BioSystems 112 (2013) 31–36

Contents lists available at SciVerse ScienceDirect

BioSystems

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

# Finding minimal generating set for metabolic network with reversible pathways

Dimitrije Jevremović*, Daniel Boley

*Computer Science & Engineering, University of Minnesota, Minneapolis 55455, MN, USA*

ARTICLE INFO

*Article history:*
Received 10 March 2012
Received in revised form 7 October 2012
Accepted 6 February 2013

*Keywords:*
Metabolic networks
Elementary flux modes
Minimal generating set
Metabolic pathways

ABSTRACT

Elementary flux modes give a mathematical representation of metabolic pathways in metabolic networks satisfying the constraint of non-decomposability. The large cost of their computation shifts attention to computing a minimal generating set which is a conically independent subset of elementary flux modes. When a metabolic network has reversible reactions and also admits a reversible pathway, the minimal generating set is not unique. A theoretical development and computational framework is provided which outline how to compute the minimal generating set in this case. The method is based on combining existing software to compute the minimal generating set for a "pointed cone" together with standard software to compute the Reduced Row Echelon Form.



## 1. Introduction

One approach to analyze cellular metabolic networks of biochemical reactions is by means of the constraint-based stoichiometry network analysis. In this analysis, the quasi-steady state of a network is represented by a matrix of coefficients $\mathbf{S}$ in which the element $s_{ij}$ in the $i$-th row and $j$-th column is the molar amount of metabolite $i$ produced by a unit flux of reaction $j$ (a negative entry denotes consumption). The reactions can be reversible or irreversible; where the latter are constrained to a single direction only.

All the reactions in the network are characterized by their *flux rate* values which correspond to the speed of reaction execution. A set of reaction fluxes is collected into a *metabolic flux vector* $\mathbf{x}$, whose non-zero entries represent active reactions. In order to be feasible, the flux rates of irreversible reactions must be non-negative. The metabolic flux vector quantitatively corresponds to and describes a concept of a *metabolic pathway*. The quasi-steady state of the cellular metabolism imposes the constraint of mass-balance on the internal metabolites: net consumption must match net production. This leads to the equality constraints $\mathbf{S}\cdot\mathbf{x}=0$, i.e., the flux vector must lie in the right nullspace of the stoichiometry matrix.

The solutions of the stoichiometry equation $\mathbf{S}\cdot\mathbf{x}=0$, which also satisfies the non-negativity constraints for the flux of its irreversible reactions, describe all possible metabolic states in which the metabolic network may be found. Geometrically this solution space corresponds to the polyhedral cone (Schrijver, 1988; Wagner and Urbanczik, 2005), and it may be fully generated by means of its extreme rays (Fukuda and Prodon, 1996). Extreme rays are conically independent set of vectors and in the convex analysis are also known as a *minimal generating set*. In the stoichiometry network analysis, alongside with the concept of the minimal generating set, stand the extreme pathways and elementary flux modes. It is important to say that the *minimal generating set*, *extreme pathways* and *elementary flux modes* are computed using the standard Double Description Method for the enumeration of extreme rays (i) when no reversible reactions are split, (ii) only internal reversible reactions are split, and (iii) all of the reversible reactions are split, into two irreversible components, respectively (Schuster and Hilgetag, 1994; Wagner and Urbanczik, 2005; Schilling et al., 2000; Llaneras and Pico, 2010; Jevremović et al., 2010). Unlike the minimal generating set, in the original reaction space the extreme pathways and elementary flux modes are not necessarily conically independent which depends on the existence and number of the reversible reactions.

In the absence of reversible reactions, the *minimal generating set*, *extreme pathways* and *elementary flux modes* coincide, are uniquely defined, and correspond to the *extreme rays* of the polyhedral cone. Regarding the directionality of the metabolic pathways which the metabolic network accepts we distinguish two cases.

In the first case, if the metabolic network admits only irreversible pathways (i.e., every pathway contains at least one

* Corresponding author.
*E-mail address:* jevrem@cs.umn.edu (D. Jevremović).

0303-2647/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2013.02.003

irreversible reaction), then the minimal generating set is unique, and the corresponding polyhedral cone is said to be *pointed*. On the other side, in the second case, if the metabolic network admits reversible pathways, the minimal generating set is no longer unique and the polyhedral cone is *not pointed*.

Regardless if the cone is pointed or not, the set of the elementary flux modes (or extreme pathways) is a superset of any minimal generating set, and some of the elementary flux modes (or extreme pathways) may lie in the interior of the cone. In addition, putatively exponential hardness and high computational cost of the algorithm used to compute elementary flux modes (Acuña et al., 2009, 2010) is another reason to shift the attention from extreme pathways and elementary flux modes to the minimal generating set.

Answering many questions requires the use of extreme pathways (Schilling and Palsson, 2000; Papin et al., 2002; Wiback et al., 2002; Wiback and Palsson, 2002; Price et al., 2002) and elementary flux modes (Stelling et al., 2002; Trinh et al., 2008, 2009; Trinh and Srienc, 2009; Pérès et al., 2011; Flynn et al., 2012), however there are several applications one can answer with minimal generating sets. This situation especially arises in the case of genome-scale metabolic networks where the computation of elementary modes is prohibitively expensive (Larhlimi, 2008). Some simple structural properties may be observed, such as whether any reversible reaction appears only in one direction or only in irreversible pathways, or whether some reaction appears in no pathway at all. Flux coupling analysis, a procedure of determining dependencies between network reactions, can be accomplished using the minimal generating set vectors (Larhlimi and Bockmayr, 2006). Control-effective analysis of individual reactions in the network was initially proposed on the basis of computed elementary flux modes (Stelling et al., 2002). However, minimal generating sets can be used to obtain an analogous control-effective metric, used in the regulatory network analysis and reaction importance assessment (Larhlimi, 2008). Minimal metabolic behaviors are exposed by the minimal generating set (Larhlimi and Bockmayr, 2009), but a simple method to compute it is still needed. In large genome-scale networks, where computation of entire minimal generating sets may be impractical, efforts have been made to compute the K-shortest minimal generating vectors (Rezola et al., 2011) (i.e., pathways involving as few reactions as possible). This was accomplished by means of solving several linear optimization problems and using existing methods for the computation of K-shortest elementary flux modes (de Figueiredo et al., 2009).

The problem of computing the minimal generating set for the metabolic network which admits reversible pathways is the topic of this manuscript. An earlier analysis of the metabolic networks with reversible pathways by means of two subnetworks, one with no reversible pathways and one with all reversible pathways, can be found in (Larhlimi and Bockmayr, 2009). The computation of the unique minimal generating set for a pointed cone can be accomplished using existing algorithms (von Kamp and Schuster, 2006; Terzer and Stelling, 2008; Jevremović et al., 2011) or using the general paradigm in (Jevremović et al., 2010). But this is considerably more difficult when the cone is not pointed (i.e., there are reversible pathways). This situation can be recognized by computing the rank of the submatrix of **S** consisting of the reversible reactions (Jevremović et al., 2010).

The major contribution of this paper is to provide a simple procedure to compute the minimal generating set for a stoichiometric network which has reversible pathways. The method is based on combining two existing algorithms: a method to compute the minimal generating set for a pointed cone, and a method to compute a nullspace of a matrix based on the Reduced Row Echelon Form, a classical method in linear algebra. All this is carried out without the necessity to compute all the elementary flux modes for any network. This paper is organized as follows. Section 2 gives a theoretical treatment of the representation of reversible and irreversible pathways and the decomposition of the original metabolic network into two subnetworks. Section 3 outlines the algorithm for the computation of the minimal generating set using two subnetworks. Section 4 uses a simple example to illustrate the method and show how the method exposes some of the structure of the network.

## 2. Theory

Let $\mathbf{S}=(\mathbf{A},\mathbf{B},\mathbf{C})$ be an $m\times n$ stoichiometry matrix with the $n$ columns (reactions) ordered so that **A** consists of the irreversible reactions (of which there are $n_i$) and **B**, **C** consists of the reversible reactions (of which there are $n_r$). We assume the reversible reactions (**B**, **C**) form a matrix of rank $k_r$ and that **B** consists of $k_r$ columns which are independent, while **C** consists of $n_r-k_r$ columns. This implies that all the columns of **C** can be written as linear combinations of the columns **B**: $\mathbf{C}=\mathbf{B}R$ for some $k_r\times(n_r-k_r)$ coefficient matrix $R$. We remark that the columns of **B** can be found by a variety of methods such as the Reduced Row Echelon Form (RREF) (Lay, 2012) where they appear as the "pivot" columns, while the columns **C** appear as the "non-pivot" columns. Hence we will refer to **B** as the "pivot" columns. The standard RREF algorithm scans the matrix **S** left-to-right extracting independent columns **B**, hence the choice of pivot columns varies depending on the order of columns (reactions) in the original **S**, but once the latter is fixed, the former is also.

The matrix (**B**, **C**) has a nullspace of dimension $n_r-k_r$, and a suitable basis for this space is $N_R=\begin{pmatrix}-R\\ I\end{pmatrix}$. Any vector in this nullspace is a valid path for the subnetwork (**B**, **C**) and is a reversible path. By prepending zeros, we obtain

$$\hat{N}_R=\begin{pmatrix}0\\ \mathrm{NR}\end{pmatrix}=\begin{pmatrix}0\\ -R\\ I\end{pmatrix},$$

which we will show is a minimal basis for the set of all reversible paths in the original network.

A column vector **x** is a valid path of the network represented by stoichiometry matrix **S** if and only if $\mathbf{Sx}=\mathbf{0}$ and the entries of **x** corresponding to irreversible reactions are non-negative. If we split $\mathbf{x}=(\mathbf{x}_a;\mathbf{x}_b;\mathbf{x}_c)$ to conform with (**A**, **B**, **C**) (where ";" denote vertical concatenation *à la* Matlab), then **x** is a valid path if and only if $\mathbf{Ax}_a+\mathbf{Bx}_b+\mathbf{Cx}_c=0$ and $\mathbf{x}_a\geq\mathbf{0}$ (elementwise).

We have the following Lemmas:

**Lemma 1.** *Any reversible pathway* **x** *for the stoichiometry matrix* $\mathbf{S}=(\mathbf{A},\mathbf{B},\mathbf{C})$ *split as above can be written in terms of the minimal generating set for the reversible subnetwork* (**B**, **C**)*, as follows:*

$$\mathbf{x}\equiv\begin{pmatrix}\mathbf{x}_a\\ \mathbf{x}_b\\ \mathbf{x}_c\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\ \mathbf{x}_b\\ \mathbf{x}_c\end{pmatrix}=\hat{N}_R\boldsymbol{\alpha}\equiv\begin{pmatrix}\mathbf{0}\\ -R\\ I\end{pmatrix}\boldsymbol{\alpha}$$

*for some coefficient vector* $\boldsymbol{\alpha}$.

Since there are no sign constraints in the subnetwork represented by (**B**, **C**), the basis $N_R$ is the minimal generating set for all possible reversible paths. In fact any basis for the nullspace of (**B**, **C**) would be a minimal generating set, but we choose this specific one because each column in this basis has a minimal set of non-zeros, i.e., each is also an elementary flux mode. In this sense, we call this a "minimal basis" or "minimal generating set." There is still freedom