Modes and Cuts in Metabolic Networks:
Complexity and Algorithms*

Flavio Chierichetti[1], Vincent Lacroix[2,3], Alberto Marchetti-Spaccamela[1], Marie-France Sagot[2,3], Leen Stougie[4,5]

[1] Universitá di Roma "La Sapienza", Via Eudossiana 18, 00184 Rome, Italy
[2] Équipe BAOBAB, Laboratoire de Biométrie et Biologie Evolutive (UMR 5558); CNRS; Univ. Lyon 1, 43 bd du 11 nov 1918, 69622, Villeurbanne Cedex, France
[3] Projet Helix, INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France
[4] Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, The Netherlands
[5] Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098SJ Amsterdam, The Netherlands

Abstract

Constraint-based approaches recently brought new insight into our understanding of metabolism. By making very simple assumptions such as that the system is at steady-state and some reactions are irreversible, and without requiring kinetic parameters, general properties of the system can be derived. A central concept in this methodology is the notion of an elementary mode (EM for short). The computation of EMs still constitutes a limiting step in metabolic studies and several algorithms have been proposed to address this problem leading to increasingly faster methods. However, although a theoretical upper bound on the number of elementary modes that a network may possess has been established, surprisingly, the complexity of this problem has never been systematically studied.

In this paper, we first establish results regarding network consistency. Most consistency problems can be solved in polynomial time (are easy). We then establish the complexity of finding and counting elementary modes. We show in particular that finding one elementary mode is easy but that this task becomes hard when a specific EM (*i.e.* an EM containing some specified reactions) is sought. We also examine a number of EM related problems and establish their complexity. We emphasise that the easy problems can be solved by existing software.

We then analyse the complexity of a closely related task which is the computation of so-called minimal reaction cut sets. In this case again, we show that this problem is hard. We then present two positive results which both allow to avoid computing EMs as a prior to the computation of reaction cuts. The first one is a polynomial approximation algorithm for finding a minimum cut set. The second one is a test for verifying if a set of reactions constitutes a reaction cut; this test could be readily included in existing algorithms to improve their performance. Finally, we discuss the complexity of other cut-related problems.

1 Introduction

Metabolism is usually defined as the union of two processes: anabolism (synthesis of molecules through the use of energy and reducing power) and catabolism (degradation of molecules yielding energy and reducing power). From a modeller's perspective, metabolism can be seen as a network of interconnected reactions, each reaction corresponding to the transformation of metabolites into other metabolites. This network can then be studied either from a structural perspective, or from a dynamic perspective.

Studying the dynamics of metabolic networks is usually performed using models related to differential equations whereas structural analyses are mainly based on graph-related formalisms or, as far as metabolism

is concerned, on constraint-based modelling. The choice of a particular model heavily depends on the type of question one wishes to address (structural or dynamic) but also on the type of data that is available (qualitative or quantitative). Another type of criterion that may be taken into account is the computational cost of a given analysis, and therefore its scalability to large datasets (such as genome-scale metabolic networks).

The limited amount of available quantitative data (such as kinetic parameters) has led to an increase in popularity of the models which do not use such information. Graphs and constraint-based models fall into this category. However, while the first typically scales well to large datasets, the latter still needs improvements to be more widely used.

In a constraint-based approach, only admissible flux distributions are studied. An admissible flux distribution corresponds to a set of reactions, which, when taken together in given proportions, perform the transformation of available substrates into removable products with the special property that all intermediate compounds are balanced (steady-state assumption) and irreversible reactions are taken in the appropriate direction (thermodynamic constraint). Such an admissible flux distribution is called a mode.

Even though each mode is potentially interesting, not all of them are generally considered. Classically, two major sub-problems have been introduced. The first one is known as flux balance analysis. It consists in searching for a mode that optimises a given objective function. Examples of objective functions include biomass (usually represented as a pseudo-reaction of the network, in general determined from experimental data) or ATP production. This optimisation problem has several applications [17] and can be solved using linear programming (LP).

The second sub-problem is the one we discuss in this paper. In the case where no particular function is to be optimised, all modes are equally interesting. A sensible strategy is then to try to find a set that could generate them all. Such a generating set has been proposed and called the set of elementary modes [20], EM for short. Intuitively, an elementary mode is a special mode that has the property of not containing any other mode.

Instead of elementary modes, extreme pathways may also be considered. Both notions are very close, the difference being mainly due to the way reversible reactions are handled. Indeed, extreme pathways and elementary modes appear to be identical in the special case where all reactions in the network are irreversible. For a detailed comparison of both approaches, see [14].

Elementary modes have been said to represent a formalised definition of a biological pathway. Indeed, a biological interpretation can be given to such flux vectors: a mode is a set of enzymes that operate together at steady state [19] and a mode is elementary when the removal of one enzyme causes it to fail.

The computation of elementary modes is far from trivial and led to the development of several algorithms. The major ones that have been proposed are by Schuster [21], Urbanczik and Wagner [23], Klamt and Gagneur [7].

Another concept we use in the paper is closely connected to the notion of elementary mode. This is the concept of a minimal reaction cut set, recently introduced in [12]. In order to avoid any confusion with other types of cuts in graphs or hypergraphs that may be found in the literature [22], we explicitly choose here to use the term *reaction* cut. An elementary mode may be seen as a set of reactions that, when used together, perform a given task while a minimal reaction cut set is a set of reactions one needs to inhibit to prevent a given task, also called *target reaction*, from being performed. As mentioned in [10], the task to be silenced can be a combination of reactions. Minimal reaction cut sets have been operationally defined as corresponding to a set of reactions whose deletion from the network stops each elementary mode that contains the target reaction(s).

In this paper, we first establish results regarding network consistency (Section 2.1). Most consistency problems can be solved in polynomial time (are easy). We then establish the complexity of finding and counting elementary modes (Sections 3.1 and 3.2). We show in particular that finding one elementary mode is easy but that this task becomes hard when a specific EM (*i.e.* an EM containing some specified reactions) is sought. We also examine a number of EM related problems and establish their complexity. We emphasize that the easy problems can be solved by existing software.

We then analyse the complexity of a closely related task which is the computation of so-called minimal reaction cut sets. In this case again, we show that this problem is hard (Sections 4.1 and 4.3). We then present two positive results which both allow to avoid to compute EMs as a prior to the computation of reaction cuts. The first one (Section 4.2) is a polynomial approximation algorithm for finding a minimum

cut set. The second one (Section 4.3 using a result of Section 4.1) is a test for verifying if a set of reactions constitutes a reaction cut; this test could be readily included in existing algorithms for enumerating minimal reaction cuts to improve their performance. Finally, we discuss the complexity of other cut-related problems (Section 4.4).

2 Modes

In the following, we define more precisely several objects, classically used in constraint-based modelling.

The **stoichiometric matrix** S of a network is a matrix with n rows and m columns, n being the number of internal metabolites and m the number of reactions. A cell $S(i, j)$ of the matrix takes value k if reaction j produces k units of metabolite i. It takes value $-k$ if reaction j consumes k units of metabolite i. Otherwise, it takes value 0. The value k corresponds to the stoichiometric coefficient of metabolite i in reaction j. The stoichiometric matrix summarises the structure of the metabolic network. Observe that the reversibility of the reactions has to be inserted as separate information.

The set of reactions is partitioned into two subsets: Rev and $Irrev$, the set of, respectively, reversible and irreversible reactions.

A **mode** is a flux vector $v \in \mathbb{R}^m$ such that:

1. $Sv = 0$

2. $v_j \geq 0 \ \forall j \in Irrev$

We introduce the support of the solution v, denoted by $R(v) = \{j \mid v_j > 0\}$, *i.e.*, the set of reactions participating (with non-zero flux) in v.

An **elementary mode** is a vector v that satisfies conditions 1 and 2 and

3. there is no non-trivial flux vector r such that: $R(r) \subset R(v)$.

Modes and elementary modes can be given a geometrical interpretation. Indeed, the set of vectors $\{v \geq 0 \mid Sv = 0\}$ defines a convex cone in the flux space. When all reactions are irreversible, the elementary modes exactly correspond to the extreme rays of this cone. Klamt and Gagneur showed [7] that even when some reactions are reversible, one can define a pointed cone in a higher dimensional space (all reversible reactions are split into two irreversible reactions) with the rays corresponding to a superset of the original EMs.

This geometrical interpretation has been particularly fruitful since most algorithms used for computing EMs are based on the double description method, a method borrowed from computational geometry [6].

From now on we assume that all reactions are irreversible unless explicitly stated otherwise.

2.1 Consistency of the stoichiometric matrix

One of the applications of constraint-based modelling is in checking the consistency of reconstructed metabolic networks [19]. A network is said to be consistent if all its reactions belong to at least one elementary mode. When a network is consistent, we say equivalently that its stoichiometric matrix is consistent: the stoichiometric matrix S is **consistent** if $Sv = 0$ has a solution $v_j > 0 \ \forall j$, or equivalently, each reaction is part of some mode (elementary mode).

We give some statements about the consistency of a stoichiometric matrix. Clearly, if a matrix S is not consistent, then there must be a mismodelling of the metabolic network. In that sense, detecting inconsistency is a valuable tool for finding deficiencies in the metabolic network description.

Theorem 1. *Given a stoichiometric matrix S, checking the consistency of S can be done using LP.*

Proof. Consider the following LP, where we insert a bound on the sum of the values of the v_j's to avoid unboundedness of the problem.

$$
\begin{aligned}
\max \quad & z \\
\text{s.t.} \quad & v_j \geq z \ \forall j \\
& Sv = 0 \\
& \textstyle\sum_j v_j \leq 1
\end{aligned}
$$

3

S is consistent if the optimal value is strictly positive, otherwise it is not. □

In case of inconsistency, it is also easy to find a consistent submatrix containing a maximum number of reactions.

Theorem 2. *Given a stoichiometric matrix S, detecting a minimum number of reactions to be deleted to make S consistent can be done using LP.*

Proof. For each reaction h, solve the LP

$$
\begin{aligned}
\max \quad & z \\
\text{s.t.} \quad & v_h \geq z \\
& Sv = 0 \\
& \textstyle\sum_j v_j \leq 1 \\
& v \geq 0
\end{aligned}
$$

If for reaction h, the optimal value is strictly positive, then h is part of some mode, and one such a mode is given by the optimal solution. Otherwise there is no mode in which reaction h appears. □

A question complementary to the previous one is hard.

Theorem 3. *Given a stoichiometric matrix S, and some other set of reactions represented by a stoichiometric matrix S', find a subset of reactions of S' of minimum cardinality such that the corresponding submatrix added to S yields a consistent matrix is NP-hard.*

Proof. Taking for S an empty matrix and for S' the stoichiometric matrix of the network, the problem is a special case of finding an elementary mode with a minimum number of reactions in its support. NP-hardness of the latter problem will be established in Theorem 6. □

2.2 Difference between hypergraph and stoichiometric matrix

The stoichiometric matrix enables to represent the structure of a metabolic network. In some cases, particularly for visualization, hypergraphs may also be used. An hypergraph representation of a metabolic network can be done as follows: metabolites are represented as nodes and there is a (directed) hyperedge for each reaction going from its substrates to its products. In fact, this hypergraph can on its turn be represented by its vertex-edge incidence matrix, which is very similar to the stoichiometric matrix; the former matrix has a 1 at each entry where the latter has a positive integer, a -1 where the latter has a negative integer, and their 0 entries coincide.

The hypergraph description does not take into account all parameters of the stoichiometric matrix as can be seen by the following toy example in which two different networks are presented having the same hypergraph description.

Network 1	**Network 2**
input: a,b	input: a,b
output: f	output: f
Reaction 1: a+b → c+d	Reaction 1: a+b → c+2d
Reaction 2: c+ d → f	Reaction 2: c+ 3d → f

Observe that the first network is consistent while the second one is not. Therefore, consistency of a network cannot be checked using a hypergraph (regardless of the stoichiometry). Thus, hypergraphs need to be supplemented with weights on vertices-edge combinations if one wants to use them interchangeably with the stoichiometric matrix.

3 Elementary modes

As mentioned in the introduction, we may see an elementary mode as an extreme ray of the cone $\{v \geq 0 \mid Sv = 0\}$, due to the observation in [7] that an elementary mode is characterised completely by its set of reactions, *i.e.*, given S and the support $R(v)$ of an elementary mode v, up to scalar multiplication, v is uniquely determined.

In this section, we assume consistency of the stoichiometric matrices of the problem instances we consider.

3.1 Finding elementary modes

Surprisingly few results have been established on the complexity of problems concerning detection, counting and enumeration of elementary modes. In their paper, Klamt and Stelling [13] mainly focus on finding an upper bound on the number of elementary modes.

In fact, as mentioned in [6], the complexity of the general problem, given a description of a cone (or polytope) in terms of its facets (inequalities), find a description in terms of (enumerate all) its extreme rays (vertices), as a function of the length of the output (number of rays or vertices) is a long-standing open question in computational geometry.

In this section, we show some difficult aspects of computing elementary modes. In particular, we try to show where the hardness comes from when enumerating elementary modes. We show that the three following tasks are easy: (i) finding a mode, (ii) finding an elementary mode, (iii) finding a mode that contains one specified reaction. However, the last task is hard: (iv) finding an elementary mode that contains a specified set of reactions.

As observed already in [11], standard linear algebra teaches us how to check that $Sv = 0$ in order to decide if $v \geq 0$ is a mode. It is also easy to decide if a given mode $v \geq 0$ is an elementary mode by Gaussian elimination on the submatrix of S consisting of the reaction in the support of v.

Theorem 4. *Given a stoichiometric matrix S, an elementary mode can be found in polynomial time.*

Proof. We "slice" the cone $Sv = 0$ by the inequality $\sum_j v_j \leq 1$ and solve the LP:

$$
\begin{aligned}
\max \quad & z \\
\text{s.t.} \quad & v_h \geq z \\
& Sv = 0 \\
& \textstyle\sum_j v_j \leq 1 \\
& v \geq 0.
\end{aligned}
\tag{1}
$$

In case of a consistent matrix, the optimal solution v^* is a non-all-0 vertex of the polytope $\{v \geq 0 \mid Sv = 0, \sum_j v_j \leq 1\}$ satisfying the inequality $\sum_j v_j \leq 1$ with equality. Thus, $\{\lambda v^* \mid \lambda \geq 0\}$ is an extreme ray of the cone $\{v \geq 0 \mid Sv = 0\}$. $\qquad\square$

The optimal solution of the LP in the proof of the lemma gives an elementary mode that contains reaction h. In general, it is easy to detect if there exists a *mode* whose support contains a given set of reactions T_{IN}, and does not contain any of the reactions of another set T_{OUT}: simply add the restrictions:

$$
v_j = 0 \ \forall j \in T_{OUT}
\tag{2}
$$

to LP (1), replace the first restriction of LP (1) by

$$
v_j \geq z \ \forall j \in T_{IN},
$$

and check if the optimal solution is positive or 0.

If we could answer the same question for the existence of an *elementary mode* with the same properties for any set T_{IN} in polynomial time, then we would be able to enumerate elementary modes in time polynomial in their number. However, unfortunately this decision problem is NP-hard in general, which may (partly) explain the difficulties we encounter in enumerating elementary modes.

Theorem 5. *Given a stoichiometric matrix S, sets of reactions T_{IN} and T_{OUT}, deciding if an elementary mode v exists that has positive value in all its coordinates corresponding to T_{IN}, and has value 0 in all its coordinates corresponding to the set T_{OUT} is*
(i) polynomial solvable if $|T_{IN}| = 1$,
(ii) NP-complete in the general case.

Proof. If $|T_{IN}| = 1$ the proof follows from adding (2) to LP (1).

The proof of the NP hardness in the general case is by a reduction from HAMILTONIAN CIRCUIT. Given a directed graph G, for each vertex u in G, create two compounds u_1, u_2 and create a reaction from u_1 to u_2. For each edge (u, w) of G, create a reaction from u_2 to w_1. Choose T_{IN} to be the set of all reactions corresponding to (derived from) vertices in G and $T_{OUT} = \emptyset$. The only elementary mode that contains all the reactions in T_{IN} corresponds to a Hamiltonian circuit and vice versa. □

As we have seen, the problem of the lemma is easy if $|T_{IN}| = 1$. This leaves open the complexity of the problem if $|T_{IN}| = k$ for any fixed k. Also the proof leaves open the complexity of the problem if the hypergraph underlying the stoichiometry is acyclic, or if it is known that each elementary mode describes a path in the hypergraph.

Theorem 6. *Given a matrix S and a number k, deciding the existence of an elementary mode with at most k reactions in its support is NP-complete.*

Proof. The proof is a reduction from the NP-complete 3-DIMENSIONAL MATCHING problem (3DM) (see [8]): Given a set of elements $X = \{x_1, \ldots, x_{3n}\}$ and given a collection of subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$, each containing exactly 3 elements of X, does there exist a subcollection of $\mathcal{S}$ of n subsets that cover all elements of X?
For each element and each 3-element set of the 3DM instance, a compound vertex is created. The first reaction is an input reaction that has as output all elements of the 3DM instance; *i.e.*, the first column of the stoichiometric matrix has 1-entries at all element compounds and 0 at all element set compounds. The elements of each 3-element set of the 3DM instance are input to a reaction with output the element set compound; *i.e.*, a column in the stoichiometric matrix with -1-entries at the three element compounds, 1 at the element set compound and 0's elsewhere. For each element set there is also an output reaction that has the element set compound as its only input. Finally we choose $k = 2n + 1$.

The vector of reactions which has a 1 at the positions of the first reaction and the two reactions corresponding to each element set of any 3-dimensional matching and 0's elsewhere, clearly forms an elementary mode with $2n + 1$ reactions in its support. On the other hand, any mode must contain the first reaction. Hence, any elementary mode must have a positive value in the first position, and therefore has as output exactly one copy of each element, all of which must have the same value. For every 3-element-set-reaction that we choose, we have to add the corresponding output reaction. Thus to cover all 3n element from the first reaction, we have to choose exactly n reactions that correspond to 3-element sets. Such a set of reactions correspond to a 3-dimensional matching. □

This theorem shows that finding the shortest elementary mode (the one with a minumum number of reactions) is NP-hard. Note that in the theorem, k is considered to be part of the input. For fixed values of k, the problem is trivially solvable in polynomial time by complete enumeration.

As a final example to illustrate the intricacies in detecting elementary modes, we define the notion of a *simple elementary mode* as an elementary mode v such that $\forall j \; v_j \in \{0, 1\}$. The reduction in the proof of Theorem 6 shows that it is hard to find simple elementary modes. Though it is unlikely that any biological relevance will ever be found for the notion of simple elementary mode, the result shows again the subtlety of elementary mode computations, even more so, since the hardness can be extended to any fixed interval of integers.

Corollary 7. *Given a matrix S, deciding the existence of a simple elementary mode is NP-complete.*

3.2 Counting elementary modes

System biologists are interested in enumerating all elementary modes of a metabolic network. Before turning to that problem, we show that merely counting elementary modes is hard. In [13] the authors show that

the number of elementary modes can be bounded by $\binom{m}{n+1}$, but they did not give the complexity of computing the exact number.

Counting elementary modes is essentially a problem of counting the rays of a polyhedral cone, which in its turn is equivalent to a problem of counting vertices of a polytope, which is known to be $\sharp$P-complete [5][1]. Therefore not surprisingly, counting elementary modes turns out to be $\sharp$P-complete. $\sharp$P-complete is a class of computationally hard counting problems (for precise definitions we refer to [16]).

Theorem 8. *Given a matrix S counting the number of elementary modes is $\sharp$P-complete.*

Proof. The proof follows by a reduction from COUNTING PERFECT MATCHINGS IN A BIPARTITE GRAPH, which was shown to be $\sharp$-P hard in [24]. Given a bipartite graph $G = (U, V, E)$ with two color classes U and V, each of size n, we construct hypergraph H. First, we create an input compound vertex s, which we connect to each vertex in U by an ordinary edge, which we direct from s to the U vertex. We direct all edges of E from U to V. Finally, we create an output compound vertex t which we connect with one hyperedge to all vertices of V, and direct this hyperedge from V into t. This relates in the obvious way to a $\{-1, 0, +1\}$-stoichiometric matrix. It is easy to see that an elementary mode corresponds one-to-one to a perfect matching in G. $\qquad\square$

3.3 Enumerating elementary modes

In case all reactions are reversible, an elementary mode corresponds to a minimally dependent set of columns of the stoichiometric matrix. Hence the elementary modes are exactly the *circuits* of a *linear matroid* (for definitions of matroids and circuits we refer to [15] or [18]). In [3] it has been shown how to enumerate circuits of matroids *with polynomial delay*, *i.e.*, the time needed between the consecutive generation of any two circuits is polynomial in the number of elements in the ground set of the matroid, in our case the number of reactions, columns of the stoichiometric matrix. As a result, circuits of a matroid, hence elementary modes of a completely reversible network, can be enumerated in time polynomial in their number. In fact, the modes of the cone form a linear subspace.

Theorem 9. *In case all reactions in a metabolic network are reversible, the elementary modes can be enumerated with polynomial delay.*

The enumeration task becomes dramatically more difficult if the reactions are irreversible. In this case, the modes of the network form a cone, and the elementary modes are the rays of the cone. Indeed, the question if elementary modes can be enumerated with polynomial delay touches a basic open problem in computational geometry (see *e.g.* [6]): given a polyhedral description of a cone, can the rays be enumerated with polynomial delay, or the even the weaker question if the description in terms of its rays can be found in time polynomial in the number of rays.

4 Reaction cuts

In this section, we focus on Reaction Cut Sets. The notion of minimal cut sets in a reaction graph was first introduced by Klamt and Gilles [12]. The motivation is to study so-called "failure modes" that render the functioning of a given target reaction x impossible. A minimal cut set is a set of reactions that cut reaction x. Operationally, this has been defined as a set of reactions whose deletion from the network stops each elementary mode that contains x.

In what follows, we study more particularly two problems: finding a reaction cut of minimum cardinality, which we call MIN REACTION CUT, and enumerating all minimal reaction cuts. We prove that MIN REACTION CUT is APX-hard. For definition of this complexity class we refer to [1]: we observe that APX-hardness implies that there exists a constant c such that finding a solution that is at most a factor c away from the optimum is a NP-hard problem.

Building on results obtained in the previous section, we propose an approximation algorithm. The algorithm runs in polynomial time as it does not require enumeration of all elementary modes containing the target reaction to be cut.

---

[1] In fact, [5] only claims NP-hardness, but the proof establishes #P-completeness.

We then propose significant improvements over existing algorithms for enumerating all minimal reaction sets.

4.1 Finding minimal reaction cuts

The first basic problem about reaction cuts is recognizing them.

Theorem 10. *Given a stoichiometric matrix S, some target reaction r, and a subset F of reactions, deciding if F is a reaction cut of r can be done using LP.*

Proof. Consider the following LP:

$$
\begin{aligned}
\max \quad & \sum_{j \in \mathcal{J}} v_j \\
\text{s.t.} \quad & Sv = 0 \\
& v_r = 1 \\
& v_j = 0 \ \forall j \in F \\
& \sum_j v_j \leq 1 \\
& v_j \geq 0 \ \forall j \notin F \cup r,
\end{aligned}
$$

where $\mathcal{J}$ denotes the set of input reactions of the metabolic network. The optimal solution value is positive if and only if F is not a reaction cut of r. $\qquad \square$

Finding the optimal cut is a lot more difficult.

Theorem 11. MIN REACTION CUT *is APX-hard.*

Proof. We first show a reduction from the NP-hard problem HITTING SET (see [8]): Given a set of elements $X = \{x_1, \ldots, x_n\}$ and given a collection of subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$, find a minimum cardinality subset of elements $Y \subset X$ such that $S_i \bigcap Y \neq \emptyset \ \forall i = 1, \ldots, m$.

For each element x_j and for each set S_i, we create a compound vertex, which we also denote by x_j and S_i, respectively. We create an input compound vertex s, a compound vertex t and an output compound vertex t'. For each x_j, we create a *single reaction* one s to one x_j. Similarly, for each S_i, we create a *single reaction* one S_i to one t. We also create the *single reaction* from one t to one t'. For each set $S_i = \{x_{i_1}, \ldots, x_{i_k}\}$, we create a *composed reaction* with input compounds one of each of $x_{i_1}, \ldots, x_{i_k}$ and output compound one S_i. Thus, the stoichiometric matrix contains only entries with value $-1, 0,$ or $+1$.

To each set $S_i = \{x_{i_1}, \ldots, x_{i_k}\}$ corresponds an elementary mode consisting of the reactions $(s \rightarrow x_{i_1}), \ldots, (s \rightarrow x_{i_k}), (x_{i_1}, \ldots, x_{i_k} \rightarrow S_i), (S_i \rightarrow t), (t \rightarrow t')$. Indeed, it is easy to check that the vector that assigns a 1 to each of these reactions and a 0 otherwise is indeed a mode. Removing any reaction from this set gives a submatrix which does not have any mode.

Moreover, suppose that some elementary mode would have $v(S_i \rightarrow t) = a_i > 0$ and $v(S_j \rightarrow t) = a_j > 0$, respectively. Then this mode should also have $v(x_{i_1}, \ldots, x_{i_k} \rightarrow S_i) = a_i$ and $v(x_{j_1}, \ldots, x_{j_h} \rightarrow S_j) = a_j$, and also $v(t \rightarrow t') = a_i + a_j$ and $v(s \rightarrow x_\ell) = a_i \ \forall x_\ell \in (S_i \backslash S_j), v(s \rightarrow x_\ell) = a_j \ \forall x_\ell \in (S_j \backslash S_i), v(s \rightarrow x_\ell) = a_i + a_j$ $\forall x_\ell \in (S_i \cap S_j)$, and $v(s \rightarrow x_\ell) = 0$ otherwise. Hence this is the linear combination of two elementary modes of the above type, and therefore by itself not an elementary mode.

Thus, the set-related elementary modes are exactly all the elementary modes, and from each of them some reaction must be selected in the reaction cut. Selecting $(s \rightarrow x_\ell)$ cuts all the elementary modes whose corresponding set contains x_ℓ. Thus, the reactions from s to the x's of a hitting set cut all elementary modes. On the other hand, any composed reaction $(x_{i_1}, \ldots, x_{i_k} \rightarrow S_i)$ or (S_i, t) reaction in the reaction cut can be replaced by one (s, x_j) reaction (with $x_j \in S_i$), giving another reaction cut. Thus, there exists a minimum reaction cut consisting only of reactions of type (s, x_j), hence corresponding to a hitting set. This completes the proof of NP-hardness.

To prove APX-hardness, we observe that the reduction is approximation preserving: minimal reaction cuts and hitting sets have a one-to-one correspondence and that the reduction is indeed an AP-reduction (see for the definition *e.g.* [1]). Since HITTING SET is APX-hard optimization problem this observation implies that MIN REACTION CUT is APX-hard. $\qquad \square$

One may argue that the reduction yields reactions in the metabolic network that contain too many different compounds. Typically, the number of different compounds involved in metabolic reactions is no more than 5 or 6. However, this restriction will unfortunately not be of great help: in a technical report version of this paper, we prove that the problem remains NP-hard for networks with only one target reaction, and where all compounds (vertices) have in-degree at most 2, all vertices have out-degree at most 2, and each reaction has 2 input compounds and 1 output compound.

4.2 Approximation algorithm for finding a minimum reaction cut

On the positive side, we design an approximation algorithm for finding minimum reaction cuts, even for a weighted version of the problem. We assume that a weight function w associates to each reaction r a positive weight $w(r)$. Given a stoichiometric matrix S and a weight function w, we are interested in finding a reaction cut F^* of minimum total weight.

The algorithm consists of two phases: in the first phase, a set F of reactions is constructed by starting from the empty set and adding reactions until a reaction cut of the target reaction x is obtained. The set F is not necessarily a minimal reaction cut. In the second phase, minimality is obtained by removing reactions from F.

Given a stoichiometric matrix S and a set of reactions F, we denote by S_F the stoichiometric matrix obtained from S by removing the columns corresponding to all reactions in F; with a slight abuse of notation, we denote the sum of the weights of reactions in a set G by $w(G)$.

Algorithm RC (Reaction Cut)
    input:
        a stoichiometric matrix S, a weight function w, a reaction x to be cut;
    phase 1
        $F = \emptyset$;
        while F is not a reaction cut of x
        do begin
            let C be the set of reactions defining an elementary mode in S_F that includes x
            let $\bar{w} = \min_{r \in C} w(r)$
            for each reaction r in C
            do begin
                $w(r) = w(r) - \bar{w}$
                if $w(r) = 0$ **then** $F = F \bigcup \{r\}$
            end
        end
    phase 2
        let $r_1, r_2, \ldots, r_k$ be the reaction in F
        for $j = 1$ **to** k **do**
            if $F - r_j$ is a reaction cut of x **then** $F = F - r_j$
    output: F

For the performance analysis of the solution found by the algorithm we exploit the local ratio technique, a general technique for proving performance ratios of approximation algorithms [2]. It is based on decomposing the weight function associated to each reaction.

Lemma 12. *Let S be a stoichiometric matrix, and let F^*, F_1^* and F_2^* be the minimum reaction cuts of x with respect to three different weight functions w, w_1 and w_2, respectively, such that $w(r) = w_1(r) + w_2(r)$ for each reaction r. Then*

$$w(F^*) \geq w_1(F_1^*) + w_2(F_2^*)$$

Proof.

$$w(F^*) = w_1(F^*) + w_2(F^*) \geq w_1(F_1^*) + w_2(F_2^*)$$

<div align="right">□</div>

This technique has been applied to a number of combinatorial optimization problems arising in several areas (scheduling, graph, packing, etc.). Inspired by [4], we prove the following theorem.

Theorem 13. *Given a stoichiometric matrix S and a reaction x, Algorithm* REACTION CUT *runs in polynomial time and returns a reaction cut F of x such that $w(F) \leq \lambda w(F^*)$, where F^* is the minimum reaction cut of x and λ is the maximum number of reactions in an elementary mode in S including x.*

Proof. Assume that S contains n reactions. In Phase 1, the algorithm performs the test of checking whether a set of reactions is a reaction cut of x at most n times. It also computes an elementary mode including reaction x for n times at most. Analogously, Phase 2 of the algorithm performs at most n times the test of deciding whether a set is a reaction cut of x. By Theorems 5(i) and 10, it follows that the running time of the algorithm is polynomial.

The proof proceeds by induction on the number of reactions, with the basis of a stoichiometric matrix of only 1 reaction clearly being true. Suppose it is true for n reactions and consider a stoichiometric matrix S with $n + 1$ reactions. Let F be the reaction cut set returned by RC.

Let C be the elementary mode detected in the first call on Phase 1 and $\delta = \min_{r \in C} w(r)$.

We define two new weight functions w_1 and w_2:

$$w_1(r) \quad = \quad \delta \text{ if } r \text{ belongs to } C \text{ and } w_1(r) = 0 \text{ otherwise}$$
$$w_2(r) \quad = \quad w(r) - w_1(r).$$

Let F_1^* and F_2^* be minimum reaction cut sets under weight functions w_1 and w_2, respectively. Since $w(r) \geq w_1(r) \geq 0$, we have $0 \leq w_2(r) \leq w(r)$ and, therefore, the conditions of Lemma 12 apply. Let F_1^* and F_2^* be minimum reaction cut sets under weight functions w_1 and w_2, respectively.

Claim 1. $w_1(F) \leq \lambda w_1(F_1^*)$
Observe that $w_1(F_1^*) = \delta$, because for cutting elementary mode C, one reaction of C is sufficient and necessary, while for any other elementary mode, a reaction with weight 0 can be selected in the reaction cut. Moreover, the weight of $w_1(F) \leq m\delta$, where m denotes the number of reactions in C, because all the reactions not belonging to C have cost 0. This together with $m \leq \lambda$ proves the claim.

Claim 2. $w_2(F) \leq \lambda w_2(F_2^*)$
Let F_1 be the set of reactions selected after passing Phase 2 for the first time, *i.e.* the set of reactions that cut C. Notice that in fact F_1 contains one reaction with weight δ only. Let $F_2 = F \setminus F_1$, which is, by definition of the algorithm, the RC solution for the problem with stoichiometric matrix S_{F_1} obtained by deleting the columns of reaction set F_1 from S and weight function w_2. Let $\mathcal{F}_2^*$ be the optimal solution to the latter problem.

Since $w_2(F_1) = 0$, any reaction cut for S_{F_1} w.r.t. w_2 can be supplemented to a reaction cut for S w.r.t. w_2, by adding F_1 at no extra cost, if necessary. In particular, $w_2(F) = w_2(F_1) + w_2(F_2) = w_2(F_2)$, and $w_2(F_2^*) = w_2(\mathcal{F}_2^*)$. Application of the induction hypothesis to the performance of RC to S_{F_1} with weight function w_2 proves that $w_2(F_2) \leq \lambda w_2(\mathcal{F}_2^*)$ and therefore $w_2(F) \leq \lambda w_2(F_2^*)$.

Both claims together with Lemma 12 yields

$$w(F) = w_1(F) + w_2(F) \leq \lambda w_1(F_1^*) + \lambda w_2(F_2^*) \leq \lambda w(F^*).$$

$\square$

We finally observe that the above result can be easily extended to the case when more than one reaction should be cut. Given S, assume we are interested in finding a cut of reactions $x_1, x_2, \ldots$; two problems arise: we might be interested in either the problem of cutting all reactions $x_1, x_2, \ldots$ or in cutting at least one.

The result of Theorem 13 can be easily extended to both problems above, by adding compounds and reactions to the stoichiometric matrix. Namely, if we are interested in cutting all reactions in $x_1, x_2, \ldots$ we may add one compound y to the output of each reaction x_i, $i = 1, 2, \ldots$ and add a new reaction $\bar{r}$ that transforms y in an output z. Clearly, cutting $\bar{r}$ requires to cut each reaction in $x_1, x_2, \ldots$. Note that the above transformation might not be feasible because it is not mass balanced; however a slight modification ensures the mass balance and feasibility properties. Details will be given in the full paper.

A similar transformation applies to the problem in which we are interested in cutting at least one reaction in $x_1, x_2, \ldots$.

4.3 Enumerating reaction cuts

Beyond the question of finding a reaction cut, or a minimum reaction cut, the question of enumerating all reaction cuts may also be interesting. As for modes, one can concentrate on minimal sets [12].

Minimality refers to reaction cuts from which no reaction can be removed without destroying the cutting property. Klamt and Gilles [12] propose an algorithm based on enumerating all possible subsets of reactions starting from singleton sets, then all pair sets, then all triples, and so on. For each candidate set F, they propose to test whether all elementary modes are cut by F. Clearly this test is theoretically, and many times also practically, very inefficient. We propose as an alternative to use Theorem 10.

It remains an intriguing open problem if we can do essentially better in case of irreversible reactions. In case all reactions are reversible, a minimal reaction cut is a co-basis of the linear matroid constituted by the columns of the stoichiometric matrix. Bases of matroids and therefore co-bases of matroids can be enumerated with polynomial delay [9].

4.4 Other cuts

One may argue that a reaction cut that kills too many elementary modes is not desirable. As an alternative one may therefore be interested in finding a reaction cut which cuts the target reaction but leaves as many elementary modes intact as possible.

MAX SURVIVING REACTION CUT: Given a stoichiometric matrix and a reaction, find the reaction cut isolating that reaction and leaving a maximum number of elementary modes intact.

Given hardness of MIN REACTION CUT it is not surprising that also this problem is NP-hard. Taking the reduction from HITTING SET in the proof of Theorem 11, and introducing for each x_j-vertex an extra *output* vertex t_j and a reaction $(x_j \to t_j)$ the NP-hardness can be verified easily.

Theorem 14. MAX SURVIVING REACTION CUT *is NP-hard.* □

Other types of cuts in reaction graphs can be formulated.

5 Conclusion

Elementary modes and minimal reaction cuts are common tools in metabolic network analysis. Their computation is not trivial and poses a computational challenge. Several algorithms have been proposed to bring solutions to this problem but no systematic complexity analysis had been carried out.

We show here that some problems, like checking the consistency of a network, finding one elementary mode or checking that a set of reactions constitutes a cut, are easy problems and we emphasise that "easy" also means that they can readily be solved using existing software.

We also prove the hardness of central problems like finding an elementary mode containing a specified set of reactions, counting elementary modes or finding a minimum reaction cut.

Furthermore, we propose an approximation algorithm for computing the minimum reaction cut as well as a significant optimisation of the original algorithm used for computing minimal cut sets. Both results are based on the idea of avoiding to compute the elementary modes for obtaining the reaction cuts.

References

[1] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and approximation – Combinatorial optimization problems and their approximability properties.* Springer-Verlag, Berlin, 1999.

[2] A. Bar-Noy, R. Bar-Yehuda, A. Freund, J. Naor, and B. Schieber. A unified approach to approximating resource allocation and scheduling. *J. ACM*, 48(5):1069–1090, 2001.

[3] E. Boros, K. M. Elbassioni, V. Gurvich, and L. Khachiyan. Algorithms for enumerating circuits in matroids. In *ISAAC*, pages 485–494, 2003.

[4] C. Demetrescu and I. Finocchi. Combinatorial algorithms for feedback problems in directed graphs. *Inf. Process. Lett.*, 86(3):129–136, 2003.

[5] M. E. Dyer. The complexity of vertex enumeration methods. *Mathematics of Operations Research*, 8:381–402, 1983.

[6] K. Fukuda and A. Prodon. Double description method revisited. In *Combinatorics and Computer Science*, number 1120 in Lecture Notes in Computer Science, pages 91–111, 1996.

[7] J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5:175, 2004.

[8] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness.* Freeman, 1979.

[9] L. G. Khachiyan, E. Boros, K. M. Elbassioni, V. Gurvich, and K. Makino. On the complexity of some enumeration problems for matroids. *SIAM J. Discrete Math.*, 19(4):966–984, 2005.

[10] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.

[11] S. Klamt, J. Gagneur, and A. von Kamp. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *Syst Biol (Stevenage)*, 152(4):249–255, Dec 2005.

[12] S. Klamt and E. D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, Jan 2004.

[13] S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*, 29(1-2):233–236, 2002.

[14] S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends Biotechnol*, 21(2):64–69, Feb 2003.

[15] J. G. Oxley. *Matroid theory*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1992.

[16] C. H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994.

[17] J. A Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–405, 2004.

[18] A. Schrijver. *Combinatorial optimization: Polyhedra and Efficiency*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003.

[19] S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–332, 2000.

[20] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, 2:165182, 1994.

[21] S. Schuster, C. Hilgetag, J. H. Woods, and D. A. Fell. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J Math Biol*, 45(2):153–181, 2002.

[22] P.D. Seymour. The matroids with the max-flow min-cut property. *J. Comb. Theory Ser. B*, 23(7):189–222, 1977.

[23] R. Urbanczik and C. Wagner. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, 21(7):1203–1210, 2005.

[24] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.