BioSystems 99 (2010) 210–214

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

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

ARTICLE INFO

Article history:
Received 26 August 2009
Received in revised form
16 November 2009
Accepted 26 November 2009

Keywords:
Metabolic networks
Computational complexity
Elementary modes
Enumeration

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.

1. Introduction

We study some problems related to extreme rays of the cone $\{x \in \mathbb{R}^n | Ax = 0, x \geq 0\}$, 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 $\{x \in \mathbb{R}^n | Ax = 0, \underline{1}^T x = 1, x \geq 0\}$, 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 Fig. 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

* Corresponding author at: Université de Lyon, F-69000 Lyon; Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France. Tel.: +33 4 72 43 12 84; fax: +33 4 72 43 13 88.
E-mail address: viacuna@biomserv.univ-lyon1.fr (V. Acuña).

0303-2647/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2009.11.004

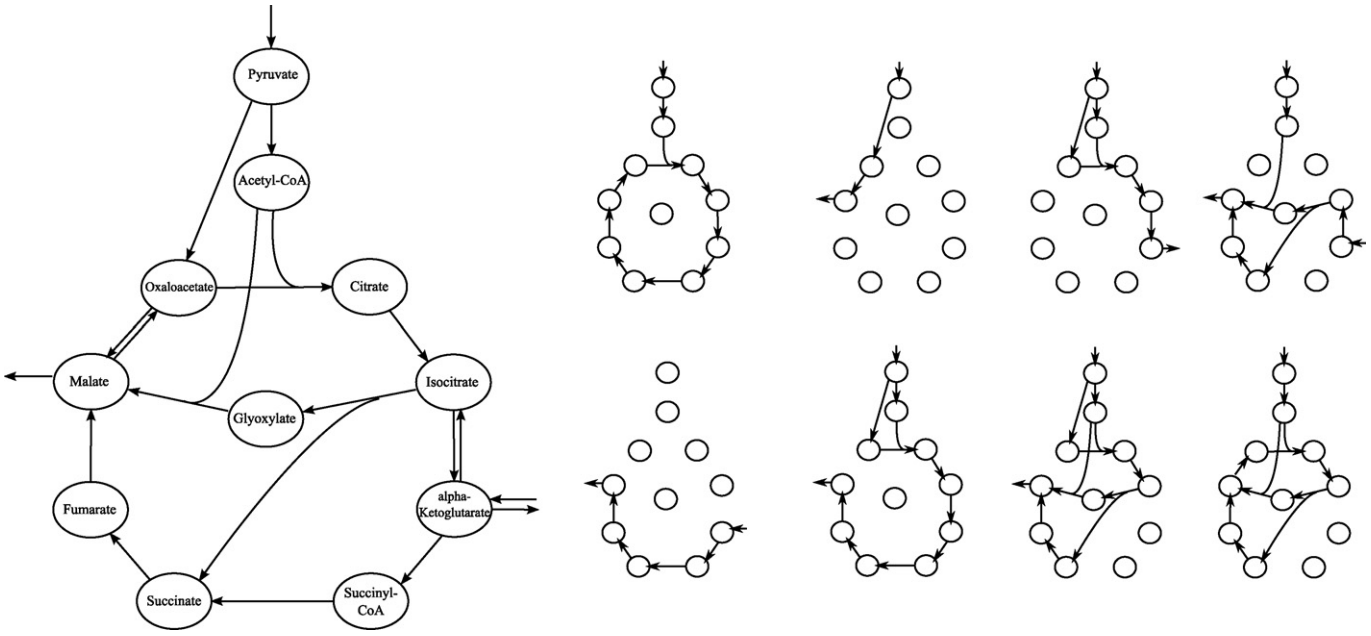

Fig. 1. An example of elementary modes analysis. *Left:* A simplified model of the Citric Acid Cycle (including some anaplerotic reactions and the glyoxylate cycle). Some ubiquitous compounds were excluded from the model. The stoichiometric matrix has values −1, 1 and 0. *Right:* The eight elementary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

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 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 Local Reverse Search method (Dyer, 1983). Cones corresponding to real-life stoichiometric matrices appear to be highly degenerate (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,