Elsevier Editorial System(tm) for BioSystems
Manuscript Draft

Manuscript Number: BIO-D-09-00175R1

Title: A note on the complexity of finding and enumerating elementary modes

Article Type: Full Length Article

Section/Category:

Keywords: Metabolic networks; Computational complexity; Elementary modes; Enumeration

Corresponding Author: Mr Vicente Acuña, M.D. (Ph.D. student)

Corresponding Author's Institution: INRIA and Université de Lyon

First Author: Vicente Acuña, M.D. (Ph.D. student)

Order of Authors: Vicente Acuña, M.D. (Ph.D. student); Alberto Marchetti-Spaccamela, Ph.D.; Marie-France Sagot, Ph.D.; Leen Stougie, Ph.D.

Abstract: In the context of the study into elementary modes of metabolic networks, we prove two complexity results. Enumerating elementary modes containing a specific reaction is hard in an enumeration complexity sense. The decision problem if there exists an elementary mode containing two specific reactions is NP-complete. The complexity of enumerating all elementary modes remains open.

Dear Editor,

We are pleased to resubmit our manuscript "A note on the complexity of finding and enumerating elementary modes". We include in this version the changes made to the manuscript according to the comments of the reviewers. Style of bibliographic references have been modified according to the Biosystems format.

Thanks in advance, with best regards,

Vicente Acuña.

A note on the complexity of finding and enumerating elementary modes

Vicente Acuña[a,b], Alberto Marchetti-Spaccamela[c], Marie-France Sagot[a,b], Leen Stougie[d,e]

[a] *Université de Lyon, F-69000, Lyon ; Université Lyon 1 ; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France.*
[b] *INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France*
[c] *Universitá di Roma "La Sapienza",Via Eudossiana 18, 00184 Rome, Italy*
[d] *Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands*
[e] *Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098SJ Amsterdam, The Netherlands*

Abstract

In the context of the study into elementary modes of metabolic networks, we prove two complexity results. Enumerating elementary modes containing a specific reaction is hard in an enumeration complexity sense. The decision problem if there exists an elementary mode containing two specific reactions is NP-complete. The complexity of enumerating all elementary modes remains open.

Key words:
Metabolic networks, Computational complexity, Elementary modes, Enumeration

1. Introduction

We study some problems related to extreme rays of the cone $\{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}$, for some $m \times n$ matrix A. An extreme ray of a cone is a vector of the cone that cannot be expressed as a convex combination of any two other vectors of the cone. The cone is pointed in the origin 0 of $\mathbb{R}^n$. Therefore, its extreme rays correspond one-to-one to the vertices of the bounded polyhedron $\{x \in \mathbb{R}^n \mid Ax = 0, \underline{1}^T x = 1, x \geq 0\}$, with $\underline{1}$ denoting the all-1 vector in $\mathbb{R}^n$. As a result, enumerating the extreme rays of the cone is not harder than enumerating the vertices of a bounded polyhedron

(polytope). Since the number of objects to be enumerated can be exponential in the size of the input, the complexity in terms of running time is measured as a function of the size of the input and of the output (we give precise definitions of enumeration complexity later).

The complexity of enumerating vertices of polytopes is a famous and long-standing open question (see e.g., Dyer and Proll, 1977). We do not solve this question but present an intriguing related result: given a coordinate i, enumerating all extreme rays r of the cone that have $r_i > 0$ cannot be done *in polynomial total time* (that is, polynomial in the size of the input and of the output) unless P=NP.

Our second complexity result, using essentially the same reduction, is: it is NP-complete to decide if there exists an extreme ray r of the cone that has both $r_i > 0$ and $r_j > 0$ for two given coordinates i and j.

Both results are based on a reduction to the decision problem on the existence of negative simple cycles in directed graphs and are inspired by the work of Khachiyan et al. (2008), who proved that enumerating vertices of any (possibly unbounded) polyhedron cannot be achieved in polynomial total time unless P=NP. Of course, Khachiyan et al.'s result does not apply to *polytopes*, which could still be easier than the general case.

Both questions appeared in computational biology studies of metabolic networks (Acuña et al., 2009; Larhlimi and Bockmayr, 2009; Schuster and Hilgetag, 1994; Terzer and Stelling, 2008; Terzer, 2009; Urbanczik and Wagner, 2005). In this context A is the so-called *stoichiometric matrix*. Each row of this matrix represents a chemical compound and each column an irreversible chemical reaction: a_{ik} is a positive integer if compound i is a product (i.e. output) of reaction k and a negative integer if it is a substrate (i.e. input) of the reaction. It is 0 if it is not involved in the reaction. The equation $Ax = 0$ indicates that the metabolic network is in steady state, in the sense that all (internal) compounds that are produced are also consumed.

The extreme rays of the cone are in this context called *elementary modes*, and, biologically speaking, they are minimal sequences of reactions that would "survive" if the rest of the network were cut. An example is given in Figure 1 for the Citric Acid Cycle.

In this biological context, our results show that: a) it is not possible to generate, in polynomial total time, all elementary modes that pass through a given reaction unless P=NP; and b) deciding if there exists an elementary mode that passes through two given reactions is NP-complete. The first result can have biotechnological relevance. Indeed, by knocking-out enzymes

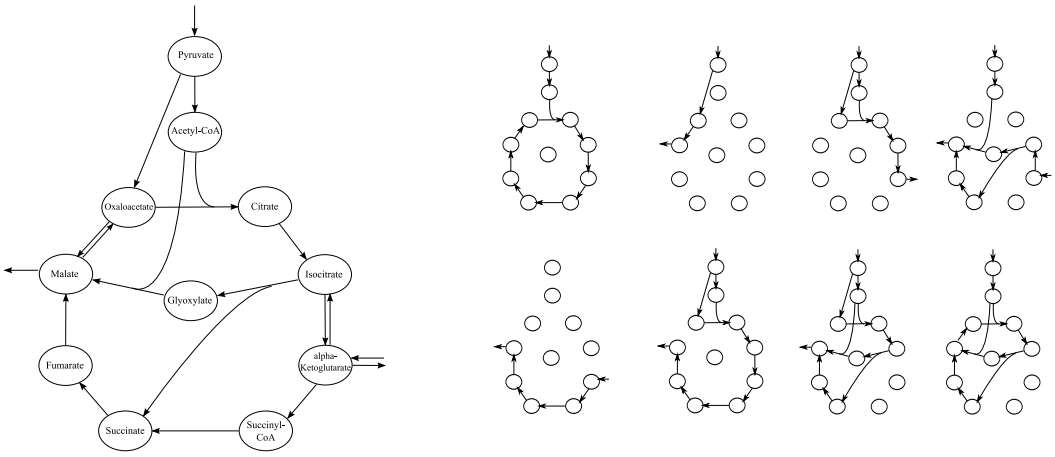

Figure 1: An example of elementary modes analysis. *Left:* A simplified model of the Citric Acid Cycle (including some anaplerotic reactions and the glyoxylate cycle). Some ubiquitous compounds were excluded from the model. The stoichiometric matrix has values -1, 1 and 0. *Right:* The eight elementary modes of this metabolic network (trivial cycles of two reactions are excluded).

and analysing the effect this has on metabolic behaviour, one can identify whether and where a metabolic network is robust or fragile, and ultimately arrive at a better understanding of cellular phenotypes and of their link with the genotype. Enumerating all elementary modes that pass through a given reaction would thus allow to determine all possible steady-state behaviours this reaction enables to block. The decision problem of our second result is more of academic interest. In Acuña et al. (2009), we investigated computational complexity issues related to the analysis of metabolic networks and found that several questions concerning the cone $\{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}$ can be answered by appropriate linear programming formulations, such as finding some extreme ray and finding an extreme ray with one given coordinate positive, whereas other questions are NP-hard, such as finding an extreme ray with a minimum number of positive coordinates and (related to it) finding an extreme ray with a given set of k positive coordinates. Here k is regarded as part of the input of the problem. The complexity of the latter problem for fixed k was posed as an open question. Our second result settles this question by showing that it is NP-complete already if $k = 2$. The main question about the complexity of enumerating all elementary modes (a particular case of enumerating vertices of a polytope) remains open.

Although we are supposing that all reactions are irreversible, both complexity results remain valid if we consider elementary modes (or extreme pathways) in networks where some reactions are *reversible*. Indeed, our formulation is a particular instance (empty set of reversible reactions) and therefore cannot be harder than the general case.

The complexity of the first problem remains the same if we consider enumeration of elementary modes passing through a given *compound* instead of reaction. Indeed, both problems are equivalent: we can reduce one formulation to the other by just breaking the given reaction (respectively compound) in two steps and putting an extra compound (respectively reaction) connecting both. Analogously, deciding if there is an elementary mode that passes through two given compounds (or through a given compound and a given reaction) is also NP-hard.

Biologists have been interested in finding the biological pathways in a metabolic network that produce a specific output, e.g. chemical compounds related to growth, (see e.g. Becker et al., 2007; Nielsen, 1998; Nielsen and Olsson, 2002; Pharkya et al., 2004; Price et al., 2004; Rocha et al., 2008; Teusink and Smid, 2006; van der Werf, 2005) for surveys and two well used methods on the topic, and more in general, work by the Nielsen, Palsson, and Teusink groups plus some others (Senger and Papoutsakis, 2008a,b) on specific applications.

Modelling biological pathways with elementary modes, i.e. extreme rays of the cone, leads to the enumeration problem that we address in our first result. It may seem strange that enumerating the elementary modes passing through a given reaction is hard while the complexity of enumerating *all* elementary modes remains unknown. This apparent contradiction comes from the fact that time is measured in terms of the output size. Given the "normalisation" effect introduced by this, enumerating a smaller subset of objects could therefore be harder than enumerating the whole set. Nevertheless, the hardness of enumerating a specific subset of the elementary modes gives some intuition on the difficulty of enumerating the whole set of them. Our first result is in fact rather surprising. Although nobody has enough confidence to call it a conjecture, most people who have done theoretical research in this field guess that enumerating vertices of polytopes should be achievable in *polynomial total time* (see for a definition Section 2). If, contrary to this guess, enumerating vertices of polytopes will appear to be hard, it will be caused by degeneracy, since enumerating vertices of non-degenerate polytopes can be done in polynomial total time by a Lo-

cal Reverse Search method (Dyer, 1983). Cones corresponding to real-life stoichiometric matrices appear to be highly degenerate, see Terzer (2009). Therefore, for enumerating extreme rays of the cone, variations of the *double description method* of Motzkin et al. (1953) are the most popular ones in the analysis of stoichiometric metabolic networks (Terzer, 2009). Where local reverse search methods suffer from degeneracy, double description methods suffer from generating intermediate (candidate) vectors that do not appear in the output.

In the following section we present our results. We start by defining notions of the complexity of enumeration problems. Essentially we show that enumerating negative cycles of weighted directed graphs, shown unlikely in polynomial total time by Khachiyan et al. (2008), is a special case of our enumeration problem of extreme rays with a given coordinate in the support. To this end, we need to provide essential ingredients from the proof by Khachiyan et al. (2008) that enumerating negative cycles in weighted directed graphs is hard. We conclude with some remarks about related open questions raised by metabolic network analysis in Section 3. Among these, the complexity of enumerating vertices of polytopes remains *the* challenging open question in this field.

2. Complexity results

To define the complexity of enumeration problems in terms of functions of the size of the input is not suitable because the number of solutions can be exponential in the input size. Definitions of enumeration complexity classes have been proposed in Johnson et al. (1988). We use the largest of these classes here.

Definition 2.1. *An enumeration problem can be solved in* polynomial total time *if an algorithm exists with running time bounded by a polynomial function of the combined size of the input and the output. We call this class of problems PT.*

Given a directed graph, or network, $G = (V, E)$, each column of its node-arc incidence matrix M corresponds to an arc $(u, v) \in E$ and contains exactly one -1 in the row of its tail-node u, and exactly one $+1$ in the row of its head-node v, and otherwise 0 entries. We call a cycle simple if it does not contain subcycles.

Lemma 2.2. *Let $G = (V, E)$ be a directed graph with the node-arc incidence matrix M, then the extreme rays of the cone $\{x \in \mathbb{R}^{|E|} \mid Mx = 0, x \geq 0\}$ correspond one-to-one to the directed simple cycles of G.*

Proof. In graph optimisation, a vector $x \geq 0$ that satisfies $Mx = 0$ is called a *flow circulation*. The flow decomposition lemma (see e.g. Schrijver, 1986) states that any such vector x can be written as a positive linear combination of characteristic vectors of directed simple cycles of G. Thus, the set of all characteristic vectors of directed simple cycles of G contains all extreme rays of the cone. It is also obvious that the characteristic vector of any simple cycle cannot be written as a positive linear combination of vectors of other simple cycles. □

In a *weighted directed graph*, a function $w : E \to \mathbb{R}$ assigns weights to arcs. The total weight of a set of arcs is the sum of the weights of each of the arcs. We say that a cycle C is negative if its total weight is negative. We create a matrix M' related to a directed graph $G(V, E)$ by appending an extra row to the node-incidence matrix M of G. In the column corresponding to arc $e \in E$, the entry in the extra row is $w(e)$. The extra row could be seen as corresponding to a dummy node d, and a column as representing a directed hyperarc with a weight on the extra node. In terms of stoichiometry, this would correspond to a reaction transforming 1 molecule of compound u into 1 molecule of compound v and $w(u, v)$ molecules of compound d in case $w(u, v) > 0$, or transforming 1 molecule of compound u and $w(u, v)$ molecules of compound d into 1 molecule of compound v in case $w(u, v) < 0$. We append two extra columns to M': the first one is the unit vector of the dummy node and the second is its negative. In stoichiometric terms, these can be regarded, respectively, as a reaction that produces 1 molecule of compound d from external nutrients and a reaction that excretes 1 molecule of compound d. To facilitate the exposition, we denote the two arcs between the dummy node and "some invisible external node", respectively, e^+ and e^-. We call the resulting matrix M^+. In Figure 2, we present an example of the matrix M^+.

Given a vector x we denote by $S(x)$ its support, i.e., the set of non-zero coordinates of x. As observed by many researchers, the extreme rays of the cone $\{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}$ are exactly the vectors in the cone with minimal support and they are uniquely characterised by their support, up to a positive scalar multiplication.

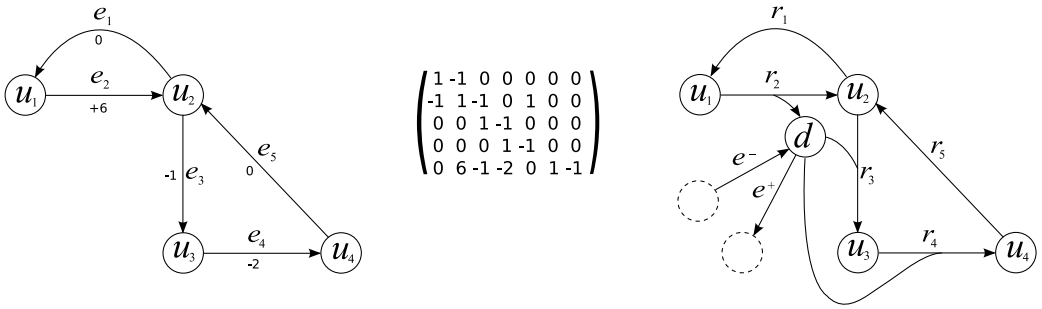

Figure 2: Transformation of a weighted directed graph into a metabolic network with stoichiometry. *Left:* A weighted directed graph G. *Center:* The matrix M^+ of G. *Right:* The set of reactions that represent the stoichiometric matrix M^+ (stoichiometry is not shown).

In the context of the cone related to directed graphs, we index the coordinates of the vectors by the arcs to which they correspond and write the support of a vector as a subset of arcs. The following relations exist between directed simple cycles of G and extreme rays of the cone $\Gamma =: \{x \in \mathbb{R}^{|E|+2} \mid M^+x = 0, x \geq 0\}$.

Lemma 2.3. *Let $G = (V, E, w)$ be a directed weighted graph. Every extreme ray x of the cone Γ has either $S(x) = \{e^+, e^-\}$, or $S(x) \setminus \{e^+, e^-\}$ is the union of simple directed cycles of G.*

Proof. Let x be an extreme ray of the cone Γ such that $S(x) \neq \{e^+, e^-\}$. Let $x' \in \mathbb{R}^{|E|}$ be the truncated vector x without the values corresponding to the arcs e^+ and e^-. Then, $x' \neq 0$ and $Mx' = 0$ where M is the node-arc incidence matrix of G. The vector x' belongs to the cone $\{x \in \mathbb{R}^{|E|} \mid Mx = 0, x \geq 0\}$ and therefore $x' \neq 0$ is a positive linear combination of extreme rays of this cone. By Lemma 2.2, the support of x' is the union of simple directed cycles of G. Therefore, $S(x) \setminus \{e^+, e^-\}$ is the union of simple directed cycles of G. □

Lemma 2.4. *Let $G = (V, E, w)$ be a directed weighted graph. Let C be a directed simple cycle.*

- *if C is negative then $C \cup \{e^-\}$ is the support of an extreme ray of Γ;*
- *if C is positive then $C \cup \{e^+\}$ is the support of an extreme ray of Γ;*
- *otherwise C is the support of an extreme ray of Γ.*

Proof. If C is negative, that is $w(C) = \sum_{e \in C} w(e) < 0$, we consider the vector x defined by

$$x_i = \begin{cases} 1 & \text{if} \quad i \in C \\ -w(C) & \text{if} \quad i = e^- \\ 0 & \text{otherwise} \end{cases}$$

Clearly, x is in the cone $\Gamma =: \{x \in \mathbb{R}^{|E|+2} \mid M^+ x = 0, x \geq 0\}$ and has support $S(x) = C \cup \{e^-\}$. We show that x is an extreme ray of Γ. Suppose it is not. Then it is a positive linear combination of extreme rays of Γ. Thus, there must exist an extreme ray x' such that $S(x') \subseteq S(x)$ and $S(x') \ni e^-$. By Lemma 2.3, $S(x') \setminus \{e^-\}$ is the union of directed simple cycles of G. But $S(x') \setminus \{e^-\} \subseteq S(x) \setminus \{e^-\} = C$, a directed simple cycle C. We conclude that $S(x') = S(x) = C \cup \{e^-\}$. The proof is analogous for the case that C is positive.

Now, if the cycle has weight 0, then we choose in $\Gamma(G)$ the vector x defined as

$$x_i = \begin{cases} 1 & \text{if} \quad i \in C; \\ 0 & \text{otherwise} \end{cases}$$

Arguing in a similar way as in the previous case, we conclude that x is an extreme ray of Γ with $S(x) = C$. □

Lemma 2.5. *Let $G = (V, E, w)$ be a directed weighted graph. Let x be an extreme ray of Γ. Then exactly one of the following possibilities is true:*

- $S(x) = \{e^+, e^-\}$;
- $S(x)$ *is a union of directed simple cycles;*
- $S(x) = C \cup \{e^-\}$ *where C is a negative directed simple cycle of G;*
- $S(x) = C \cup \{e^+\}$ *where C is a positive directed simple cycle of G.*

Proof. If $S(x)$ contains both e^+ and e^- then, because of minimality of support, $S(x) = \{e^+, e^-\}$. If $S(x)$ and $\{e^+, e^-\}$ are disjoint sets, then by Lemma 2.3 the set $S(x)$ is a union of simple cycles.

Suppose that $S(x)$ contains e^- but does not contain e^+. Let us consider $F = S(x) \setminus \{e^-\}$. By Lemma 2.3, $S(x) \setminus \{e^-\}$ is a union of simple cycles. Let C be one of them. We show that C is negative and $C = S(x) \setminus \{e^-\}$.

If not and C has weight 0, then by Lemma 2.4, C is the support of an extreme ray that does not contain e^-. This contradicts the fact that x is an extreme ray.

If not and C is positive, $C \cup \{e^+\}$ is the support of an extreme ray, say x'. Let $\alpha = x'_{e^+} > 0$ be the value of the coordinate e^+ of x'. Let $\beta = x_{e^-} > 0$ be the value of the coordinate e^- of x. Let $z \in \mathbb{R}^{|E+2|}$ be the vector such that $z_i = 1$ if $i \in \{e^-, e^+\}$ and $z_i = 0$ otherwise. The vector $y = (1/\alpha)x' + (1/\beta)x - z$ is positive and $M^+y = 0$. Therefore, $y \in \Gamma$ and $S(y) = C \subseteq S(x) \setminus \{e^-\}$. This contradicts the fact that x is an extreme ray. We conclude that C is a negative simple cycle and therefore $C \cup \{e^-\} = S(x)$.

Analogously, if $S(x)$ contains e^+ but does not contain e^-, then the support of x is $C \cup \{e^+\}$ for some positive simple cycle C. □

As a corollary, we obtain the following crucial observation for our results.

Theorem 2.6. *Let $G = (V, E, w)$ be a directed weighted graph and let $F \subset E$. Then the following two statements are equivalent:*

- *F is a negative simple directed cycle;*
- *$F \cup \{e^-\}$ is the support of an extreme ray of Γ.*

Our first result follows directly from this theorem in combination with a result from Khachiyan et al. (2008).

Theorem 2.7. *Given a cone $\{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}$ and a coordinate i, enumerating all extreme rays that contain i in their support is not in PT unless P=NP.*

Proof. Khachiyan et al. (2008) showed that enumerating negative cycles of a weighted directed graph G is not in PT unless P=NP. By Theorem 2.6, searching for negative cycles in G is equivalent to searching for extreme rays of Γ having e^- in their support. Since, given G, we can construct Γ in polynomial time, the theorem follows. □

The proof of the hardness of enumerating all negative cycles in a directed weighted graph made by Khachiyan et al. (2008) is done by a reduction from the CNF satisfiability problem: a well-know NP-complete problem of deciding if a boolean expression ϕ in conjunctive normal form has a positive assignment. The proof also shows (this is not explicitly mentioned in the paper) that the problem of *deciding if there exists a negative cycle that uses*

a given arc u is NP-hard. In fact, from an instance ϕ of the CNF satisfiability problem, the authors construct a weighted directed graph G such that there is a one-to-one correspondence between a positive assignment of ϕ and a negative cycle that uses a particular arc of G (called (u_{m+n}, u_0) in the proof). We use this result to prove the following theorem.

Theorem 2.8. *Given a cone* $\{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}$ *and two coordinates* i *and* j*, deciding if there exists an extreme ray of the cone that has both* r_i *and* r_j *in its support is NP-complete.*

Proof. Verifying that a vector $x \in \mathbb{R}^n$ is an extreme ray can be done in polynomial time, hence the problem is in NP. On the other hand, we know from Khachiyan et al. (2008) that the problem of *deciding if there exists a negative cycle in a graph that uses a given arc u* is NP-hard. By Theorem 2.6, this is equivalent to deciding if there exists an extreme ray of Γ that contains e^- and u in its support. Since, given G, we can construct Γ in polynomial time, the theorem follows. □

3. Conclusion

Elementary modes analysis has often been used as a way to understand the cellular characteristics of a metabolic network (Stelling et al., 2002). An elementary mode can be seen as a minimal set of reactions that can work in steady state independently of the other reactions in the network. It has therefore served as a mathematical model for the possible metabolic pathways of a cell.

In this paper, we proved some complexity results related to the search and enumeration of elementary modes. Theorem 2.7 implies that is not possible to enumerate in polynomial total time all the elementary modes having a given reaction in their support unless P=NP. In biological terms, based only on the stoichiometry of the network, this means that enumerating all possible pathways that contain a given reaction is a hard task. Theorem 2.8 further shows that, given two reactions, it is hard to decide if there exists an elementary mode whose support contains both reactions.

Both results provide some idea about the complexity of enumerating all elementary modes. Although the question of whether this enumeration can be done in polynomial total time remains unanswered and is indeed a major open issue, the results presented here give some insights of which strategies could be useful to answer to this question. Moreover, enumeration of all

elementary modes corresponds to a special case of the enumeration of vertices of a bounded polyhedron, whose complexity is, in its turn, one of the major open questions in computational geometry.

Beyond that, Schwartz and Kanehisa (2006) have shown that all elementary modes are not equal contributors to physiological cellular states. It remains an open biological question of how to identify elementary modes that are physiologically relevant to which Schwartz and Kanehisa (2006) provide some pointers. Once this is fully answered, it will then become an open question whether such subset of all elementary modes can be more efficiently enumerated.

4. Acknowledgement

This work was funded by the French project ANR MIRI BLAN08-1 335497, the French-UK project ANR-BBSRC MetNet4SysBio ANR-07-BSYS 003 02, the Dutch BSIK-BRICKS project, and the INRIA Associated Team SIMBIOSI.

References

Acuña, V., Chierichetti, F., Lacroix, V., Marchetti-Spaccamela, A., Sagot, M.-F., Stougie, L., 2009. Modes and cuts in metabolic networks: Complexity and algorithms. BioSystems 95, 51–60.

Becker, S., Feist, A., Mo, M., Hannum, G., Palsson, B., Herrgard, M., 2007. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA toolbox. Nat Protoc 2, 727–738.

Dyer, M., 1983. The complexity of vertex enumeration methods. Mathematics of Operations Research 8 (3), 381–402.

Dyer, M., Proll, L., 1977. An algorithm for determining all extreme points of a convex polytope. Mathematical Programming 12, 81–96.

Johnson, D., Yannakakis, M., Papadimitriou, C., 1988. On generating all maximal independent sets. Information Processing Letters 27 (3), 119–123.

Khachiyan, L., Boros, E., Borys, K., Elbassioni, K., Gurvich, V., 2008. Generating all vertices of a polyhedron is hard. Discrete and Computational Geometry 39, 174–190.

Larhlimi, A., Bockmayr, A., 2009. A new constraint-based description of the steady-state flux cone of metabolic networks. Discrete Applied Mathematics 157 (10), 2257–2266.

Motzkin, T., Raiffa, H., Thompson, G., Thrall, R., 1953. The double description method. In: Kuhn, H., Tucker, A. (Eds.), Contributions to the Theory of Games. Vol. II. Princeton University Press, pp. 51–73.

Nielsen, J., 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. Biotechnol Bioeng 58, 125–132.

Nielsen, J., Olsson, L., 2002. An expanded role for microbial physiology in metabolic engineering and functional genomics: Moving towards systems biology. FEMS Yeast Res 2, 175–181.

Pharkya, P., Burgard, A., Maranas, C., 2004. Optstrain: A computational framework for redesign of microbial production systems. Genome Res 14, 2367–2376.

Price, N., Reed, J., Palsson, B., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2, 886–897.

Rocha, I., Frster, J., Nielsen, J., 2008. Design and application of genome-scale reconstructed metabolic models. Methods Mol Biol 416, 409–431.

Schrijver, A., 1986. Theory of Linear and Integer Programming. John Wiley & Sons.

Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems 2 (2), 165–182.

Schwartz, J., Kanehisa, M., 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7, 186.

Senger, R., Papoutsakis, E., 2008a. Genome-scale model for C*lostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. Biotechnol Bioeng 101, 1036–1052.

Senger, R., Papoutsakis, E., 2008b. Genome-scale model for C*lostridium acetobutylicum*: Part II. Development of specific proton flux states and numerically determined sub-systems. Biotechnol Bioeng 101, 1053–1071.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E., 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420, 190–193.

Terzer, M., 2009. Large Scale Methods to Enumerate Extreme Rays and Elementary Modes. PhD-Thesis, ETH Zürich.

Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24 (19), 2229–2235.

Teusink, B., Smid, E., 2006. Modelling strategies for the industrial exploitation of lactic acid bacteria. Nat Rev Microbiol 4, 46–56.

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

van der Werf, M., 2005. Towards replacing closed with open target selection strategies. Trends Biotechnol 23, 11–16.

***Detailed Response to Reviewers**

Dear Editor,

We are pleased to resubmit our manuscript "A note on the complexity of finding and enumerating elementary modes". We include in this version the changes made to the manuscript according to the comments of the reviewers. Style of bibliographic references have been modified according to the Biosystems format.

Thanks in advance, with best regards,

Vicente Acuña.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Answers to reviewer's comments:

Reviewer #1:

> 1) In the introduction and also in the conclusion section the authors
argue several times that enumerating EMs containing a given reaction
cannot be done in polynomial time. Clearly, in section 2 they explain
what is meant with polynimial time (namely with respect to input/output
size) but it would be good to have it mentioned at least once also in
the introduction section.

Answer: We have modified the text according to the suggestion (label #1-1 in .tex file).

> 2) Throughout the manuscript, the authors assume that the flux cone is
given by {Ax=0; x>=0} implying that all reactions are irreversible.
Clearly, one can transform any reaction network with
reversible reactions into a network with only irreversible reactions
(by splitting up reversible reactions) but this should be mentioned
explicitly. Moreover, is it clear at all (or has it been proven elsewhere)
that complexity results obtained for the transformed network also hold
for the algorithms that operate on the original one?

Answer: In fact, that transformation is not needed for the result.
Considering all reaction reversible is a special case of having some
reactions irreversible. We agree that this was not clearly explained so we have modified the text to clarify this point. (label #1-2) in .tex file)

> 3) The analogy between cycles, elementary modes and extreme rays
was also discussed/shown in a recent paper that could be cited:
Klamt S and von Kamp A. 2009. Computing paths and cycles in biological
interaction graphs. BMC Bioinformatics, 10:181.

Answer: We do not think that this reference is relevant to our
article. While Klamt et von Kamp describe methods to compute cycles
having a biological sense, in our paper we use enumeration of negative
cycles as a problem that is related to the enumeration of elementary modes only by its complexity.

> 4)The meaning of lowercase gamma is not mentioned in the manuscript.

Answer: We have modified the text accordingly (lowercase gamma deleted).

5)The meaning of SAT (Satisfiability ...) on page 9 could be explained.

Answer: We have modified the text accordingly (label #1-5 in .tex file).

> 6) On pages 9/10 in the conclusions the authors write:

> "Theorem 2.7 implies that is not possible to enumerate in polynomial
total time all the elementary modes having a given reaction in their
support unless P=NP. In biological terms, based only on the stoichiometry
of the network, this means that enumerating all possible pathways that
contain a given reaction is a hard task. Theorem 2.8 further shows that,
given two reactions, it is hard to decide if there exists an elementary
mode whose support contains both reactions.
Both results provide some idea about the complexity of enumerating all
elementary modes. Although the question of whether this enumeration can
be done in polynomial total time remains unanswered and is indeed a major
open issue, the results presented here give some insights of which strategies
could be useful to answer to this question."

> For some readers not experienced in the field (and I have to confess that
I'm also not an expert in complexity theory) it might be
confusing that enumerating all EMs having a given reaction is a hard
problem whereas for computing all EMs it is not clear yet whether it
is polynmial or not. This might suggest that computing all EMs could
be simpler (faster) than computing only a special subset of them (which
is obviuosly not the case).
It would be helpful if the authors could clarify this point - as far
as I understand, the case of computing the EMs with a given reaction
could be harder than computing all EMs only when normalizing it to the
output size.

Answer: We have modified the text accordingly (label #1-6 first issue in .tex file).

> Related to this, in the introduction section the authors mention the
relatedness between EMs and vertex enumeration. The title of the
reference [6] might suggest that vertex (and thus EM) computation
is hard - but this is not true as the results of [6] refer only to
unbounded polyhedra.

Answer: We have modified the text accordingly (label #1-6) last issue in .tex file).

Reviewer #2:

> MINOR ISSUES
> (1) page 7, proof of Lemma 2.4, ff
Maybe I overlooked it, but I could not find an explanation of the <gamma>(C) notation.

Answer: We have modified the text accordingly (lowercase gamma deleted).

> (2) page 8, second to last section of the proof
current: C U {q+}
probably meant: C U {e+}
second occurrence in last section

Answer: We have modified the text accordingly.

> (3) page 9, Conclusion, line 4
typo: netowrk (instead of network)

Answer: We have modified the text accordingly.