $C \subseteq \mathbb{R}^n$ is called *l-simple* for some $l \geq 1$, if every k-face of C is contained in at most $l \cdot (\dim(C) - k)$ facets of C, for all $k = \dim(\mathrm{lin.\,space}(C)), \ldots, \dim(C)$. Assuming that a flux cone is l-simple leads to another bound on the degree of EFMs.

Proposition 3.6.7. *Let* $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$ *be an* l*-simple cone with no redundant inequalities or implicit equalities in* $v_{\mathrm{Irr}} \geq 0$. *Then for each EFM* $e \in C$

$$\deg(e) \leq \dim(C) - \frac{|\mathrm{Irr}| - (\operatorname{rank}(S_{\star,\mathrm{Irr}}) + 1)}{l}.$$

Proof. By Prop. 3.6.3, $|\text{irr.supp}(e)| \leq \text{rank}(S_{\star,\text{Irr}}) + 1$ for each EFM $e \in C$ and thus at least $|\text{Irr}| - (\text{rank}(S_{\star,\text{Irr}}) + 1)$ entries of v_{Irr} are equal to zero. Hence e is contained in at least $|\text{Irr}| - (\text{rank}(S_{\star,\text{Irr}}) + 1)$ facets of C. Suppose $\deg(e) = k$ and let $e \in F$, where F is a k-face of C. Since C is l-simple, F is contained in at most $l \cdot (\dim(C) - k)$ facets. It follows

$$|\text{Irr}| - (\text{rank}(S_{\star,\text{Irr}}) + 1) \leq l \cdot (\dim(C) - k)$$

or

$$\deg(e) = k \ \leq \ \dim(C) - \frac{|\text{Irr}| - (\text{rank}(S_{\star,\text{Irr}}) + 1)}{l}.$$

□

Note that this bound is mainly theoretical because for the computation of l all faces of the flux cone have to be considered. Nevertheless $|\text{Irr}| \geq \text{rank}(S_{\star,\text{Irr}})$ and $|\text{Irr}|$ is typically significantly larger than $\text{rank}(S_{\star,\text{Irr}})$ (cf. Table 3.2).

3.7 The cardinality of minimal metabolic behaviors

Minimal metabolic behaviors have been introduced and studied by Larhlimi and Bockmayr (2009). With our observations in this chapter, we establish an upper bound on the cardinality of MMBs and examine the effects of removing redundant constraints from the flux cone's description.

We begin by proving an upper bound on the cardinality of MMBs.

Proposition 3.7.1. *Let C be the flux cone of a metabolic network $\mathcal{N}$ with lineality space L. Then for each MMB D*

$$|D| \ \leq \ |\text{Irr}| - (\dim(C) - \dim(L)) + 1.$$

Proof. By definition of an MMB, there exist a minimal proper face G of C, $\dim(G) = \dim(L) + 1$, and a flux vector $g \in G \setminus L$ such that $D = \text{irr.supp}(g)$.

It follows that g is contained in at least $\dim(C) - (\dim(L) + 1)$ facets of C. Therefore at least $\dim(C) - (\dim(L) + 1)$ inequalities in $v_{\text{Irr}} \geq 0$ are satisfied by g with equality, which implies $|\text{irr.supp}(g)| = |D| \leq |\text{Irr}| - (\dim(C) - (\dim(L) + 1))$. □

In general, MMBs often contain irreversible reactions for which the corresponding non-negativity constraint is redundant. Removing redundant non-negativity constraints (i.e., shifting the corresponding reactions from Irr to Rev) until the description contains no more redundant inequalities, can lead to much smaller cardinalities of MMBs.

For our example networks `E.coli core`, `Pentose Phosphate Pathway` and `Pyruvate`, the number of MMBs are 1421, 19, and 37 respectively. In Figure 3.5 we compare the cardinalities of the MMBs in the original description of the flux cone and after removing redundant irreversibility constraints. If we start from the original description, Prop. 3.7.1 provides the upper bounds 18, 5, and 4 respectively, while the actual maximal sizes of the MMBs are 17, 4, and 3. If we remove redundant non-negativity constraints in lexicographical order (i.e., the redundant non-negativity constraint corresponding to the irreversible reaction with the smallest index is removed first), the bounds become sharp, i.e., we get the bounds 9, 3, and 1 respectively, and these bounds coincide with the actual maximal cardinalities of the MMBs.

Proposition 3.7.2. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\text{Irr}} \geq 0\}$ be the flux cone of a metabolic network with lineality space L and no redundant inequalities or implicit equalities in $v_{\text{Irr}} \geq 0$. Then the number of facets of C is equal to $\dim(C) - \dim(L)$ if and only if each MMB has cardinality 1.*

Proof. By Prop. 3.3.2, the number of facets is equal to $|\text{Irr}|$. Thus if $|\text{Irr}| = \dim(C) - \dim(L)$, then by Prop. 3.7.1, $1 \leq |D| \leq |\text{Irr}| - (\dim(C) - \dim(L)) + 1 = 1$, for each MMB D in C.

Conversely, if each MMB has cardinality 1, then each minimal proper face is the intersection of all facets of C but one. For all $i \in \text{Irr}$, $G^i = \{v \in$

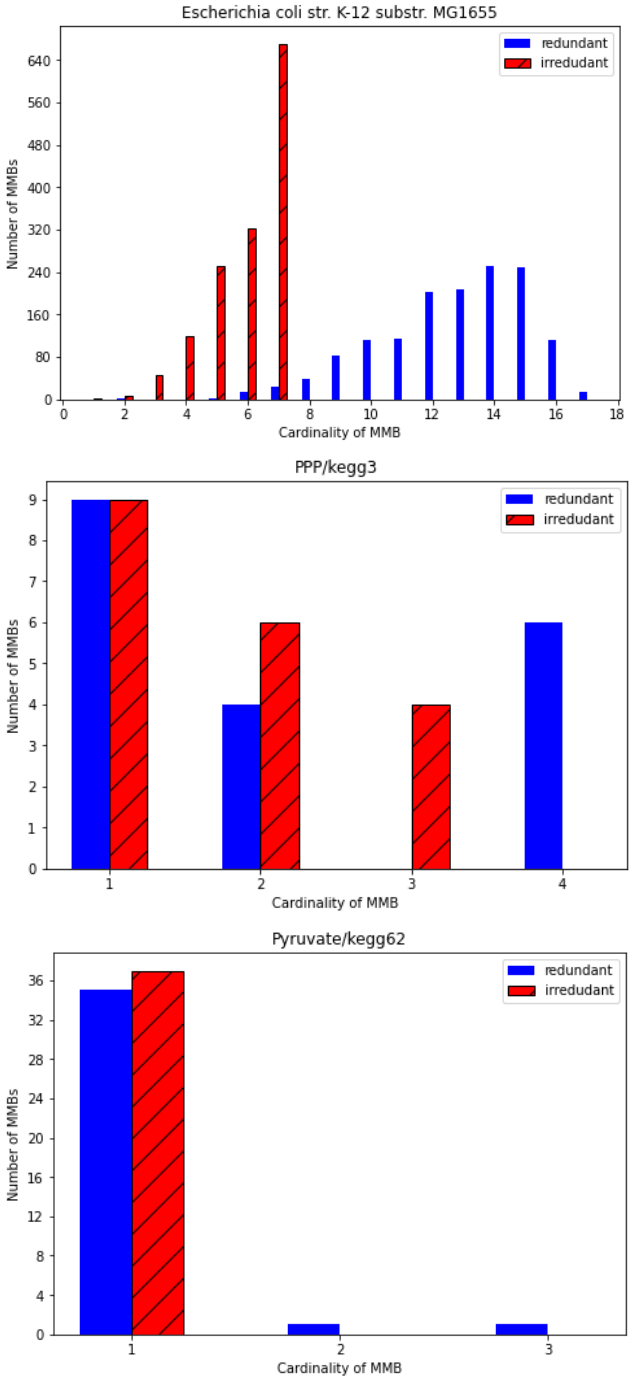

Figure 3.5: Cardinalities of MMBs with and without redundant irreversibility constraints.

$\mathbb{R}^n \mid Sv = 0, v_{\text{Irr}\setminus\{i\}} = 0, v_i \geq 0\}$ is a face of C, with $\dim(G^i) \leq \dim(L) + 1$. Since $v_i \geq 0$ is not an implicit equality, there exists $g^i \in C$ with $g^i_i > 0$. Let $\bar{D}^i = \{j \in \text{Irr} \mid g^i_j > 0\}$ be the metabolic behavior defined by g^i. By Prop. 3.5.1, $\bar{D}^i, i \in \bar{D}^i$, is the union of MMBs, which by hypothesis all have cardinality 1. Thus, for all $i \in \text{Irr}$, $D^i = \{i\}$ is an MMB with the corresponding minimal proper face G^i, where $G^i \setminus L = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\text{Irr}\setminus\{i\}} = 0, v_i > 0\}$ and $\dim(G^i) = \dim(L) + 1$. We conclude that the number of minimal proper faces of C is equal to the number of facets, which by Prop. 3.3.2 is equal to $|\text{Irr}|$.

It remains to prove that $|\text{Irr}| = \dim(C) - \dim(L)$. Let $U = \{u \in \mathbb{R}^n \mid Su = 0\}$ and $W = \{w \in \mathbb{R}^n \mid w_{\text{Irr}} = 0\}$. Then $U \cap W = L$ and since by hypothesis there are no implicit equalities, $\dim(C) = \dim(\text{aff}(C)) = \dim(U)$. From the dimension formula, we get $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W) = \dim(C) + (n - |\text{Irr}|) - \dim(L)$ or $\dim(C) - \dim(L) = |\text{Irr}| - (n - \dim(U + W))$.

We claim $\dim(U + W) = n$, i.e., $U + W = \mathbb{R}^n$. For each minimal proper face $G^i, i \in \text{Irr}$, choose $e^i \in G^i \setminus L$ with $e^i_i = 1$. Then e^i_{Irr} is a unit vector, for all $i \in \text{Irr}$. Given $v \in \mathbb{R}^n$, let $u = \sum_{i \in \text{Irr}} v_i \cdot e^i$ and

$$w = \begin{pmatrix} v_{\text{Rev}} - u_{\text{Rev}} \\ 0 \end{pmatrix}.$$

Since $Se^i = 0, i \in \text{Irr}$, we get $Su = \sum_{i \in \text{Irr}} v_i \cdot Se^i = 0$ and thus $u \in U$. By definition, $w \in W$. For all $j \in \text{Irr}$, we have $u_j = \sum_{i \in \text{Irr}} v_i \cdot e^i_j = v_j e^j_j = v_j$, and thus $u_{\text{Irr}} = v_{\text{Irr}}$. Altogether, we get

$$u + w = \begin{pmatrix} u_{\text{Rev}} \\ u_{\text{Irr}} \end{pmatrix} + \begin{pmatrix} v_{\text{Rev}} - u_{\text{Rev}} \\ 0 \end{pmatrix} = \begin{pmatrix} v_{\text{Rev}} \\ v_{\text{Irr}} \end{pmatrix} = v,$$

which shows $U + W = \mathbb{R}^n$. □

3.8 Conclusion

In this chapter, we studied geometric properties of flux cones of metabolic networks with a focus on elementary flux modes. To structure the set of EFMs, we introduced the degree of flux vectors and studied the degrees of EFM. EFMs of a smaller degree, can be seen as more elementary than those of a higher degree, since the EFMs of higher degree can be decomposed into EFMs of smaller degree. We will further investigate decompositions of flux vectors into EFMs in Chapter 4 and decompositions of EFMs into other EFMs in Chapter 5.

In previous work, see e.g. (Wagner and Urbanczik, 2005), EFMs that are not extreme rays of a pointed flux cone were simply considered to be rays in the (relative) interior of the flux cone. We showed that EFMs belonging to the (relative) interior of the cone (in the sense of Prop. 3.6.1) occur only rarely. Our bounds on the degree of EFMs, as well as our computational results, indicate that EFMs are primarily contained in lower-dimensional faces of the flux cone.

Regarding future research, our new insights into the distribution of EFMs in the face lattice of the flux cone raise the question of how they can be algorithmically exploited. For example, one could focus on enumerating EFMs of lower degrees and omit the enumeration of EFMs with higher degrees in order to make EFM analysis for genome-scale metabolic network reconstructions more tractable.

In Section 3.7, we analyzed the cardinality of minimal metabolic behaviors and studied the effect of removing redundant irreversibility constraints. Although irredundant descriptions of flux cones are not uniquely determined, further research is needed to determine the optimal handling of redundant constraints. Some problems may benefit from the removal of redundant constraints, while others might benefit from adding them. For example, exploring the faces of the flux cone is easier when there are no redundant irreversibil-

ity constraints, as every irreversible reaction then corresponds to a facet of the flux cone. Conversely, the reconfiguration method to determine EFMs by splitting reversible reactions may benefit from adding redundant irreversibility constraints. In that case, fewer reversible reactions have to be split, which decreases the dimension of the reconfigured cone.

Chapter 4

Low-degree decompositions of flux vectors

Most of the results presented in this chapter are derived from a joint work with Alexander Bockmayr. A preprint currently under review can be found at (Wieder and Bockmayr, 2024). The remaining results were published in the Journal of Mathematical Biology (Wieder, Henk and Bockmayr, 2023)(Reproduced with permission from Springer Nature).

4.1 Introduction

EFMs can be viewed as minimal functional units of a metabolic network. Therefore, *decomposing a flux vector* into a positive combination of EFMs is an important task in metabolic network analysis (Poolman *et al.,* 2004; Schwartz and Kanehisa, 2005; Chan and Ji, 2011; Jungers *et al.,* 2011; Rügen *et al.,* 2012; Kelk *et al.,* 2012; Chan, Solem, *et al.,* 2014; Maarleveld *et al.,* 2015; Oddsdóttir *et al.,* 2015; Gerstl *et al.,* 2016; Chen, Huang and Zhong, 2023). Flux distributions that optimize the flux through some given reaction, typically biomass production, have been studied extensively as target vectors for decomposition into EFMs. These optimizing vectors can be obtained by performing a flux balance analysis (FBA), i.e., solving a linear optimization problem maximizing or minimizing an objective function under a given set of linear constraints (Orth, Thiele and Palsson, 2010).

In Chapter 3, we studied the facial structure of the steady-state flux cone of a metabolic network and introduced the degree of a flux vector as the dimension of the inclusionwise minimal face of the flux cone containing it. We suggested that EFMs of lower degrees should be viewed as more elementary than those

of higher degrees.

Here, we use the mathematical results from Chapter 3 to develop a novel method for decomposing flux vectors in genome-scale metabolic networks. We illustrate this method by decomposing optimal solutions v^* of flux balance analysis (FBA) problems. However, this method can also be applied to arbitrary flux vectors. After computing v^*, we first determine the minimal face F^* of the flux cone C to which v^* belongs. Next, we enumerate the EFMs in the face F^*, which typically has a much lower dimension than the full flux cone. With this approach, we are able to determine all EFMs that may participate in a decomposition of v^*. We solve a mixed-integer linear program (MILP) to compute different decompositions of v^* into these EFMs. Taking into account their degree allows us to explore the interplay of the different EFMs that can participate in the decomposition of a given flux vector, highlighting the particular importance of low-degree decompositions.

We illustrate our method on various genome-scale metabolic networks from the BiGG database (Norsigian *et al.,* 2020). For all these networks (with the exception of e_coli_core), we are not able to compute a full representative set of the EFMs, while the EFMs in the face F^* containing the solution to the FBA problem can be obtained easily. Their number is typically rather small, allowing us to explore the different possible EFM decompositions by solving MILPs. While there often exist many different shortest EFM decompositions of a given target vector v^*, it turns out that *low-degree decompositions*, using only EFMs of degree at most $t + 1$, where t is the dimension of the lineality space of the flux cone, are unique in many cases and serve as the building blocks for the other decompositions.

4.2 Decomposition of flux vectors

By Prop. 2.5.5, any vector $v^* \in C$ in the flux cone C of a metabolic network is a positive combination

$$v^* = \sum_{i=1}^{k} \lambda_i e^i, \text{ with } \lambda_i > 0, \tag{4.1}$$

of elementary flux modes $e^1, \ldots, e^k \in C$ (Schuster, Hilgetag, *et al.,* 2002). Sometimes, we will call such a positive combination a *(conic) decomposition* of v^* into $e^1, \ldots, e^k$. By requiring $\lambda_i > 0$, we can define the *length* of a decomposition as k and a *shortest decomposition* as a decomposition of minimal length. Note that shortest decompositions are not unique is many cases (cf. Figure 4.4).

We will show that the length of a decomposition of v^* crucially depends on the degree of v^*. More precisely, we prove in Prop. 4.2.2 that any flux vector of degree k is a positive combination of at most k EFMs, each of which has degree at most k.

First, we show that a decomposition of a flux vector v, contained in some face $F \subseteq C$, only uses vectors that also contained in the face F.

Proposition 4.2.1. *Let $C = \{v \in \mathbb{R}^n \mid Sv = 0, v_{\text{Irr}} \geq 0\}$ be the flux cone of a metabolic network and let $F = \{v \in C \mid v_I = 0\}$, for some $I \subseteq \text{Irr}$, be a face of C.*

If $v = \sum_{i=1}^{m} \lambda_i v^i, \lambda_i > 0$, is a decomposition of a flux vector $v \in F$ into flux vectors $v^1, \ldots, v^m \in C$, then $v^1, \ldots, v^m \in F$.

Proof. Suppose $v^l \notin F$, for some $l \in \{1, \ldots, m\}$. Then, $v^l_j > 0$, for some $j \in I$. It follows $v_j = \sum_{i=1}^{m} \lambda_i v^i_j \geq \lambda_l v^l_j > 0$, and therefore $v \notin F$. □

Next we show that any flux vector $v \in C$ of degree k can be decomposed into a positive combination of at most k EFMs of degree at most $t+1$, where t

is the dimension of the lineality space L of C. In other words, this means that any flux vector of degree k can be written as a positive combination of at most k EFMs, all belonging to the lineality space or to some minimal proper face of the flux cone. From each minimal proper face, at most one EFM is needed for the decomposition and all EFMs belonging to the same minimal proper face share the same minimal metabolic behavior (cf. Prop. 3.5.1). In the special case $k = t$, the EFMs in the decomposition all have degree t and belong to the lineality space of the flux cone (cf. Prop. 4.2.1).

Proposition 4.2.2. *Let $C \subseteq \mathbb{R}^n$ be the flux cone of a metabolic network and let $v \in C$ be a flux vector with $\deg(v) = k$. Let $F \subseteq C, \dim(F) = k$, be the inclusionwise minimal face containing v. Then*

i) *v can be decomposed into a positive combination $v = \sum_{i=1}^{m} \lambda_i e^i, \lambda_i > 0$, of at most $m \leq k$ EFMs $e^i \in F$, with $\deg(e^i) \leq t + 1$, for $i = 1, \ldots, m$, where t is the dimension of the lineality space L of C.*
In the special case $k = t$ and $F = L$, we have $e^i \in L$ and $\deg(e^i) = t$, for $i = 1, \ldots, m$.

ii) *with probability 1, v does not allow for a conic decomposition into $m < k$ EFMs.*

Proof. i): By Prop. 2.5.5, there exists a conic combination such that $v = \sum_{i=1}^{m} \lambda_i e^i, \lambda_i > 0$, for some EFMs $e^1, \ldots, e^m \in C$. By Prop. 3.2.2 we can assume that $e^1, \ldots, e^m$ belong to the lineality space or to the minimal proper faces of C, i.e., $\deg(e^i) \leq t+1$, for $i = 1, \ldots, m$. Using Carathéodory's theorem, see e.g. (Lauritzen, 2013), we can also assume that the EFMs $e^1, \ldots, e^m$ in the conic decomposition (4.1) are linearly independent. By Prop. 4.2.1, it follows that $e^1, \ldots, e^m \in F$. Since $\dim(F) = k$, there can be at most k linearly independent vectors in F, thus $m \leq k$ and the result follows.

ii): Using $k - 1$ EFMs one can generate at most a $(k - 1)$-dimensional subset of the k-dimensional face F. Since there are only finitely many ways

of choosing $k-1$ EFMs in the set of all EFMs, the set of flux vectors $v \in F$ that can be decomposed into a combination of $k-1$ EFMs is the finite union of sets of k-dimensional volume 0 and therefore itself a set of k-dimensional volume 0 in F. This implies the result. □

Note that Prop. 4.2.2 also applies to EFMs in higher-dimensional faces, for which we get the following result:

Corollary 4.2.3. *Let $e \in C$ be an EFM with $\deg(e) = k \geq t+2$, where t is the dimension of the lineality space L of C. Then e can be decomposed into a conic combination $e = \sum_{i=1}^{m} \lambda_i e^i, \lambda_i > 0, 2 \leq m \leq k$, of at least 2 and at most k EFMs of degree strictly lower than k.*

Since every k-face $F \subseteq C$ is the conic hull of the EFMs contained in F, we get the following lower bound on the number of EFMs in F.

Proposition 4.2.4. *Let $F \subseteq C$ be a face of the flux cone C with $\dim(F) = k$. Then F contains at least k EFMs. In particular, the cone C itself contains at least $\dim(C)$ EFMs.*

By Prop. 4.2.2 any flux vector $v^* \in C$ of degree $d = \deg(v^*) > 0$ has a decomposition $v^* = \sum_{i=1}^{k} \lambda_i e^i, \lambda_i > 0$, into at most $k \leq d$ EFMs $e^1, \ldots, e^k$ of degree at most $t+1$, where $t = \dim(L)$ is the dimension of the lineality space L of C. For the rest of this chapter we will call such a decomposition a *low-degree decomposition*. Note that the EFMs of degree t or $t+1$ participating in a low-degree decomposition are elements of a MEMo as defined in Chapter 2. Therefore, EFMs of degree t and $t+1$ will also be called *MEMo-EFMs*.

The following small example as well as the results in Table 4.2 show that shortest decompositions of flux vectors typically contain EFMs of degrees higher than $t+1$. In general shortest decompositions are not low-degree decompositions.

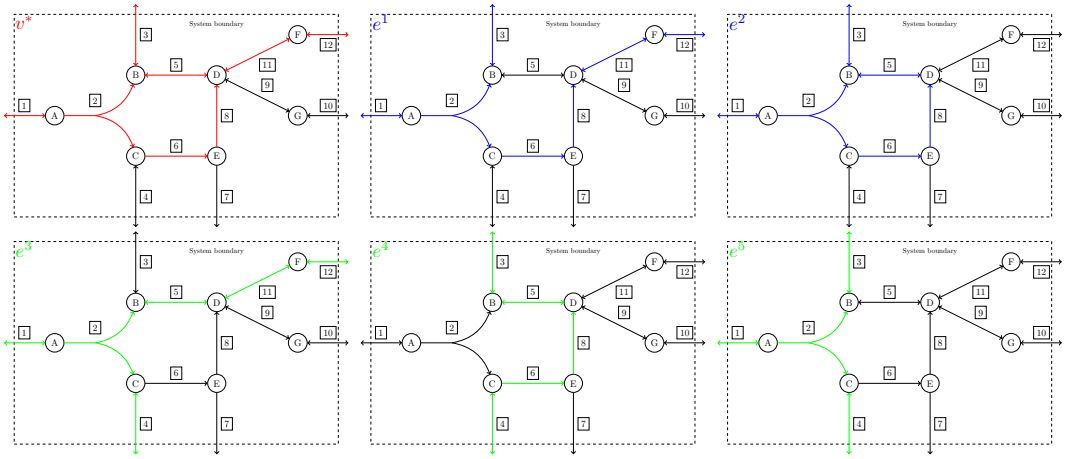

Figure 4.1: Vector v^* to decompose (red), EFMs e^1, e^2 of degree 4 (blue) and EFMs e^3, e^4, e^5 of degree 3 (green) for a small example network.

Example 4.2.5. *Consider again the metabolic network in Figure 2.4. In Figure 4.1 the vector v^* (in red) is the one to be decomposed and has degree 4. It could be either a given steady-state flux vector or the optimal solution of some FBA problem (e.g. maximizing flux rates of reactions 3 and 12).*

In a first step, a shortest decomposition of v^* into EFMs is computed. The resulting EFMs e^1 and e^2 (in blue) have degree 4 and $v^* = e^1 + e^2$. In the next steps, e^1 and e^2 are decomposed into EFMs of lower degree. The resulting EFMs e^3, e^4, e^5 (in green) have degree 3 and cannot be further decomposed into EFMs of lower degrees. One can verify $e^1 = e^3 + e^4$ and $e^2 = e^4 + e^5$, and thus $v^* = e^3 + 2e^4 + e^5$. The corresponding vectors are (the symbol $\cdot^T$ denotes

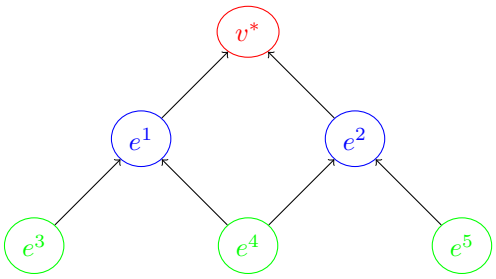

Figure 4.2: Decomposition of v^* into two EFMs of degree 4 (blue), which can be further decomposed into 3 MEMo-EFMs of degree 3 (green).

transposition)

$$
\begin{aligned}
v^* &= (2,2,3,0,-1,2,0,2,0,0,1,1)^T,\\
e^1 &= (1,1,1,0,0,1,0,1,0,0,1,1)^T,\\
e^2 &= (1,1,2,0,-1,1,0,1,0,0,0,0)^T,\\
e^3 &= (1,1,0,1,1,0,0,0,0,0,1,1)^T,\\
e^4 &= (0,0,1,-1,-1,1,0,1,0,0,0,0)^T,\\
e^5 &= (1,1,1,1,0,0,0,0,0,0,0,0)^T.
\end{aligned}
$$

In this simple example, the weights λ_i in the above decompositions are all small integers. In general, this need not be the case. In particular, the weights depend on the normalization of the EFMs, which are determined only up to multiplication by a positive scalar.

We conclude this section by discussing an example that demonstrates how the degree of an FBA solution changes when reactions are added or removed from the network.

Example 4.2.6. *Consider again the e_coli_core model from the BiGG-database (Norsigian* et al., *2020). Using* `cobrapy` *(Ebrahim* et al., *2013), a Python wrapper for the CoBra toolbox (Heirendt* et al., *2019), we perform a flux balance analysis (FBA) to determine the optimal growth rate while glucose uptake*

Ex_glc__D_e *is limited by a lower bound of* -10 *(the number is negative because exchange reactions in this model are oriented outwards). The remaining reactions have an upper bound of* 1000 *and lower bounds of* -1000 *resp.* 0 *for the reversible resp. irreversible reactions. The optimal growth rate of* 0.917 *is attained by a flux vector of degree* 1*, which is an extreme ray of the pointed flux cone and thus, it is an EFM.*

Adding the lower bound of 8.39 *for the ATP-maintenance reaction* ATPM *leads to the standard version of the model, i.e., the version of the model that can be downloaded from the BiGG database. Now, the optimal growth rate is* 0.874 *and the flux vector achieving this growth rate has degree* 2*. Figure 4.5 shows that there is a unique decomposition into the two MEMo-EFMs of degree* 1 *that span the* 2*-face containing this optimal solution.*

If we perform a single knock-out of the gene b1761 (gdhA)*, which blocks the reaction* GLUDy*, the new optimal growth rate is* 0.851*. The degree of the flux vector achieving this optimal growth rate changes to* 3 *and the* 3*-face containing this optimal solution is spanned by* 3 *EFMs of degree* 1*. Therefore, there is only one decomposition of this optimal solution into three MEMo-EFMs.*

After knocking out a second gene, b3236 (mdh)*, which blocks the reaction* MDH*, the new optimal growth rate is* 0.801*. The degree of the flux vector achieving this optimal growth rate now is* 4 *and the* 4*-face containing this optimal solution is spanned by six EFMs of degree* 1*. In this case, there are four different decompositions of the optimal solution into four MEMo-EFMs.*

It should be noted that the optimal solutions to FBA problems are generally not uniquely determined. In this example, we chose the solution that was returned by `cobrapy`*. Different solutions with the same optimal value can lie in different faces of the flux cone. Therefore, they can have a different degree and thus also a different number of EFMs needed in their decomposition into MEMo-EFMs.*

4.3 Algorithm for decomposition of flux vectors

In Chapter 3, we showed that for any $J \subseteq \text{Irr}$, the set $F_J = \{v \in C \mid v_j = 0 \text{ for all } j \in J\}$ is a face of the polyhedral cone C. Given $v^* \in C$, the inclusion-wise minimal face of C containing v is $F_{J(v^*)}$ with $J(v^*) = \{j \in \text{Irr} \mid v^*_j = 0\}$. We call $F^* = F_{J(v^*)} \subseteq C$ the *face defined by* v^*. Note that by definition $\dim(F_{J(v^*)}) = \deg(v^*)$.

It follows from Prop. 4.2.1 that for decomposing a flux vector $v^* \in C$ into EFMs, it suffices to consider those EFMs that are contained in the face $F^* = F_{J(v^*)} \subseteq C$ defined by v^*. The face F^* is again a polyhedral cone, which informally is the flux cone of the metabolic network $\mathcal{N}' = (\mathcal{M}, \mathcal{R} \setminus J(v^*), S', \text{Irr}')$, obtained from $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Irr})$ by deleting the reactions in $J(v^*)$.

By computing the EFMs of $\mathcal{N}'$ using standard tools like `efmtool` (Terzer, 2009),*EFMlrs* (Buchner and Zanghellini, 2021) or a MILP-based approach (de Figueiredo *et al.,* 2009) (we used `efmtool`), and adding zero components $v_j = 0$ for the deleted reactions $j \in J(v^*)$, we obtain exactly the EFMs of $\mathcal{N}$ contained in F^*. For a formal proof, we refer to Lemma 4 in (Marashi and Bockmayr, 2011).

Our workflow for decomposing the optimal solution of an FBA problem into EFMs is shown in Figure 4.3. The main steps are as follows:

Choose model. After downloading an SBML (Hucka *et al.,* 2003) file from the BiGG database, the software package `cobrapy` (Ebrahim *et al.,* 2013) offers the function `cobra.read_sbml_model()` to create a model from the file containing all metabolites, reactions, bounds and further information from the chosen metabolic network.

Standardize model. In their standard (downloadable) form, some models in the BiGG database are not suitable for EFM computation with `efmtool`.

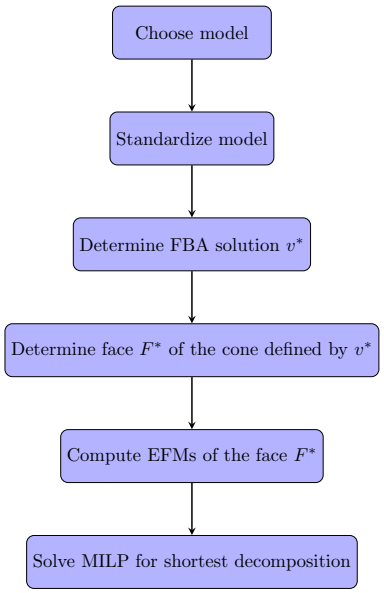

Figure 4.3: Workflow for determining decompositions of an FBA solution into EFMs.

This is due to the fact that for some models irreversible uptake reactions are oriented as negative output reactions (i.e., an irreversible reaction has a negative flux, which is not in line with the requirements of `efmtool`). This can be addressed by changing the orientation of these reactions. We did this by replacing these reactions with oppositely oriented copies. Furthermore, we removed all reactions for which the lower and upper bound are both equal to zero.

Determine FBA solution v^*. Typically, the standard objective is to maximize biomass production. By calling the function `cobra.optimize()`, a flux vector v^* achieving the optimal growth rate is computed. We did not change any objective function and just worked with the model exactly the way it can be downloaded (up to changing the orientation of some irreversible exchange reactions and removing reactions with lower and upper bound equal to 0, as mentioned above). If one is interested in decomposing a flux vector that is not

obtained by solving an FBA problem, this step can be skipped and the given vector can be used as v^* instead of the FBA solution.

Determine the face F^* defined by v^*. By calling the function `cobra.util.create_stoichiometric_matrix()`, the stoichiometric matrix defining the flux cone C can be constructed. Adding unit row vectors to the stoichiometric matrix corresponding to the constraints $v_j = 0, j \in J(v^*)$, together with a binary vector containing information about the reversibility of reactions, leads to a description of the face F^* containing all candidate EFMs for decomposing the target vector v^*.

Compute EFMs of the face F^*. The modified stoichiometric matrix and the reversibility vector are used as input for EFM computation with `efmtool` (Terzer, 2009). The EFMs computed in this step represent all EFMs in the face F^* of the flux cone that contains the optimal FBA solution v^*. They form a subset of a representative set of all EFMs in the flux cone.

Solve MILP for shortest decomposition. A shortest decomposition of a vector v^*, given a set of candidate vectors (in our case EFMs in F^* or MEMo-EFMs in F^*), can be found by solving the following mixed-integer linear program (MILP):

$$\begin{aligned} \min \quad & a_1 + ... + a_k \\ \text{subject to} \quad & \\ E\lambda &= v^*, \\ 0 &\leq \lambda \leq \mathcal{M}a, \\ \lambda &\in \mathbb{R}^k, \\ a &\in \{0,1\}^k. \end{aligned} \tag{4.2}$$

Here, $E = (e^1, \ldots, e^k) \in \mathbb{R}^{n \times k}$ is the matrix containing the k candidate EFMs $e^1, \ldots, e^k$ in the face F^* as columns. The vector $\lambda \in \mathbb{R}^k$ contains the coefficients of the EFMs in the decomposition $E\lambda = \lambda_1 e^1 + \cdots + \lambda_k e^k = v^*$. The binary variables a_j indicate whether or not EFM e^j takes part in the decomposition, i.e., whether $\lambda_j > 0$ or $\lambda_j = 0$. By minimizing the sum $a_1 + \cdots + a_k$, we find a shortest decomposition using as few EFMs as possible. The bigM constant $\mathcal{M}$

is an integer number that has to be chosen sufficiently large (we used 10^6 for the calculations in Table 4.2). The constraint $0 \leq \lambda \leq \mathcal{M}a$ implies that $\lambda_j = 0$ if $a_j = 0$ and that λ_j is unbounded (from above) if $a_j = 1$ (more precisely, λ_j is bounded by $\mathcal{M}$, that is why choosing $\mathcal{M}$ large enough is crucial). [1]

In general, different decompositions can be computed with the set of EFMs contained in the face F^*, depending on their degree. We illustrate this in Section 4.4 and Section 4.5 by presenting and discussing not only shortest decompositions considering all EFMs in F^*, but also shortest decompositions into MEMo-EFMs. For shortest decompositions into MEMo-EFMs, the set of candidates in the decomposition is further restricted to EFMs of degree $t := \dim(L)$ or $t + 1$.

4.4 Computational results

In this section, we present computational results for decompositions of optimizing vectors of FBA problems using metabolic networks from the BiGG database (Norsigian *et al.,* 2020). All computations were performed on a Thinkpad T14s with an AMD Ryzen 7 PRO 4750U processor and 32GB of RAM.

Table 4.1 shows, for a large selection of genome-scale metabolic networks from the BiGG database, the number of EFMs that have to be considered for decomposing the optimal FBA solution of the given model and the time needed to compute these EFMs. The running times range between a few seconds and a couple of minutes. This is remarkable because for all but one of these networks, we are unable to compute a full representative set of the EFMs on our computer due to memory limitations. The only exception is e_coli_core,

[1] We used the "max" normalization option offered by `efmtool`, which scales the computed EFMs to have a largest absolute value of 1. Using the default "min" normalization option that scales the smallest absolute value to 1 leads to numerical issues when solving the MILP for some models.

which has a total number of 100274 EFMs.

The first column in Table 4.1 contains the identifier of the model in the BiGG database. The second column represents the size of the metabolic network (after our standardization) by stating the number of metabolites and reactions in the network. Column 3 gives the dimension t of the lineality space, which is relatively small for all the networks, ranging between 0 and 6 (if $t = 0$, the corresponding flux cone is pointed). The degree of the optimal FBA solution v^* provided by `cobrapy` (Ebrahim *et al.,* 2013) is given in column 4. Again these degrees are relatively low and range between 2 and 12. Note that the dimension of the full flux cone is typically much higher. Due to the low degree of v^*, which equals the dimension of the face F^* defined by v^*, it becomes possible to compute a representative set of the EFMs contained in this face. These numbers are given in column 5. As explained in the beginning of this chapter, there are no other EFMs needed to decompose v^*. Finally, column 6 contains the running time (in seconds) to compute all these EFMs, which is less than 10 minutes for all but 1 network (iECIAI1_1343 with 25095 EFMs in the face F^*).

Network ID	$(\|\mathcal{M}\|,\|\mathcal{R}\|)$	$\dim(L)$	$\deg(v^*)$	|EFMs|	Time
STM_v1_0	(1802, 2528)	1	6	77	44
e_coli_core	(72, 95)	0	2	5	8
iAB_RBC_283	(342, 469)	2	10	120	11
iAF692	(628, 690)	5	7	135	13
iAF1260b	(1668, 2384)	1	6	168	33
iBWG_1329	(1949, 2727)	4	9	371	45
ic_1306	(1936, 2712)	4	12	3902	418
iE2348C_1286	(1919, 2689)	6	14	5782	480
iEC042_1314	(1926, 2700)	3	9	972	67
iEC1344_C	(1934, 2716)	3	9	1412	108
iEC1349_Crooks	(1946, 2745)	3	11	2576	102
iEC1364_W	(1927, 2754)	3	11	5774	327
iEC1368_DH5a	(1951, 2769)	3	8	322	44
iEC1372_W3110	(1918, 2748)	3	10	3756	121
iEC55989_1330	(1953, 2742)	4	10	947	65

Network ID	$(\|\mathcal{M}\|,\|\mathcal{R}\|)$	$\dim(L)$	$\deg(v^*)$	|EFMs|	Time
iECABU_c1320	(1942, 2717)	4	10	733	55
iECB_1328	(1951, 2734)	4	10	733	70
iECBD_1354	(1952, 2734)	4	10	373	55
iECD_1391	(1943, 2727)	4	12	2894	101
iEcDH1_1363	(1949, 2736)	4	10	975	75
iECDH1ME8569_1439	(1950, 2741)	4	12	3759	117
iECDH10B_1368	(1947, 2728)	4	9	479	56
iEcE24377_1341	(1972, 2750)	6	11	378	46
iECED1_1282	(1929, 2692)	4	10	493	54
iECH74115_1262	(1918, 2681)	4	12	3857	127
iEcHS_1320	(1963, 2740)	3	8	548	79
iECIAI1_1343	(1968, 2751)	3	13	25095	1299
iECIAI39_1322	(1953, 2707)	4	11	3252	194
iECNA114_1301	(1927, 2704)	4	10	947	65
iECO26_1355	(1965, 2766)	4	11	1937	142
iECO103_1326	(1958, 2744)	4	9	552	51
iECO111_1330	(1959, 2746)	3	11	1934	102
iECOK1_1307	(1941, 2715)	4	11	1454	82
iEcolC_1368	(1969, 2754)	4	12	2895	97
iECP_1309	(1941, 2725)	4	10	734	64
iECS88_1305	(1942, 2715)	4	12	3757	169
iECSE_1348	(1957, 2754)	4	10	975	69
iECSF_1327	(1951, 2728)	4	9	372	47
iEcSMS35_1347	(1947, 2732)	4	11	3133	132
iECSP_1301	(1920, 2698)	4	10	732	74
iECUMN_1333	(1935, 2726)	5	12	3763	131
iECW_1372	(1973, 2768)	4	9	371	47
iEK1008	(998, 1224)	0	9	1965	337
iEKO11_1354	(1972, 2764)	4	12	5776	260
iETEC_1333	(1962, 2742)	4	8	190	43
iG2583_1286	(1919, 2691)	4	10	733	55
iHN637	(698, 773)	2	5	22	13
iNRG857_1313	(1943, 2721)	6	10	196	36
iPC815	(1552, 1960)	3	9	404	34
iS_1188	(1914, 2606)	5	9	281	36
iSB619	(655, 729)	0	3	11	12
iSbBS512_1146	(1910, 2580)	3	11	6491	370
iSBO_1134	(1908, 2579)	3	11	2319	69
iSDY_1059	(1888, 2529)	3	7	241	50

Network ID	($\|\mathcal{M}\|$,$\|\mathcal{R}\|$)	dim(L)	deg(v^*)	|EFMs|	Time
iSF_1195	(1917, 2617)	5	10	373	44
iSFV_1184	(1917, 2608)	5	10	374	43
iSFxv_1172	(1918, 2626)	4	9	131	37
iSSON_1240	(1936, 2680)	5	9	74	37
iUMN146_1321	(1942, 2721)	4	12	5630	303
iUMNK88_1353	(1969, 2763)	4	11	2820	140
iUTI89_1310	(1940, 2711)	4	11	2173	109
iY75_1357	(1953, 2745)	4	11	1886	110
iYL1228	(1658, 2262)	1	6	58	30
iYO844	(990, 1250)	2	4	17	18
iYS1720	(2436, 3339)	3	7	22	53
iZ_1308	(1923, 2707)	4	8	191	43

Table 4.1: EFMs in the face F^* defined by the optimal solution v^* of FBA problems in the BiGG database. Network Id: Identifier of the network; ($|\mathcal{M}|$,$|\mathcal{R}|$): Number of metabolites and reactions; dim(L): Dimension of the lineality space; deg(v^*): Degree of the optimal FBA solution v^*. |EFMs|: Number of EFMs in the face F^* defined by v^*. Time: Running time in seconds to compute the EFMs in F^*.

Table 4.2 provides more detailed information about the EFMs and the shortest decompositions for a subset of the models in Table 4.1. The first four columns are the same as before. Column 5 lists the degrees of the EFMs in the shortest decomposition computed by solving the MILP (4.2), while column 6 lists the degrees of EFMs in the shortest decomposition into MEMo-EFMs. For example, in the first row of Table 4.2, the optimal solution to the FBA problem for STM_v1_0 has degree 6 and it can be decomposed into two EFMs of degree 5 or five EFMs of degree 2. Finally, the last column contains a list summarizing information about the degrees of all EFMs in the face F^* defined by v^*. Looking again at the first row, the list [0,2,5,12,26,24,8] should be read in the following way: The face containing the optimal solution contains no EFMs of degree 0 (this immediately follows from the fact $t := \dim(L) = 1$). There are two EFMs of degree 1, five EFMs of degree 2, 12 of degree 3, 26 of

Network ID	($\|\mathcal{M}\|$,$\|\mathcal{R}\|$)	t	deg(v^*)	deg([EFMs])	deg([MEMo-EFMs])	EFM distribution in F^*
STM_v1_0	(1802, 2528)	1	6	[5,5]	[2,2,2,2,2]	[0, 2, 5, 12, 26, 24, 8]
e_coli_core	(72, 95)	0	2	[1,2]	[1,1]	[0, 2, 3]
iAB_RBC_283	(342, 469)	2	10	[5,4,4,4,4,4,4]	[3,3,3,3,3,3,3,3]	[0, 0, 4, 18, 44, 54, 0, 0, 0, 0, 0]
iAF1260b	(1668, 2384)	1	6	[6,5]	[2,2,2,2,2]	[0, 2, 11, 32, 59, 49, 15]
iLJ478	(570, 652)	2	5	[5]	[3,3,3]	[0, 0, 6, 15, 21, 7]
iSB619	(655, 729)	0	3	[2,2]	[1,1,1]	[0, 3, 6, 2]
iSSON_1240	(1936, 2680)	5	9	[9,8]	[6,6,6,6]	[0, 0, 0, 0, 0, 12, 8, 21, 24, 9]

Table 4.2: Decompositions of FBA solutions v^* for a selection of networks from the BiGG database. ($|\mathcal{M}|$,$|\mathcal{R}|$): Number of metabolites and reactions, t: Dimension of the lineality space, deg([EFMs]): Degrees of EFMs in shortest decomposition, deg([MEMo-EFMs]): Degrees of EFMs in shortest MEMo-EFM decomposition, EFM distribution in F^*: Number of EFMs with degree equal to their index in the list.

degree 4, 24 of degree 5 and 8 of degree 6.

4.5 Discussion of computational results

In this section, we discuss the computational results from Section 4.4. Consider again the first row of Table 4.2. In total, there are only 7 MEMo-EFMs and the shortest decomposition into MEMo-EFMs we computed uses all five EFMs of degree 2. Figure 4.4 visualizes how the two EFMs of degree 5 in the shortest decomposition can be further decomposed into the five EFMs of degree 2 in the shortest MEMo-decomposition. In Figure 4.4, the arrows pointing to an EFM start at the EFMs that are used in a decomposition of it. So the two arrows pointing from the two EFMs of degree 5 to the FBA solution of degree 6 indicate that the optimal solution of degree 6 can be decomposed into two EFMs of degree 5. Furthermore, each of the EFMs of degree 5 can be decomposed into one EFM of degree 4 and one of degree 2 and so on.

In Table 4.1 and Table 4.2, we can immediately see that by determining the face F^* defined by the optimal solution v^*, we are able to find all candidate

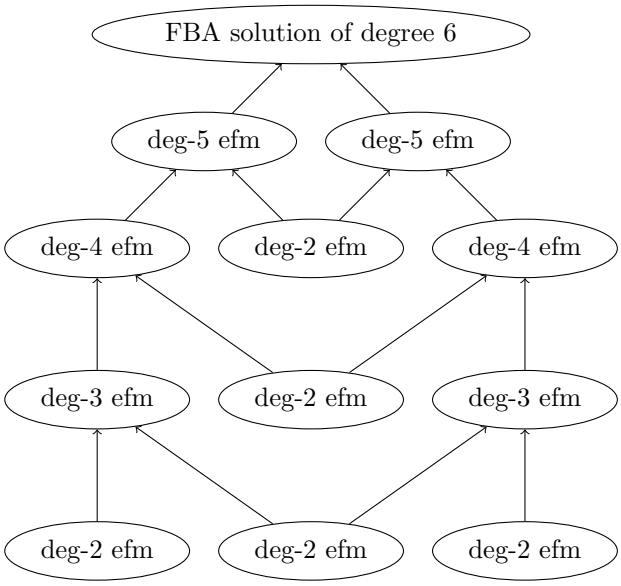

Figure 4.4: Decomposition of the optimal FBA solution v^* of STM_v1_0 into MEMo-EFMs.

EFMs for a decomposition in metabolic networks where the enumeration of all EFMs is not possible in practice.

The advantage of determining shortest decompositions into MEMo-EFMs lies in the fact that these EFMs cannot be further decomposed into EFMs of lower degrees. Thus, MEMo-EFMs in low-degree decompositions are the building blocks for decompositions into higher degree EFMs.

Furthermore, the number of different possible shortest decompositions becomes much smaller when using MEMo-EFMs, as Table 4.3 shows. For all the networks in Table 4.3, the shortest decomposition into MEMo-EFMs turns out to be unique, while there exist multiple shortest decompositions into EFMs of higher degree (with the exception of iLJ478).

To determine all different shortest decompositions we can solve iteratively

Network ID	# shortest decompositions	# shortest decompositions into MEMo-EFMs
STM_v1_0	5	1
e_coli_core	4	1
iAB_RBC_283	12	1
iAF1260b	4	1
iLJ478	1	1
iSB619	2	1
iSSON_1240	4	1

Table 4.3: Number of different shortest decompositions into EFMs and MEMo-EFMs.

the MILP (4.2), adding after each step a so-called *no-good cut* of the form

$$\sum_{i \in I} a_i < k, \tag{4.3}$$

where $I = \{i \mid a_i^* = 1\}$ is the set of indices and k the length of the shortest decomposition (λ^*, a^*) found in the previous step (cf. EFM enumeration with MILP (de Figueiredo *et al.,* 2009)). This allows us to enumerate all shortest decompositions, as we use a different set of EFMs in each step. When there are no more shortest decompositions of length k, the solution of the MILP is a decomposition of length greater than k.

A visual representation of why there can be multiple shortest decompositions is given in Figure 4.5. Consider the model e_coli_core, where the degree of the optimal FBA solution v^* is 2 and only 5 EFMs are contained in the face F^* defined by v^*. The figure shows a projection (not to scale) onto the face F^*. The optimal solution v^* can be decomposed into the EFM e^2 and any one of the other four EFMs. A decomposition of v^* without e^2 is not possible. Furthermore, the EFMs e^3, e^4 and e^5 of degree 2 can be decomposed into the extreme rays of the cone, which are the EFMs e^1 and e^2 of degree 1.

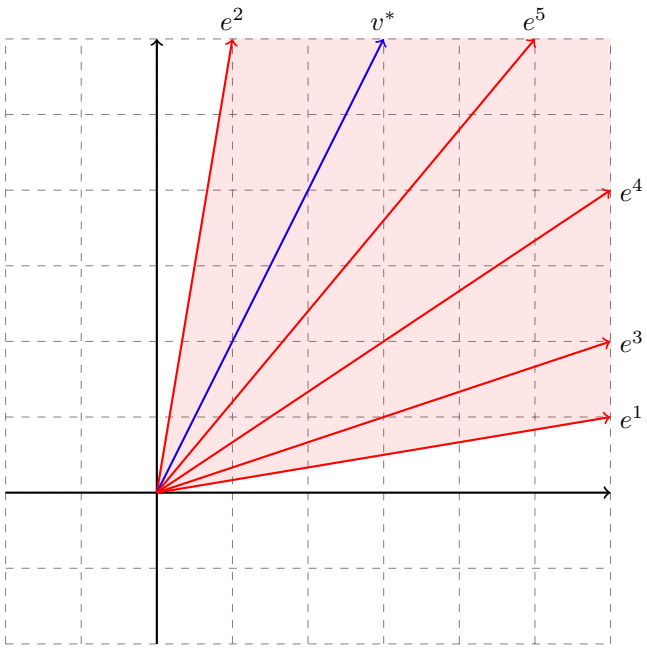

Figure 4.5: Visualization of the 2-dimensional face F^* defined by the optimal FBA solution v^* of e_coli_core and the 5 EFMs contained in it.

In summary, any vector in Figure 4.5 that is positioned between two others can be decomposed into those vectors.

Although in Table 4.3 all shortest decompositions into MEMo-EFMs are unique, this need not always be the case.

To see this, consider a slightly different version of iAF_1260b, namely iAF_1260, also from the BiGG database. After performing an FBA, the resulting flux vector v^* has degree 7 and can be decomposed into two EFMs with degrees 6 and 7. The 7-dimensional face defined by v^* contains 546 EFMs. Of those, only 2 are in the lineality space (have degree 1), and 16 have degree 2, resulting in a total number of 18 MEMo-EFMs. Using these 18 MEMo-EFMs

as candidates, there exist two different shortest decompositions of the FBA solution v^*, which both have length 6.

4.6 Conclusion

Based on geometric insights into the facial structure of the flux cone of a metabolic network established in Chapter 3, we developed a novel method for decomposing flux vectors into elementary flux modes (EFMs). While our approach is applicable to any given flux vector in the cone, we focused on common decomposition targets, namely solutions of FBA problems. By reducing the search space for EFM computation to the face of the flux cone defined by the target vector, the number of EFMs that have to be computed becomes much smaller. Therefore, our method can also be applied to genome-scale metabolic networks where computing all EFMs is not feasible in practice.

Taking into account the degree of the EFMs, we obtained additional insights into the structure of different decompositions and highlighted the special role of low-degree decompositions, consisting of MEMo-EFMs. These EFMs cannot be further decomposed into EFMs of lower degree and thus, they are the building blocks for decompositions of higher degree.

We illustrated our approach on various genome-scale metabolic networks from the BiGG database. For all these networks, the EFMs in the face defined by the optimal FBA solution can be obtained easily, allowing for the enumeration of different shortest decompositions into either MEMo-EFMs or shortest decompositions into EFMs of higher degrees. In many cases, a low-degree decomposition in MEMo-EFMs turns out to be unique, although this need not be the case in general. Future research should address the biological implications of low-degree decompositions and the MEMo-EFMs they are based on.

Chapter 5

Decompositions of elementary flux modes

5.1 Introduction

In this chapter, we will discuss a conjecture about decompositions of EFMs into other EFMs. In real examples, we observed that every EFM either had a shortest decomposition of length 2 or was not decomposable at all. Our conjecture claims that every decomposable EFM has a shortest decomposition of length 2.

We formalize the conjecture and make some observation related to it in Section 5.2. In Section 5.3, we discuss an example that initially appears to be a counterexample and explain why it is not. In Section 5.4, we discuss the relationship between matroids and metabolic networks where all reactions are reversible. Next, we use these observations to prove a weaker version of our conjecture. In Section 5.5, we present two new algorithms. One to determine a decomposition of length 2 for a given EFM if it exists and one to generate new EFMs when a starting set of EFMs is given. In the final section of this chapter, we examine approaches to find a counterexample to our conjecture. We present a method to generate all metabolic networks of a given size with limitations to the stoichiometric coefficients. These networks are used to computationally verify our conjecture and demonstrate that there does not exist a counterexample of small size with a predefined set of possible stoichiometric coefficients.

5.2 Conjecture about decompositions of EFMs

In Chapter 4, we explored decompositions of flux vectors into EFMs. In this section, we will formalize our conjecture about decompositions of EFMs (into other EFMs) and discuss some facts related to this conjecture. For the remainder of this section, let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$ be a metabolic network, C the flux cone of $\mathcal{N}$, and $\mathcal{E} \subseteq C$ a representative set of the EFMs of $\mathcal{N}$. We will abbreviate $a \neq b$, $b \neq c$ and $a \neq c$ and just write $a \neq b \neq c$ for better readability. For a set of vectors $v^1, \ldots, v^k \in \mathbb{R}^N$, we call

$$\sum_{i=1}^{k} \lambda_i v^i$$

a *positive combination* of $v^1, \ldots, v^k$ if $\lambda_i > 0$ for $i = 1, \ldots, k$. For an EFM $e \in \mathcal{E}$, we call a positive combination

$$e = \sum_{i=1}^{k} \lambda_i e^i, \lambda_i > 0, e^i \in \mathcal{E}, i = 1, \ldots, k \tag{5.1}$$

a *non-trivial decomposition of e into k other EFMs* if $\text{supp}(e^j) \neq \text{supp}(e^i) \neq \text{supp}(e)$ for $i, j \in \{1, \ldots, k\}$. Specifically a non-trivial decomposition cannot be the scaling of a vector and no reversible EFMs with equal support combined to 0 appear, so

$$e = \lambda_i e^i \text{ and}$$
$$e = (\sum_{i=1}^{k} \lambda_i e^i) + e' + \lambda' e'',$$

where $e' = -\lambda' e''$ are oppositely oriented reversible EFMs, are not non-trivial decompositions of e. For better readability, we will sometimes just write e is *decomposable into k EFMs* or *e can be decomposed into k EFMs* if there exists

a non-trivial decomposition of e into k other EFMs. Finally, we will often just say e is *decomposable* if there exists some k, such that e is decomposable into k EFMs and since we only consider decompositions into EFMs in this chapter we will sometimes just write "there exists a decomposition" when we mean "there exists a decomposition into EFMs". This notation allows for a compact formulation of our conjecture about decompositions of EFMs.

Conjecture 5.2.1. *If $e \in \mathcal{E}$ is decomposable, then e can be decomposed into two EFMs from $\mathcal{E}$.*

In particular, this conjecture states that there cannot be an EFM of a metabolic network that can be decomposed into $k \geq 3$ EFMs but not into two. Note that every EFM e with $\deg(e) > t + 2$, where $t := \dim(\text{lin. space}(C))$ is the dimension of the lineality space of C, is decomposable (cf. Chapter 4). This means that Conj. 5.2.1 claims that every EFM that is not a MEMo-EFM (as well as every decomposable MEMo-EFM) is decomposable into two EFMs.

An important observation about decompositions of EFMs, in contrast to decompositions of arbitrary flux vectors, is the *cancellation of reversible reactions.*

Definition 5.2.2. *Let $v^1, \ldots, v^k \in C$ be flux vectors in the flux cone C of a metabolic network $\mathcal{N}$ and $r \in \text{Rev}$ a reversible reaction. A positive combination $v = \sum_{i=1}^{k} \lambda^i v^i$ is called a cancellation of the reversible reaction $r \in \text{Rev}$ if and only if*

$$v_r = 0, \text{ and } v_r^i \neq 0 \text{ for at least two } i \in \{1, \ldots, k\}.$$

(i.e., the coefficients $\lambda_i > 0$ are chosen in such a way that the flux through reaction r is equal to zero in the positive combination of $v^1, \ldots, v^k$ but it is not equal to zero for at least two of the vectors $v^1, \ldots, v^k$.)

Specifically, if a positive combination $v = \sum_{i=1}^{k} \lambda_i v^i$ is a cancellation of a reversible reaction $r \in \text{Rev}$, then $r \notin \text{rev. supp}(v)$ and $r \in \text{rev. supp}(v^i)$ for at

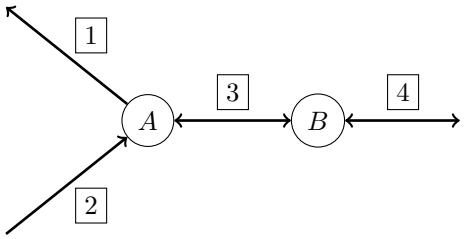

Figure 5.1: Example of a metabolic network for which there exists a cancellation of 2 reversible reactions at once.

least two $i \in \{1, \ldots, k\}$. Clearly, a positive combination can be the cancellation of multiple reversible reactions at the once, as the following example shows.

Example 5.2.3. *Consider the metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$, visualized in Figure 5.1. Assuming all stoichiometric coefficients are in $\{1, 0, -1\}$ and all reactions are oriented from left to right, the stoichiometric matrix S and the set $\mathrm{Rev} \subseteq \mathcal{R}$ of reversible reactions, are*

$$S = \begin{pmatrix} -1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \text{ and } \mathrm{Rev} = \{3, 4\}.$$

A representative set $\mathcal{E}$ of the EFMs of $\mathcal{N}$ is given by

$$\begin{aligned} e^1 &= (1, 0, -1, -1)^T, \\ e^2 &= (0, 1, 1, 1)^T, \\ e^3 &= (1, 1, 0, 0)^T. \end{aligned}$$

Clearly $e^3 = e^1 + e^2$ is a cancellation of the two reversible reactions 3 and 4.

The next lemma shows that only reversible reactions can be canceled in a positive combination of flux vectors.

Lemma 5.2.4. *If a flux vector $v \in C$ is a positive combination*

$$v = \sum_{i=1}^{k} \lambda_i v^i, \lambda_i > 0, i = 1, \ldots, k,$$

of flux vectors $v^1, \ldots, v^k \in C$, *then* $\operatorname{irr.supp}(v) = \bigcup_{i=1}^{k} \operatorname{irr.supp}(v^i)$ *and* $\operatorname{rev.supp}(v) \subseteq \bigcup_{i=1}^{k} \operatorname{rev.supp}(v^i)$.

Proof. Direct consequence of $v_i \geq 0$ for all $i \in \mathrm{Irr}$. □

Note that Lemma 5.2.4 is also true for positive combinations of EFMs.

Next, we demonstrate that cancellations of reversible reactions are crucial for decompositions of EFMs.

Lemma 5.2.5. *If*

$$e = \sum_{i=1}^{k} \lambda_i e^i, \lambda_i > 0$$

is a non-trivial decomposition of an EFM $e \in \mathcal{E}$ *into EFMs* $e_1, \ldots, e_k \in \mathcal{E}$, *then it is a cancellation of at least one reversible reaction.*

Proof. Let $e = \sum_{i=1}^{k} \lambda_i e^i$ be a non-trivial decomposition that is not a cancellation of a reversible reaction. We have

(i) $\operatorname{rev.supp}(e) = \bigcup_{i=1}^{k} \operatorname{rev.supp}(e^i)$ (because $e = \sum_{i=1}^{k} \lambda_i e^i$ is not a cancellation of a reversible reaction, cf. Def. 5.2.2),

(ii) $\operatorname{irr.supp}(e) = \bigcup_{i=1}^{k} \operatorname{irr.supp}(e^i)$ (Lemma 5.2.4).

(iii) $\operatorname{supp}(e^i) \neq \operatorname{supp}(e), i = 1, \ldots, k.$ (definition of non-trivial decompositions)

In total, we have $\operatorname{supp}(e) = \bigcup_{i=1}^{k} \operatorname{supp}(e^i)$ which implies $\operatorname{supp}(e^i) \subseteq \operatorname{supp}(e)$ for $i = 1, \ldots, k$. Because of (iii), $\operatorname{supp}(e^i) \subsetneq \operatorname{supp}(e)$ for $i = 1, \ldots, k$ and e does not have inclusionwise minimal support and thereby cannot be an EFM of $\mathcal{N}$. □

5.3 Definition of a counterexample

Often, the easiest way to disprove a conjecture is to find a counterexample. Before we discuss a candidate for a counterexample to Conj. 5.2.1, we want to formalize what exactly makes a metabolic network a counterexample to our conjecture. By doing so, we can say for certain that the conjecture does not hold as soon as a metabolic network with the following properties is found.

Definition 5.3.1. *A counterexample to Conj. 5.2.1 is a metabolic network* $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$ *with a representative set* $\mathcal{E}$ *of the EFMs of* $\mathcal{N}$*, if there exists* $e \in \mathcal{E}$*, such that:*

(i) e *is decomposable and*

(ii) e *cannot decomposed into 2 EFMs.*

It is important to note that a counterexample consists of a metabolic network together with a representative set of its EFMs. We will discuss an example of a set of vectors that would be a counterexample, if it was a representative set of the EFMs of a metabolic network. Since we cannot find a network with exactly those EFMs, it is not a counterexample after all.

Before we discuss this example, we make an important observation about cancellations of reversible reactions in decompositions of EFMs.

Lemma 5.3.2. *Let* $\mathcal{N} = \{\mathcal{M}, \mathcal{R}, S, \text{Rev}\}$ *be a metabolic network and* $\mathcal{E}$ *a representative set of the EFMs of* $\mathcal{N}$*. If* $e \in \mathcal{E}$ *is decomposable into* $k \geq 3$ *EFMs but cannot be decomposed into 2 EFMs, then at least 2 reversible reactions are canceled in every decomposition of* e *into* k *other EFMs.*

Proof. Assume the opposite, i.e., only one reversible reaction r is canceled in a decomposition of e into 3 other EFMs (the proof is analog if e can be decomposed into $k > 3$ EFMs):

$$e = \sum_{i=1}^{3} \lambda_i e^i, \lambda_i > 0, \text{supp}(e^i) \neq \text{supp}(e^j) \neq \text{supp}(e), i, j \in \{1, 2, 3\}. \tag{5.2}$$

Then $e_r = 0$ and $e_r^1, e_r^2, e_r^3 \neq 0$. Otherwise if $e_r^i = 0$ for some $i \in \{1,2,3\}$, then $\text{supp}(e^i) \subseteq \text{supp}(e)$ (because only one reversible reaction r is canceled) and we get $\text{supp}(e^i) \neq \text{supp}(e)$ by our definition of non-trivial decompositions. But then e^i has inclusionwise smaller support than e and e is not an EFM of $\mathcal{N}$. Note that e^1, e^2, e^3 cannot all have flux in the same direction on reaction r. Otherwise r cannot be canceled in (5.2), i.e., we would get $e_r \neq 0$. We can assume, without loss of generality, $e_r^1, e_r^2 > 0$,$e_r^3 < 0$, and define a positive combination

$$e' := \lambda_1 e^1 + \lambda_3 e^3, \lambda_1, \lambda_3 > 0,$$

such that it is a cancellation of the reversible reaction r, i.e., $e'_r = 0$. We get

$$\begin{aligned}\text{supp}(e') &\subseteq (\text{supp}(e^1) \cup \text{supp}(e^3)) \setminus \{r\} \\ &\subseteq (\text{supp}(e^1) \cup \text{supp}(e^2) \cup \text{supp}(e^3)) \setminus \{r\} = \text{supp}(e).\end{aligned}$$

Thus either e was not an EFM to begin with (if $\text{supp}(e') \subsetneq \text{supp}(e)$, e does not have inclusionwise minimal support) or e is a positive combination of e^1 and e^3 (if $\text{supp}(e') = \text{supp}(e)$) which implies that e is decomposable into 2 EFMs. □

Now we discuss the example mentioned at the beginning of this chapter, which at first glance seems to be a counterexample to Conj. 5.2.1.

Example 5.3.3. *Assume there is a metabolic network $\mathcal{N}$ and that*

$$\mathcal{E} = \left\{ e^1 = \begin{pmatrix}1\\0\\0\\1\\1\end{pmatrix}, \quad e^2 = \begin{pmatrix}-1\\1\\1\\0\\1\end{pmatrix}, \quad e^3 = \begin{pmatrix}0\\-1\\1\\1\\0\end{pmatrix}, \quad e^4 = \begin{pmatrix}0\\0\\2\\2\\2\end{pmatrix} \right\}$$

is a representative set $\mathcal{E}$ of the EFMs of $\mathcal{N}$. Let us also assume that only the first two reactions are reversible (since they appear with negative flux rates in e^2 and e^3). Thus, these four EFMs are not reversible and the flux cone

containing them is pointed. Considering the inclusionwise minimal support property 2.3, these vectors could all be EFMs of the same metabolic network since $\operatorname{supp}(e^i) \not\subset \operatorname{supp}(e^j), i \neq j \in \{1,2,3,4\}$. *Moreover, it is easy to see that* $e^4 = e^1 + e^2 + e^3$ *is a positive combination and the cancellation of the 2 reversible reactions* 1 *and* 2. *Also note that* e^4 *is not a positive combination of any two of* e^1, e^2 *and* e^3 *and that no other positive combination of these vectors has inclusionwise smaller support than all of the four. For*

$$e^1 + e^2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 2 \end{pmatrix} \quad \textit{and} \quad e^2 + e^3 = \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \\ 1 \end{pmatrix},$$

we get $\operatorname{supp}(e^3) \subset \operatorname{supp}(e^1 + e^2)$ *and* $\operatorname{supp}(e^1) \subset \operatorname{supp}(e^2 + e^3)$.

For this to be a counterexample to Conj. 5.2.1, we still have to find the metabolic network $\mathcal{N}$ *that has exactly these four EFMs. The most intuitive way to do this is to determine an outer description of* $C := \operatorname{cone}(\{e^1, e^2, e^3, e^4\})$. *This can be done using the double description method. We compute the following outer description of* C*:*

$$\begin{aligned} 2x_1 + x_2 + x_3 &\geq 0 \\ -x_2 + x_3 &\geq 0 \\ x_2 + x_3 &\geq 0 \\ x_1 + x_3 - x_4 &= 0 \\ x_1 + 3x_2 + x_3 + 2x_4 - 3x_5 &= 0 \\ x_3, x_4, x_5 &\geq 0. \end{aligned}$$

Our definition of the flux cone of a metabolic network only allows for homogeneous linear equations and non-negativity constraints for a subset of the variables. To transform this outer description of C *into a flux cone, we in-*

troduce three slack variables s_1, s_2, s_3. *This allows us to transform the non-homogeneous inequalities to homogeneous equations:*

$$\begin{aligned} 2x_1 + x_2 + x_2 - s_1 &= 0 \\ -x_2 + x_3 - s_2 &= 0 \\ x_2 + x_3 - s_3 &= 0 \\ x_1 + x_3 - x_4 &= 0 \\ x_1 + 3x_2 + x_3 + 2x_4 - 3x_5 &= 0 \\ s_1, s_2, s_3, x_3, x_4, x_5 &\geq 0. \end{aligned}$$

Note that the introduced slack variables correspond to additional irreversible reactions in a modified network $\mathcal{N}'$. *This network is described by the stoichiometric matrix* S *and the set of reversible reactions* Rev *defined as follows:*

$$S = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 2 & -3 & 0 & 0 & 0 \end{pmatrix}, \text{Rev} = \{1, 2\},$$

where reactions $6, 7$ *and* 8 *correspond to the introduced slack variables and are represented in the last three columns of* S. *A visualization of the corresponding metabolic network is given in Fig 5.2. Using* `efmtool`, *we determine a matrix* E *that contains the EFMs of* $\mathcal{N}$ *as rows:*

$$E = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 & 2 & 0 & 2 \\ -1 & 1 & 1 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 2 & 1 & 1 & 0 & 2 & 2 \\ 0 & -1 & 1 & 1 & 0 & 0 & 2 & 0 \end{pmatrix}$$

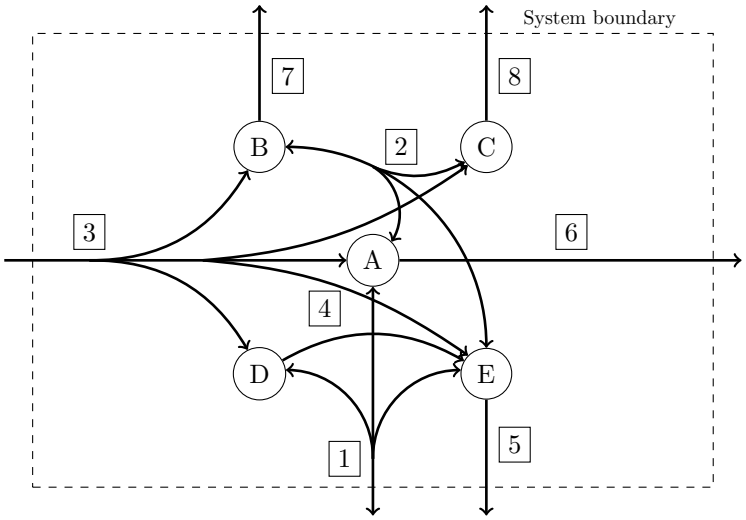

Figure 5.2: Visualization of the metabolic network of the counterexample candidate.

The dimension of the lineality space of this cone is zero. If we denote $E_{i,\star}$ *the i-th row of* E*,* $E_{1,\star}$ *corresponds to* e^1*,* $E_{3,\star}$ *to* e^2 *and* $E_{6,\star}$ *to* e^3*, before adding the slack variables, while* $E_{4,\star}$ *corresponds to* e^4 *(they have equal support when only considering the first 5 reactions and disregarding the additional irreversible reactions* $6, 7, 8$ *resulting from the introduction of the slack variables* s_1, s_2, s_3*). The degree of* $E_{1,\star}, E_{3,\star}$ *and* $E_{6,\star}$ *is 1. Thus* $E_{1,\star}, E_{3,\star}$ *and* $E_{6,\star}$ *are the extreme rays of* C' *and are not decomposable. The remaining EFMs are all decomposable into 2 EFMs:*

$$
\begin{aligned}
E_{2,\star} &= E_{1,\star} + E_{3,\star} \\
E_{4,\star} &= \frac{1}{2}(E_{2,\star} + E_{6,\star}) \\
E_{5,\star} &= E_{3,\star} + E_{6,\star}.
\end{aligned}
$$

Example 5.3.3 illustrates the main difficulty in the search for a counterexample to Conj. 5.2.1. While it is possible (at least without time or memory limitations) to compute a representative set of the EFMs for each metabolic

network defined by a stoichiometric matrix and a set of reversible reactions, it is hard (and potentially impossible) to determine a metabolic network that has a given set of EFMs. It should be noted that we did not prove that there is no metabolic network that has exactly the EFMs e^1, e^2, e^3 and e^4 from Example 5.3.3. Another metabolic network with exactly these EFMs could very well exist, but to the best of our knowledge, there is no way (other than our approach in Example 5.3.3) to find it or to prove that it does not exist.

5.4 Connections to matroid theory

In this section, we will explore the relationship between matroids and metabolic networks where all reactions are reversible. There are several different ways to define matroids in general, as well as different types of matroids. For more details and applications of matroids we refer to (Oxley, 2006) and (Welsh, 2010). We begin with a common definition of matroids via independent sets and an equivalent definition via circuits. Next, we show that the circuits of a matroid correspond to EFMs of metabolic networks where all reaction are reversible. Khachiyan *et al.* (2005) introduced an algorithm to enumerate the circuits of a matroid in incremental polynomial time (cf. Capelli and Strozecki, 2021, for more information about incremental polynomial time). We will discuss their proof and consequently show that EFMs in metabolic networks, where all reactions are reversible, can also be enumerated in incremental polynomial time. Finally, we will apply these insights to prove a weaker version of Conj. 5.2.1 for the special case that all reactions are reversible. Our notation and definitions are derived from (Oxley, 2006).

Definition 5.4.1. *Let E be a finite set of n elements, i.e. $|E| = n$, and let $\mathcal{I} \subseteq 2^E$ be a collection of subsets of E. $M = (E, \mathcal{I})$ is a matroid if and only if:*

(i) $\emptyset \in \mathcal{I}$,

(ii) if $X \in \mathcal{I}$ *and* $Y \subseteq X$ *then* $Y \in \mathcal{I}$,

(iii) if $X, Y \in \mathcal{I}$ *and* $|Y| > |X|$*, then there exists* $y \in Y \setminus X$ *such that* $X \cup \{y\} \in \mathcal{I}$*.*

The elements of $\mathcal{I}$ are called *independent sets of* M. Subsets of E that are not in $\mathcal{I}$ are then called *dependent sets*. We denote $\mathcal{C}(M)$ or just $\mathcal{C}$ the set of all minimal dependent sets of M and call it the set of *circuits* of M. A set $X \in 2^E$ is minimal dependent if removing any element from X leads to an independent set (i.e. $\forall X \in \mathcal{C}(M)$ and all $x \in X : X \setminus \{x\} \in \mathcal{I}$) (Oxley, 2006). Analogously, maximal independent sets in $\mathcal{I}$ are called *bases of* M and the set of all bases is denoted by $\mathcal{B}(M)$ or just $\mathcal{B}$.

The following definition of a matroid using circuits is equivalent to Def. 5.4.1 (Oxley, 2006).

Definition 5.4.2. *Let* $M := (E, \mathcal{C})$*, where* E *is a finite set and* $\mathcal{C} \subseteq 2^E$ *a set of subsets of* E*, which we will call the circuits of* M*.* $M = (E, \mathcal{C})$ *is a matroid if and only if:*

(C1) $\emptyset \notin \mathcal{C}$*,*

(C2) If $C_1, C_2 \in \mathcal{C}$ *and* $C_1 \subseteq C_2$*, then* $C_1 = C_2$*,*

(C3) If $C_1, C_2 \in \mathcal{C}$ *and* $e \in C_1 \cap C_2$*, there exists* $C_3 \in \mathcal{C}$ *such that* $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\}$*.*

Note the similarities to our definition of elementary flux modes. In the definition of EFMs we also excluded 0 similar to (C1) and the inclusionwise minimal support property (2.3) is similar to (C2). Before we take a closer look at the similarity of the *circuit axiom*(C3) and our definition of cancellations of reversible reactions (Def. 5.2.2) we show that every metabolic network where all reactions are reversible defines a matroid where the supports of the EFMs are the circuits of that matroid. Note that if all reactions of a metabolic network $\mathcal{N}$ are reversible, then for all $e \in \mathcal{E}$ there exists $\lambda > 0$ such that $e' = -\lambda e \in \mathcal{E}$ and $\text{supp}(e) = \text{supp}(e')$, i.e., the cardinality of $\mathcal{E}$ is twice the cardinality $\mathcal{U} := \{\text{supp}(e) \mid e \in \mathcal{E}\}$.

Proposition 5.4.3 (cf. Reimers, 2014)**.** *Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev} = \mathcal{R})$ be a metabolic network where all reactions are reversible and define $\mathcal{U} := \{\text{supp}(e) \mid e \in \mathcal{E}\}$ where $\mathcal{E}$ is a representative set of the EFMs of $\mathcal{N}$. Then, $M := (\mathcal{R}, \mathcal{U})$ is a matroid and $\mathcal{U}$ is the set of circuits of M.*

Proof.
(C1): By definition $v = 0$ is not an EFM of $\mathcal{N}$ and thus, $\text{supp}(v) = \emptyset \notin \mathcal{U}$.
(C2): Let $U_1, U_2 \in \mathcal{U}$ be the supports of two EFMs in $\mathcal{E}$ and let $U_1 \subseteq U_2$. Then $U_1 = U_2$ directly follows from equation (2.3).
(C3): Let $U_1, U_2 \in \mathcal{U}$ and $r \in U_1 \cap U_2$. Since all reactions in $\mathcal{R}$ are reversible there exist $e^1, e^2 \in \mathcal{E}$ such that $e^1_r < 0, e^2_r > 0$ and $\text{supp}(e^1) = U_1, \text{supp}(e^2) = U_2$. Define $e := \lambda_1 e^1 + \lambda_2 e^2$ as the cancellation of the reversible reaction r ($r \notin U := \text{supp}(e)$). Now if e is an EFM of $\mathcal{N}$, we are done because $U \subseteq (U_1 \cup U_2) \setminus \{r\}$. Otherwise, if $e \notin \mathcal{E}$, then $U \notin \mathcal{U}$. But then some subset $U' \subsetneq U$ is an element of $\mathcal{U}$ and we get $U' \subseteq (U_1 \cup U_2) \setminus \{r\}$. □

For the proof of (C3) we used the fact that the support of every steady-state flux vector that is not a EFM contains the support of an EFM which is a direct consequence of the inclusionwise minimal support property (2.3). If all reactions of a metabolic network $\mathcal{N}$ are reversible, the matroid $M = (\mathcal{R}, \mathcal{U})$ defined in Lemma 5.4.3 is called the *flux matroid* of $\mathcal{N}$ (Reimers, 2014).

The following two lemmas explain how one can check for a given set X, whether it is a circuit of a matroid (Lemma 5.4.4) and how a circuit C_3 fulfilling the circuit axiom (C3) can be found (Lemma 5.4.5).

Lemma 5.4.4. *Let f be an independence oracle of a matroid $M(E, \mathcal{C})$, i.e., $f : 2^E \mapsto \{True, False\}$ such that $f(X) = True$ if $X \in \mathcal{I}$ and $f(X) = False$ if $X \notin \mathcal{I}$. For $X \in 2^E$ it takes $|X| + 1$ calls to f to verify $X \in \mathcal{C}$.*

Proof. Let $f : 2^E \mapsto \{True, False\}$ be an independence oracle for $M = (E, \mathcal{C})$. By definition $X \in \mathcal{C}$ is a circuit of M, if X is a minimal dependent set in 2^E. We need to verify that $f(X) = False$ and for each $x \in X$ we need to verify that $f(X \setminus \{x\}) = True$. We call f a total of $|X| + 1$ times. □

Lemma 5.4.5. *Let $M = (E, \mathcal{C})$ be a matroid and $f : 2^E \mapsto \{True, False\}$ an independence oracle as defined in Lemma 5.4.4. Given two circuits $C_1, C_2 \in \mathcal{C}$ and an element $e \in C_1 \cap C_2$, a circuit $C_3 \in (C_1 \cup C_2) \setminus \{e\}$ can be found by calling f fewer than $\frac{(n+1)(n+2)}{2}$ times, where $n = |(C_1 \cup C_2) \setminus \{e\}|$.*

Proof. Define $C' := (C_1 \cup C_2) \setminus \{e\}$ and $n := |C'| = |(C_1 \cup C_2) \setminus \{e\}|$. By the circuit axiom (C3) there exists a circuit $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\} = C'$ and thus, C' cannot be independent. By Lemma 5.4.4, we can check whether C' is a circuit with $n + 1$ calls to f. If it is not, C' is not minimal dependent and we choose $x \in C'$ such that $f(C' \setminus \{x\}) = False$ (i.e., $C' \setminus \{x\}$ is dependent). Again by Lemma 5.4.4, it now takes $|C' \setminus \{x\}| + 1 = |C'| = n$ calls to f to test whether $C' \setminus \{x\} \in \mathcal{C}$. Repeating this process and successively removing elements for which the resulting set is still dependent, we find a circuit $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\}$ after fewer than $(n + 1) + n + (n - 1) + \cdots + 1 = \sum_{k=1}^{n+1} k = \frac{(n+1)(n+2)}{2}$ calls to f. □

Note that the bound $\frac{(n+1)(n+2)}{2}$ in Lemma 5.4.5 is not optimal. We would not have to check whether $C' \setminus \{x\}$ is dependent again after the first iteration (and subsequent sets).

Theorem 5.4.6 (Khachiyan *et al.,* 2005)**.** *The circuits $\mathcal{C}$ of a matroid $M = (E, \mathcal{C})$ can be enumerated in incremental polynomial time.*

In their proof, Khachiyan *et al.* (2005) start with a *fundamental system of circuits* $\mathcal{F}(B) := \{C(B, x) \mid x \in E \setminus B\}$ for some basis B of M. A *fundamental circuit* $C(B, x)$ of x for the basis B is defined as the unique circuit C such that $x \in C \subseteq B \cup \{x\}$ for $x \in E \setminus B$ (Oxley, 2006).

To enumerate all circuits of M they begin by initializing $\mathcal{C}' = \mathcal{F}(B)$ and repeatedly check whether $\mathcal{C}'$ is closed with respect to the circuit axiom. For each violation of the circuit axiom, a new circuit is produced and added to $\mathcal{C}'$ (cf. Lemma 5.4.5). They finalize their proof by showing that $\mathcal{C}'$ is the set of all circuits of M as soon as it is closed with respect to the circuit axiom.

We will refer to this method of generating new circuits as *circuit enumeration method* for the remainder of this chapter.

As a direct consequence we get:

Corollary 5.4.7. *(cf. Theorem 10 in (Acuña* et al., *2009)) The supports of the EFMs of a metabolic network where all reactions are reversible can be enumerated in incremental polynomial time.*

The following example is an application of the circuit enumeration method to a concrete example.

Example 5.4.8. *It is a well-known fact (cf. (Oxley, 2006)) that given a matrix $A \in \mathbb{R}^{m\times n}$, the set of column labels E is the ground set of a matroid $M = (E, \mathcal{I})$, where $\mathcal{I}$ is the set of subsets X of E for which the multiset of columns labeled by X is linearly independent. This matroid is called a* vector matroid. *Consider the matrix*

$$A = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 & -1 \end{pmatrix} \end{array} \in \mathbb{R}^{2\times 5},$$

with column labels $E := \{1, 2, 3, 4, 5\}$. One can easily verify that

$$\begin{aligned} \mathcal{I} = \{&\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \\ &\{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\}. \end{aligned}$$

The dependent sets are the elements of $2^E \setminus \mathcal{I}$. Furthermore the set of bases of M is given by

$$\mathcal{B} = \{\{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\},$$

and the set of circuits of M by

$$\mathcal{C} = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}, \{3, 4, 5\}\}.$$

If the circuit enumeration method described in (Khachiyan et al., *2005) chooses the basis* $B^0 = \{1, 3\}$ *in the first step, the system* $\mathcal{F}(B^0)$ *of fundamental circuits for* B^0 *is given by*

$$\mathcal{F}(B^0) = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}\},$$

because $\{1, 2\} \subseteq B^0 \cup \{2\}$, $\{1, 3, 4\} \subseteq B^0 \cup \{4\}$ *and* $\{1, 3, 5\} \subseteq B^0 \cup \{5\}$. *To check whether* $\mathcal{F}(B^0)$ *is closed with respect to the circuit axiom, we compare each pair of circuits that share an element. We easily find* $1 \in C_1 := \{1, 2\}$ *and* $1 \in C_2 := \{1, 3, 4\}$ *but no circuit* $C_3 \subseteq (C_1 \cup C_2) \setminus \{1\} = \{2, 3, 4\}$. *This produced the new circuit* $\{2, 3, 4\}$. *Analogously, we find*

$$\begin{aligned}
\{2, 3, 5\} &\subseteq (\{1, 2\} \cup \{1, 3, 5\}) \setminus \{1\},\\
\{3, 4, 5\} &\subseteq (\{1, 3, 4\} \cup \{1, 3, 5\}) \setminus \{1\},\\
\{1, 4, 5\} &\subseteq (\{1, 3, 4\} \cup \{1, 3, 5\}) \setminus \{3\}.
\end{aligned}$$

The newly produced circuits lead to another violation of the circuit axiom ($\{2, 3, 4\}$ *and* $\{2, 3, 5\}$*) producing the circuit* $\{2, 4, 5\}$. *Note that we have now enumerated all circuits in* $\mathcal{C}$ *and that there are no more violations of the circuit axiom.*

Let us take a look at the metabolic network that has Example 5.4.8 as flux matroid. To do this we define a metabolic network where the matrix A from Example 5.4.8 is the stoichiometric matrix S and assume that all reactions are reversible.

Example 5.4.9. *Consider the metabolic network* $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S\,\mathrm{Rev})$, *defined by*

$$A = S = \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 & -1 \end{pmatrix} \text{ and } \mathrm{Rev} = \{1, 2, 3, 4, 5\}.$$

We compute a matrix E *that contains a representative set of EFMs of* $\mathcal{N}$

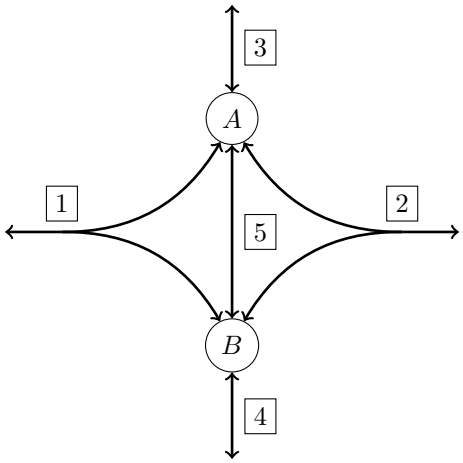

Figure 5.3: Metabolic network defined in Example 5.4.9.

as rows:

$$E = \begin{pmatrix} 0 & -1 & 0 & 4 & 2 \\ 0 & 0 & -1 & 1 & 1 \\ -1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & -1 & 4 & 0 & -2 \\ 0 & 1 & -4 & 0 & 2 \\ -1 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & -4 & -2 \\ 0 & -1 & 2 & 2 & 0 \\ -1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -2 & -2 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 1 \\ 1 & 0 & 0 & -2 & -1 \\ 1 & 0 & -1 & -1 & 0 \end{pmatrix}$$

Note that reversible EFMs appear in both orientations and have equal support. The set $\mathcal{U}$ of supports of EFMs of $\mathcal{N}$ is given by

$$\mathcal{U} = \{\{2,4,5\},\{3,4,5\},\{1,4,5\},\{2,3,5\},\{1,3,5\},\{2,3,4\},\{1,3,4\},\{1,2\}\}$$

and coincides with the set of circuits $\mathcal{C}$ from Example 5.4.8. Furthermore, we mention that all EFMs of $\mathcal{N}$ have degree 3 and can be non-trivially decomposed into two other EFMs, confirming Conj. 5.2.1:

$$\begin{aligned}
\begin{pmatrix} 2 & -1 & 0 & 0 & 0 \end{pmatrix} &= \begin{pmatrix} 0 & -1 & 2 & 2 & 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 1 & 0 & -1 & -1 & 0 \end{pmatrix} \\
\begin{pmatrix} 1 & 0 & 0 & -2 & -1 \end{pmatrix} &= \begin{pmatrix} -1 & 0 & 2 & 0 & -1 \end{pmatrix} + 2 \cdot \begin{pmatrix} 1 & 0 & -1 & -1 & 0 \end{pmatrix} \\
\begin{pmatrix} 1 & 0 & -1 & -1 & 0 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & -2 & 0 & 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 & 0 & 1 & -1 & -1 \end{pmatrix} \\
\begin{pmatrix} 1 & 0 & -2 & 0 & 1 \end{pmatrix} &= \begin{pmatrix} 0 & 0 & -1 & 1 & 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 & 0 & -1 & -1 & 0 \end{pmatrix} \\
\begin{pmatrix} -1 & 0 & 1 & 1 & 0 \end{pmatrix} &= \begin{pmatrix} 0 & 0 & -1 & 1 & 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 & 0 & 2 & 0 & -1 \end{pmatrix} \\
\begin{pmatrix} 0 & -1 & 2 & 2 & 0 \end{pmatrix} &= \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 & 0 & 1 & 1 & 0 \end{pmatrix} \\
\begin{pmatrix} -1 & 0 & 2 & 0 & -1 \end{pmatrix} &= \begin{pmatrix} -1 & 0 & 1 & 1 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 & 0 & 1 & -1 & -1 \end{pmatrix} \\
\begin{pmatrix} 0 & -1 & 4 & 0 & -2 \end{pmatrix} &= \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 & 0 & 2 & 0 & -1 \end{pmatrix} \\
\begin{pmatrix} 0 & 0 & 1 & -1 & -1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 & -2 & -1 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 & 0 & 1 & 1 & 0 \end{pmatrix} \\
\begin{pmatrix} 0 & 0 & -1 & 1 & 1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & -1 & -1 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 & 0 & 0 & 2 & 1 \end{pmatrix} \\
\begin{pmatrix} -1 & 0 & 0 & 2 & 1 \end{pmatrix} &= \begin{pmatrix} -1 & 0 & 1 & 1 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 & 0 & -1 & 1 & 1 \end{pmatrix} \\
\begin{pmatrix} 0 & -1 & 0 & 4 & 2 \end{pmatrix} &= \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 & 0 & 0 & 2 & 1 \end{pmatrix} \\
\begin{pmatrix} 0 & 1 & -4 & 0 & 2 \end{pmatrix} &= \begin{pmatrix} 0 & -1 & 0 & 4 & 2 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 & 1 & -2 & -2 & 0 \end{pmatrix} \\
\begin{pmatrix} 0 & 1 & -2 & -2 & 0 \end{pmatrix} &= \begin{pmatrix} 0 & 1 & -4 & 0 & 2 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 & 0 & 1 & -1 & -1 \end{pmatrix} \\
\begin{pmatrix} 0 & 1 & 0 & -4 & -2 \end{pmatrix} &= \begin{pmatrix} 0 & -1 & 4 & 0 & -2 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 & 1 & -2 & -2 & 0 \end{pmatrix} \\
\begin{pmatrix} -2 & 1 & 0 & 0 & 0 \end{pmatrix} &= \begin{pmatrix} 0 & 1 & -4 & 0 & 2 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 & 0 & 2 & 0 & -1 \end{pmatrix}.
\end{aligned}$$

Next, we further analyze metabolic networks in the context of matroid theory to prove a weaker version of Conj. 5.2.1. First we establish a relationship between decomposability of EFMs and fundamental circuits of the flux matroid. Then we prove a result about the uniqueness of circuits fulfilling the circuit axiom. Combining these results lets us prove Theorem 5.4.13, as a weaker version of Conj. 5.2.1, that applies to metabolic networks where all reactions are reversible.

Proposition 5.4.10. *Let $M = (\mathcal{R},\mathcal{U})$ be the flux matroid of a metabolic network $\mathcal{N}$ where all reactions are reversible and $\mathcal{B}$ the set of bases of M. If $U \in \mathcal{U}$ is the support of a decomposable EFM e of $\mathcal{N}$, then there exists a basis $B \in \mathcal{B}$ such that U is not in the system of fundamental circuits $\mathcal{F}(B)$ for B.*

Proof. We will show that an EFM e, for which $\text{supp}(e)$ is in the system of fundamental circuits for every basis of the flux matroid, cannot be decomposable.

First, recall the definition of the system of fundamental circuits for a basis $B \in \mathcal{B}$:

$$\mathcal{F}(B) := \{C(B,x) \mid x \in \mathcal{R} \setminus B\},$$

where $C(B,x)$ is the unique fundamental circuit of x for B such that $C(B,x) \subseteq B \cup \{x\}$ (Oxley, 2006). Let $\mathcal{N}$ be a metabolic network where all reactions are reversible and M the flux matroid of $\mathcal{N}$. Furthermore let $\mathcal{B}$ be the set of all bases of M and let e be an EFM of $\mathcal{N}$ such that for every $B \in \mathcal{B}$ we have $\text{supp}(e) =: U \in \mathcal{F}(B)$. Secondly, recall that every circuit is a fundamental circuit for some basis B (cf. (Oxley, 2006)). Specifically, for all $C \in \mathcal{U}$ and for every $x \in C$, there exists $B \in \mathcal{B}$ such that $C = C(B,x)$ (cf. (Oxley, 2006)). With this, we can show that U is uniquely determined by any of its elements, i.e.,

$$\forall U' \in \mathcal{U} : U' \cap U \neq \emptyset \Rightarrow U' = U. \tag{5.3}$$

To see this, suppose there exists a circuit $U' \neq U$ with $U' \cap U \neq \emptyset$ and choose $r \in U' \cap U$. Then $r \in U'$ implies that there there exists a basis B

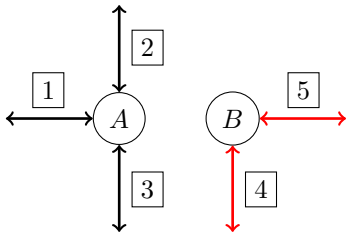

Figure 5.4: Example of an EFM (in red) that is not decomposable in a network where all reactions are reversible.

such that $U' = C(B, r)$ is the fundamental circuit of r for B and $r \notin B$ follows immediately. By our assumption $U \in \mathcal{F}(B)$ is also in the system of fundamental circuits for B. Since $r \in U, r \notin B$ we can conclude $U = C(B, r)$ is the unique fundamental circuit of r for B and thus, $U = U'$. But then

$$e = \sum_{i=1}^{k} \lambda_i e^i \tag{5.4}$$

cannot be a non-trivial decomposition of e into other EFMs. Define $U_i := \text{supp}(e^i), i = 1, \ldots, k$. By construction we get

$$U = \text{supp}(e) \subseteq \bigcup_{i=1,\ldots,k} \text{supp}(e^i) = \bigcup_{i=1,\ldots,k} U_i.$$

For every $r \in U$ there exists $i \in \{1, \ldots, k\}$ such that $r \in U_i$ and because of (5.3) we get $\text{supp}(e) = U = U_i = \text{supp}(e^i)$ and therefore, (5.4) is not a non-trivial decomposition of e into other EFMs. □

Note that EFMs that are not decomposable can occur in metabolic networks where all reactions are reversible. The following example illustrates this.

Example 5.4.11. *Assuming all stoichiometric coefficients in the metabolic network visualized in Figure 5.4 are in* $\{-1, 0, 1\}$ *and reactions are oriented*

from left to right and top to bottom, the stoichiometric matrix S is given by

$$S = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \end{array},$$

the vector matroid defined by S has 6 bases

$$\mathcal{B} = \{\{1,4\},\{1,5\},\{2,4\},\{2,5\},\{3,4\},\{3,5\}\}$$

and 4 circuits

$$\mathcal{C} = \{\{1,2\},\{1,3\},\{2,3\},\{4,5\}\}.$$

One can easily verify that $\{4,5\}$ is a fundamental circuit for every basis $B \in \mathcal{B}$:

$$\begin{aligned} \mathcal{F}(\{1,4\}) &= \{\{1,2\},\{1,3\},\{4,5\}\}, \\ \mathcal{F}(\{1,5\}) &= \{\{1,2\},\{1,3\},\{4,5\}\}, \\ \mathcal{F}(\{2,4\}) &= \{\{1,2\},\{2,3\},\{4,5\}\}, \\ \mathcal{F}(\{2,5\}) &= \{\{1,2\},\{2,3\},\{4,5\}\}, \\ \mathcal{F}(\{3,4\}) &= \{\{1,3\},\{2,3\},\{4,5\}\}, \\ \mathcal{F}(\{3,5\}) &= \{\{1,3\},\{2,3\},\{4,5\}\}. \end{aligned}$$

Proposition 5.4.12. *Let $M = (E, \mathcal{C})$ be a matroid with set of circuits $\mathcal{C}$ and set of bases $\mathcal{B}$. If a circuit $C \in \mathcal{C}$ is not contained in the system of fundamental circuits $\mathcal{F}(B)$ for some $B \in \mathcal{B}$, then there exist distinct circuits $C_1, C_2 \in \mathcal{C}$ and $e \in C_1 \cap C_2$ such that C is the unique circuit with $C \subseteq (C_1 \cup C_2) \setminus \{e\}$.*

Proof. Let $C \in \mathcal{C}$ be a circuit and B be a basis of M such that $C \notin \mathcal{F}(B)$. By Khachiyan *et al.* (2005), if some system $\mathcal{C}' \subseteq \mathcal{C}$ of circuits is closed with respect to the circuit axiom (i.e., for all distinct $C_1, C_2 \in \mathcal{C}'$ and $e \in C_1 \cap C_2$ there exists $C_3 \in \mathcal{C}'$ such that $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\}$) and $\mathcal{F}(B) \subseteq \mathcal{C}'$, then $\mathcal{C}' = \mathcal{C}$.

Suppose there do not exist distinct circuits C_1, C_2 and $e \in C_1 \cap C_2$ such that C is the unique circuit contained in $(C_1 \cup C_2) \setminus \{e\}$, i.e., for all $C_1, C_2 \in \mathcal{C}$ and $e \in C_1 \cap C_2$ with $C \subseteq (C_1 \cup C_2) \setminus \{e\}$ there exists another circuit $C' \neq C \in \mathcal{C}$ such that $C' \subseteq (C_1 \cup C_2) \setminus \{e\}$. But then $\mathcal{C}' := \mathcal{C} \setminus \{C\}$ is closed with respect to the circuit axiom and $\mathcal{F}(B) \subseteq \mathcal{C}'$ because $C \notin \mathcal{F}(B)$, but $\mathcal{C}' \neq \mathcal{C}$, which is a contradiction. □

With Prop. 5.4.12 and Prop. 5.4.10 we can prove our weaker version of Conj. 5.2.1 for networks where all reactions are reversible.

Theorem 5.4.13. *Let $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \mathrm{Rev})$ be a metabolic network where all reactions are reversible and $\mathcal{E}$ a representative set of the EFMs of $\mathcal{N}$.*

If $e \in \mathcal{E}$ is decomposable, then there exist EFMs $e^1, e^2 \in \mathcal{E}$ with $\mathrm{supp}(e^1) \neq \mathrm{supp}(e^2) \neq \mathrm{supp}(e)$, $\lambda_1, \lambda_2 > 0$ and $r \in \mathrm{supp}(e^1) \cap \mathrm{supp}(e^2)$, such that $v := \lambda_1 e^1 + \lambda_2 e^2$ yields $v_r = 0$, $e^1_r \neq 0, e^2_r \neq 0$ and

$$\mathrm{supp}(e) \subseteq \mathrm{supp}(v) \subseteq (\mathrm{supp}(e^1) \cup \mathrm{supp}(e^2)) \setminus \{r\}. \tag{5.5}$$

Proof. Let $M = (\mathcal{R}, \mathcal{U})$ be the flux matroid of $\mathcal{N}$ with set of bases $\mathcal{B}$, set of circuits $\mathcal{U}$ and let $e \in \mathcal{E}$ be decomposable. By Prop. 5.4.10 there exists a basis $B \in \mathcal{B}$ such that $U := \mathrm{supp}(e)$ is not in the system of fundamental circuits $\mathcal{F}(B)$.

By Prop 5.4.12 there exist distinct circuits $U_1, U_2 \in \mathcal{U}$ and $r \in U_1 \cap U_2$ such that $U \subseteq (U_1 \cup U_2) \setminus \{r\}$ is the unique circuit contained in $(U_1 \cup U_2) \setminus \{r\}$.

Since all reactions are reversible we can choose $e^1, e^2 \in \mathcal{E}$ such that $e^1_r > 0, e^2_r < 0$ and $\mathrm{supp}(e^1) = U_1, \mathrm{supp}(e^2) = U_2$. Define $\lambda_1, \lambda_2 > 0$ such that $v := \lambda_1 e^1 + \lambda_2 e^2$ yields $v_r = 0$ and thus $V := \mathrm{supp}(v) \subseteq (U_1 \cup U_2) \setminus \{r\}$.

By construction $v \neq 0$ is a feasible flux vector and by (2.3) there exists $e' \in \mathcal{E}$ with $\mathrm{supp}(e') \subseteq V$ which implies $U' := \mathrm{supp}(e') \in \mathcal{U}$. Since U is the unique circuit contained in $(U_1 \cup U_2) \setminus \{r\}$ we get $U = U'$ and (5.5) follows.

□

Note that v was constructed as a cancellation of the reversible reaction r. This result is weaker than Conj. 5.2.1 because we proved that the support of every decomposable EFM is contained in the support of a vector that results from a cancellation of a reversible reaction by two other EFMs. If we had equality instead of set inclusions in (5.5), this would prove Conj. 5.2.1 for metabolic networks where all reactions are reversible.

We refer to Arne Reimers' PhD thesis (Reimers, 2014) for a more detailed discussion of metabolic networks that include irreversible reactions and their representation via oriented matroids.

As of now, we only know that Conj. 5.2.1 holds for metabolic networks with fewer than two reversible reactions (cf. Lemma 5.3.2) as well as for EFMs of degree at most 2 by Prop. 4.2.2.

5.5 Algorithms for composition and decomposition

In this section, we discuss two algorithms based on our observations from the previous sections. In Section 5.4, we discussed the circuit enumeration method, which begins with the system of fundamental circuits for a basis and repeatedly checks whether the set is closed with respect to the circuit axiom. Whenever a violation of the circuit axiom is found, a new circuit is produced.

We begin with Algorithm 1, which generates EFMs from a starting set by repeatedly checking if new EFMs can be generated by positive combinations of two EFMs that are cancellations of reversible reactions. For better readability we will call a positive combination of two EFMs e^1, e^2 that results in a new EFM e, i.e.

$$e = \lambda_1 e^1 + \lambda_2 e^2, \lambda_1, \lambda_2 > 0, \operatorname{supp}(e^1) \neq \operatorname{supp}(e^2) \neq \operatorname{supp}(e),$$

a *composition of e by the EFMs e^1 and e^2*. We can then say that a set of

EFMs $\mathcal{E}'$ of a metabolic network $\mathcal{N}$ is *closed with respect to compositions by two EFMs*, if for all $e^1, e^2 \in \mathcal{E}'$ for which there exist $\lambda_1, \lambda_2 > 0$ such that $e := \lambda_1 e^1 + \lambda_2 e^2$ is an EFM of $\mathcal{N}$, we have $\lambda e \in \mathcal{E}'$ for some $\lambda > 0 \in \mathbb{R}$. This means that no two EFMs in $\mathcal{E}'$ can be positively combined to an EFM that is not represented by an EFM in $\mathcal{E}'$. Note that trivially, a representative set $\mathcal{E}$ of the EFMs of a metabolic network $\mathcal{N}$ is closed with respect to compositions by two EFMs.

Algorithm 1 EFM composition

1: **Input**: A subset $\mathcal{E}'$ of EFMs of a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$.
2: **OUTPUT**: Set of EFMs $\mathcal{E}'$ that is closed with respect to compositions by 2 EFMs.
3: **while** $\mathcal{E}'$ is not closed with respect to compositions by 2 EFMs **do**
4: **for** each reversible reaction $r \in \text{Rev}$ **do**
5: Determine $P := \{e \in \mathcal{E}' \mid e_r > 0\}$ and $N := \{e \in \mathcal{E}' \mid e_r < 0\}$.
6: **for** each pair of EFMs $e^+ \in P$ and $e^- \in N$ **do**
7: Define new vector $e' = \lambda_1 e^+ + \lambda_2 e^-$ such that $e'_r = 0$.
8: **if** $\lambda e' \notin \mathcal{E}'$ for all $\lambda \in \mathbb{R}$ and e' is an EFM of $\mathcal{N}$ **then**
9: Add e' to $\mathcal{E}'$.
10: **end if**
11: **end for**
12: **end for**
13: **end while**

Algorithm 1 is initialized with a starting set $\mathcal{E}'$ of EFMs of a metabolic network $\mathcal{N}$ and generates new EFMs that are added to $\mathcal{E}'$ until the set is closed with respect to compositions by 2 EFMs. The algorithm can be viewed as a variant of the circuit enumeration method for matroids discussed in Section 5.4. The main difference is that instead of working with circuits of a matroid (i.e. the supports of EFMs) we are working with EFMs in the flux cone C of the metabolic network. We determine new EFMs by performing cancellations of

reversible reactions instead of finding a new circuit for each violation of the circuit axiom (cf. Lemma 5.4.5).

In line 4, we iterate over each reversible reaction $r \in \text{Rev}$. In line 5, we determine two subsets P, N of the set of EFMs $\mathcal{E}'$ that we currently have. We sort EFMs with positive flux on r to P and EFMs with negative flux on r to N. EFMs with no flux on reaction r (every EFM $e \in \mathcal{E}'$ with $e_r = 0$) are omitted in this step. In lines 6 and 7 we positively combine every pair of one EFM from P and one EFM from N such that this positive combination is a cancellation of the reversible reaction r. Finally, in lines 8 and 9 we check whether this cancellation leads to a new EFM of $\mathcal{N}$ and if so, we add it to $\mathcal{E}'$. We verify $\lambda e' \notin \mathcal{E}'$ so that we don't have multiple EFMs with equal support. By Prop. 2.5.1, it suffices to compare the support of e' and every EFM $e \in \mathcal{E}'$. To verify whether e' is an EFM of $\mathcal{N}$, the rank test (cf. (2.5)) can be applied.

First, we mention that Algorithm 1 is a rather theoretical approach. It is not supposed to compete with the state-of-the-art tools in terms of running time. The main disadvantage is the large amount of pairs of EFMs that are combined in each iteration. Furthermore, the algorithm requires a subset of the EFMs as a starting set. A viable candidate would be a MEMo, which can be computed for networks where the enumeration of all EFMs is not possible (Röhl and Bockmayr, 2019). If Conj. 5.2.1 holds, the algorithm could be optimized to enumerate representative sets of EFMs and tested against the other methods mentioned in Section 2.6. Clearly, if the conjecture does not hold and a metabolic network $\mathcal{N}$ is a counterexample, EFMs that can only be decomposed into 3 or more EFMs will not be found by Algorithm 1. If this is the case, Algorithm 1 can still be used to find additional EFMs, when some starting subset of EFMs is given.

Independently of whether the conjecture holds or not, we also mention a couple of advantages of the algorithm. First of all, in contrast to approaches based on the double description method it is not a pass or fail algorithm. This means, the algorithm can be interrupted at any point and $\mathcal{E}'$ represents a subset

of EFMs. Secondly, the parallelization possibilities exceed those of the MILP approach (de Figueiredo *et al.,* 2009). For example, in line 4 every reversible reaction could be handled simultaneously. Additionally each pair of one EFM from P and one from N in line 6 could be combined to a new candidate simultaneously. In the MILP approach, finding the k-th shortest EFM requires finding the $(k-1)$-th shortest EFM first. Apart from the parallelization aspect, it should be noted that the algorithm does not have to be initialized with a MEMo. It can also be used to determine targeted subsets of EFMs. To name just one example, the algorithm could be initialized with a generating set of some specific face F of the flux cone (cf. Chapter 4).

Several steps of the computation can and should be optimized when implementing the algorithm for real application. For example in lines 4 and 5 every combination is reproduced in every iteration. This means that every new EFM that is found in some iteration, will be found again in each iteration afterwards. This can be optimized by only combining pairs of EFMs from P and N that have not been considered before. The rank test (cf. Chapter 2) that is performed in line 8 is computationally expensive and repeated for each newly generated vector. One possible improvement here could be a support comparison of the newly found vector e' with the current set of EFMs $\mathcal{E}'$. If there already is an EFM with inclusionwise smaller support in $\mathcal{E}'$, e' cannot be an EFM of $\mathcal{N}$ and the rank test does not need to be executed. Another possible optimization is to check whether an EFM is reversible before it is added to $\mathcal{E}'$ and if it is, one could directly add both orientations.

Finally, by Theorem. 5.4.13, the algorithm can be modified such that a representative set of the EFMs of a metabolic network $\mathcal{N}$, where all reactions are reversible, is enumerated when initialized with a MEMo. We need to adjust lines 8 and 9 with finding an EFM with inclusionwise smaller support than e'.

Corollary 5.5.1. *If Algorithm 1 is initialized with a MEMo $\mathcal{W}$ of a metabolic network $\mathcal{N}$ where all reactions are reversible and lines 8,9 are replaced with*

8: *Find EFM e such that* $\text{supp}(e) \subseteq \text{supp}(e')$.

9: *Add e to $\mathcal{E}'$.*

the algorithm terminates with a representative set $\mathcal{E}$ of the EFMs of $\mathcal{N}$.

The support U of such an EFM e in line 8 can be found as described in Lemma 5.4.5. The rank test (cf. Prop 2.5.3) can be applied instead of the calls to an independence oracle. Finally by solving a system of linear equations (cf. Cor. 2.5.2) an EFM with support U can be found and added to $\mathcal{E}'$. Alternatively, a shortest EFM e with $\text{supp}(e) \subseteq \text{supp}(e')$ can be found by solving a mixed integer linear program (de Figueiredo *et al.,* 2009).

Proof. Every EFM that is not decomposable is already in $\mathcal{E}_{start} = \mathcal{W}$ by the definition of a MEMo.

Define $\mathcal{U}' := \{\text{supp}(e) \mid e \in \mathcal{E}'\}$ as the set of supports of the EFMs in $\mathcal{E}'$ after the modified algorithm terminated. It suffices to show that $\mathcal{U}'$ is closed with respect to the circuit axiom. Suppose it is not. Then there exist $U_1, U_2 \in \mathcal{U}$ and $r \in U_1 \cap U_2$ such that $(U_1 \cup U_2) \setminus \{r\}$ does not contain a circuit. Since every reaction is reversible there exist EFMs $e^1, e^2 \in \mathcal{E}'$ such that $\text{supp}(e^1) = U_1, \text{supp}(e^2) = U_2$ and $e^1_r > 0$,$e^2_r < 0$. Now $\lambda_1, \lambda_2 > 0$ can be chosen such that $v := \lambda_1 e^1 + \lambda_2 e^2$ yields $v_r = 0$ and thus $\text{supp}(v) \subseteq (U^1 \cup U^2) \setminus \{r\}$. In the modified lines 8 and 9 an EFM e with inclusionwise smaller support than v was added to $\mathcal{E}'$ in an iteration over reaction r and thus $U := \text{supp}(e) \subseteq \text{supp}(v) \subseteq (U_1 \cup U_2) \setminus \{r\}$ and $U \in \mathcal{U}$ contradicts our assumption. □

The next example illustrates Algorithm 1 on our network from Figure 2.4.

Example 5.5.2. *Consider the metabolic network in Figure 2.4 again. A*

MEMo $\mathcal{W}$ of this network is given by the rows of the matrix

$$W = \begin{array}{c} \\ W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \\ W_7 \end{array} \begin{pmatrix} 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where W_i denotes the i-th row of W. When we iterate over the reversible reactions $\mathrm{Rev} = \{1, 3, 4, 5, 9, 10, 11, 12\}$ *we find that reaction* 1 *cannot be canceled by any combinations of rows of W. Leaving out cancellations of reactions that do not lead to new EFMs, we find that reaction* 3 *can be canceled by*

$$W_3 + W_7 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} =: W_8,$$

which is another EFM of the metabolic network. For reaction 4 we find

$$W_5 + W_7 = \begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} =: W_9,$$
$$W_6 + W_7 = \begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} =: W_{10}.$$

We find for reactions 9 and 10

$$W_4 + W_5 = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} =: W_{11},$$

and for reactions 11 and 12

$$W_1 + W_4 = \begin{pmatrix} 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \end{pmatrix} =: W_{12},$$
$$W_2 + W_3 = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} =: W_{13}.$$

We now append to W the new EFMs $W_8, \ldots, W_{12}$ and repeat the process. Again reaction 1 cannot be canceled but we find new cancellations for reactions 3,4,11 and 12:

$$
\begin{aligned}
W_2 + W_8 &= \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} =: W_{14}, \\
W_3 + W_{10} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} =: W_{15} \\
W_6 + W_8 &= \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} =: W_{16} \\
W_7 + W_{11} &= \begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} =: W_{17} \\
W_1 + W_{11} &= \begin{pmatrix} 0 & 0 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} =: W_{18}
\end{aligned}
$$

Again after appending these to our current set of EFMs we find in the next iterations

$$
\begin{aligned}
W_1 + W_{17} &= \begin{pmatrix} 1 & 1 & -2 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} =: W_{19}, \\
W_2 + W_{15} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} =: W_{20}, \\
W_3 + W_{17} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 2 & 2 \end{pmatrix} =: W_{21},
\end{aligned}
$$

and obtain a representative set of the EFMs of the metabolic network (cf. Röhl and Bockmayr, 2019 or Larhlimi and Bockmayr, 2008).

In Chapter 4, we discussed decompositions of flux vectors into EFMs. We presented an approach to determine the minimal face F a given flux vector v^* is contained in. We also discussed how a representative set $\mathcal{E}_F$ of the EFMs of F can be determined and showed that non-trivial decompositions of v^* only use EFMs from $\mathcal{E}_F$. Due to our observations in this chapter, we derive an algorithm based on the cancellation of reversible reactions to decompose EFMs into two other EFMs. Algorithm 2 can be seen as the inverse algorithm to Algorithm 1. While in Algorithm 1 we generate new EFMs by combining pairs from a set of given EFMs, Algorithm 2 searches for decompositions of a given EFM by identifying the cancellation of a reversible reversible that leads to it. This

means that, other than in Chapter 4, no mixed-integer linear programs have to be solved in the decomposition of an EFM by Algorithm 2. To the best of our knowledge it is a novel method to decompose a given EFM into two other EFMs.

Algorithm 2 Decomposition of one EFM into two EFMs

1: **INPUT**: Representative set $\mathcal{E}$ of a metabolic network $\mathcal{N}$, set of reversible reactions Rev of $\mathcal{N}$, and target EFM e that is to be decomposed.
2: **OUTPUT**: Pair (e^+, e^-) of EFMs that can be positively combined to e if such a pair exists.
3: **for** every reversible reaction $r \in \text{rev.supp}(e)$ **do**
4: Separate $\mathcal{E}$ into P and N depending on their flux on reaction r (cf. Algorithm 1)
5: **for** each pair of one EFM $e^+ \in P$ and one EFM $e^- \in N$ **do**
6: Determine $\lambda_+, \lambda_- > 0$ such that $e' := \lambda_+ e^+ + \lambda_- e^-$ yields $e'_r = 0$.
7: **if** $\text{supp}(e') = \text{supp}(e)$ **then return** (e^+, e^-).
8: **end if**
9: **end for**
10: **end for**

In lines 4 and 5, we iterate over every reversible reaction $r \in \text{rev.supp}(e)$ and separate $\mathcal{E}$ into P and N depending on their flux on reaction r like in Algorithm 1. We then perform each possible cancellation of the reversible reaction r in lines 5 and 6 and finally, we check if any resulting vector e' has support equal to the support of e in line 7. Since EFMs are uniquely determined by their support (cf. Prop. 2.5.1), this implies that e' can be decomposed into e^+ and e^-.

Since every decomposition of an EFM e is a cancellation of at least one reversible reaction (cf. Lemma 5.2.5) and we test every cancellation of each reversible reaction by two other EFMs, Algorithm 2 finds one of these decompositions, if there exists one. The algorithm can be adjusted to not terminate

in line 7, when the first decomposition of length 2 is found and collect all decompositions of length 2 instead. Independently of whether Conj. 5.2.1 holds or not, Algorithm 2 finds all decompositions of length 2 for a given EFM e. If the algorithm does not find one, we know there is none, giving us a candidate for a counterexample for Conj. 5.2.1. We still have to verify that the EFM for which we did not find a decomposition of length 2 is decomposable into 3 or more EFMs. In that case one can determine a shortest decomposition of e with the MILP approach (cf. Chapter 4). If a decomposition of length larger than 2 is found this way, we would have found a counterexample that disproves Conj. 5.2.1.

Note that Algorithm 2 can be significantly improved by only checking cancellations of reversible reactions by 2 EFMs that are in the same face of the flux cone that e is contained in. This can be done by finding a subset $\mathcal{E}_e \subseteq \mathcal{E}$ defined as $\mathcal{E}_e := \{e' \in \mathcal{E} \mid \text{irr.supp}(e') \subseteq \text{irr.supp}(e)\}$.

With this improvement, we were able to find decompositions of length 2 for all decomposable EFMs of the e_coli_core network from the BiGG (Norsigian *et al.,* 2020) database as well as for all networks from the KEGG (Kanehisa and Goto, 2000) database that are discussed in (Larhlimi and Bockmayr, 2009). It should be noted that Algorithm 2 is not suitable to find decompositions of flux vectors that are not EFMs, since there does not have to be a cancellation of a reversible reaction. Furthermore, the relatively small numbers of EFMs in specific faces (cf. Table 4.1) indicate how much of an improvement the reduction of the set of all EFMs to EFMs from the same face can be.

5.6 Searching for counterexamples

Since we have not been able to prove Conj. 5.2.1 we have to consider the possibility that it is not true and that there exists a counterexample as defined in Def. 5.3.1. In this section, we briefly discuss our approaches to find a counterexample to Conj. 5.2.1. We begin by presenting an Algorithm to

test whether the conjecture holds for a given metabolic network $\mathcal{N}$. Next, we discuss a method to generate all networks of a predefined size (number of metabolites and reactions) with restrictions to the possible stoichiometric coefficients. Clearly, if we do not impose restrictions on the coefficients that can appear in the stoichiometric matrix, i.e., if we allow the entries to be any number in $\mathbb{R}$, then the number of metabolic networks of any size is infinite.

Algorithm 3 Check conjecture

1: **INPUT**: Metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$.
2: **OUTPUT**: **True** if Conj. 5.2.1 holds for $\mathcal{N}$, **False** otherwise.
3: Compute representative set $\mathcal{E}$ of the EFMs of $\mathcal{N}$.
4: **for** each EFM $e \in \mathcal{E}$ **do**
5: Find shortest decomposition of e into other k EFMs.
6: **if** $k \geq 3$ **then return False**
7: **end if**
8: **end for**
9: **return True**

Algorithm 3 returns **False** as soon as the length of a shortest decomposition of an EFM e into other EFMs is 3 or larger. We then found an EFM that is decomposable but that cannot be decomposed into 2 EFMs. If the algorithm returns **True** in line 9, the length of each shortest decomposition of an EFM into other EFMs was 2, or it was not decomposable. In this case Conj. 5.2.1 holds for the metabolic network $\mathcal{N}$ that the algorithm was initialized with. We found it to be advantageous to replace line 5 of Algorithm 3 with Algorithm 2 with the discussed improvement: only if no decomposition of length 2 was found by Algorithm 2 a shortest decomposition is determined by solving a MILP.

To define a metabolic network, a stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ and a set $\text{Rev} \subseteq \{1, \ldots, n\}$ suffice. The next algorithm generates all possible $(m \times n)$-matrices with entries from a given subset $L \subseteq \mathbb{R}$ and combines them with every

possible subset of reversible reactions. For each of these networks, Algorithm 3 checks whether Conj. 5.2.1 holds.

Algorithm 4 Generate all networks

1: **INPUT**: $m, n \in \mathbb{N}$, $L \subseteq \mathbb{R}$.
2: **OUTPUT**: **False** when a counterexample to Conj. 5.2.1 is found.
3: Generate the set $\mathcal{S} := L^{m \times n}$ that contains all $(m \times n)$- matrices with entries in L and the power set $R := 2^{\{1,\dots,n\}}$ containing every subset of $\{1, \dots, n\}$
4: **for** each matrix $S \in \mathcal{S}$ **do**
5: **for** each set Rev $\in R$ **do**
6: Apply Algorithm 3 to check whether Conj. 5.2.1 holds for the metabolic network $\mathcal{N}$ defined by S and Rev.
7: **end for**
8: **end for**

At this point it should be noted that Algorithm 4 generates a lot of metabolic networks even for small input parameters. Assume we choose $m = n = 3, L = \{0, 1, -1\}$, i.e., we are checking the conjecture for all metabolic networks with 3 metabolites, 3 reactions and stoichiometric coefficients in $\{0, 1, -1\}$. We would already have $3^9 = 19683$ matrices in $\mathcal{S}$ and $2^3 = 8$ subsets of reversible reactions. In total, we have to check $3^9 \cdot 2^3 = 19683 \cdot 8 = 157464$ metabolic networks. In general the number of possible metabolic networks can be derived by the input parameters and is given by

$$(m \cdot n)^{|L|} \cdot 2^n.$$

A few improvements can be made. For example, we know that the conjecture holds for metabolic networks with at most 1 reversible reaction. Thus, we do not have to check networks where $|\,\text{Rev}\,| \leq 1$. Furthermore $\mathcal{S}$ contains

every ordering of the columns of any matrix $S \in \mathcal{S}$, i.e., the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

are all in $\mathcal{S}$, but describe the same metabolic network (up to relabeling the reactions). We can drastically reduce the number of metabolic networks that have to be checked if we account for these observations.

Algorithm 5 Generate all networks without parallel reactions

1: **INPUT**: $m, n \in \mathbb{N}$, $L \subseteq \mathbb{R}^n$.
2: **OUTPUT**: **True** if Conj. 5.2.1 holds, **False** otherwise.
3: Generate a set $cols := L^{m \times 1}$ that contains all length m vectors with entries in L and the set $R := \{\text{Rev} \in 2^{\{1,\dots,n\}} : |\,\text{Rev}\,| \geq 2\}$ containing all subsets of $\{1, \dots, n\}$ with cardinality larger than 1.
4: **for** each cardinality n subset of $cols$ **do**
5: Define a matrix S with the n vectors in this subset as columns.
6: **for** each set Rev $\in R$ **do**
7: Apply Algorithm 3 to check whether Conj. 5.2.1 holds for the metabolic network $\mathcal{N}$ defined by S and Rev.
8: **end for**
9: **end for**

We choose combinations without repetition, i.e., we also do not allow parallel reactions. The number of metabolic networks with 3 metabolites, 3 reactions and stoichiometric coefficients in $\{0, 1, -1\}$ is significantly smaller if we disregard models with parallel reactions and consider networks with relabeled reactions to be the same. We now have $3^3 = 27$ vectors in $cols$ leading to $\binom{27}{3} = 2925$ matrices and 4 subsets of $\{1, 2, 3\}$ of cardinality larger than 1. Now, the total number of metabolic networks we have to check is only $2925 \cdot 4 = 11700$. In general, with our improvements the number of metabolic

networks we check reduces to

$$\binom{|L|^m}{n} \cdot (2^n - n - 1).$$

This number can still quickly become very large. If we wanted to generate all networks such that one of them is our illustrative example in Figure 2.4, we would have to choose $m = 7$ and $n = 12$. Even with our reductions we would be generating $\binom{3^7}{12} \cdot (2^{12} - 12 - 1) \approx 9.9 \cdot 10^{34}$ metabolic networks. Clearly, only small stoichiometric matrices with very limited coefficients can be ruled out as counterexamples by Algorithm 5 in reasonable amounts of time. Nonetheless, if there exists a counterexample to Conj. 5.2.1, it is not unthinkable that there also exists a small counterexample. The implications of a counterexample having to be at least of some size would open interesting research questions. We have tested Algorithm 5 with the input parameters listed in Table 5.1 without finding a counterexample.

For some of the input parameters no EFMs were decomposed. This is due to the fact that no models with those input parameters have decomposable EFMs. We conclude this section by mentioning that we also increased the input parameters m and n as well as the list of possible stoichiometric coefficients and randomly chose combinations of matrices and reversible reactions without finding a counterexample.

5.7 Conclusion

In this chapter, we discussed a conjecture about decompositions of EFMs. We observed that all decomposable EFMs we tested had a shortest decomposition of length 2 leading to the conjecture that this is always the case. We presented an example that appears to be a counterexample to this conjecture and showed that it is, in fact, none. Using matroid theory, we were able to prove a weaker version of this conjecture, namely Theorem 5.4.13 which states that the

(m, n)	L	|Models|	|EFMS|	max(|EFMS|)
$(2, 4)$	$\{0, 1, -1\}$	1386	0	8
$(2, 5)$	$\{0, 1, -1\}$	3276	3800	16
$(2, 6)$	$\{0, 1, -1\}$	4788	19942	28
$(2, 7)$	$\{0, 1, -1\}$	4320	36574	46
$(2, 8)$	$\{0, 1, -1\}$	2223	39066	50
$(2, 9)$	$\{0, 1, -1\}$	502	13116	74
$(2, 4)$	$\{0, 1, -1, 2, -2\}$	139150	594	12
$(2, 5)$	$\{0, 1, -1, 2, -2\}$	1381380	2266868	20
$(3, 3)$	$\{0, 1, -1\}$	3828	0	6
$(3, 3)$	$\{0, 1, -1, 2, -2\}$	172504	0	6
$(3, 4)$	$\{0, 1, -1\}$	168692	428	12
$(3, 5)$	$\{0, 1, -1\}$	2022341	107920	20
$(3, 6)$	$\{0, 1, -1\}$	16667274	17991018	30
$(4, 4)$	$\{0, 1, -1\}$	4912260	0	8

Table 5.1: Parameters tested with Algorithm 5. (m, n): Number of metabolites and reactions, L: Set of possible stoichiometric coefficients, |Models|: Number of models that have EFMs, |EFMs|: Total number of EFMs that are decomposable, max(EFMs|): Largest number of EFMs in a single model.

support of every EFM in a network where all reactions are reversible is contained in the support of a vector resulting from the cancellation of a reversible reaction. In Section 5.5 we presented algorithms based on our observations to generate new EFMs from a starting subset of EFMs and to decompose an EFM into 2 EFMs if such a decomposition exists. While we were not able to prove Conj. 5.2.1 for all EFMs of all metabolic networks (except for those of degree 2 or smaller), Algorithm 1 can still be used to enumerate EFMs and modified to be a novel algorithm that enumerates all EFMs of a metabolic network where all reactions are reversible. Algorithm 2 is a novel method to determine a length 2 decomposition of an EFM. In the final section of this chapter, we briefly discussed attempts to find a counterexample to Conj. 5.2.1. As of now, we have not found one.

Chapter 6

Toolbox for flux cone analysis

In this chapter, we present the `fluxcones` Python package that we developed to calculate geometric properties of flux cones of metabolic networks. The package aims to provide an easy-to-use toolbox for calculating geometric properties of flux cones and is capable of reproducing the data presented in this thesis. The software is open source and available on GitHub at `https://github.com/fwieder/fluxcones`. In Section 6.1, we describe the general structure and basic usage of the `fluxcones` package and the functionality it provides. In Section 6.2, we briefly describe three options to enumerate a representative set of EFMs that come with the package. Finally, in Section 6.3, we discuss examples of algorithms from Section 5.5 and their implementation.

6.1 Overview

Figure 6.1 gives an overview of the main functionality of the `fluxcones` package and should be viewed from top to bottom. An instance of the `FluxCone` class can be created in two different ways. One way is to call `FluxCone(stoich,rev)` where `stoich` is a 2-dimensional `numpy.ndarray` that is a stoichiometric matrix and `rev` is a 1-dimensional `numpy.ndarray` that is a binary vector with ones at indices of reversible reactions and zeros at indices of irreversible reactions. Alternatively, given an SBML file (e.g. from the BiGG-database), an instance of the `FluxCone` class can be created by calling `FluxCone.from_sbml(<path>)`, where `<path>` is the path to the SBML file. `fluxcones` then extracts the stoichiometric matrix and the binary vector `rev` from the SBML file and creates the `FluxCone` object.

If needed, the description of the flux cone can be transformed to an irredundant description (cf. Section 3.3) by calling `FluxCone.make_irredundant()`.

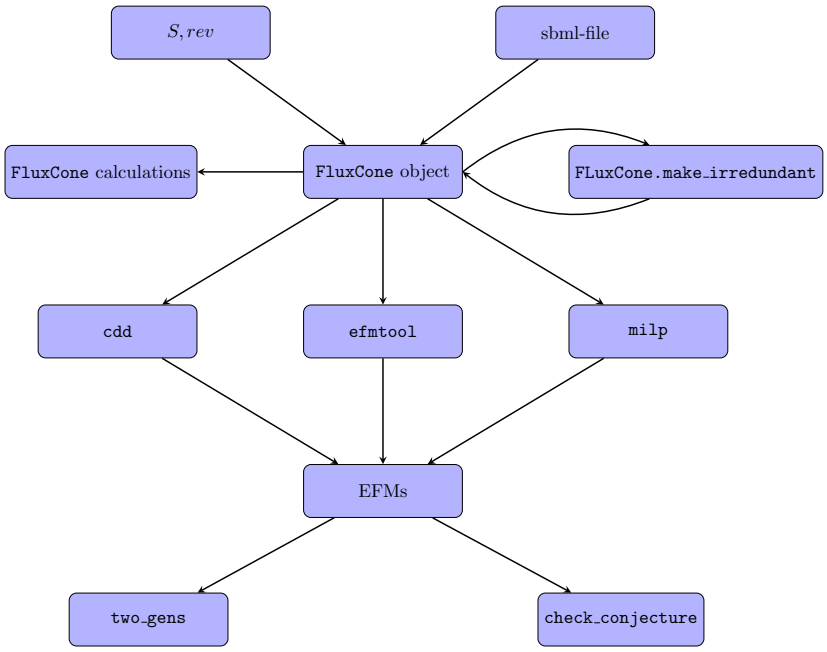

Figure 6.1: Structural overview of the `fluxcones` package.

The function identifies redundant non-negativity constraints by solving linear optimization problems and moves the corresponding reactions from the set of irreversible reactions to the set of reversible reactions by changing corresponding entries in the binary vector `FluxCone.rev`. Note that this does not change the flux cone and that an irredundant description of the flux cone is not uniquely determined.

The `FluxCone` class offers several functions for polyhedral computations, listed in Table 6.1. Three different options to enumerate a representative set of the EFMs are provided, using the Python packages `pycddlib`, `efmtool` and `Python-MIP`. We will provide more details on the implementations with these different methods in the next section. For now we mention that each of these methods, if it terminates, returns a matrix containing the EFMs as rows. Computing EFMs of the same flux cone with different methods can lead to different orders and different scaling of the EFMs. The computed

representative set of EFMs can be further analyzed with functions from the `fluxcones` package.

For example, the method `FluxCone.degree(vector)` returns the degrees of `vector` as defined in Chapter 3. The method assumes that the `numpy.ndarray` it is called with is a steady-state flux vector in the flux cone (i.e. it fulfills the steady-state constraints and thermodynamic irreversibility constraints). For EFMs computed by one of the provided methods this is guaranteed.

Some of the algorithms presented in Section 5.5 are implemented by the package and can be applied to the representative set of EFMs. The functions `two_gens` and `check_conjecture` are implementations of Algorithms 2 and 3 respectively. They are implemented in the `algorithms` module of the `fluxcones` package.

It has to be noted that many of the calculations done by the `fluxcones` package require floating point comparisons. This is not only the case whenever the support of a vector is determined, but also when a face F defined by a vector v is created as a new `FluxCone` object by calling `FluxCone.face_defined_by(v)`. This is due to the fact that a face is defined by homogeneous linear inequalities that are fulfilled with equality. Numerical inaccuracies can lead to false results if not handled with care and the tolerances must be adjusted to match the respective problem.

6.2 EFM enumeration methods

In this section, we describe the implementation of three different methods of the `FluxCone` class to enumerate a representative set $\mathcal{E}$ of the EFMs of a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, S, \text{Rev})$. The methods are all written to return a 2-dimensional `numpy.ndarray` (i.e., a matrix) such that each row of this matrix is one of the EFMs in $\mathcal{E}$ (as opposed to `efmtool`, which returns a matrix where the columns are the EFMs). Furthermore, each of the methods can be called with an optional parameter `only_reversible = True` to enumerate a

Method name	Return
`get_lin_dim()`	dimension of lin. space(C)
`get_cone_dim()`	dimension of flux cone C
`is_in(vector)`	`True` if `vector` $\in C$
`is_efm(vector)`	`True` if `vector` is EFM in C
`get_efms_efmtool()`	EFMs computed with `efmtool`
`get_efms_cdd()`	EFMs computed with `cdd`
`get_efms_milp()`	EFMs computed with MILP approach
`degree(vector)`	degree of `vector`
`irr_supp(vector)`	irreversible support of `vector`
`irr_zeros(vector)`	irreversible zero entries of `vector`
`rev_supp(vector)`	reversible support of `vector`
`rev_zeros(vector)`	reversible zero entries of `vector`
`make_rev(index)`	move reaction at `index` to Rev
`make_irr(index)`	move reaction at `index` to Irr
`get_redundants()`	list of redundant irreversibility constraints
`make_irredundant()`	irredundant description of `FluxCone` object
`blocked_irr_reactions()`	list of blocked irreversible reactions
`blocked_rev_reactions()`	list of blocked reversible reactions
`face_defined_by(vector)`	face of C defined by `vector`

Table 6.1: Methods of the `FluxCone` class and their output. C is the flux cone defined by a stoichiometric matrix `stoich` and the binary vector `rev` that are used as input for a `FluxCone` class object.

representative set of reversible EFMs of $\mathcal{N}$. The EFMs returned in this case are all reversible and can often be computed for metabolic networks where a representative set of all EFMs cannot.

6.2.1 efmtool

The method `get_efms_efmtool()` is using a Python wrapper for `efmtool` that can be found at `https://gitlab.com/csb.ethz/efmtool`. This wrapper allows us to call `efmtool`, which was originally implemented by Terzer (2009) in Java, directly from Python. The options that `efmtool` is called with by the `fluxcones` package were adjusted to suppress console output and normalize the returned vectors to have a maximal absolute value of 1, for the entries of the returned EFMs.

6.2.2 Direct cdd approach

`get_efms_cdd()` uses a Python wrapper for Komei Fukuda's `cddlib`, which is available at `https://github.com/mcmtroffaes/pycddlib`. When the method is called, the `fluxcones` package splits each reversible reaction into two irreversible reactions of opposite orientation and determines the extreme rays of the resulting augmented cone with the double description method. The extreme rays are then processed to return a representative set of the EFMs of the original cone by removing cycles that are 2-cycles of the two oppositely oriented irreversible reactions that replaced one reversible reaction. Although `efmtool` is optimized for EFM calculation and uses the same algorithm, for very small examples `get_efms_cdd()` outperforms `get_efms_efmtool()` due to smaller initial overhead.

6.2.3 MILP approach

The `fluxcones` package is also able to enumerate EFMs using the mixed-integer linear programming approach described in (de Figueiredo *et al.*, 2009) by calling the method `get_efms_milp()`. The Python package `Python-Mip` (`https://www.python-mip.com/`) is used to solve the MILPs.

Although the approach is outperformed by `efmtool` and `cdd` for medium-sized networks, the function can be adjusted to consecutively enumerate shortest EFMs for metabolic networks of any size and especially for genome-scale networks that are too large for `efmtool` and the direct `cdd` approach.

6.3 Applications

The `fluxcones` package comes with the functions `MILP_shortest_decomp`, `check_conjecture` and `two_gens` which implement algorithms applicable to `FluxCone` objects.

`MILP_shortest_decomp` takes a target vector and a set of candidates as input and returns the coefficients of a shortest decomposition of the target vector into the candidate vectors, computed by solving a mixed-integer linear program (cf. equation (4.2)).

In Chapter 4 we developed an approach to determine a face defined by a given flux vector and the EFMs contained in it (cf. Figure 4.3). This approach can be reproduced with the `fluxcones` package as follows.

Given a vector `v` as `numpy.ndarray`, `F = FluxCone.face_defined_by(v)` instantiates a new `FluxCone` instance `F` which is the face of the original flux cone defined by `v`. `F.get_efms_efmtool()` (or one of the other two provided EFM enumeration methods) returns a matrix containing the EFMs in `F` as rows and `MILP_shortest_decomp(v,EFMs)` can be called to determine the coefficients of a shortest decomposition of `v` into the EFMs of `F`. This approach was used to find faces defined by FBA solutions and shortest decompositions

in Chapter 4.

`check_conjecture` is an implementation of Algorithm 3. It uses `two_gens`, an implementation of Algorithm 2, because it is faster for very small (in the sense of number of reactions and metabolites) models.

`all_networks`, an implementation of Algorithm 5 (cf. Chapter 5) that uses `two_gens` and `check_conjecture` is also included in the `fluxcones` package. The data in Table 5.1 was obtained using this function.

Chapter 7

Conclusion

This thesis studies the geometric properties of flux cones of metabolic networks and elementary flux modes. The main contribution of Chapter 3 is the introduction of the degree of a flux vector as the dimension of the face it is contained in. The degree gives structure to the set of EFMs and allows us to explore their distribution among the face lattice of the flux cone. We illustrate this structure by determining the degree distribution of EFMs in the flux cone of three well-known, medium-sized metabolic networks and observe that EFMs tend to be contained in lower-dimensional faces. We formally prove upper bounds for the degree of EFMs and thereby show that EFMs occur in the relative interior of the flux cone only in very specific cases. The distribution of EFMs among lower-dimensional faces of the flux cone raises the question of how these observations can be exploited computationally in EFM enumeration algorithms. Furthermore, in Chapter 3, we generalize the result of the 1-1 correspondence between minimal metabolic behaviors and minimal proper faces to metabolic behaviors and higher-dimensional faces of the flux cone. This generalization allows us to define faces of the flux cone by sets of active irreversible reactions in a flux vector. The chapter concludes by establishing a relationship between combinatorial properties of the flux cone and the cardinality of minimal metabolic behaviors. Further research should address the biological implications of the degree of an EFM or flux vectors in general.

In Chapter 4, we study decompositions of flux vectors into EFMs. Based on the relationship between faces of the flux cone and metabolic behaviors established in Chapter 3, we develop a method to determine EFMs in the face of the flux cone defined by a given flux vector and show that a decomposition of this flux vector only uses EFMs in this face. Enumerating only EFMs in a given face of the flux cone significantly reduces the search space and the number of EFMs. We illustrate the scalability of our approach by determining

EFMs in faces defined by solutions of FBA problems in a large selection of genome-scale metabolic networks. By determining shortest decompositions of these optimizing flux vectors, we observe that the participating EFMs have relatively high degrees. Thus, they can be further decomposed into EFMs of lower degrees. This observation leads us to the definition of low-degree decompositions, where only EFMs that cannot be further decomposed into EFMs of lower degrees are considered. We expect interesting results from further research into the biological relevance of low-degree decompositions, especially with respect to additional thermodynamic constraints (cf. Gerstl *et al.,* 2016).

The decomposability of higher-degree EFMs observed in Chapter 4 is the main topic of Chapter 5. An empirical observation that the length of a shortest decomposition of an EFM into other EFMs never exceeds 2 leads to the conjecture that this is always the case. We discuss several approaches to prove or disprove the conjecture. After we formally state the conjecture and define what a counterexample to the conjecture is, we prove that a counterexample, if it exists, has to be a metabolic network with at least two reversible reactions. Next, we explore the relationship between matroids and metabolic networks, where all reactions are reversible. In this context, we prove a weaker version of the conjecture, namely that the support of every EFM in a metabolic network where all reactions are reversible is contained in the support of a positive combination of two other EFMs. From these observations we derive novel algorithms to generate new EFMs from a starting set by positively combining them pairwise. This algorithm can be viewed as a modification of the circuit enumeration method (cf. Khachiyan *et al.,* 2005). We also describe an inverse algorithm to find length-2 decompositions of EFMs without solving mixed-integer linear programs. The chapter concludes with a description of algorithms to generate all metabolic networks of a given size and to test whether the conjecture holds for them. These allow us to verify the conjecture for specific sets of small networks (computationally). More research is required

to determine whether the conjecture holds for all metabolic networks. We also suggest a study comparing the efficiency of our EFM generation algorithm and state-of-the-art EFM enumeration methods.

In Chapter 6, we briefly introduce a Python package that was developed over the course of this thesis. The software is open source and available on GitHub. We describe the structure of the software, available functions, provided EFM enumeration methods, and explain how the software can be applied to implement algorithms described in Chapters 4 and 5.

From a computer science perspective, improving our algorithms' efficiency is an interesting task. For future work, extensions and more efficient implementations are planned for the `fluxcones` package.

Finally, we plan to investigate our conjecture from Chapter 5 further in the context of oriented matroids and metabolic networks containing irreversible reactions.

Bibliography

Acuña, V., Chierichetti, F., Lacroix, V., Marchetti-Spaccamela, A., Sagot, M.-F. and Stougie, L. Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems* **95,** 51–60. https://doi.org/10.1016/j.biosystems.2008.06.015 (2009).

Bordbar, A., Monk, J. M., King, Z. A. and Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics* **15,** 107–120. https://doi.org/10.1038/nrg3643 (2014).

Buchner, B. A. and Zanghellini, J. EFMlrs: a Python package for elementary flux mode enumeration via lexicographic reverse search. *BMC Bioinformatics* **22,** 547. https://doi.org/10.1186/s12859-021-04417-9 (2021).

Capelli, F. and Strozecki, Y. *Geometric Amortization of Enumeration Algorithms* 2021. arXiv: 2108.10208[cs]. https://doi.org/10.48550/arXiv.2108.10208.

Chan, S. H. J. and Ji, P. Decomposing flux distributions into elementary flux modes in genome-scale metabolic networks. *Bioinformatics* **27,** 2256–2262. https://doi.org/10.1093/bioinformatics/btr367 (2011).

Chan, S. H. J., Solem, C., Jensen, P. R. and Ji, P. Estimating biological elementary flux modes that decompose a flux distribution by the minimal branching property. *Bioinformatics* **30,** 3232–3239. https://doi.org/10.1093/bioinformatics/btu529 (2014).

Chen, J., Huang, Y. and Zhong, C. Minimizing enzyme mass to decompose flux distribution for identifying biologically relevant elementary flux modes. *Biosystems* **231,** 104981. https://doi.org/10.1016/j.biosystems.2023.104981 (2023).

Clarke, B. L. *Stability of Complex Reaction Networks* in *Advances in Chemical Physics* 1–215 (John Wiley & Sons, Ltd, 1980). https://doi.org/10.1002/9780470142622.ch1.

De Figueiredo, L. F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J. E., Schuster, S. and Planes, F. J. Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* **25,** 3158–3165. https://doi.org/10.1093/bioinformatics/btp564 (2009).

Diestel, R. *Graph Theory* https://doi.org/10.1007/978-3-662-53622-3 (Springer, Berlin, Heidelberg, 2017).

Ebrahim, A., Lerman, J. A., Palsson, B. O. and Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7,** 74. https://doi.org/10.1186/1752-0509-7-74 (2013).

Fang, X., Lloyd, C. J. and Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology* **18,** 731–743. https://doi.org/10.1038/s41579-020-00440-4 (2020).

Fukuda, K. and Prodon, A. *Double description method revisited* in *Combinatorics and Computer Science* (Springer, Berlin, Heidelberg, 1996), 91–111. https://doi.org/10.1007/3-540-61576-8_77.

Gagneur, J. and Klamt, S. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* **5,** 175. https://doi.org/10.1186/1471-2105-5-175 (2004).

Gerstl, M. P., Jungreuthmayer, C., Müller, S. and Zanghellini, J. Which sets of elementary flux modes form thermodynamically feasible flux distributions? *The FEBS Journal* **283,** 1782–1794. https://doi.org/10.1111/febs.13702 (2016).

Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols* **14,** 639–702. https://doi.org/10.1038/s41596-018-0098-2 (2019).

Henk, M., Richter-Gebert, J. and Ziegler, G. M. *Basic Properties of Convex Polytopes* in *Handbook of Discrete and Computational Geometry* 3rd ed. (Chapman and Hall/CRC, 2017). https://doi.org/10.1201/9781315119601.

Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19,** 524–531. https://doi.org/10.1093/bioinformatics/btg015 (2003).

Jevremovic, D., Trinh, C. T., Srienc, F. and Boley, D. On Algebraic Properties of Extreme Pathways in Metabolic Networks. *Journal of Computational Biology* **17,** 107–119. https://doi.org/10.1089/cmb.2009.0020 (2010).

Jevremović, D. and Boley, D. Finding minimal generating set for metabolic network with reversible pathways. *Biosystems* **112,** 31–36. https://doi.org/10.1016/j.biosystems.2013.02.003 (2013).

Jungers, R. M., Zamorano, F., Blondel, V. D., Wouwer, A. V. and Bastin, G. Fast computation of minimal elementary decompositions of metabolic flux vectors. *Automatica. Special Issue on Systems Biology* **47,** 1255–1259. https://doi.org/10.1016/j.automatica.2011.01.011. (2011).

Kanehisa, M. and Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28,** 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).

Kelk, S. M., Olivier, B. G., Stougie, L. and Bruggeman, F. J. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Scientific Reports* **2,** 580. https://doi.org/10.1038/srep00580 (2012).

Khachiyan, L., Boros, E., Elbassioni, K., Gurvich, V. and Makino, K. On the Complexity of Some Enumeration Problems for Matroids. *SIAM Journal on Discrete Mathematics* **19,** 966–984. https://doi.org/10.1137/S0895480103428338 (2005).

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. and Lewis, N. E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **44,** D515–D522. https://doi.org/10.1093/nar/gkv1049 (D1 2016).

Kitano, H. Systems biology: a brief overview. *Science (New York, N.Y.)* **295,** 1662–1664. https://doi.org/10.1126/science.1069492 (2002).

Larhlimi, A. and Bockmayr, A. A new constraint-based description of the steady-state flux cone of metabolic networks. *Discrete Applied Mathematics. Networks in Computational Biology* **157,** 2257–2266. https://doi.org/10.1016/j.dam.2008.06.039 (2009).

Larhlimi, A. and Bockmayr, A. *On Inner and Outer Descriptions of the Steady-State Flux Cone of a Metabolic Network* in *Computational Methods in Systems Biology* (2008). https://doi.org/10.1007/978-3-540-88562-7_22.

Lauritzen, N. *Undergraduate Convexity: From Fourier and Motzkin to Kuhn and Tucker* 298 pp. https://doi.org/10.1142/8527 (World Scientific, 2013).

Maarleveld, T. R., Wortel, M. T., Olivier, B. G., Teusink, B. and Bruggeman, F. J. Interplay between Constraints, Objectives, and Optimality for Genome-Scale Stoichiometric Models. *PLOS Computational Biology* **11,** e1004166. https://doi.org/10.1371/journal.pcbi.1004166 (2015).

Marashi, S.-A. and Bockmayr, A. Flux coupling analysis of metabolic networks is sensitive to missing reactions. *Biosystems* **103,** 57–66. https://doi.org/10.1016/j.biosystems.2010.09.011 (2011).

Mardis, E. R. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* **9,** 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359 (Volume 9 2008).

Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O. and King, Z. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research* **48,** D402–D406. https://doi.org/10.1093/nar/gkz1054 (D1 2020).

Oddsdóttir, H. Æ., Hagrot, E., Chotteau, V. and Forsgren, A. On dynamically generating relevant elementary flux modes in a metabolic network using optimization. *Journal of Mathematical Biology* **71,** 903–920. https://doi.org/10.1007/s00285-014-0844-1 (2015).

Orth, J. D., Thiele, I. and Palsson, B. Ø. What is flux balance analysis? *Nature Biotechnology* **28,** 245–248. https://doi.org/10.1038/nbt.1614 (2010).

Oxley, J. G. *Matroid Theory* 550 pp. https://doi.org/10.1093/acprof:oso/9780198566946.001.0001 (Oxford University Press, 2006).

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S. and Palsson, B. O. Comparison of network-based pathway analysis methods. *Trends in Biotechnology* **22,** 400–405. https://doi.org/10.1016/j.tibtech.2004.06.010 (2004).

Poolman, M., Venkatesh, K., Pidcock, M. and Fell, D. A method for the determination of flux in elementary modes, and its application to Lactobacillus rhamnosus. *Biotechnology and Bioengineering* **88,** 601–612. https://doi.org/10.1002/bit.20273 (2004).

Reimers, A. C. Metabolic Networks, Thermodynamic Constraints, and Matroid Theory. http://dx.doi.org/10.17169/refubium-15314 (2014).

Rezola, A., de Figueiredo, L. F., Brock, M., Pey, J., Podhorski, A., Wittmann, C., Schuster, S., Bockmayr, A. and Planes, F. J. Exploring metabolic pathways in genome-scale networks via generating flux modes. *Bioinformatics* **27,** 534–540. https://doi.org/10.1093/bioinformatics/btq681 (2011).

Röhl, A. and Bockmayr, A. Finding MEMo: minimum sets of elementary flux modes. *Journal of Mathematical Biology* **79,** 1749–1777. https://doi.org/10.1007/s00285-019-01409-5 (2019).

Rügen, M., Bockmayr, A., Legrand, J. and Cogne, G. Network reduction in metabolic pathway analysis: Elucidation of the key pathways involved in the photoautotrophic growth of the green alga *Chlamydomonas reinhardtii*. *Metabolic Engineering* **14,** 458–467. https://doi.org/10.1016/j.ymben.2012.01.009 (2012).

Schilling, C. H., Letscher, D. and Palsson, B. Ø. Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *Journal of Theoretical Biology* **203,** 229–248. https://doi.org/10.1006/jtbi.2000.1073 (2000).

Schneider, R. *Convex Bodies: The Brunn–Minkowski Theory* 2nd ed. https://doi.org/10.1017/CBO9781139003858 (Cambridge University Press, Cambridge, 2013).

Schrijver, A. *Theory of Linear and Integer Programming* 488 pp. (John Wiley & Sons, 1998).

Schuster, S. and Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems* **02,** 165–182. https://doi.org/10.1142/S0218339094000131 (1994).

Schuster, S., Hilgetag, C., Woods, J. and Fell, D. Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *Journal of mathematical biology* **45,** 153–81. https://doi.org/10.1007/s002850200143 (2002).

Schwartz, J.-M. and Kanehisa, M. A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. *Bioinformatics* **21,** ii204–ii205. https://doi.org/10.1093/bioinformatics/bti1132 (suppl_2 2005).

Terzer, M. *Large scale methods to enumerate extreme rays and elementary modes* Doctoral Thesis (ETH Zurich, 2009). https://doi.org/10.3929/ethz-a-005945733.

Urbanczik, R. and Wagner, C. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics* **21,** 1203–1210. https://doi.org/10.1093/bioinformatics/bti127 (2005).

Wagner, C. and Urbanczik, R. The Geometry of the Flux Cone of a Metabolic Network. *Biophysical Journal* **89,** 3837–3845. https://doi.org/10.1529/biophysj.104.055129 (2005).

Welsh, D. J. A. *Matroid Theory* 450 pp. (Courier Corporation, 2010).

Wieder, F. and Bockmayr, A. *Low-degree decompositions of flux vectors in faces of the flux cone* 2024. https://doi.org/10.1101/2024.05.22.595286.

Wieder, F., Henk, M. and Bockmayr, A. On the geometry of elementary flux modes. *Journal of Mathematical Biology* **87,** 50. https://doi.org/10.1007/s00285-023-01982-w (2023).

Ziegler, G. M. *Lectures on Polytopes* https://doi.org/10.1007/978-1-4613-8431-1 (Springer, New York, NY, 1995).

Deutsche Zusammenfassung

Systembiologie befindet sich an der Schnittstelle von Biologie, Informatik und Mathematik und basiert auf der mathematischen Modellierung biologischer Systeme. Ziel ist es, das Verhalten biologischer Systeme vorherzusagen, um zeit- und kostenintensive Forschungsarbeiten im Labor effizienter zu gestalten.

In dieser Arbeit konzentrieren wir uns auf die Beschreibung und das Verständnis von metabolischen Netzwerken, welche die metabolischen Prozesse in einer Zelle modellieren.

Die stöchiometrischen und thermodynamischen Bedingungen, die in einem metabolischen Netzwerk im stationären Zustand gelten, definieren den stationären Flusskegel. Ein wichtiges Konzept zur Analyse dieser Flusskegel auf mathematisch und biologisch sinnvolle Weise sind elementare Flussmodi, die als minimale funktionelle Einheiten von metabolischen Netzwerken betrachtet werden können. Sie entsprechen Vektoren mit inklusionsweise minimaler Trägermenge im Flusskegel.

Wir konzentrieren uns auf die geometrischen Aspekte von Flusskegeln metabolischer Netzwerke und elementarer Flussmodi. Die Anzahl der elementaren Flussmodi kann schon für mittelgroße metabolische Netzwerke sehr groß sein. Wir untersuchen die Struktur der Seiten und die Verteilung der elementaren Flussmodi auf die Seiten des Flusskegels und beobachten, dass diese hauptsächlich in Seiten relativ niedriger Dimensionen enthalten sind. Mit dieser Beobachtung entwickeln wir eine Methode zur Aufzählung von Teilmengen elementarer Flussmodi, nämlich denen, die in einer bestimmten Seite des Flusskegels enthalten sind und wenden dies auf Zerlegungen von Flussvektoren an.

Dabei haben wir beobachtet, dass zerlegbare elementare Flussmodi immer eine positive Summe von genau 2 anderen waren. Darauf basierend kommen wir zu der Vermutung, dass dies immer der Fall ist. Wir disktutieren Ansätze diese Vermutung zu beweisen und sie zu widerlegen.

Darüber hinaus stellen wir ein Python software Paket vor, dass die Daten, die wir verwendet haben um unsere theoretischen Ergebnisse zu veranschaulichen, reproduzieren kann.

Declaration

Declaration of authorship

Name: Wieder
First names: Frederik Laszlo

I declare to the Freie Universität Berlin that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. The present thesis is free of plagiarism. I have marked as such all statements that are taken literally or in content from other writings. This dissertation has not been submitted in the same or similar form in any previous doctoral procedure.

I agree to have my thesis examined by a plagiarism examination software.

Date:_____ Signature: _____‘

Acknowledgements

First of all, I want to thank my supervisor, Alexander Bockmayr for supporting me over the last years, for his guidance, valuable advice, and for his patience. Furthermore, I thank Francis J. Planes for agreeing to review this thesis on such a short notice.

Another important role in my research was played by Martin Henk, who has been a part of my scientific journey for quite some time now. I want to thank him for our collaboration and all his help over the years.

Thanks to all my colleagues at FU from the Mathematics in Life Sciences and Discrete Biomathematics groups at the FU for the inspiring, interesting, and entertaining conversations.

I also want to thank my mother, and my brothers, Nico and Ben, for their support, for their friendship, and for always being there when I needed them the most.

Lena and Emmi, thank you for all your support, for all the wonderful moments we have together, for being the best, and for reminding me that "in real math, you add numbers".

Finally, I want to thank my dad for everything and all the support he would have given me, if he was still here. I miss you.