Modes and Cuts in Metabolic Networks: Complexity and Algorithms*

Vicente Acuña[2,3], Flavio Chierichetti[1], Vincent Lacroix[2,3,6], Alberto Marchetti-Spaccamela[1], Marie-France Sagot[2,3], Leen Stougie[4,5]

[1] Universitá di Roma "La Sapienza", Via Ariosto 25, 00184 Rome, Italy
[2] Université de Lyon, F-69000, Lyon ; Université Lyon 1 ; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, 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
[6] Centre de Regulació Genòmica, Genome Bioinformatics Research Group - CRG, PRBB, Aiguader 88, 08003 Barcelona, Spain

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 forms 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 give a systematic overview of the complexity of optimisation problems related to modes. We first establish results regarding network consistency. Most consistency problems are easy, i.e., they can be solved in polynomial time. 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 then show that counting the number of elementary modes is ♯P-complete. We emphasize that the easy problems can be solved using currently existing software packages.

We then analyse the complexity of a closely related task which is the computation of so-called minimum reaction cut sets and 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 reaction cut set. The second one is a test for verifying whether a set of reactions constitutes a reaction cut; this test can 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,

*Research was partially supported by the Dutch BSIK-BRICKS project, by the Dutch mathematics cluster DIAMANT, the Project II-Vert of the ACI Nouvelles Interfaces des Mathématiques of the French Ministry of Research, the Project IBN of the Actions Concertées Coopératives of the INRIA, the Project REGLIS of the French Agence Nationale de la Recherche Number ANR-05-NT05-3_45205, the EU FET Integrated Project AEOLUS, IST- 15964

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 based on differential equations whereas structural analyses are mainly based on graph-related formalisms or, as far as metabolism is concerned, on a constraint-based modelling. The latter term is commonly employed in the bioinformatics community following two papers by Palsson (2000) and Covert and Palsson (2003). In the constraint-based framework, the network may still be modelled as an edge-labelled hypergraph, but several types of constraints (stoichiometric, thermodynamic and in some cases regulatory) are added to restrict the possible fluxes through the network. 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).

In a constraint-based approach, only admissible flux distributions are of interest. 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 (Edwards et al., 2001; Fong and Palsson, 2004) 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 (Schuster and Hilgetag, 1994), EM for short. Intuitively, an elementary mode is a special mode that has the property of not containing any other mode.

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 (Schuster et al., 2000) and a mode is elementary when the removal of one enzyme causes it to fail.

The related concept of extreme pathway has also been introduced in the field (Schilling et al., 2000). Extreme pathways are actually a subset of elementary modes. Both notions coincide in the case where all exchange reactions (reactions connecting some metabolite with the surrounding of the model) are irreversible. For a detailed comparison of both approaches, see Klamt and Stelling (2003).

As outlined by Schuster et al. (2002b), the concept of minimal T-invariant used in Petri Nets is also closely related to the concept of elementary mode. Both notions coincide in the case where all reactions are irreversible. For completeness sake, we can also mention that the extreme currents defined by Clarke (1981) also coincide with elementary modes in the irreversible case. Unlike extreme pathways and elementary modes, minimal T-invariants and extreme currents have only been defined in the case of a network of irreversible reactions. Clearly, there are links between the algorithms for enumerating elementary modes and the ones for minimal T-invariants since, as we shall see, they all boil down to enumerating the extreme rays of a convex cone. We will not discuss the techniques in detail here. Interested readers may refer to (Colom and Silva, 1991) for algorithms for enumerating minimal T-invariants and to Schuster et al. (2002a); Urbanczik and Wagner (2005); Gagneur and Klamt (2004) for enumerating elementary modes. More generally, the usefulness of Petri-Net approaches to the study of metabolic pathways is presented in (Voss et al., 2003).

Another concept we study here is closely related to the notion of elementary mode. This is the concept of a reaction cut set, recently introduced by Klamt and Gilles (2004). In order to avoid any confusion with other types of cuts in graphs or hypergraphs that may be found in the literature (see e.g. (Seymour, 1977)), 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 by Klamt (2006), the task to be silenced can be a combination of reactions. 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).

The main contribution of this paper is in giving a systematic overview of the complexity of optimisation problems related to modes. We first establish results regarding network consistency (Section 2.1). Most consistency problems can be solved in polynomial time (are easy). Most, if not all, of these results have been stated before in the literature. It is in fact easy to formulate these problems as LP-problems, which has the side advantage that computer packages are available to solve them.

We then establish the complexity of finding and enumerating 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 computational complexity of problems concerning reaction cuts. We prove that finding a *minimum* reaction cut set, one that contains a minimum number of reactions, is hard (Sections 4.1) 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.

2 Modes

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

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. Entry $S(i,j)$ of the matrix takes value k if reaction j produces k units of metabolite i, in which we say i is output of reaction j, and $-k$ if reaction j consumes k units of metabolite i, in which case we say that i is input of reaction j; otherwise, $S(i,j)$ 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.

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

In the work of Klamt et al. (2005) it is already observed that standard linear algebra teaches us how to check that $Sv = 0$ in order to decide if $v \geq 0$ is a mode.

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

An **elementary mode** is a non-trivial flux vector $v \neq 0$ that satisfies conditions 1 and 2 and

3. there is no non-trivial flux vector $w \neq 0$ satisfying conditions 1 and 2 such that : $R(w) \subset R(v)$.

Modes and elementary modes can be given a geometrical interpretation. Indeed, the set of vectors $\{v \geq 0 \mid Sv = 0\}$ defines a convex polyhedral cone in the flux space. When all reactions are irreversible, the elementary modes exactly correspond to the *extreme rays* of this cone. An extreme ray is a ray of the cone that can not be expressed as a convex combination of other rays of the cone. We refer the reader to (Schuster and Hilgetag, 1994) for the earliest known proof, which relies on basic linear algebra.

Lemma 1. *If all reactions are irreversible, then the set of EMs corresponds one-to-one to the set of extreme rays of the cone* $\{v \geq 0 \mid Sv = 0\}$.

Gagneur and Klamt (2004) observed that when some reactions are reversible, one can define a pointed cone in a higher dimensional space by representing each reversible reaction by two irreversible reactions, in the obvious way: suppose the reaction r is represented in S by the column s_r then we add the column $s_{\overleftarrow{r}} = -s_r$ to S, yielding S^+, and require both v_r and $v_{\overleftarrow{r}}$ to be non-negative. The matrix S^+ has extreme rays that consist of those of S and the vectors v with $v_r = v_{\overleftarrow{r}} = 1$ and $v_i = 0$ otherwise, corresponding to length-2 cycles consisting of the two reaction making up for a reversible reaction. We can easily detect and simply ignore these length-2 cycles. A consequence of this observation is that we can analyse the complexity and propose algorithms in the irreversible case without loss of generality.

In the other extreme case in which all reactions are reversible ($Irrev = \emptyset$), 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 (Oxley, 1992) or (Schrijver, 2003)).

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 (Schuster et al., 2000). A network is said to be consistent if all its reactions belong to at least some mode, or equivalently, in terms of Petri-net terminology, if the network is covered by T-invariants (Heiner and Koch, 2004). 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 an overview of some problems related to the consistency of stoichiometric matrices. If a matrix S is not consistent, this may indicate a case of incomplete modelling of the metabolic network. In that sense, detecting inconsistency is a valuable tool for finding deficiencies in the metabolic network description.

In the following theorems we explicitly state that the problems can be solved using LP. Since the LP-formulations have a size that is bounded by a polynomial function of the stoichiometric matrix, we implicitly state that these problems are easy (in P). We chose for stating solvability through LP to emphasize that they are not only theoretically tractable but that in fact off-the-shelf computer packages can be used to solve the problems.

Theorem 2. *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.

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

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 3. *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

$$\begin{array}{ll} \max & v_h \\ \text{s.t.} & Sv = 0 \\ & \sum_j v_j \leq 1 \\ & v \geq 0 \end{array}$$

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 and it must be deleted to make S consistent. This is a safe operation: since h belongs to no mode, eliminating h will not eliminate any existing mode. For the same reason, the order of elimination is indifferent. □

Unfortunately, a problem complementary to the previous one is hard.

Theorem 4. *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.*

This is of practical interest as in general, when a stoichiometric matrix is not consistent, it is because some enzymes, and therefore some reactions, were not detected as present due to the lack of a strong enough similarity with the enzymes in a known network, usually that of *Escherichia coli*, from which the one for a newly sequenced organism was inferred.

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 7. □

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 visualisation, hypergraphs may also be used. A 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**
External input: a,b	External input: a,b
External output: f	External 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).

3 Elementary modes

As mentioned above in Section 2, we may see an elementary mode as an extreme ray of the cone $\{v \geq 0 \mid Sv = 0\}$. The solution methods for the easy problems related to finding EMs rely on this equivalence. It is consistent with the observation in (Gagneur and Klamt, 2004) 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 (2002) mainly focus on finding an upper bound on the number of elementary modes.

In fact, as mentioned in (Fukuda and Prodon, 1996), 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 following tasks are easy: finding an EM and finding an EM that contains one specified reaction. However, the following task is hard: finding an EM that contains a specified set of reactions.

As observed already in (Klamt et al., 2005), 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 calculating the rank of the submatrix of S consisting of the reaction in the support of v. If this is equal to the rank of S minus 1 the vector v represents an elementary mode (Klamt et al., 2005). But also finding some EM is easy.

Theorem 5. *Given a stoichiometric matrix S, an elementary mode can be found in polynomial time.*

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

$$\begin{array}{ll} \max & v_h \\ \text{s.t.} & Sv = 0 \\ & \sum_j v_j \leq 1 \\ & v \geq 0. \end{array} \tag{1}$$

In case of a consistent matrix, there is an optimal solution which 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. Let v_h^* be the optimal solution value.

Using interior point methods for finding an optimal solution, does not necessarily yield a vertex of the polytope. However, if a vertex is not found the objective function is set equal to the optimal value, $v_h = v_h^*$, and is added as a constraint. As an auxiliary objective, maximization of one of the decision variables (other than v_h) is chosen. In this way, iteratively applying an interior point method, in each such iteration the dimension of the optimal solution set is diminished by at least one. Thus, after a number of iterations less than the number of variables a vertex will be obtained, and we conclude that an elementary mode can be found in polynomial time. □

Of course, we can use also any simplex method-based LP-package for solving LP (1), since, though being worst-case not a polynomial time method, it is very fast in practice. Moreover, it has the advantage that it will always produce directly a non-all-0 vertex of the polytope as an optimal solution.

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.

The existence of an *elementary mode* with the same properties for any set T_{IN} is NP-complete in general, which may (partly) explain the difficulties we encounter in enumerating elementary modes.

Theorem 6. *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) solvable in polynomial time if $|T_{IN}| = 1$,
(ii) NP-complete in the general case.

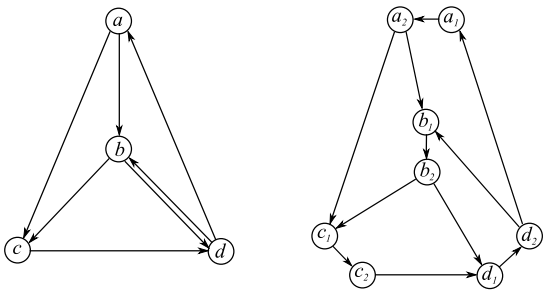

Figure 1: Graphical illustration of the HC reduction.

Proof. We start by observing that this decision problem is in NP because, if we give a flux vector as certificate, we can check in polynomial time if it is an elementary mode with the desired properties. We observe also that set T_{OUT} has no influence on the complexity of the problem. Indeed, we just need to delete from S the columns corresponding to the reactions in T_{OUT} and solve the problem on the reduced matrix. If $|T_{IN}| = 1$ the proof follows from selecting h in the LP (1) as the only reaction in T_{IN}.

NP-completeness in the general case is proved by a reduction from Directed Hamiltonian Circuit. A Hamiltonian circuit of an directed graph is a circuit (loop) that visits each node of the graph exactly once. Deciding if a (directed) graph contains a Hamiltonian Circuit is a well known NP-complete problem. The intuition behind the proof is to build, in polynomial time, from a general instance of the Directed Hamiltonian Circuit problem, a specific instance of our problem, that is a network, with the following characteristic: each elementary mode in the network that contains all reactions in T_{IN} corresponds to a Hamiltonian circuit in the graph and vice versa. A solution to our problem therefore provides a solution to the Directed Hamiltonian Circuit problem.

Given a directed graph G that is a general instance of the Directed Hamiltonian Circuit problem, 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 (see Figure 1). Let H be this network that can be built in linear time from G. The corresponding stoichiometric matrix is simply the $\{-1, 0, +1\}$ incidence matrix of this directed bipartite graph. Choose T_{IN} to be the set of all reactions corresponding to (derived from) vertices in G and $T_{OUT} = \emptyset$.

Notice that any circuit C in H corresponds to an elementary mode. Just set all values for the reactions corresponding to the arcs on C equal to 1 and the rest to 0 gives a mode. It is clear that no subset of the arcs can give rise to a mode, hence it must be an elementary mode. Notice also that because of the absence of outputs in this network any mode in H has to contain a circuit in its support. But a circuit is, in fact, the support of an elementary mode and therefore any elementary mode must be a single circuit of H. Since circuits in H and G have a one-to-one correspondence, any elementary mode corresponds to a circuit in G and vice versa.

In particular there is a one-to-one correspondence between elementary modes of H that contain all of T_{IN} and circuits in G that contain all vertices of G, i.e., Hamilton Circuits. □

As we have seen, the problem is easy if $|T_{IN}| = 1$. We can observe that it becomes trivial when $|T_{IN}| \geq rank(S) + 1$. Indeed, according to Lemma 4 in (Schuster et al., 2002a), no elementary mode can have as many non-zero elements as that. This leaves an interesting and rather fundamental open problem:

Open problem: What is the complexity of the problem if $|T_{IN}| = k$ for any fixed $k, 1 < k < rank(S) + 1$.

In fact, we conjecture that it is hard already if $|T_{IN}| = 2$.

Theorem 7. *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.*

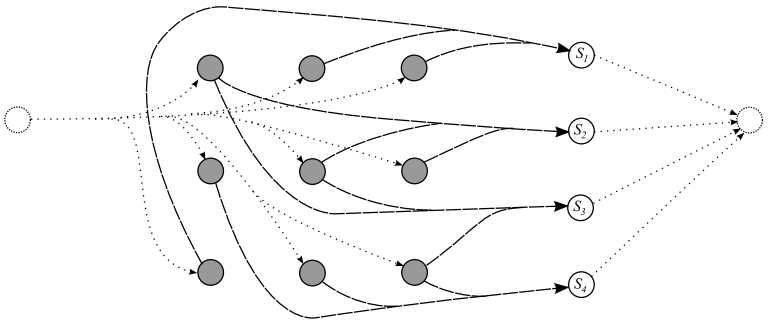

Figure 2: Graphical illustration of the 3DM reduction.

Proof. Clearly this decision problem is in NP because, if we give a flux vector as certificate, we can check in polynomial time if it is an elementary mode with at most k reactions in its support. The proof of the hardness is a reduction from the NP-complete 3-Dimensional Matching problem (3DM) (see (Garey and Johnson, 1979)): Given a set of elements $X = \{x_1, \ldots, x_{3n}\}$ partitioned into three sets of n elements each, and given a collection of 3-element-subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$ each subset containing exactly one element from each set of the partition, does there exist a subcollection of $\mathcal{S}$ of n subsets that covers all elements of X?

The reduction is depicted in Figure 2. 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, the grey vertices in Figure 2; *i.e.*, the first column of the stoichiometric matrix has 1-entries at all element compounds and 0 at all element set compounds. For each 3-element set of the 3DM instance a reaction is created with input the compounds corresponding to the three elements of the set and output the compound corresponding to the 3-element set, the s_i-nodes in Figure 2; *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 3-element set there is also an output reaction that has the 3-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 3-element set of any 3-dimensional matching and 0's elsewhere, clearly forms an EM with $2n + 1$ reactions in its support. On the other hand, any mode must contain the first reaction. Hence, any EM 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 corresponds to a 3-dimensional matching. □

This theorem shows that finding the *shortest* elementary mode (the one with a minimum 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. In practice, enumerating elementary modes with at most k reactions may therefore be possible for small values of k. To the best of our knowledge, there is no current application of short elementary modes, but it should become interesting if size were considered as a relevant criterion to classify elementary modes. Short elementary modes may also be seen as good seeds for a motif detection algorithm such as in (Lacroix et al., 2006): two (or more) short EMs that represent connected sets of equivalent chemical transformations (equivalent enzymatic functions) may help to understand how metabolism evolved. In any case, Theorem 7 is also interesting in itself for the further insight it provides in the hardness of elementary mode computations.

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 7 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 EM computations, even more so, since the hardness can be extended to any fixed interval of integers.

Corollary 8. *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 their work, Klamt and Stelling (2002) 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 ♯P-complete (Dyer, 1983)[1] in general.

♯P is a complexity class of counting problems associated with decisions problems in NP, for instance, counting the number of hamiltonian circuits in a graph is in ♯P. Since if we can count objects, we can decide the existence of at least one of them, a counting problem in ♯P must be at least as hard as the corresponding decision problem in NP. Like the class NP also ♯P has complete problems; the hardest problems within the class. Solving any of the ♯P-complete problems in polynomial time would prove that any problem in ♯P can be solved in polynomial time, and therefore P=NP. For precise definitions, we refer the reader to (Papadimitriou, 1994).

Not surprisingly, given its connection with the extreme rays of a convex cone, counting elementary modes turns out to be also ♯P-complete.

Theorem 9. *Given a matrix S counting the number of elementary modes is ♯P-complete.*

Proof. The proof follows by a reduction from the ♯P-hard problem Counting perfect matchings in a bipartite graph (Valiant, 1979). A bipartite graph G is a graph whose set of nodes may be divided into two sets V_1 and V_2 such that every edge in G links a node in V_1 to one in V_2. A perfect matching in a bipartite graph is a set of edges such that no two of them share a common node (in either V_1 or V_2) and all nodes in V_1 and V_2 are covered (by exactly one edge). Given a bipartite graph $G = (U, V, E)$ with two color classes U and V, each of size n, we construct the following hypergraph H. First, we create an input compound vertex s, which we connect with one hyperedge to all vertices in U, and direct this hyperedge from s into U. 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 EM corresponds one-to-one to a perfect matching in G. □

3.3 Enumerating elementary modes

Listing all feasible solutions of a combinatorial problem is a fundamental problem in combinatorics. Typical cases of interest that have been considered in the literature are enumerating the spanning trees of a graph, enumerating the vertices and the facets of a convex polyhedron or an arrangement of hyperplanes given by a system of linear inequalities.

Since the number of feasible solutions to be enumerated may be exponential in the size of the input description the efficiency of an enumeration algorithm is measured in both the input and output sizes (see e.g., (Lawler et al., 1980)). Namely, an enumeration problem is said to be solvable in polynomial total time if the output can be generated in time polynomial in the input and output size. Usually the stricter requirement of *polynomial delay* is required. In this case we require that, given a feasible set of solutions S, the time required for generating a new feasible solution (not in S) is polynomial in the input size. Clearly, if an enumeration problem can be solved with polynomial delay then it is also solvable in polynomial total time.

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 (Oxley, 1992) or (Schrijver, 2003)). In (Boros et al., 2003) it has been shown how to enumerate circuits of matroids *with polynomial delay*, *i.e.*, the time

[1] In fact, (Dyer, 1983) only claims NP-hardness, but the proof establishes #P-completeness.

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 10. *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.

Open question: Can elementary modes be enumerated with polynomial delay if $Irrev \neq \emptyset$.

Indeed, this touches a basic open problem in computational geometry (see *e.g.* (Fukuda and Prodon, 1996)): given a polyhedral description of a cone, can the rays be enumerated with polynomial delay, or the even the weaker question whether the description in terms of its rays can be found in time polynomial in the number of rays. The enumeration methods proposed in the literature are all based on the double description method introduced in (Fukuda and Prodon, 1996). The fastest one at this moment is by Terzer and Stelling (2006).

4 Reaction cuts

In this section, we focus on Reaction Cut Sets. The notion of minimal cut sets in a metabolic network represented as a hypergraph was first introduced by Klamt and Gilles (2004). The motivation is to study so-called "failure modes" that render the functioning of a given target reaction r^o impossible. A minimal cut set is a set of reactions that must be cut (removed) in order to prevent a flux through the target reaction r^o. Operationally, this has been defined as a set of reactions whose deletion from the network stops each elementary mode that contains r^o.

Before proceeding we mention that the notion of s,t-cut of a hypergraph, i.e., a cut that separates nodes s and t, has been proposed and studied for directed hypergraphs. In (Gallo et al., 1998) it has been observed that finding s,t-cuts in unweighted directed hypergraphs can be done in polynomial time if all hyperdges are defined by a subset of input nodes and a single destination node; in the context of metabolic networks this would model the situation in which each reaction is irreversible and produces a single metabolite. We also refer to (Ausiello et al., 2001) for a survey of related results on directed hypergraphs.

In what follows, we study 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 cannot be approximated within any constant approximation ratio unless P=NP. Building on results obtained in the previous section, we propose an approximation algorithm of ratio λ corresponding to the maximum number of reactions in an elementary mode in S including the target reaction. The algorithm runs in polynomial time as it does not require enumeration of all elementary modes containing the target reaction to be cut.

We then notice how as a consequence of Theorem 11 an easy improvement over existing algorithms for enumerating all minimal reaction sets can be obtained. The improvement affects not the number of candidate sets that needs to be checked, this is the same for both, but the efficiency of each check that is thereby greatly improved. Indeed, the approach in (Klamt and Gilles, 2004) requires computing all elementary modes in order to check, for each candidate set, whether it represents a minimal set cutting all possible routes for producing the target reaction. Using Theorem 11, this check can be done for each set directly.

4.1 Finding minimal reaction cuts

The first basic problem about reaction cuts is recognising them.

Theorem 11. *Given a stoichiometric matrix S, some target reaction r^o, and a subset F of reactions, deciding if F is a reaction cut of r^o can be done using LP.*

Proof. Consider the following LP:

$$\begin{aligned} \max \quad & v_{r^o} \\ \text{s.t.} \quad & Sv = 0 \\ & v_j = 0 \ \forall j \in F \\ & \textstyle\sum_j v_j \leq 1 \\ & v_j \geq 0 \ \forall j \notin F \cup r^o. \end{aligned}$$

The optimal solution value is positive if and only if F is not a reaction cut of r^o. □

Finding the optimal cut is a lot more difficult.

Theorem 12. Min Reaction Cut *is NP-hard.*

Proof. We prove NP-completeness of the decision version of Min Reaction Cut. By the previous theorem this problem is in NP. Completeness is proved through a reduction from Hitting set (see (Ausiello et al., 1999)): Given a set of elements $X = \{x_1, \ldots, x_n\}$, a collection of subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$, and an integer K, does there exist a subset $Y \subset X$ of at most k elements 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. To facilitate the exposition we create three additional compounds s, t and t'. s and t' are considered as external compounds and we create the reaction $(1t \rightarrow 1t')$ as the target reaction r^o. For each x_j, we create a reaction $(1s \rightarrow 1x_j)$. Similarly, for each S_i, we create a $(1S_i \rightarrow 1t)$. For each set $S_i = \{x_{i_1}, \ldots, x_{i_k}\}$, we create a reaction with multiple input compounds $1x_{i_1}, \ldots, 1x_{i_k}$ and output compound $1S_i$. Thus, the stoichiometric matrix contains only entries with value -1,0, or $+1$, and we suppress the coefficient 1 in the description of reactions from here on. We select for the decision version of Min Reaction Cut the same integer K as in the Hitting Set instance.

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 mode would contain reactions corresponding to two sets, that is, $v(S_i \rightarrow t) = a_i > 0$, $v(S_j \rightarrow t) = a_j > 0$ and $v(S_l \rightarrow t) = 0 \ \forall l \notin \{i, j\}$. Then this mode should also have $v(x_{i_1}, \ldots, x_{i_k} \rightarrow S_i) = a_i$ and $v(x_{j_1}, \ldots, x_{j_l} \rightarrow S_j) = a_j$, 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. Clearly, the same reasoning holds if a mode were to correspond to more than two sets. If we suppose the existence of an elementary mode containing k set nodes, with $2 < k \leq m$, we can similarly show that it can be written as a linear combination of the k corresponding elementary modes of the above type.

Thus, the elementary modes corresponding to the sets of $\mathcal{S}$ 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. This immediately implies that given a hitting set of size at most K, the reactions from s to the x's of this hitting set cut all elementary modes and therefore forms a reaction cut of size at most K.

On the other hand, any reaction $(x_{i_1}, \ldots, x_{i_k} \rightarrow S_i)$ or $(S_i \rightarrow t)$ reaction in a reaction cut can be replaced by one reaction $(s \rightarrow x_j)$ (with $x_j \in S_i$), giving another reaction cut. Thus for any reaction cut of size at most K there exists a reaction cut of the same size consisting only of reactions of type $(s \rightarrow x_j)$, hence corresponding to a hitting set of size at most K. □

The above reduction yields a one-to-one correspondence between minimal reaction cuts of size K and hitting sets of size K. Therefore it is approximation preserving (see for a precise definition of an approximation preserving reduction *e.g.* (Ausiello et al., 1999)). Because of its equivalence to Set Cover in which elements have to be covered by sets, no polynomial time algorithm for Hitting set can have approximation ratio $o(\log n)$ unless P=NP (Raz and Safra, 1997), with n the number of elements. The following inapproximability result follows directly.

Theorem 13. *Any polynomial time approximation algorithm for* Min Reaction Cut *cannot have approximation ratio* $o(\log n)$*, with* n *the number of reactions, unless P=NP.* □

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 r^o is obtained. The set F is not necessarily a minimal reaction cut. In the second phase, minimality is obtained by removing reactions from F. The algorithm *Reaction Cut* (RC) is described below.

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 r^o to be cut;
phase 1
$F = \emptyset$;
while F is not a reaction cut of r^o
do begin
let C be the set of reactions defining an elementary mode in S_F that includes r^o
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 r^o **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 devised in (Bar-Noy et al., 2001). Translated into terms of weighted reaction cut, it is based on decomposing the weight function associated to each reaction.

Lemma 14. *Let* S *be a stoichiometric matrix, and let* F^**,* F_1^* *and* F_2^* *be the minimum reaction cuts of* r^o *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.

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

□

The local ratio technique has been applied to a number of combinatorial optimisation problems arising in several areas (scheduling, graph, packing, etc.). Inspired by (Demetrescu and Finocchi, 2003), we prove the following theorem.

Theorem 15. *Given a stoichiometric matrix S and a target reaction r^o, Algorithm* Reaction Cut *runs in polynomial time and returns a reaction cut F of r^o such that $w(F) \leq \lambda w(F^*)$, where F^* is the minimum reaction cut of r^o and λ is the maximum number of reactions in an elementary mode in S including r^o.*

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. Therefore by Theorems 6(i) and 11, the algorithm requires to solve $O(n)$ linear programming problems whose size is linear in n and m; it follows that the running time of the algorithm is polynomial.

The proof of the approximation bound 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_1 be the elementary mode detected in the first iteration of the while loop in Phase 1 and $\delta = \min_{r \in C_1} w(r)$.

We define two new weight functions w_1 and w_2:

$$\begin{aligned} w_1(r) &= \delta \text{ if } r \text{ belongs to } C_1 \text{ and } w_1(r) = 0 \text{ otherwise} \\ w_2(r) &= w(r) - w_1(r). \end{aligned}$$

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 14 apply.

Claim 1. $w_1(F) \leq \lambda w_1(F_1^*)$
Observe that $w_1(F_1^*) = \delta$, because for cutting elementary mode C_1, one reaction 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_1, because all the reactions not belonging to C_1 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 the first iteration of the while loop in Phase 1. Notice that in fact F_1 contains one reaction with weight δ only that cuts C_1. 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 14 yields:

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

□

Note that λ can be linear in the number of reactions; it follows that in the worst case the bound provided by the theorem is linear. This leaves open the problem of finding an algorithm with logarithmic approximation ratio.

We finally observe that the above result can easily be 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 15 can easily be 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. 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 (Klamt and Gilles, 2004).

Minimality refers to reaction cuts from which no reaction can be removed without destroying the cutting property. Klamt and Gilles (2004) 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 11.

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 (Khachiyan et al., 2005).

5 Conclusion

Elementary modes and minimal reaction cuts are common tools in metabolic network analysis. Their computation is not trivial and poses computational challenges. Several algorithms have been proposed for solving these problems 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 LP software. It also implies that many problems in flux balance analysis can be done using LP 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 an improvement for enumerating minimal cut sets. Both results are based on the idea of avoiding to compute the elementary modes for obtaining the reaction cuts.

One may argue that a reaction cut that disables 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 or a reaction cut that leaves some prespecified set of reactions intact. Almost every variation of the minimum reaction cut that emerges in this way is NP-hard.

At present, pathway analysis is still confronted with a problem of scalability to genome-wide models. This paper provides a basis on the complexity of the underlying computational tasks. Such an analysis should help in deciding which tasks can be tackled.

References

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

Ausiello, G., Franciosa, P. G., and Frigioni, D., 2001. Directed hypergraphs: Problems, algorithmic results, and a novel decremental approach. In Restivo, A., Rocca, S. R. D., and Roversi, L., editors, *ICTCS*, volume 2202 of *Lecture Notes in Computer Science*, pages 312–327. Springer.

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

Boros, E., Elbassioni, K. M., Gurvich, V., and Khachiyan, L., 2003. Algorithms for enumerating circuits in matroids. In *ISAAC*, volume 2906 of *Lecture Notes in Computer Science*, pages 485–494. Springer.

Clarke, B. L., 1981. Complete set of steady states for the general stoichiometric dynamical system. J. Chem. Phys., 75:4970–4979.

Colom, J. M. and Silva, M., 1991. Convex geometry and semiflows in P/T nets. a comparative study of algorithms for computation of minimal P-semiflows. In *Proceedings of the 10th International Conference on Applications and Theory of Petri Nets*, pages 79–112, London, UK. Springer-Verlag.

Covert, M. W. and Palsson, B. O., 2003. Constraints-based models: Regulation of gene expression reduces the steady-state solution space. J. Theor. Biol., 221:309–325.

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

Dyer, M. E., 1983. The complexity of vertex enumeration methods. Math. Operations Res., 8:381–402.

Edwards, J. S., Ibarra, R. U., and Palsson, B. O., 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat. Biotechnol., 19(2):125–130.

Fong, S. S. and Palsson, B. O., 2004. Metabolic gene-deletion strains of *escherichia coli* evolve to computationally predicted growth phenotypes. Nat. Genet., 36(10):1056–1058.

Fukuda, K. and Prodon, A., 1996. Double description method revisited. In *Combinatorics and Computer Science*, volume 1120 of *Lecture Notes in Computer Science*, pages 91–111. Springer.

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

Gallo, G., Gentile, C., Pretolani, D., and Rago, G., 1998. Max Horn SAT and the minimum cut problem in directed hypergraphs. Math. Program., 80:213–237.

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

Heiner, M. and Koch, I., 2004. Petri net based system validation in systems biology. In *Proc. ICATPN*, volume 3099 of *LNCS*, pages 216–237, Bologna.

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

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

Klamt, S., Gagneur, J., and von Kamp, A., 2005. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. IEE Proc.-Syst. Biol., 152(4):249–255.

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

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

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

Lacroix, V., Fernandes, C. G., and Sagot, M.-F., 2006. Motif search in graphs: application to metabolic networks. IEEE/ACM Trans. Comput. Biol. Bioinform., 3(4):360–368.

Lawler, E. L., Lenstra, J. K., and Kan, A. H. G. R., 1980. Generating all maximal independent sets: NP-hardness and polynomial-time algorithms. SIAM J. Comput., 9(3):558–565.

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

Palsson, B. O., 2000. The challenges of *in silico* biology. Nat. Biotechnol., 18:1147–1150.

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

Raz, R. and Safra, S., 1997. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing (STOC)*, pages 475–484.

Schilling, C. H., Letscher, D., and Palsson, B. O., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol., 203(3):229–248.

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

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

Schuster, S. and Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst., 2:165–182.

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

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I., and Dandekar, T., 2002b. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *mycoplasma pneumoniae*. Bioinformatics, 18(2):351–361.

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

Terzer, M. and Stelling, J., 2006. Accelerating the computation of elementary modes using pattern trees. In *WABI*, volume 4175 of *Lecture Notes in Computer Science*, pages 333–343. Springer.

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

Valiant, L. G., 1979. The complexity of computing the permanent. Theor. Comp. Sci., 8:189–201.

Voss, K., Heiner, M., and Koch, I., 2003. Steady state analysis of metabolic pathways using Petri nets. In Silico Biol., 3(3):367–387.