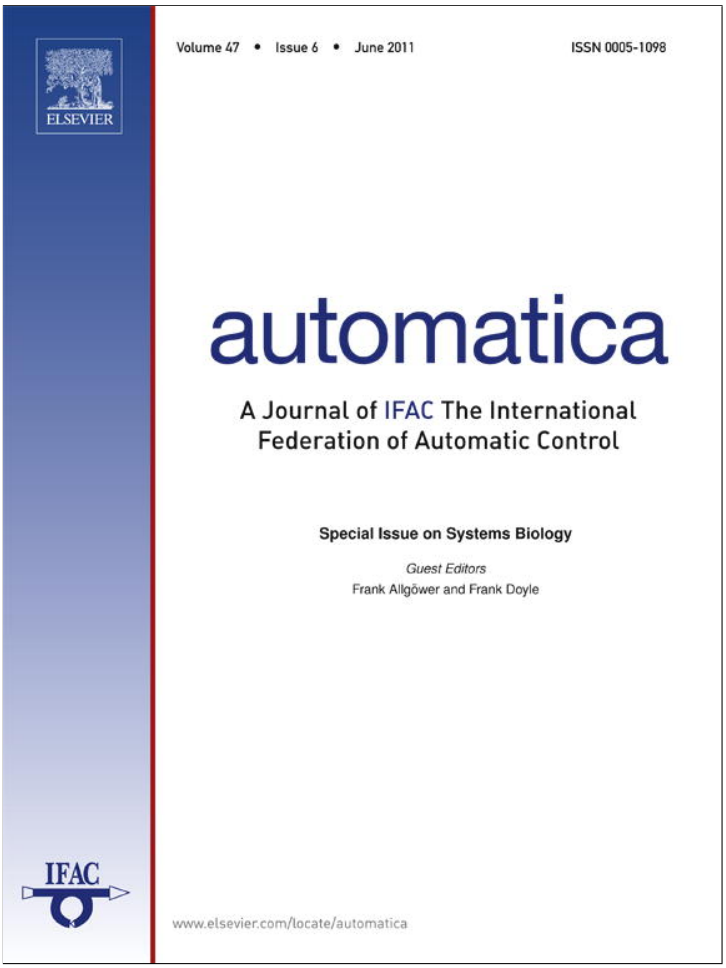

Automatica 47 (2011) 1255–1259

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Fast computation of minimal elementary decompositions of metabolic flux vectors☆

Raphael M. Jungers [a,*], Francisca Zamorano [c], Vincent D. Blondel [b], Alain Vande Wouwer [c], Georges Bastin [b]

[a] Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 32D-668. Cambridge, MA 02139, USA

[b] Center for Systems Engineering and Applied Mechanics (CESAME), Department of Mathematical Engineering, Batiment Euler, Avenue Georges Lemaitre, 4, 1348 Louvain-La-Neuve, Belgium

[c] Service d'Automatique, Université de Mons, 31 Boulevard Dolez, 7000 Mons, Belgium

ARTICLE INFO

Article history:
Received 28 January 2010
Received in revised form
23 July 2010
Accepted 14 October 2010
Available online 8 February 2011

Keywords:
Metabolic engineering
Modeling and identification
Elementary flux modes
Animal cells

ABSTRACT

The concept of elementary flux vector is valuable in a number of applications of metabolic engineering. For instance, in metabolic flux analysis, each admissible flux vector can be expressed as a non-negative linear combination of a small number of elementary flux vectors. However a critical issue concerns the total number of elementary flux vectors which may be huge because it combinatorially increases with the size of the metabolic network. In this paper we present a fast algorithm that randomly computes a decomposition of admissible flux vectors in a minimal number of elementary flux vectors without explicitly enumerating all of them.

1. Introduction

The intracellular metabolism of living cells is usually represented by a metabolic network under the form of a directed hypergraph that encodes a set of biochemical reactions taking place within the cell. In this hypergraph, the nodes represent the metabolites and the edges represent the metabolic fluxes.

According to the quasi steady-state paradigm of metabolic flux analysis (MFA) (e.g. Stephanopoulos, Nielsen, & Aristidou, 1998), it is assumed that the fluxes are balanced at each internal node. This means that the net sum of production and consumption fluxes, weighted by their stoichiometric coefficients, is zero for each internal metabolite of the network. This is expressed by the algebraic relation:

$$\mathbf{N}\mathbf{v} = \mathbf{0} \quad \mathbf{v} \geqslant \mathbf{0} \tag{1}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_m)^T$ is the m-dimensional column vector of fluxes and $\mathbf{N} = [n_{ij}]$ is the $n \times m$ stoichiometric matrix of the metabolic network (m is the number of fluxes and n the number of internal nodes of the network). More precisely, a flux v_j denotes the rate of reaction j and a non-zero n_{ij} is the stoichiometric coefficient of the metabolite i in reaction j.

For a given metabolic network, the set S of possible flux distributions is the set of vectors $\mathbf{v}$ that satisfy the linear system (1). This set S is the pointed polyhedral cone resulting from the intersection of the kernel of $\mathbf{N}$ with the non-negative orthant. By the well-known *Caratheodory's theorem,* this implies that there exists a set of elementary flux vectors $\mathbf{e}_i$ (Urbanczik, 2007) which are the extreme rays of the polyhedral cone and such that any flux distribution $\mathbf{v}$ can be expressed as a non-negative linear combination of the vectors $\mathbf{e}_i$:

$$\mathbf{v} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + \cdots + w_q\mathbf{e}_q \quad w_i \geqslant 0. \tag{2}$$

The $m \times q$ non-negative matrix $\mathbf{E}$ with column vectors $\mathbf{e}_i$ obviously satisfies $\mathbf{NE} = 0$ and (2) can be written in matrix form as

$$\mathbf{v} = \mathbf{E}\mathbf{w} \quad \text{with } \mathbf{w} \triangleq (w_1, w_2, \ldots, w_q)^T. \tag{3}$$

Thus, the elementary flux vectors are a way of representing the set of possible flux distributions. A well-known issue related to this representation is that typically, the number of such vectors grows exponentially with the size of the problem. In our case-study Jungers, Zamorano, Blondel, Wouwer, and Bastin (2009), we apply the methodology presented in this paper to a metabolic network of CHO cells which involves 82 reactions and 53 metabolites.

☆ This research is supported by the Scientific Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Belgian Programme on Interuniversity Attraction Poles (IAP VI/4). The material in this paper was partially presented at Mathmod 2009, February 11–13, 2009, Vienna, Austria. This paper was recommended for publication in revised form by Associate Editor Michael A. Henson under the direction of Editor Francis J. Doyle III.

* Corresponding author. Tel.: +1 32 474 36 13 27; fax: +1 32 10 47 2180.
E-mail addresses: raphael.jungers@uclouvain.be (R.M. Jungers), francisca.zamorano@umons.ac.be (F. Zamorano), vincent.blondel@uclouvain.be (V.D. Blondel), Alain.VandeWouwer@umons.ac.be (A.V. Wouwer), Georges.Bastin@uclouvain.be (G. Bastin).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.01.011

Though the dimension of this network is rather limited, METATOOL calculates 65 329 elementary flux modes. Manipulating such a number of elementary flux vectors is not an easy task. For certain specific applications (e.g., recombinant protein production, media optimization, etc.) more detailed metabolic networks involving several hundreds of reactions can be of major interest (for instance Famili, Forster, Nielsen, & Palsson, 2003, deals with a network having 1175 fluxes and 733 internal metabolites). However, it is generally acknowledged that the enumeration of the elementary flux modes can already be out of reach for systems with more than $\approx$100 reactions and metabolites (see e.g. Klamt, Gagneur, & von Kamp, 2005; Terzer & Stelling, 2006).

In this paper, we address the issue of decomposing a flux distribution $\mathbf{v}$ in the convex basis of elementary flux vectors $\mathbf{e}_i$. In general, this decomposition does not necessitate the whole enumeration of the convex basis but requires only the knowledge of a few elementary vectors. The information needed for computing these elementary vectors can be obtained directly from the stoichiometric matrix $\mathbf{N}$ which is much smaller than the matrix $\mathbf{E}$.

Motivated by practical applications including the analysis of metabolic pathways (Klamt & Stelling, 2003; Papin et al., 2004; Price, Papin, & Palsson, 2002), the derivation of dynamic macroscopic models (Provost, 2006; Provost & Bastin, 2004) and the characterization of metabolic phenotypes (Famili et al., 2003), our objective in this paper will precisely be to determine *minimal* such decompositions.

Well-known theoretical results exist, that give upper bounds on the minimal number of vectors of the convex basis that are necessary to decompose the flux distribution. Namely, this number is $m - n$, which is typically very small in comparison with the number of elementary flux modes. As an example, in the case study presented in Jungers et al. (2009), the flux distribution can be expressed with 22 elementary flux modes, while the convex basis counts 65 329 such vectors. We show how to compute this decomposition without actually computing the whole convex basis, thanks to convex programming techniques.

When the vector $\mathbf{v}$ is the solution of an underdetermined metabolic flux analysis problem, the situation is more complex and it may be possible to find a decomposition with even less elementary flux modes. The problem is more complex because it is not known a priori which vector, among all admissible flux distributions, is the one that can be decomposed in the minimal number of elementary flux modes. For this more difficult problem, we also provide a method that allows one to compute this distribution in polynomial time, without computing the whole convex basis.

2. Metabolic flux analysis and decomposition in elementary flux modes

2.1. Metabolic Flux Analysis

Metabolic flux analysis (MFA) is the exercise of calculating the admissible flux distributions $\mathbf{v}$ that satisfy the steady state balance equation $\mathbf{Nv} = \mathbf{0}$ together with an additional set of linear constraints added by using experimental measurements. Here we consider the case where the measurements are collected in a vector $\mathbf{v}_m$ which is a linear function of the unknown flux distribution $\mathbf{v}$ and is expressed as

$$\mathbf{Pv} = \mathbf{v}_m \tag{4}$$

where $\mathbf{P}$ is a given $p \times m$ full-rank matrix. Then, from Eqs. (1)–(4), we have the following fundamental equation of metabolic flux analysis

$$\begin{pmatrix} \mathbf{N} \\ \mathbf{P} \end{pmatrix} \mathbf{v} = \begin{pmatrix} \mathbf{0} \\ \mathbf{v}_m \end{pmatrix} \quad \mathbf{v} \geqslant \mathbf{0}. \tag{5}$$

For a given metabolic network and a given set of measurements, the solution of the MFA problem is defined as the set $\mathcal{F}$ of admissible flux distributions, i.e. the set of non-negative vectors $\mathbf{v}$ that satisfy the linear system (5). Hence, as emphasized in Provost (2006, Chapter 4), Provost and Bastin (2004) and Provost and Bastin (2006), the set $\mathcal{F}$ is a polytope in the positive orthant $\mathbb{R}^m_+$. This means that any admissible flux distribution $\mathbf{v}$ can be expressed as a non-negative linear combination of a set of non-negative basis vectors $\mathbf{f}_i$ which are the vertices of this polytope and form therefore a *unique* convex basis of the flux space $\mathcal{F}$. In other words, the solution of the MFA problem is the *admissible flux space* $\mathcal{F}$ defined as

$$\mathcal{F} \triangleq \left\{ \mathbf{v} : \mathbf{v} = \sum_i \alpha_i \mathbf{f}_i,\ \alpha_i \geqslant 0 \right\}. \tag{6}$$

2.2. Minimal decomposition in elementary vectors

For any admissible flux vector $\mathbf{v}$ in the polytope $\mathcal{F}$ satisfying equation (5), it must be emphasized that the decomposition of $\mathbf{v}$ in the convex basis $\{\mathbf{e}_i\}$ is *not* unique. As mentioned above, our aim is to determine such a decomposition in a minimal number of $\{\mathbf{e}_i\}$.

Using (3), system (5) is equivalent to the system:

$$\begin{pmatrix} \mathbf{NE} \\ \mathbf{PE} \end{pmatrix} \mathbf{w} = \begin{pmatrix} \mathbf{0} \\ \mathbf{v}_m \end{pmatrix} \quad \mathbf{w} \geqslant 0. \tag{7}$$

We observe that the first equation $\mathbf{NEw} = \mathbf{0}$ is trivially satisfied independently of $\mathbf{w}$ since by definition $\mathbf{NE} = 0$. Hence, system (7) may be reduced to the second equation:

$$\mathbf{PEw} = \mathbf{v}_m \quad \mathbf{w} \geqslant \mathbf{0}. \tag{8}$$

In this form, it is clear that the set of admissible weighting vectors $\mathbf{w}$ that satisfy (8) again constitutes a convex polytope that we denote $\mathcal{H}$. Therefore there exists a set of appropriate edge vectors $\mathbf{h}_i$ such that any arbitrary convex combination of the form:

$$\mathbf{w} = \sum_i \beta_i \mathbf{h}_i \qquad \beta_i \geqslant 0 \qquad \sum_i \beta_i = 1 \tag{9}$$

is necessarily an admissible $\mathbf{w}$ satisfying (8). The convex basis vectors $\mathbf{h}_i$ have a critical property: the number of non-zero entries in these vectors is equal to the number of measurements p.

Theorem 1 (*Fukuda & Prodon, 1996; Provost, 2006*)**.** *Any admissible flux distribution $\mathbf{v} \in \mathcal{F}$ can be expressed as a convex combination of $m - n$ elementary flux vectors $\mathbf{e}_i$.*

Moreover, if $p < m - n$, then, there is at least one vector $\mathbf{v}^* \in \mathcal{F}$, such that $\mathbf{v}^*$ can be expressed as a convex combination of only p elementary flux vectors.

From a metabolic viewpoint, each vector $\mathbf{h}_i$ is a solution $\mathbf{w}$ of (8) corresponding to a particular admissible flux distribution $\mathbf{v}$:

$$\mathbf{v} = \mathbf{Eh}_i \quad \mathbf{v} \in \mathcal{F}. \tag{10}$$

An important issue concerns the number of distinct extreme rays or vertices that are generated when computing the cone S or the polytopes $\mathcal{F}$ and $\mathcal{H}$. This number may become very large because it combinatorially increases with the size of the underlying metabolic network. More precisely, if n is the number of metabolites and m is the number of reactions, there is no polynomial function $p(m, n)$ such that the number of elementary flux modes q satisfies $q \leq p(m, n)$ for all metabolic networks (Grötschel, Lovász, & Schrijver, 1988). It is also the case for the number of vectors h_i that are vertices of the polytope $\mathcal{H}$.

In practical applications of MFA, the enumeration of all extreme rays is not necessarily a critical objective. In many applications it is sufficient to know only one minimal decomposition of some vectors $\mathbf{v} \in \mathcal{F}$ in terms of elementary vectors $\mathbf{e}_i$. According to Theorem 1, there necessarily exists an admissible $\mathbf{v}$ having a decomposition that involves only p terms. Computing this decomposition may be very expensive at first sight since the dimension of $\mathbf{E}$ is not bounded by a polynomial in the sizes of $\mathbf{N}$ and $\mathbf{P}$.

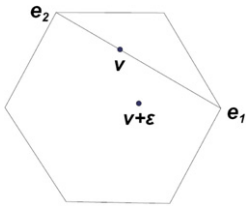

Fig. 1. A situation where $\mathbf{v}$ can be decomposed in fewer vertices than the number provided by Theorem 1. However this situation is clear a singularity which is not robust with respect to small perturbations.

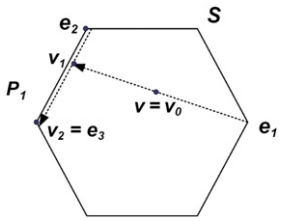

Fig. 2. Decomposition of $\mathbf{v}$ in the convex basis $\{e_i\}$.

3. The algorithm

In this section we present efficient methods to decompose a flux distribution in a minimal number of elementary flux modes. We first comment on the term "minimal". We are interested here in finding a decomposition of the point $\mathbf{v}$ in $m - n$ elementary flux modes, or, if $p < m - n$, in p elementary flux modes. That is, we are interested in matching the theoretical bounds of Theorem 1. This does not preclude the possibility that $\mathbf{v}$ could be expressed with even less elementary flux modes. This can be seen in Fig. 1, where the point $\mathbf{v}$ is "by chance" exactly on a line passing through two vertices of the polytope. In this situation, while the theoretical bound tells us that $\mathbf{v}$ can be decomposed as a convex combination of 3 vertices, it actually appears that such a decomposition exists with 2 vertices. However, we are not interested in finding such decompositions for mainly two reasons:

First, we are looking for a robust decomposition, while, as can be seen in Fig. 1, a decomposition in less than 3 vertices is not robust: a slight perturbation $\boldsymbol{\epsilon}$ of $\mathbf{v}$ can move it out of the line $\mathbf{e}_1\mathbf{e}_2$, so that the previous decomposition is not possible anymore. Since $\mathbf{v}_m$ is a vector of measurements (typically retrieved from a regression), it does not make sense to look for a decomposition which is sensitive to very small perturbations.

Second, this kind of minimal decomposition in even fewer vertices than theoretically foreseen is NP-hard to find, so that one cannot hope to develop a polynomial time algorithm to obtain such a decomposition (Donoho & Tanner, 2005).

3.1. Decomposing a vector in a convex basis

Let us first consider the following simpler problem, which will be solved as a subroutine of our method: We are given a vector $\mathbf{v}$ that belongs to a cone S, and we would like to express this vector as a linear combination of a few extreme rays of S. The cone S is given via its defining matrix $\mathbf{M}$ (in our case, this will be the stoichiometric matrix N).

The geometric intuition behind the algorithm can be understood in Fig. 2. Similar algorithms can be found in the literature Grötschel et al. (1988, Thm. 6.5.11). All the optimization problems we have to solve in order to do that can be cast as linear programs and can then be solved in polynomial time with classical linear programming methods (e.g. Boyd & Vandenberghe, 2004).

We now present the algorithm with its technical details. Let us denote $a = \mathbf{1}^T\mathbf{v}$ the sum of the entries in $\mathbf{v}$ ($\mathbf{1}$ denotes the vector whose all entries are equal to one). In the following we will consider without loss of generality the slightly different problem where we are looking for extreme rays $\mathbf{e}_i$ such that $\mathbf{1}^T\mathbf{e}_i = a$. Geometrically speaking, we cut the cone with a plane passing through $\mathbf{v}$ such that the intersection is a bounded polytope whose vertices correspond to extreme rays of the initial cone S. We are thus given a (bounded) polytope $\mathcal{P}$, and a vector $\mathbf{v}$ in this polytope and we want to express this vector $\mathbf{v}$ as a convex combination of vertices of the polytope.

We first pick up (at random) a vertex of the polytope $\mathcal{P}$. The problem of finding a vertex of the polytope defined by the equations

$$\mathbf{M}\mathbf{x} = \mathbf{0}, \qquad \mathbf{1}^T\mathbf{x} = a, \qquad \mathbf{x} \geqslant \mathbf{0}$$

can be solved in time polynomial in the number of constraints and the dimension. Indeed, consider the following linear program:

$$\begin{aligned} &\min \mathbf{d}^T\mathbf{x} \\ &\text{s.t.} \\ &\mathbf{M}\mathbf{x} = \mathbf{0}, \\ &\mathbf{x} \geqslant \mathbf{0}, \\ &\mathbf{1}^T\mathbf{x} = a. \end{aligned} \tag{11}$$

If $\mathbf{d}$ is not parallel with a constraint of the program (11), then, the solution is a vertex of the corresponding polytope. So in practice, if $\mathbf{d}$ is a random direction, the probability that the solution is not an extreme ray is zero.

We will proceed iteratively, by projecting $\mathbf{v}$ on faces $\mathcal{P}_i$ of the polytope $\mathcal{P}$ described by the constraints of the program (11). Since the dimension of the face $\mathcal{P}_i$ strictly decreases at each step, the algorithm will provably take at most $k - 1$ steps, where k is the dimension of the polytope $\mathcal{P}$.

Take any vertex $\mathbf{e}_1$ of $\mathcal{P}$ (for instance by solving the linear program (11)); then the vector $\mathbf{v}$ can be written as the convex combination of $\mathbf{e}_1$ and of a vector $\mathbf{v}_1$, which belongs to a face $\mathcal{P}_1$ of $\mathcal{P}$: $\mathbf{v} = \gamma_1\mathbf{e}_1 + (1 - \gamma_1)\mathbf{v}_1$ (see Fig. 2). These quantities $\mathbf{v}_i$, γ_i are easy to compute, as $\mathbf{v}_1$ is the solution $\mathbf{x}^*$ of the linear program

$$\begin{aligned} &\max \mu \\ &\text{s.t.} \\ &\mathbf{M}\mathbf{x} = \mathbf{0}, \\ &\mathbf{x} \geqslant \mathbf{0}, \\ &\mathbf{1}^T\mathbf{x} = a, \\ &\mathbf{v} + \mu(\mathbf{v} - \mathbf{e}_1) = \mathbf{x}. \end{aligned} \tag{12}$$

The geometric meaning of this linear program is as follows: starting from the vector $\mathbf{v}$ one tries to find a point $\mathbf{x}$ which is diametrically opposite to $\mathbf{e}_1$ and as far as possible from $\mathbf{v}$, and μ represents the distance from $\mathbf{v}$ to $\mathbf{x}$. Clearly this point will be on a face $\mathcal{P}_1$ of the polytope.

Now, at each step $i = 1, \ldots, k'$, $\mathcal{P}_i$ is a new polytope, and, since $\mathbf{v}_i \in \mathcal{P}_i$, we still can express $\mathbf{v}_i$ as a convex combination of a vertex of $\mathcal{P}_i$ (which is also a vertex of $\mathcal{P}$) and a point $\mathbf{v}_{i+1}$ that belongs to a face $\mathcal{P}_{i+1}$ of $\mathcal{P}_i$ (which is also a face of $\mathcal{P}$, but of dimension strictly smaller than dim $\mathcal{P}_i$). Thus, after $k' \leq k$ steps, the dimension of $\mathcal{P}_{k'}$ is equal to 0, which means that $\mathbf{v}_{k'}$ is actually a vertex of $\mathcal{P}$ which we denote $\mathbf{e}_{k'+1}$. Thus, $\mathbf{v}_{k'-1} = \gamma_{k'}\mathbf{e}_{k'} + (1 - \gamma_{k'})\mathbf{e}_{k'+1}$. Finally, by successively decomposing $\mathbf{v}_{k'}$ we can write:

$$\mathbf{v} = \sum_1^{k'+1} w_i\mathbf{e}_i, \qquad \sum w_i = 1.$$

Thus, since the dimension of the cone S is equal to $k + 1 = m - n$, we obtain at most $m - n$ extreme vectors $\mathbf{e}_i$. We have thus found the decomposition in polynomial time, which is a dramatic

improvement compared to the naive brute force approach that requires the enumeration of all vectors $\mathbf{e}_i$.

3.2. Finding the minimal decomposition if $p < m - n$

We now suppose that $p < m - n$, which implies that there are several vectors $\mathbf{v}$ that satisfy Eq. (5) (the system is underdetermined). By Theorem 1, we know that at least one of them can be expressed as a convex combination of only p vertices of S.[1] In other words, there are admissible vectors $\mathbf{w}$ (the extreme rays of $\mathcal{H}$), that only contain at most p nonzero values. However, if one does not want to compute the matrix $\mathbf{E}$ of extreme rays of S, this is not an easy task a priori to find such a minimal representation. Indeed, the dimension of $\mathbf{w}$ is exponential in the size of the problem. Thus, we are facing a more complex problem, since we want to decompose a point as a convex combination of vertices of the cone, but we even do not know a priori which is this point, among the whole set of points that satisfy Eq. (5).

In order to compute such a "good" vector $\mathbf{v}$ and its corresponding decomposition, we introduce yet another cone $\mathcal{K} \subset \mathbb{R}^p$. This cone is the projection of S by the matrix $\mathbf{P}$: $\mathcal{K} = \{\mathbf{y} = \mathbf{P}\mathbf{v} : \mathbf{v} \geqslant \mathbf{0}, \mathbf{N}\mathbf{v} = \mathbf{0}\}$.

The idea of the algorithm is as follows: We know that the vector $\mathbf{v}_m$ is in $\mathcal{K}$ (see Eq. (5)). So, it can be expressed as a convex combination of p extreme rays of $\mathcal{K}$ (because $\mathcal{K}$ has dimension p). Now, the extreme rays of $\mathcal{K}$ are the projections of extreme rays $\mathbf{e}_i$ of $\mathcal{S}$ under the matrix $\mathbf{P}$. This implies that the corresponding convex combination of the $\mathbf{e}_i$ gives us the required $\mathbf{v}$. We start from an extreme ray $\mathbf{e}_1$ of the cone $\mathcal{S}$, for instance by applying the linear program (11) (if $\mathbf{P}\mathbf{e}_1 = 0$, one can for instance add a constraint of the type $(P\mathbf{e}_1)_j \geq \epsilon$ to prevent this from happening). The ray $\mathbf{v}_m = \mathbf{y}_0$ can be written as the convex combination of $\mathbf{P}\mathbf{e}_1$ and a ray $\mathbf{y}_1$, which belongs to a face $\mathcal{P}_1$ of $\mathcal{K}$: $\mathbf{v}_m = \alpha_1\mathbf{P}\mathbf{e}_1 + (1 - \alpha_1)\mathbf{y}_1$. This vector $\mathbf{y}_1$ is easy to find with a line search in the cone $\mathcal{K}$ as in Program (12). Now, at each step, find an extreme ray $\mathbf{e}_i$ of $\mathcal{S}$ which is mapped to the face $\mathcal{P}_{i-1}$ of $\mathcal{K}$. Then $\mathbf{y}_{i-1}$ can be expressed as a convex combination of $\mathbf{P}\mathbf{e}_i$ and a vector $\mathbf{y}_i$ that belongs to a face $\mathcal{P}_i$ of $\mathcal{P}_{i-1}$. Since the dimension of $\mathcal{P}_i$ strictly decreases at each step, after $t \leq p - 1$ steps the point $\mathbf{y}_t$ is actually an extreme ray of $\mathcal{K}$, and is thus the projection of an extreme ray $\mathbf{e}_{(t+1)}$ of S. Finally we have the relations:

$$\mathbf{v}_m = \sum_1^{t+1} \lambda_i \mathbf{P}\mathbf{e}_i = \mathbf{P}\left(\sum_1^{t+1} \lambda_i \mathbf{e}_i\right), \tag{13}$$

and the vector between the parentheses above is a convex combination of at most p extreme vectors of S that satisfies (5).

4. Case-study

In Ref. Jungers et al. (2009), the interested reader will find an application of the algorithm presented in Section 4 to a metabolic network of CHO cells cultivated in batch mode in stirred flasks in a serum-free medium. The network involves the Glycolysis pathway, the Pentose–Phosphate pathway, the Krebs cycle, the amino-acid metabolism, the urea cycle as well as the nucleotide, protein and lipid synthesis (see Zamorano, Wouwer, & Bastin, 2009, for further motivation and details). For this network we have $m = 82$ fluxes and $n = 53$ internal metabolites (i.e. $m - n = 29$), and there are 65 329 elementary flux vectors (calculated with METATOOL). Moreover, there are $p = 22$ extra-cellular species whose degradation or accumulation rates in the culture medium are measured and collected in the vector $\mathbf{v}_m$.

The algorithm of Section 4 is then implemented with these data. In Jungers et al. (2009), we present a successful trial where the selected flux distribution is indeed expressed with a minimal set of 22 elementary flux vectors only, without actually computing the whole convex basis. Obviously the obtained flux distribution is just one possible solution among many others with a minimal decomposition. If the algorithm is re-run with the same initial data, one may find other solutions with a minimal decomposition by making use of other searching directions.

5. Conclusion

The goal of this paper was to show that even though the *enumeration* of the convex basis can be prohibitive, this does not preclude processing this convex basis in some implicit way, so that much important information about the system may be retrieved without explicitly and extensively computing the convex basis. We exemplify this in the present paper with the decomposition of the flux distribution in a minimal number of elementary flux modes, and it can be expected that many more issues on elementary flux modes will be addressed in the future by processing the convex basis in an implicit way rather than by giving an exhaustive list of its components.

Linear programming techniques may provide efficient ways to analyze bioreaction and metabolic networks. In this paper we have used only a small part of the power of these techniques. For instance, the choice of the elementary flux modes in our algorithm is done at each step at random, while one could imagine adding an objective function to optimize the decomposition with regard to a certain objective.

Acknowledgements

R. Jungers is a FNRS fellow and a BAEF fellow.

A matlab implementation of our algorithms is available on http://www.inma.ucl.ac.be/~jungers/contents/efm.zip.

References

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. New York, USA: Cambridge University Press.

Donoho, D., & Tanner, J. (2005). Sparse non-negative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9446–9451.

Famili, I., Forster, J., Nielsen, J., & Palsson, B. (2003). Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23), 13134–13139.

Fukuda, K., & Prodon, A. (1996). The double description method revisited. In R. Euler, & M. E. D. I. Manoussakis (Eds.), *Lecture notes in computation sciences*: *Vol. 1120. Combinatorics and computer science* (pp. 91–111). Springer-Verlag.

Grötschel, M., Lovász, L., & Schrijver, A. (1988). *Geometric algorithms and combinatorial optimization*. Berlin: Springer-Verlag.

Jungers, R., Zamorano, F., Blondel, V., Wouwer, A. V., & Bastin, G. (2009). A fast algorithm for computing a minimal decomposition of a metabolic flux vector in terms of elementary flux vectors. In *CD Rom proceedings 6th mathmod conference. Vienna, Austria.*

Klamt, S., Gagneur, J., & von Kamp, A. (2005). Algorithmic approaches for computing elementary modes in large biochemical networks. *IEE Proceedings—Systems Biology*, *152*, 249–255.

Klamt, S., & Stelling, J. (2003). Two approaches for metabolic pathway analysis? *Trends in Biotechnology*, *21*(2), 64–69.

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schusterand, S., & Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, *22*(8), 400–405.

Price, N., Papin, J., & Palsson, B. (2002). Determination of redundancy and systems properties in the metabolic network of helicobacter pylori using genome-scale extreme pathway analysis. *Genome Research*, 760–769.

Provost, A. (2006). Metabolic design of dynamic bioreaction models. *Ph.D. thesis*. Faculty of engineering. Université Catholique de Louvain.

[1] To see this, consider the expression (8) of the polytope $\mathcal{H}$, which describes the set of admissible values of $\mathbf{w}$. It can be defined by only p equalities, so that $\dim(\mathbf{w}) - p$ inequality constraints can be activated to define an extreme ray $\mathbf{h}_i$ of $\mathcal{H}$.

Provost, A., & Bastin, G. (2004). Dynamical metabolic modelling under the balanced growth condition. *Journal of Process Control*, *14*(7), 717–728.
Provost, A., & Bastin, G. (2006). Metabolic flux analysis: an approach for solving non-stationary underdetermined systems. In *CD-Rom proceedings 5th MATHMOD conference. Paper 207 in session SP33. Vienna, Austria.*
Stephanopoulos, G., Nielsen, J., & Aristidou, A. (1998). *Metabolic engineering: principles and methodologies*. San Diego: Academic Press.
Terzer, M., & Stelling, J. (2006). Accelerating the computation of elementary modes using pattern trees. In *Lecture notes in computer sciences*: *Vol. 4175. Algorithms in bioinformatics* (pp. 333–343). Springer-Verlag.
Urbanczik, R. (2007). Enumerating constrained elementary flux vectors of metabolic networks. *IET Systems Biology*, *1*(5), 274–279.
Zamorano, F., Wouwer, A. V., & Bastin, G. (2009). Metabolic flux interval analysis of CHO cells. In *CD Rom proceedings 6th mathmod conference. Vienna, Austria.*

Raphal Jungers is a researcher at the Universite catholique de Louvain, Belgium. His main interests lie in the fields of computer science, graph theory, optimization and control. He received a Ph.D. in mathematical engineering from the Universit catholique de Louvain (2008), and an engineering degree in applied mathematics, both from the Ecole Centrale Paris, France (2004), and from the Universit catholique de Louvain (2005). He holds a minor degree in electrical engineering from the Universit catholique de Louvain (2005).

He was a postdoctoral fellow at the Department of Computer Science of the Universit Libre de Bruxelles, Belgium (2008–2009), and at the Laboratory for Information and Decision Systems of the Massachusetts Institute of Technology, MA, USA (2009–2010). He is a FNRS fellow and a BAEF fellow. He was the recipient of the IBM Belgium 2009 award.

Francisca Zamorano is a Biochemical Engineer from the Pontifical Catholic University of Valparaso, Chili and has a graduate degree in Environmental Engineering, from the University of Santiago de Compostapi, Spain. She is currently a Ph.D. student at the Automatic Control Laboratory, University of Mons, Belgium.

Vincent D. Blondel is Professor of Applied Mathematics at the Universit catholique de Louvain (Belgium). He has held research and visiting appointments at Oxford University, at the Royal Institute of Technology in Stockholm, at INRIA France, at the Ecole Normale Suprieure in Lyon and at the University of Paris VII. He was visiting professor and Fulbright scholar at the Massachusetts Institute of Technology in 2005–2006, and again in 2010–2011. Dr. Blondel's major research interests lie in several area of mathematical control theory, discrete mathematics and theoretical computer science. He has been a recipient of a Grant from the Trustees of the Mathematics Institute of Oxford University, the Prize Agathon De Potter of the Belgian Royal Academy of Science, the Prize Paul Dubois of the Montefiore Institute, the IEEE Antonio Ruberti prize in systems and control, and the SIAM prize on control and systems theory.

Alain Vande Wouwer was born in Brussels in 1966. He graduated in electrical engineering from the Facult Polytechnique de Mons (Belgium) in 1988, and received the doctorate degree in applied sciences from the same institution in 1994, together with the European doctorate degree (in collaboration with Stuttgart University). In 1994, he achieved a post-doctoral stay in the Mechanical Engineering Department of Laval University, Qubec, Canada. Nowadays, he is full professor in the Control Department of Mons University, Belgium, and coordinator of the research pole BIOSYS, which is dedicated to various aspects of life sciences, ranging from biomedical engineering to industrial biotechnology. His research interests are in mathematical modeling, soft sensing, and control of biological systems and processes, with applications in the pharmaceutical sector, as well as in water treatment and the sustainable production of renewable energies.

Georges Bastin received the electrical engineering degree and the Ph.D. degree, both from Universit Catholique de Louvain, Louvain-la-Neuve, Belgium. He his presently Professor in the Department of Mathematical Engineering at Universit Catholique de Louvain. His main research interests are in nonlinear control of compartmental systems and boundary control of hyperbolic systems with applications in biology and environmental systems.