A Simple Rank Test to Distinguish Extreme Pathways from Elementary Modes in Metabolic Networks

Dimitrije Jevremovic[1] , Cong T. Trinh[2,3], Friedrich Srienc[2,3] and Daniel Boley*[1]

[1]Department of Computer Science and Engineering, University of Minnesota, Minneapolis MN 55455 USA
[2]Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis MN 55455 USA
[3]Biotechnology Institute, University of Minnesota, St. Paul MN 55108 USA

Email: Dimitrije Jevremovic - jevrem@cs.umn.edu; Cong T. Trinh - trin0063@umn.edu; Friedrich Srienc - srienc@umn.edu; Daniel Boley*- boley@cs.umn.edu;

*Corresponding author

Abstract

Background: Metabolic pathway analysis is a powerful tool to study the metabolic structure of a cellular metabolism that comprises an intricate network for transforming metabolites through enzyme-catalyzed reactions. The approach is based on convex analysis to solve a homogeneous system of linear equations and inequality constraints derived from the steady state operation of mass conservation of metabolites. The solutions constitute the admissible flux space known as the convex polyhedral cone. Elementary Mode and Extreme Pathway Analysis are two closely related techniques that have been developed to identify pathways spanning the admissible flux space. Both elementary modes and extreme pathways are genetically independent pathways that can support steady state operation of cellular metabolism. However, the set of extreme pathways is often a subset of elementary modes, and under certain conditions only extreme pathways are the generating edges of the polyhedral cone. Because the two techniques are closely related, it is important to develop a theoretical framework to distinguish extreme pathways from elementary modes.

Results: We have found a simple algebraic test to distinguish extreme pathways from elementary modes which requires only the stoichiometry matrix. The method has been tested with published metabolic networks that have been characterized with Elementary Mode Analysis and Extreme Pathway Analysis. The identity and number of elementary modes are not altered in networks subjected to splitting every reversible reaction into two different irreversible reactions, other than the spurious futile cycles involving the new reactions themselves. However, the set of extreme pathways depends strongly on the specific treatment of the reversible reactions of the network. The application of this algebraic test for efficient computation of elementary modes in very large networks is discussed.

Conclusions: Elementary modes are the complete set of genetically independent pathways of a cellular metabolism that supports steady state operation. With the simple algebraic test, we can easily identify whether a given pathway is an elementary mode or an extreme pathway before computing the complete set of pathways. This test provides a convenient way to analyze and interpret network topology with Metabolic Pathway Analysis. The algebraic test is also useful for improving the efficiency of computing elementary modes in very large metabolic networks.

Background

Methods for the analysis of metabolic networks have become one of the major tools in systems biology, biotechnology and metabolic engineering for elucidation of metabolic properties. With the advances in genomic science, and especially in the sequencing of genomes, it has become possible to reconstruct complete metabolic networks of various organisms [1–4]. In order to explain the metabolic networks, it is necessary to develop an analytical framework which will help in the description of cell function including growth and regulation [5, 6], as well as in predicting properties like product yield [7], network robustness [6] and rational strain design [8, 9]. In this analysis metabolic networks are described in terms of their structure and topology, neglecting kinetic effects. Metabolic pathway analysis is a method which explains pathways inherent in metabolic networks and proposes algorithms to compute and analyze them.

An admissible pathway through a metabolic network is a vector of reaction fluxes such that (a) the concentration of internal metabolites remains constant under steady-state conditions, and (b) the irreversible reactions (if present) are active in the admissible direction. Condition (a) means that the reaction fluxes are such that the net consumption and production of each internal metabolite add up to zero. A pathway can be represented by a vector $\mathbf{x}$ of reaction fluxes or rates. A admissible pathway is reversible if all the individual reactions with non-zero fluxes are reversible.

If a given pathway has the property that no other admissible pathway involves a set of reactions which are a proper subset of the set of reactions in the given pathway, then the pathway is said to be an *elementary [flux] mode* [10–12]. Furthermore, if the corresponding flux or rate vector is not a convex combination of any other admissible rate vectors, then the given pathway is said to be *extreme* [13, 14], in the case where no admissible pathway is reversible. This paper presents conditions to distinguish elementary modes and extreme pathways.

History and Related Work

All algorithms that have been proposed and implemented for the computation of extreme pathways [13] and elementary modes [7, 11, 12, 15], were based on convex analysis and the Double Description Method [16, 17] for the computation of the extreme rays of a polyhedral cone. The extreme rays of the cone correspond to the extreme pathways of a metabolic network. Examples of such algorithms include the Canonical Basis Algorithm [12] and the Nullspace Algorithm [18–20]. The latter has generally been found to be more efficient on most problems.

Several implementations of the algorithms have been proposed, most based on the Nullspace Algorithm: Expa [21] implements the algorithm of [13] for computation of extreme pathways; Metatool [10, 22], and its successors FluxAnalyzer and CellNetAnalyzer [23, 24] compute the elementary modes. The Nullspace Algorithm was improved using a binary approach [20] reducing memory requirements by up to 96%, storing zero and non-zero values of the current matrix as bit values of 1 and 0.

The set of extreme pathways is often insufficient for the analysis of biochemical networks, since some of the elementary modes which are not extreme pathways may be as equally important as extreme pathways in their biological significance, even though they may be represented as a non-negative linear combination of a few extreme pathways [12]. Therefore, there is often a need to compute elementary modes, and not only extreme pathways [25]. One can also split all reversible reactions into two irreversible reactions, and compute the set of elementary modes for such metabolic network, in which the set of extreme pathways coincides with the set of elementary modes. Hence it is desirable to compute elementary modes directly, a task which corresponds only indirectly to the Double Description method. The resulting methods also avoid the need to eliminate the spurious cycles formed by the two associated irreversible reactions [18, 19].

We have observed that the Nullspace Algorithm and its Metatool implementation, in the case where

metabolic network corresponds to the pointed polyhedral cone, can also be used to compute only the extreme pathways by simply examining only the rows corresponding to irreversible reactions. This takes much less time and space compared to that needed to compute all elementary modes [19].

A concept related to the elementary modes, is the one of minimal cut sets of biochemical networks [20, 26]. A minimal cut set is a minimal set of reactions whose removal from the network blocks the operation of an objective reaction, or in general impedes feasible steady state flux distribution in which the objective reaction is active.

Estimation of the number of elementary modes and extreme pathways was also examined [27, 28] in order to predict the complexity of the computational task to find all such metabolic pathways. Being a computationally demanding task, several approaches to parallel or distributed computation of elementary modes have been proposed through parallelization techniques [13, 29–31], or algorithmic reformulations [32–34].

Elementary modes and extreme pathways are used to analyze many aspects of metabolic networks. In [6], elementary mode analysis was applied to examine the metabolic network robustness and cellular regulation of *Escherichia coli* central metabolism. The concept of extreme pathways was used to analyze the human red blood cell metabolism and determine the steady-state solution space under the given network properties [35]. Another field of application of elementary modes, the use for identification of the most efficient pathway for the production of protein, was demonstrated again in the analysis of the metabolic network of *E. coli* [36]. In [8, 37, 38] elementary modes were used to design a more efficient bacterium with a high yield of biomass. Elimination of five reactions in the metabolic network of metabolism of *E. coli* resulted in a collapsed network consisting of a single pathway producing biomass from glucose. *In vivo* implementation of this design in the lab on a glucose substrate resulted in biomass yields up to 30% higher than wild-type bacteria, close to theoretical predictions [8]. Elimination of two additional reactions resulted in a strain with minimal metabolic functionality that is optimized for efficient ethanol production [9].

In this paper we pay special attention to variations to the rank test as used within the Nullspace algorithm. It is known that the combinatorial test can also be used within this algorithm [20, 39, 40], in some cases with greater efficiency. But the rank tests proposed here are also usable as free-standing tests given only the individual pathway being tested, together with the stoichiometry matrix.

Methods

A network of metabolic pathways consists of a set of metabolites connected by reactions. Formally, we represent the network with a stoichiometry matrix N_{full} whose (i, j) element is the amount of metabolite i produced (consumed if negative) by reaction j. If $\mathbf{x}$ is a rate vector, then $N_{full}\mathbf{x}$ is a vector of net production (consumption if negative) rates for each metabolite in the system. If we let N and N_{ext} consist of the rows of N_{full} corresponding to the internal and external metabolites, respectively, then the mass balance condition gives $N\mathbf{x} = 0$, and the vector $N_{ext}\mathbf{x}$ gives the net consumption and production of the metabolites external to the system. The above informal definitions lead to the following formal definitions.

Definition 1 (Elementary Mode and Extreme Pathway)**.** Let N be the $m \times q$ stoichiometry matrix representing the internal metabolites and the reactions connecting those metabolites. A flux vector or pathway is a q-vector $\mathbf{x}$ of reaction fluxes or rates. The vector $\mathbf{x}$ is said to be an *elementary mode* or *elementary flux mode* if it satisfies three conditions [10–12, 25]:

1. *pseudo steady-state*: $N\mathbf{x} = 0$. Metabolites are not accumulating within the metabolic network.
2. *thermodynamics*: $\mathbf{x}_i \geq 0$ if the i^{th} reaction is irreversible.

3. *non-decomposability*: there is no vector $\mathbf{v}$ ($\mathbf{v} \neq \mathbf{x}$ and $\mathbf{v} \neq 0$) fulfilling (1) and (2) such that the set of indices of non-zero elements in $\mathbf{v}$ is a proper subset of set of indices of non-zero elements in $\mathbf{x}$.

We call a pathway or flux mode *admissible* if it satisfies conditions 1 and 2. In [13], an extreme pathway is defined as a member of a set of admissible and non-decomposable pathways which are obtained when the internal reversible reactions of the metabolic network are split into irreversible reactions, and which cannot be written as a convex combination of any other admissible pathways.

However, if sufficient number of internal reversible reactions are split then the metabolic network may be represented as a pointed polyhedral cone, and in that case set of extreme pathways would coincide with the "minimal generating set" [13]. Mathematically, this set represents a *convex basis* of the network if the metabolic network observed corresponds to pointed polyhedral cone. We will instead use the following definition [14], which corresponds to the property of being an extreme ray in the polyhedral cone of admissible pathways.

4. *independence*: $\mathbf{x}$ is said to be an extreme pathway if it cannot be written as a convex combination of any other admissible pathways.

In other words, a given flux distribution is an elementary mode if and only if it involves no net production or consumption of the internal metabolites at steady-state, it is thermodynamically feasible, and there is no other non-null flux distribution that satisfies these constraints and involves a proper subset of its participating reactions. The mode is also *extreme pathway* if it cannot itself be expressed as an admissible convex combination of other elementary modes.

It turns out that in the case that there are no reversible admissible pathways, the polyhedral cone of admissible pathways is pointed, and in this case the set of extreme pathways forms the unique minimal generating set for the network. Often this condition is made to be satisfied almost automatically by splitting all the internal reversible reactions into separate irreversible reactions, leaving only reversible exchange fluxes unsplit. In the case there is a reversible admissible pathway in the network, the minimal generating set is not unique. An extreme pathway may not even exist, but is reversible if does exist. This occurs, for example, in Example 3 (below) of human red blood cell metabolism [35].

Double Description Pairs

All state of the art methods for computing elementary modes are derived from the classical theory of double description pairs [16], which we now briefly highlight.

Definition 2 (Double Description Pair [16])**.** Given a $d \times q$ matrix A with full column rank q, and a $q \times n$ matrix R, A and R form a Double Description Pair (DD pair) iff

$$\{\mathbf{x} : A\mathbf{x} \geq 0\} = \{\mathbf{x} : \mathbf{x} = R\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq 0\} \tag{1}$$

The DD pair is called minimal if there is no other $\widetilde{R}$ forming a DD pair with A with fewer than n columns.

Here, the columns of R form a set of generators for all the $\mathbf{x}$ that satisfy $A\mathbf{x} \geq 0$, where all such $\mathbf{x}$ are to be expressed as convex combinations of the columns of R. The vectors $\mathbf{x}$ such that $A\mathbf{x} \geq 0$ are called "rays," due to the fact that they coincide with the edges of an infinite polyhedral cone in $\mathbb{R}^q$ anchored at the origin. If R is minimal, the columns of R are called the extreme rays, because these are exactly the rays that cannot be expressed as a convex combination of any other admissible rays. As a result, the R forming a DD pair with A is unique up to ordering and scaling of the columns.

If A has rank less than q, then one can still define a minimal DD pair, but in this case, not all of the columns of R are extreme rays and the minimal R is no longer unique. In this case, there exists at least one ray $\mathbf{x}$ such that $-\mathbf{x}$ is also a valid ray within the cone, namely any nonzero vector in the right nullspace of A. This corresponds to a non-pointed cone.

Notation. In the following development, we use the notation adapted from [20, 40]:

$$\mathsf{Z}(\mathbf{x}) = \{i : x_i = 0\} \quad \text{and} \quad \overline{\mathsf{Z}}(\mathbf{x}) = \{i : x_i \neq 0\}$$

denote the sets of indices of elements of a given vector $\mathbf{x}$ which are zero and non-zero, respectively. When the context is clear, we simplify this to Z and $\overline{Z}$. If N is a matrix, then $N_{*,Z}$ denotes the submatrix consisting of all the rows, and only those columns indexed by the index set Z, and $N_{\overline{Z},*}$ denotes the submatrix consisting of the rows indexed by $\overline{Z}$ and all columns, also adapted from the notation used in [20, 40]. The right nullspace of a matrix N is the set $\{\mathbf{x} : N\mathbf{x} = 0\}$, and its dimension is denoted $\mathsf{nullity}(N)$. It is well known from linear algebra that if N has dimensions $m \times q$, then $\mathsf{rank}(N) + \mathsf{nullity}(N) = q$.

The following theorem gives a way to check if a given individual ray is extreme when a complete set of extreme rays is not available.

Theorem 1 (Extreme Ray Theorem [16])**.** *Let A be $d \times q$ with rank q. A q-vector $\mathbf{x}$ is an extreme ray for A if and only if the rank of $A_{\mathsf{Z}(A\mathbf{x}),*}$ is $q-1$ (where $\mathsf{Z}(A\mathbf{x})$ is the set of indices of the zero entries in the vector $A\mathbf{x}$) or equivalently, the nullity of $A_{\mathsf{Z}(A\mathbf{x}),*}$ is 1. Hence the set of extreme rays are uniquely defined up to scale factors.* □

Double Description Pairs and Stoichiometry

We now discuss rank tests for various properties of admissible pathways in a stoichiometric network, which can be derived from the related theory of double description pairs [16, 20]. We begin with the well known theorem stating how the nullity of certain submatrices of the stoichiometry matrix N indicates whether a given vector representing a pathway is elementary mode [19]. We then introduce a similar rank test for the extreme pathways.

We begin with the well-known theorem giving the rank test for being an elementary mode.

Theorem 2 (Nullity Test for Elementary Modes [19])**.** *Let N be the stoichiometry matrix for a given metabolic network and let $\overline{\mathsf{Z}} = \overline{\mathsf{Z}}(\mathbf{x})$ denote the indices of the*

non-zero entries in a given vector $\mathbf{x}$: $\overline{\mathsf{Z}}(\mathbf{x}) = \{i : x_i \neq 0\}$. If $\mathbf{x}$ is is an admissible pathway for network represented by N (in the sense of Definition 1), then $\mathbf{x}$ is an elementary [flux] mode if and only if

$$\mathsf{nullity}(N_{*,\overline{\mathsf{Z}}}) = 1.$$

Any admissible pathway can be written as a linear combination of elementary modes, where the coefficients corresponding to all flux vectors with at least one non-zero irreversible flux are non-negative. □

Standard Stoichiometry Problem as a Double Description Pair. Our new rank test for the property of a vector being an extreme pathway depends on the close connection between the stoichiometry problem (Def. 1) and the double description pair (Def. 2), which we sketch here. Given an $m \times q$ stoichiometry matrix N, let

$$A = \begin{pmatrix} N \\ -N \\ E \end{pmatrix} \tag{2}$$

where $E = I_{1,\ldots,q_i,*}$ consists of the first q_i rows of a $q \times q$ identity matrix, corresponding to the irreversible reactions $(\rho_1, \ldots, \rho_{q_i})$, where q_i is the number of irreversible reactions. Then the stoichiometry problem is equivalent to finding a matrix R with minimal number of columns such that

$$\begin{aligned}\{\mathbf{x} : A\mathbf{x} \geq 0\} &= \{\mathbf{x} : N\mathbf{x} = 0, \mathbf{x}_{1,\ldots,q_i} \geq 0\} \\ &= \{\mathbf{x} : \mathbf{x} = R\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq 0\}.\end{aligned} \tag{3}$$

In the case that A has full column rank q, R should consist of all the possible extreme rays with respect to matrix A, which correspond to the extreme pathways of the matrix N. In the case that all reactions are irreversible, then the set of extreme pathways (or "rays" in the context of a double description pair) coincides with the set of elementary modes. But if there is at least one reversible reaction, then there may be elementary modes which are not extreme pathways, i.e. which are the convex combination of two or more other admissible pathways. We call these non-extreme elementary modes.

Irreversible Extreme Pathways. The concept of extreme pathways depends on whether a polyhedral cone corresponding to metabolic network is pointed. Hence it is useful to state a simple rank test that can be applied to a stoichiometry matrix to indicate whether a given network corresponds to a pointed cone or not. In terms of the stoichiometric network, a pointed cone corresponds to the absence of any reversible admissible pathways. In the case of a pointed cone, the matrix A of eqn. (2) will have full column rank and hence will have a unique minimal R of extreme rays forming the double description pair (A, R). Following this argument to its logical conclusion yields the following simple test for "pointed cone."

Theorem 3. *Let a metabolic network have m metabolites and q reactions, represented by an $m \times q$ stoichiometry matrix N. Let N_{rev} denote the submatrix of N consisting of the columns of N corresponding to all the reversible reactions. The metabolic network has a unique set of extreme pathways, equivalent to the minimal generating set of all pathways, if and only if* $\mathsf{nullity}(N_{\text{rev}}) = 0$.

Proof:

Consider the matrix N_{rev} consisting of the columns of the stoichiometry matrix N corresponding to all the reversible reactions and N_{irrev} consist of the remaining columns. A column reordering of the corresponding matrix (2) gives the partitioning as

$$\begin{pmatrix} N_{\text{irr}} & N_{\text{rev}} \\ -N_{\text{irr}} & -N_{\text{rev}} \\ I & 0 \end{pmatrix}$$

where the I is an identity matrix with dimension equal to the number of irreversible reactions. This matrix has full column rank if and only if the block N_{rev} has full column rank. Hence we can apply Theorem 1.

But independently of Theorem 1, we can observe the following. If the submatrix N_{rev} has a non-trivial right nullspace (i.e. $\mathsf{nullity}(N_{\text{rev}}) \geq 1$), then any non-zero vector in this nullspace corresponds to an admissible reversible pathway. On the other hand, any admissible reversible pathway must lie in the nullspace of this submatrix, since only reversible reactions can have non-zero fluxes. Hence the metabolic network admits a reversible pathway if and only if $\mathsf{nullity}(N_{\text{rev}}) \geq 1$.

If an admissible reversible pathway exists, then arbitrary multiples of it could be added to any member of a given minimal generating set to obtain a new minimal generating set, which is therefore not unique.

On the other hand, suppose we had at least two different minimal generating sets $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ and $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots\}$, where we assume without loss of generality (WLOG) that $\gamma_1 \mathbf{x}_1 \notin \mathcal{Y}$ for any scalar γ. We wish to show that there must be a reversible pathway. Since $\mathcal{Y}$ is a generating set, we can write

$$\mathbf{x}_1 = \textstyle\sum_i \alpha_i \mathbf{y}_i, \text{ with } \alpha_i \geq 0, \tag{4}$$

but each $\mathbf{y}_i$ can be written as a convex combination of the vectors $\mathcal{X}$. Substituting these expressions for the $\mathbf{y}$'s into (4) yields a formula of the form

$$\mathbf{x}_1 = \textstyle\sum_i \beta_i \mathbf{x}_i, \text{ with } \beta_i \geq 0. \tag{5}$$

There are four cases. Suppose that at least one of $\beta_2, \beta_3, \ldots$ are nonzero, say $\beta_2 > 0$. Let $\mathbf{x} = \sum_{i \geq 2} \beta_i \mathbf{x}_i = (1 - \beta_1)\mathbf{x}_1$.

Case I. If $(1 - \beta_1) = 0$ then $-\beta_2 \mathbf{x}_2 = \sum_{i \geq 3} \beta_i \mathbf{x}_i$, hence $\mathbf{x}_2$ is a reversible pathway.

Case II. If $(1 - \beta_1) < 0$ then $\mathbf{x}_1$ must be reversible.

Case III. If $(1 - \beta_1) > 0$ then $\mathcal{X}$ cannot be minimal, since $\mathbf{x}_1$ is a convex combination of the other $\mathbf{x}$'s.

Case IV. The only remaining possibility is that $\beta_i = 0$ for all $i = 2, 3, \ldots$. We show this cannot happen if $\mathcal{X}$ and $\mathcal{Y}$ are really different, unless some reversible pathway exists. We are forming the convex combination (5) by taking convex combinations of convex combinations. Since all coefficients are non-negative, no cancellation can occur unless the negative of some pathway is also admissible. To be more specific, we can assume WLOG that $\alpha_1 > 0$ and $\alpha_2 > 0$, since at least two of the α's must be nonzero (otherwise a multiple of $\mathbf{x}_1$ would be in $\mathcal{Y}$). When writing $\mathbf{y}_1$ and $\mathbf{y}_2$ as convex combinations of the $\mathbf{x}$'s, the combinations must involve at least two different vectors from $\mathcal{X}$, say $\mathbf{x}_i$ and $\mathbf{x}_j$ with $j \neq 1$, with corresponding positive coefficients. When we substitute these expressions into (4) to obtain (5), the terms in $\mathbf{x}_j$ must cancel. Such cancellation cannot happen if $\mathbf{x}_j$ appears in the convex combinations with only positive coefficients, unless $-\mathbf{x}_j$ is also an admissible pathway.

Hence we conclude that the minimal generating set is not unique if and only if there exists an admissible reversible pathway, which in turn exists if and only if $\mathsf{nullity}(N_{\text{rev}}) \geq 1$.

□

Extreme Rays. In what follows, we assume the cone is pointed. We use the relationship between the stoichiometry problem and the double description pair to develop a new simple "nullity test" for extreme pathways similar to the test of Theorem 2.

Theorem 4 (Nullity Test for Extreme Pathways vs Elementary Modes)**.** *Assume N is a $m \times q$ stoichiometry matrix with full rank m, whose first q_i reactions are irreversible (without loss of generality). Let $\mathtt{R} = \{q_i + 1, \ldots, q\}$ denote the indices of the reversible reactions, and assume the corresponding columns of N, denoted $N_{*,\mathtt{R}}$, are linearly independent. Let $\mathbf{x}$ be an admissible pathway (or flux vector) with respect to N, i.e.*

$$N\mathbf{x} = 0, \quad x_1 \geq 0, \ldots, x_{q_i} \geq 0.$$

Let $\overline{\mathtt{Z}}$ denote the set of indices of the non-zero entries in $\mathbf{x}$. Then $\mathbf{x}$ is an elementary mode iff $\mathsf{nullity}(N_{*,\overline{\mathtt{Z}}}) = 1$, and is an extreme pathway iff $\mathsf{nullity}(N_{*,\overline{\mathtt{Z}} \cup \mathtt{R}}) = 1$.

Of course, the assumption that the irreversible reactions occur first is just for notational convenience. This theorem holds for any ordering of the reactions as long as the entries of the flux vector $\mathbf{x}$ have the corresponding order.

Proof: Let $\mathbf{x}$ be a q-vector, with $z = |\mathtt{Z}|$ entries equal to zero and $q - z$ entries non-zero. Without loss of generality, we can permute the entries of $\mathbf{x}$ to put all the zero entries first, and apply the same permutation to the columns of A and N. With this permutation, we have $x_1 = \cdots = x_z = 0$ and $x_i \neq 0$ for $i = z + 1, \ldots, q$, and we partition the reordered A and N as $A = (A_1, A_2)$ and $N = (N_1, N_2)$, where A_1, N_1 each have z columns (corresponding to the zero entries in the vector $\mathbf{x}$ after permutation). The condition of being elementary depends on nullity of $N_{*,z+1,\ldots,q}$ by Theorem 2. The condition of being an "extreme ray" depends on the rank of $A_{\mathtt{Z}(A\mathbf{x}),*}$ by Theorem 1. These two conditions are related. We can write the

product $A\mathbf{x}$ as

$$
\begin{aligned}
A\mathbf{x} &= \begin{pmatrix} N \\ -N \\ E \end{pmatrix} \mathbf{x} \\
&= \begin{matrix} {\scriptstyle (m)}\,\{ \\ {\scriptstyle (m)}\,\{ \\ {\scriptstyle (z_{\mathrm{irr}})}\,\{ \\ {\scriptstyle (\bar{z}_{\mathrm{irr}})}\,\{ \end{matrix} \begin{pmatrix} \overbrace{}^{(z)} & \overbrace{}^{(q-z)} \\ N_1 & N_2 \\ -N_1 & -N_2 \\ E_1 & 0 \\ 0 & E_2 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{x}_2 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \\ E_2\mathbf{x}_2 \end{pmatrix},
\end{aligned}
\tag{6}
$$

where E_1, E_2 are the z_{irr}, $\bar{z}_{\mathrm{irr}}$ rows of the identity matrix corresponding to the irreversible reactions in $\mathbf{x}_1 = 0$, $\mathbf{x}_2 \neq 0$, respectively, or equivalently among $\{\rho_1, \ldots, \rho_z\}$ and $\{\rho_{z+1}, \ldots, \rho_q\}$, respectively. If all reactions were irreversible, then we would have $z_{\mathrm{irr}} = z$ and $\bar{z}_{\mathrm{irr}} = q - z$, and the blocks E_1, E_2 would be simply identity matrices of dimensions z, $q - z$, respectively.

Then we have that none of the entries in the vector $E_2\mathbf{x}_2$ are zero. We can divide the entries in $\mathbf{x}$ into the entries corresponding to the irreversible reactions and those corresponding to the reversible reactions: $\mathbf{x} = (\mathbf{x}_{1,\mathrm{irr}}^T, \mathbf{x}_{1,\mathrm{rev}}^T, \mathbf{x}_{2,\mathrm{irr}}^T, \mathbf{x}_{2,\mathrm{rev}}^T)^T = (0, 0, *, *)^T$, where the $*$ marks the parts where no entry is zero. The dimensions of these part are, respectively, z_{irr}, z_{rev}, $\bar{z}_{\mathrm{irr}}$, $\bar{z}_{\mathrm{rev}}$. We now write the product $A\mathbf{x}$ (6) consistent with this partitioning:

$$
\begin{aligned}
A\mathbf{x} &= \begin{pmatrix} N \\ -N \\ E \end{pmatrix} \mathbf{x} \\
&= \begin{pmatrix} N_{1,\mathrm{irr}} & N_{1,\mathrm{rev}} & N_{2,\mathrm{irr}} & N_{2,\mathrm{rev}} \\ -N_{1,\mathrm{irr}} & -N_{1,\mathrm{rev}} & -N_{2,\mathrm{irr}} & -N_{2,\mathrm{rev}} \\ I_{1,\mathrm{irr}} & 0 & 0 & 0 \\ 0 & 0 & I_{2,\mathrm{irr}} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1,\mathrm{irr}} \\ \mathbf{x}_{1,\mathrm{rev}} \\ \mathbf{x}_{2,\mathrm{irr}} \\ \mathbf{x}_{2,\mathrm{rev}} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \\ E_{2,\mathrm{irr}}\mathbf{x}_{2,\mathrm{irr}} \end{pmatrix},
\end{aligned}
\tag{7}
$$

where the "I" blocks above are identity matrices of appropriate dimension: $E_1 = (I_{1,\mathrm{irr}}, 0)$, $E_2 = (I_{2,\mathrm{irr}}, 0)$. Then we follow the prescription of the Extreme Ray Theorem 1 by extracting the rows of A corresponding to the zero entries in $A\mathbf{x}$. The result is the following:

$$
\begin{aligned}
A_{\mathrm{Z}(A\mathbf{x}),*} &= \begin{pmatrix} N_{1,\mathrm{irr}} & N_{1,\mathrm{rev}} & N_{2,\mathrm{irr}} & N_{2,\mathrm{rev}} \\ -N_{1,\mathrm{irr}} & -N_{1,\mathrm{rev}} & -N_{2,\mathrm{irr}} & -N_{2,\mathrm{rev}} \\ I_{1,\mathrm{irr}} & 0 & 0 & 0 \end{pmatrix} \\
&= I_1 \begin{pmatrix} N_{1,\mathrm{irr}} & N_{1,\mathrm{rev}} & N_{2,\mathrm{irr}} & N_{2,\mathrm{rev}} \\ 0 & 0 & 0 & 0 \\ I_{1,\mathrm{irr}} & 0 & 0 & 0 \end{pmatrix} \\
&= I_1 I_2 \begin{pmatrix} 0 & N_{1,\mathrm{rev}} & N_{2,\mathrm{irr}} & N_{2,\mathrm{rev}} \\ 0 & 0 & 0 & 0 \\ I_{1,\mathrm{irr}} & 0 & 0 & 0 \end{pmatrix},
\end{aligned}
\tag{8}
$$

where we have reduced the matrix by two non-singular transformations and where

$$I_1 = \begin{pmatrix} I & 0 & 0 \\ -I & I & 0 \\ 0 & 0 & I \end{pmatrix} \text{ and } I_2 = \begin{pmatrix} I & 0 & N_{1,\text{irr}} \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

are two invertible transformations. Hence

$$\begin{aligned} &\mathsf{rank}\left(A_{\mathtt{Z}(A\mathbf{x}),*}\right) = \\ &= \mathsf{rank}\left(I_{1,\text{irr}}\right) + \left[\mathsf{rank}\left(N_{1,\text{rev}},\, N_{2,\text{irr}},\, N_{2,\text{rev}}\right)\right] \\ &= z_{\text{irr}} + \left[(q - z_{\text{irr}}) - \mathsf{nullity}\left(N_{1,\text{rev}},\, N_{2,\text{irr}},\, N_{2,\text{rev}}\right)\right] \\ &= q - \mathsf{nullity}\left(N_{1,\text{rev}},\, N_{2,\text{irr}},\, N_{2,\text{rev}}\right) \end{aligned}$$

So the condition that the rank of $A_{\mathtt{Z}(A\mathbf{x}),*}$ is $q-1$ is equivalent to the condition that the nullity of $(N_{1,\text{rev}},\, N_{2,\text{irr}},\, N_{2,\text{rev}})$ is 1. One can view this formula as saying that, regarding the determination of "extremeness", we need to select those columns of N corresponding to all reactions with non-zero fluxes, plus those columns corresponding to all remaining reversible reactions regardless of their flux.

Applying the construction of equation (8) to the entire matrix A leads to the relation

$$\mathsf{rank}(A) = \bar{z}_{\text{irr}} + \mathsf{rank}(N_{*,\text{rev}}) = q - \mathsf{nullity}(N_{*,\text{rev}}), \tag{9}$$

hence the assumption $\mathsf{rank}(A) = q$ is equivalent to $\mathsf{nullity}(N_{*,\text{rev}}) = 0$. We have thus established the theorem. □

Remark 1. If we remove the all-zero rows in the right-most matrix in equation (8), we are left with a matrix of dimensions $(m + z_{\text{irr}}) \times q$. The condition for being an extreme pathway that the rank of this matrix is $q-1$ implies that the matrix must have at least $q-1$ rows. That is, $m + z_{\text{irr}} \geq q - 1$. This yields a necessary condition that is easy to check: $|\overline{\mathtt{Z}} \cup \mathtt{R}| = q - z_{\text{irr}} \leq m + 1$. Likewise, regarding the test of being an elementary mode, the condition that $\mathsf{nullity}(N_{*,\overline{\mathtt{Z}}}) = 1$ implies that $|\overline{\mathtt{Z}}| = \bar{z} \leq m + 1$. These are low-cost tests which can be applied to eliminate candidate elementary modes before applying the more expensive tests based on Theorem 4.

Properties of the Algorithms

The Nullspace and Canonical Basis algorithms have been extensively and repeatedly described in the literature [12, 16, 18, 19]. Hence we put brief descriptions of these methods in Appendix B in order to relieve clutter in this paper. But we give some properties of these methods here since these are essential to properly interpret some of the results produced by these algorithms.

Theorem 5 (The Nullspace Algorithm extracts the extreme pathways [18, 19])**.** *Let $N_{m\times q}$ be a stoichiometric matrix, and q_{rev} is the number of reversible reactions. If $m \geq q_{\text{rev}}$, the columns of N corresponding to the reversible reactions are linearly independent, and the Nullspace Algorithm as described in [19] is configured to process the entries corresponding to the irreversible reactions first, then the modes obtained upon processing all the irreversible reactions are exactly the extreme pathways for the network.* □

This theorem states that if the irreversible reactions are processed first by the Nullspace algorithm, the extreme pathways can be extracted during the course of the computation, and then the computation can be continued to generate all the elementary modes which are not extreme pathways.

Lemma 1 (Elementary modes are the same when the reversible reactions are split into two irreversible reactions [20])**.** *If the original stoichiometric matrix is reconfigured so that every reversible reaction is split into two irreversible reactions resulting in the reconfigured matrix $N' = [N_{\text{rev}},\, N_{\text{irr}},\, -N_{\text{rev}}]$, the set of elementary modes obtained from the original matrix N is equivalent to the set of elementary modes obtained from the reconfigured matrix N'.* □

If all the reversible reactions in a stoichiometric matrix are decomposed into two irreversible components, then a set of elementary modes of the reconfigured matrix coincides with the set of its extreme pathways. As was shown, the Double Description Method finds all the extreme pathways. For the case of the reconfigured stoichiometric matrix with all reactions being made irrreversible by splitting reversible reactions, the set of elementary modes coincides with the set of extreme pathways. Therefore, running the Double Description Method results in computation of all the elementary modes for the reconfigured matrix. This reasoning and the above Lemma 1 yield the following theorem.

Theorem 6 (Completeness Theorem for the Nullspace Algorithm [19])**.** *The Nullspace algorithm (as instantiated in [19]) finds all elementary modes for a given stoichiometric matrix $N_{m\times q}$.* □

Remark 2 (Reducing the cost of the rank test)**.** The rank test (or more properly the nullity test) of Theorem 4 is used within the inner loop of the Nullspace algorithm (specifically step 2 in the algorithm as presented in Appendix B) in order to check if a prospective flux vector is an elementary mode and/or extreme pathway. The combinatorial test is also a valid alternative for this check, but relative efficiency of each depends on the specific implementations [20, 39, 40]. The cost of an individual rank test is independent of the total number of prospective elementary modes, but depends only on the size of the stoichiometry matrix. But it must be applied to almost every prospective elementary mode as it is generated, and hence the contribution to the total cost of the algorithm can be significant. In the following we show how the nullity can be obtained from a proper submatrix of that indicated by Theorem 4 when the initial stoichiometry matrix N has been transformed to reduced row echelon form, thereby reducing computation time. This reduced rank test is actually used in the Metatool algorithm [22] without much explanation.

At each stage k in the Nullspace algorithm, we are generating all pathways which are elementary modes with respect to the first k reactions, by trying combinations of pathways which were elementary modes with respect to the first $k-1$ reactions computed in the previous stage. As we generate each prospective flux vector $\mathbf{x}$, we must check if the nullity of the corresponding submatrix $N_{*,\overline{\mathtt{Z}}\cup\mathtt{I}}$ equals 1, where $\overline{\mathtt{Z}}$ is the set of nonzero indices among the first k entries of each prospective vector $\mathbf{x}$, and $\mathtt{I}$ are all the indices $k+1,\ldots,q$. In terms of the original Double Description Algorithm [16], we must check the nullity of rows selected from the first $2m+k$ rows of the matrix A in equation (2) (the entire N, $-N$ blocks plus certain rows from the first k rows of the E block in (2)):

$$\mathsf{nullity}\begin{pmatrix} N_{*,1\ldots k} & N_{*,k+1\ldots q} \\ -N_{*,1\ldots k} & -N_{*,k+1\ldots q} \\ E_{\mathtt{Z}_{1,\mathrm{q-m+k}}} & 0 \end{pmatrix}, \tag{10}$$

where we select the rows corresponding to the zero entries in $A\mathbf{x}$. Following the same construction as in (8), the nullity of (10) is the same as the nullity of the submatrix $\widetilde{N}$ of N consisting of columns $k+1,\ldots,q$ together with the columns corresponding to any all-zero columns within $E_{\mathtt{Z}_{1,\mathrm{q-m}}}$. If N has previously been reduced to row echelon form $N=(N_1, I)$, with an $m\times m$ identity matrix occupying the last m columns, then we can omit those columns of $\widetilde{N}$ which are drawn from those last m columns forming the identity matrix, and omit the rows corresponding to the nonzero entries found in those columns omitted in this way. The result is that we can obtain the nullity (10) by computing the nullity of the submatrix of N obtained by fetching the columns corresponding to nonzeros among the first $q-m$ entries of $\mathbf{x}$ and the rows corresponding to the zero entries among the last m entries of $\mathbf{x}$. The result of these considerations is the following theorem.

Theorem 7 (Reduced Rank Test)**.** *Assume the stoichiometry matrix N has been reduced by row operations to row echelon form $N=N_{\mathrm{rref}}=\begin{pmatrix}N_1 & I\end{pmatrix}$. Let the Nullspace algorithm be in its k^{th} stage of execution. A vector x is an elementary mode if $\mathsf{nullity}(N_{\mathtt{Z}_\mathrm{b},\overline{\mathtt{Z}}_\mathrm{a}})=1$, where $x=[a,b,c]$ and $a=[x_1,\ldots,x_{q-m}]$,$b=[x_{q-m+1},\ldots,x_k]$ and $c=[x_{k+1},\ldots,x_q]$, and $\mathtt{Z}_\mathrm{b}$ and $\overline{\mathtt{Z}}_\mathrm{a}$ represent indices of zero and non-zero elements in vectors a and b, respectively.* □

This theorem and associated construction is most easily illustrated with Example 5 of a simple metabolic network in the Appendix C.

We remark that while the rank tests within the Nullspace algorithm can be accelerated by this construction, it has also been found that the combinatorial test [20, 39] can also be as efficient, if not more so, in certain cases. However, recent improvements of the algorithm and use of residual arithmetic and rank updating [40] indicates that the rank test may remain as a valid alternative to the combinatorial test in the computation of the entire set of elementary modes, depending on the specific network being analysed and on the specific implementations used. In any case, the rank tests can still be applied even when one has only the one individual pathway one wishes to test, together with the stoichiometry matrix. This makes the rank test useful as a free-standing test, as well as enhancing the possibilities of parallelizing the Nullspace algorithm.

Results and Discussion

Validation of Rank/Nullity Test

Example 1 (A model of *E. coli* central metabolism)**.** An example of a metabolic network usually studied is a model of the central metabolism for *E. coli*, consisting of 70 reactions (19 reversible) and 68 metabolites (52 internal to the network, illustrated in Figure 1 [9]. In the model we consider the anapleurotic pathway converting malate to pyruvate to be NADH dependent only. In addition, the reaction FEM9 catalyzed by pyruvate decarboxylase to convert PYR to ACA is not native in *E. coli* but cloned into *E. coli* through the plasmid pLOI297 [9].

To give an application of elementary mode analysis, we use the Metatool software [22] on the *E. coli* network of Figure 1 to find a total of 38,001 elementary modes using glucose as the carbon source, of which 32,604 produce biomass and 5,010 are anaerobic. Using the theory developed in this paper, we can easily find that 2,739 of these are extreme pathways (1,191 are producing biomass and 978 are anaerobic which may or may not produce biomass). In [9] the goal was to find pathways maximizing the production of ethanol as a biofuel for a given amount of glucose, while producing sufficient biomass to allow the cells to grow by deleting the inefficient pathways. Maximizing a single linear objective function such as ethanol production subject to the set of linear constraints in definition 1 (plus the constraint that glucose consumption rate is 1 mole/L/hr) naturally leads to an extreme point in the polytope defining the feasible region, corresponding to an extreme pathway. But this pathway does not support cell growth, hence the need to trade off between the optimal solutions for two or more distinct objective functions. A resulting semi-optimal solution with a minimal number of reactions will be an elementary mode, which will generally be a convex combination of at least two extreme pathways representing the optimal solutions for each individual objective function. Having all elementary modes available allows one to explore many alternative knockouts to achieve similar performance objectives. Figure 2 shows the relative ethanol and biomass production of all the anaerobic modes, both extreme pathways and non-extreme elementary modes. Biomass yield is low because the result is shown for anaerobic growth conditions only. The investigation of engineering and biological applications of these modes is beyond the scope of this paper and will be the subject of separate papers.

Splitting Internal Reversible Reactions.

It has been proposed to split internal reversible reactions into two separate irreversible reactions of opposite directions in order to guarantee the existence of a unique set of extreme pathways forming a minimal generating set [13]. We discuss two examples showing that the nullity test proposed in Theorem 4

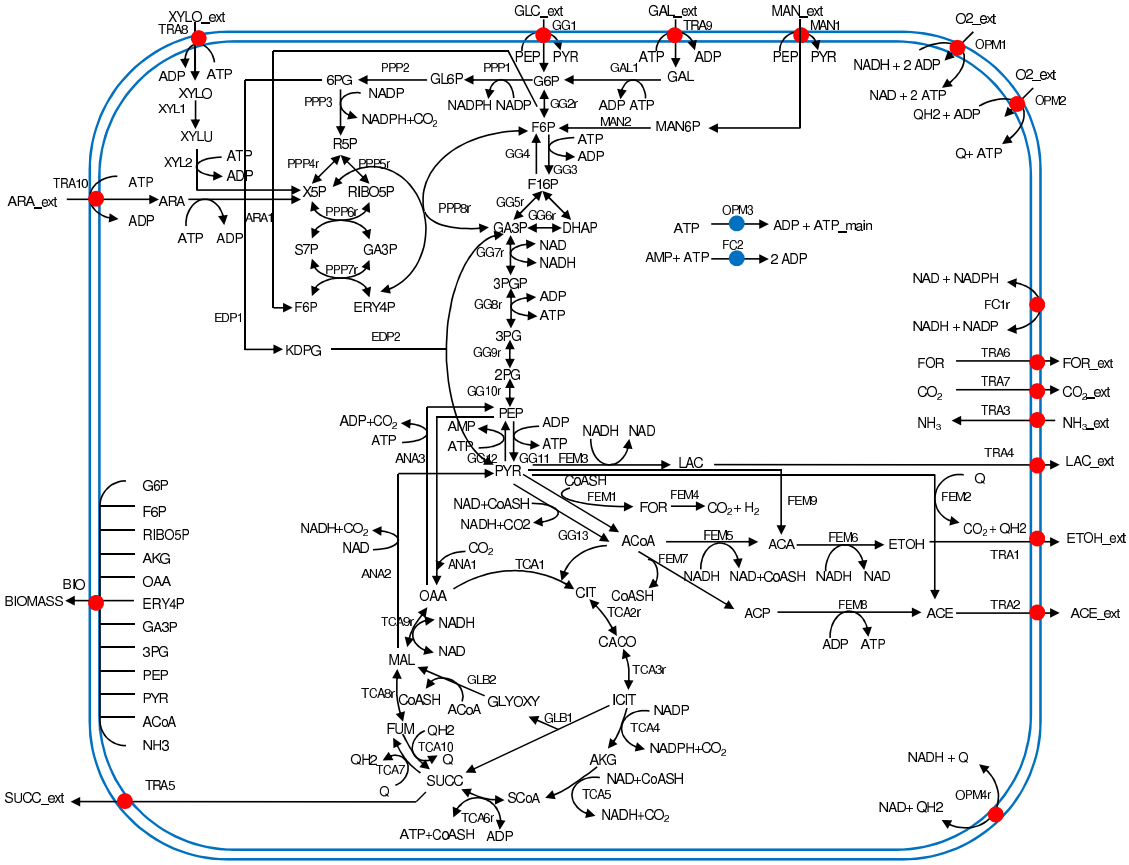

Figure 1: Metabolic map of *E. coli* central metabolism (reprinted with permission from Trinh, Unrean, Srienc [9] ©American Society for Microbiology).

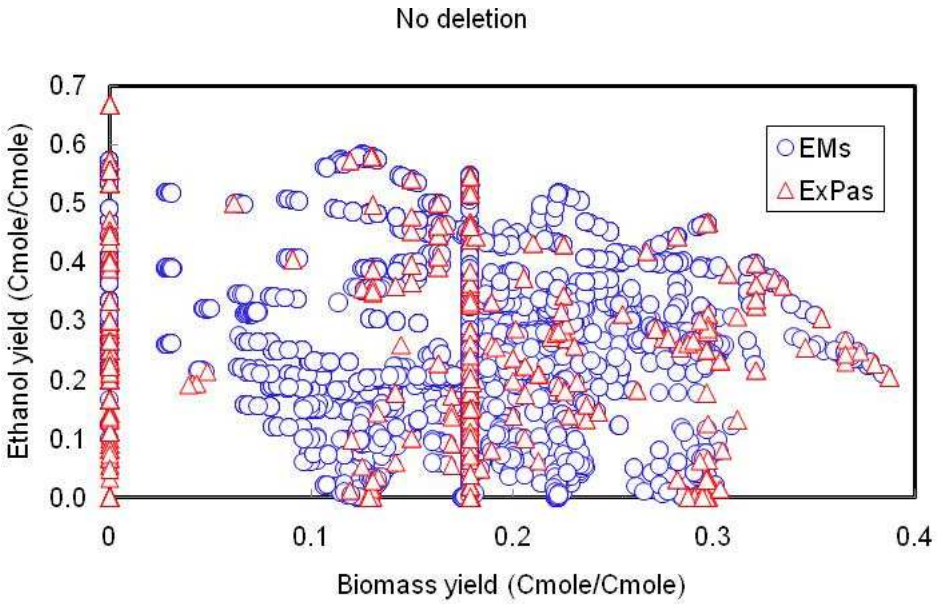

Figure 2: Relationship of ethanol and biomass yields corresponding to anaerobic elementary modes (EFMs, shown in blue circles) and extreme pathways (ExPas, shown in red triangles).

yields results that match previous computations, but also expose some discrepancies arising from this kind of modification to the network models. For example, it is observed that the set of extreme pathways can change when the network is subjected to splitting of internal reversible reactions.

Example 2 (Simple Pointed Cone)**.** We illustrate the rank test with the example from [25] in Fig. 6. We remark that in the network of [25], the elementary modes EFM1=EFM7+EMF8 and EFM2=EFM6+EFM8 each yield an overall stoichiometry of 1A = 1P, while EFM3=EFM5+EFM7 and EFM4=EFM5+EFM6 each yield an overall stoichiometry of 2A = 1P, as previously noted in [41]. To eliminate this discrepancy, we have modified reaction R8 from 1B = 1P to 2B = 1P. This change does not affect the set of reactions involved in each EFM, nor does it affect the observations we make here regarding the rank tests and extreme pathways. We observe that Klamt & Stelling [25] first split the internal reversible reaction $R7^r$ into two irreversible reactions ($R7^f$ & $R7^b$), obtaining the extreme pathways EFM3, EFM5, EFM6, EFM7, EFM8, and non-extreme elementary modes EFM1, EFM2, EFM4, plus the futile cycle $R7^f$+$R7^b$. We find that by not splitting the internal irreversible reaction, we still have a pointed cone so that it still makes sense to consider extreme pathways. With respect to this original network, our rank test indicates that EFM3 is not an extreme pathway. Indeed, one can observe that EFM3 is the sum of EFM5 and EFM7, by which the reversible reactions are cancelled. Hence the property of being an extreme vs non-extreme pathway depends very much on the specific treatment of the reversible reactions within a network. We remark that Lemma 1 indicates that, unlike the extreme pathways, the set of elementary modes is not affected by splitting the reversible reactions, other than the futile cycles involving the split reactions themselves.

Example 3 (Human red blood cell metabolism)**.** It is useful to apply the nullity test in the analysis of the Human Red Blood Cell metabolic network that has been previously analyzed using extreme pathway analysis and elementary mode analysis and is well documented in the literature [14, 35]. The published results show that there exist 6,180 EFMs and 55 ExPas (extreme pathways) [14]. It is important to note that the elementary mode analysis has been carried out with the network containing reversible reactions while the extreme pathways have been identified in a network where the internal reversible reactions have

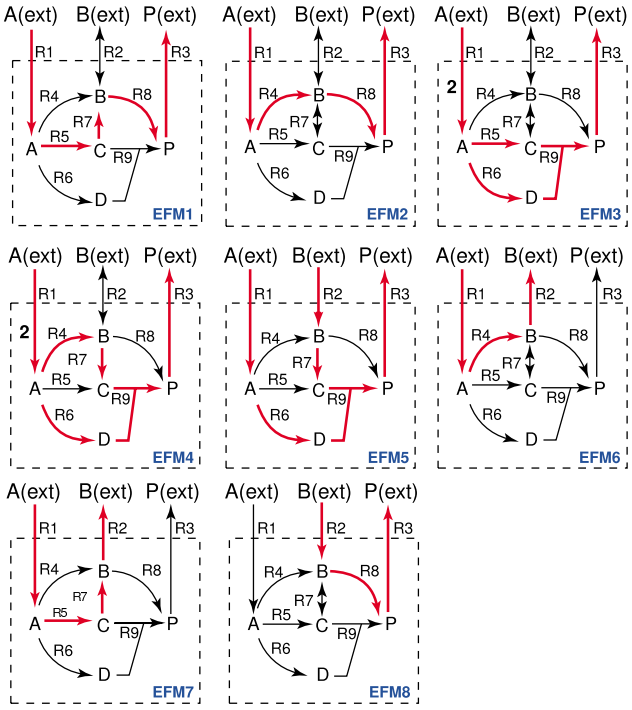

Figure 3: Elementary modes from a small network. Reprinted from Klamt & Stelling [25], ©Elsevier, used by permission.

	R1	R2^r	R3	R4	R5	R6	R7^r	R8	R9
Non-Extreme Elem. Modes									
EFM1	2	0	1	0	2	0	−2	1	0
EFM2	2	0	1	2	0	0	0	1	0
EFM3*	2	0	1	0	1	1	0	0	1
EFM4	2	0	1	1	0	1	1	0	1
Extreme Pathways									
EFM5	1	1	1	0	0	1	1	0	1
EFM6	1	−1	0	1	0	0	0	0	0
EFM7	1	−1	0	0	1	0	−1	0	0
EFM8	0	2	1	0	0	0	0	1	0

* EFM3 is extreme only when R7^r is split

Figure 4: List of elementary modes shown in Fig. 3.

	R1	R2^r	R3	R4	R5	R6	R7^r	R8	R9
internal metabolites									
A	1	0	0	−1	−1	−1	0	0	0
B	0	1	0	1	0	0	−1	*−2*	0
C	0	0	0	0	1	0	1	0	−1
D	0	0	0	0	0	1	0	0	−1
P	0	0	−1	0	0	0	0	1	1
external metabolites (not used in rank tests)									
A_ext	−1	0	0	0	0	0	0	0	0
B_ext	0	−1	0	0	0	0	0	0	0
P_ext	0	0	1	0	0	0	0	0	0

Figure 5: Stoichiometry matrix for Example shown in Fig. 3. Only the upper part corresponding to the internal metabolites is used here.

Consider the network represented by the stoichiometry matrix shown in Fig. 5, with 8 elementary modes shown in Fig. 3: where R2^r and R7^r are the only reversible reactions, and R1, R2^r, R3 are exchange fluxes. Mass balance is achieved by having reaction R8 be 2B = P (see text). To use the rank tests proposed in theorems 4, 3, we go through the following steps, ignoring the external metabolites.

1. Before using the rank test to distinguish extreme pathways, we first check that the cone is pointed by applying Theorem 3. That is, we check that there are no admissible pathways in the network by computing the nullity [dimension of the right nullspace] of the columns of Fig. 5 corresponding to the reversible reactions, R2^r & R7^r. These two columns are linearly independent, so the nullity is zero. Hence the cone is pointed.

2. We can now check each Elementary mode to see if it is an extreme pathway. For example, to check EFM6, we select the columns of Fig. 5 corresponding to the reactions with non-zero fluxes in this mode (R1, R2^r, R4) together with any remaining reversible reactions (R7^r). We find that the rank of this submatrix is 3, i.e. the nullity is 1 and hence it satisfies the conditions of Theorem 4 to be an extreme pathway. If we apply this test to every elementary mode shown in Fig. 3, we find that the extreme pathways are EFM5, EFM6, EFM7, EFM8.

Figure 6: Example illustrating how to use the rank test to distinguish extreme pathways from non-extreme elementary modes (see Example 2).

been separated into two distinct reactions operating in the opposite direction. The two approaches yield very different results and the differences are revealed when the nullity test is applied. We analyzed the identically constructed metabolic network [35] that consists of 58 metabolites (39 of which are internal) and of 51 reactions (33 of which are reversible reactions). Among the 33 reversible reactions, 17 are reversible internal reactions; 16 are reversible exchange reactions. Using Metatool, elementary mode analysis on the network that includes all reversible reactions yields 6,180 EFMs as previously reported. The nullity test applied to these elementary modes identifies only a single extreme pathway that consists only of reversible reactions. This result is obtained because the flux cone is not pointed, and corresponds to the condition $\mathsf{nullity}(N_{\mathrm{rev}}) = 1$ using the notation of Theorem 3. To perform the analysis on the exactly identical network as previously published, we have split each of the 17 reversible internal reactions into two irreversible reactions. This guarantees also that the flux cone is pointed. Elementary mode analysis on the modified network identifies 6,198 EFMs. The nullity test applied to these EFMs identifies 55 ExPas, the same number as previously published.

Inspection of the 18 additional EFMs in the network case with separated internal reversible reactions reveals that 17 of these pathways are the futile cycles consisting of the two separate reactions derived from each reversible reaction. Out of these 17 futile cycles, 16 are also extreme pathways classified as Type III extreme pathways [13]. The 18^{th} pathway is a pathway consisting only of reversible reactions matching the single extreme pathway from the unmodified network, but opposite in direction. Thus, these 18 additional EFMs will not be calculated if the internal reversible reactions are not split into two separate reactions.

These examples demonstrate that the nullity test can accurately identify ExPas from calculated EFMs, and that the obtained results are consistent with previous reports. The differences in extreme pathways identified in the two types of networks emphasizes the importance of the type of network that is subjected to the analysis. Therefore, the type of network (reversible reactions present or with reversible reactions split into two separate reactions) on which the analysis is performed should always be mentioned when the number of elementary modes or the number of extreme pathways is listed.

Conclusions

We have reviewed the connections between the classical double description method and the task of computing the elementary modes through a metabolic network, and used these connections to derive some simple new tests to distinguish elementary modes from extreme pathways, and to provide a proof and explanation for improving the efficiency of computing elementary modes. These tests make it possible to identify extreme pathways individually, and to more easily follow the effects when the network is modified by splitting reversible reactions. In examples in which it is desired to estimate how much an individual elementary mode contributes to an observed physiological flux distribution, it might be useful to know if the given mode is an extreme pathway or not, since will affect how well the same physiological flux distribution could be explained using alternate elementary modes.

It is evident from the heavy role that linear equalities and inequalities play in this development that the theories of linear programming and convex polyhedral cones are closely tied to the theory presented here. There are many analogs between some of the results presented here and techniques arising in linear programming such as the simplex method as well as convex hull problems, and these have served and can continue to foster future developments in this area.

Authors contributions

DB developed the theoretical framework. DJ derived proofs. CT validated the theoretical results. FS connected the theoretical results with the applications.

Acknowledgements

The authors would like to acknowledge the support of the National Science Foundation under grant 0534286 to DB and of the National Institute of Health under grant R01GM077529 to FS, the Minnesota Supercomputing Institute, and the University of Minnesota Rochester Biomedical Informatics and Computational Biology (BICB) Program.

References

1. Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham H, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD: **MetaCyc: a multiorganism database of metabolic pathways and enzymes**. *Nucleic Acids Research* 2006, **34**:511–516.
2. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson B: **Global reconstruction of the human metabolic network based on genomic and bibliomic data**. *Proceedings of the National Academy of Sciences of the USA* 2007, **104**(6):1777–1782.
3. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Kawashima TKS, Okuda S, Tokimatsu T, Yamanashi Y: **KEGG for linking genomes to life and the environment**. *Nucleic Acids Research* 2008, **36**:480–484.
4. Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I, Collado-Vides J, Gama-Castro S, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Bonavides-Martinez C, Ingraham J: **Multidimensional annotation of the Escherichia coli K-12 genome**. *Nucleic Acids Research* 2007, **35**(22):7577–7590.
5. Carlson R: **Metabolic systems cost-benefit analysis for interpreting network structure and regulation**. *Nucleic Acids Research* 2007, **23**(16):1258–1264.
6. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED: **Metabolic network structure determines key aspects of functionality and regulation**. *Nature* 2002, **420**(6912):190–193.
7. Schuster S, Fell D, Dandekar T: **A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks**. *Nature Biotechnology* 2000, **18**(3):326–332.
8. Trinh C, Carlson R, Wlaschin A, Srienc F: **Design, construction and performance of the most efficient biomass producing E. coli bacterium**. *Metabolic Engineering* 2006, **8**(6):628–638.
9. Trinh C, Unrean P, Srienc F: **A Minimal Escherichia coli Cell for most Efficient Ethanol Production from Hexoses and Pentoses**. *Applied and Environmental Microbiology* 2008, **74**(12):3634–3643.
10. Pfeiffer T, Sanchez-Valdenebro I, Nuno J, Montero F, Schuster S: **METATOOL: for studying metabolic networks**. *Bioinformatics* 1999, **15**(3):251–257.
11. Schuster S, Hilgetag C: **On Elementary Flux Modes in Biochemical Reaction Systems at Steady State**. *Journal of Biological Systems* 1994, **2**(2):165–182.
12. Schuster S, Hilgetag C, Woods J, Fell D: **Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism**. *Mathematical Biology* 2002, **45**(2):153–181.
13. Schilling CH, Letscher D, Palsson B: **Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective**. *Journal of Theoretical Biology* 2000, **203**(3):229–248.

14. Papin J, Stelling J, Price ND, Klamt S, Schuster S, Palsson B: **Comparison of network-based pathway analysis methods**. *Trends in Biotechnology* 2004, **22**(8):400–405.

15. Schuster S, Dandekar T, Fell D: **Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering**. *Trends in Biotechnology* 1999, **17**(2):53–60.

16. Fukuda K, Prodon A: **Double Description Method Revisited**. In *Combinatorics and Computer Science*. Edited by Deza M, Euler R, Manoussakis I, Springer 1996:91–111, [ftp://ftp.ifor.math.ethz.ch/pub/fukuda/reports/ddrev960315.p%s.gz]. [Also tech. report, Mathematics, ETH, 1995].

17. Motzkin T, Raiffa H, Thompson G, Thrall R: **The Double Description Method**. In *Contributions to theory of games*, *Volume II*. Edited by Kuhn H, Tucker A, Princeton University Press 1953:51–73.

18. Wagner C: **Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems**. *J. Phys. Chem.* 2004, **108**(7):2425–2431.

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

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

21. Bell S, Palsson B: **Expa: Program for Calculating Extreme Pathways in Biochemical Reaction Networks**. *Bioinformatics* 2005, **21**(8):1739–1740.

22. von Kamp, A, Schuster S: **Metatool 5.0: fast and flexible elementary modes analysis**. *Bioinformatics* 2006, **22**(15):1930–1931.

23. Klamt S, Stelling J, Ginkel M, Gilles E: **FluxAnalyzer: Exploring Structure, Pathways, and Flux Distributions in Metabolic Networks on Interactive Flux Maps**. *Bioinformatics* 2003, **19**(2):261–269.

24. Klamt S, Saez-Rodriguez J, Gilles ED: **Structural and Functional Analysis of Cellular Networks with CellNetAnalyzer**. *BMC Systems Biology* 2007, **1**(2).

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

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

27. Klamt S, Stelling J: **Combinatorial Complexity of Pathway Analysis in Metabolic Networks**. *Molecular Biology Reports* 2002, **29**(1-2):233–236.

28. Yeung M, Thiele I, Palsson B: **Estimation of the Number of Extreme Pathways for Metabolic Networks**. *BMC Bioinformatics* 2007, **8**(363).

29. Klamt S, Gagneur J, von Kamp A: **Algorithmic approaches for computing elementary modes in large biochemical reaction networks**. *Systems Biology, IEE Proceedings* 2005, **152**(4):249–255.

30. Samatova N, Geist A, Ostrouchov G, Melechko A: **Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways**. *Parallel and Distributed Processing Symposium., Proceedings International, IPDIPS 2002* 2002, :185–192.

31. Lee L, Varner J, Ko K: **Parallel Extreme Pathway Computation for Metabolic Networks**. In *SAS'95, Static Analysis Symposium*, *Volume 983 of* Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004). Edited by Mycroft A, Springer 2004:33–50.

32. Urbanczik R, Wagner C: **Functional Stoichiometric Analysis of Metabolic Networks**. *Bioinformatics* 2005, **21**(22):4176–4180.

33. Urbanczik R: **Enumerating Constrained Elementary Flux Vectors of Metabolic Networks**. *Systems Biology IET* 2007, **1**(5):274–279.

34. Poolman M, Sebu C, Pidcoc M, Fell D: **Modular Decomposition of Metabolic Systems via Null-space Analysis**. *Journal of Theoretical Biology* 2007.

35. Wiback S, Palsson B: **Extreme Pathway Analysis of Human Red Blood Cell Metabolism**. *Biophysics Journal* 2002, **83**(2):808–818.

36. Vijayasankaran N, Carlson R, Srienc F: **Metablic pathway structures for recombinant protein synthesis in Escherichia Coli**. *Applied Microbiology and Biotechnology* 2005, **68**(6):737–746.

37. Carlson R, Srienc F: **Fundamental Escherichia coli Biochemical Pathways for Biomass and Energy Production: Creation of Overall Flux States**. *Biotechnology and bioengineering* 2004, **86**(2):149–162.

38. Carlson R, Srienc F: **Fundamental Escherichia coli Biochemical Pathways for Biomass and Energy Production: Identification of Reactions**. *Biotechnology and bioengineering* 2004, **85**:1–19.

39. Urbanczik R: **SNA - a Toolbox for the Stoichiometric Analysis of Metabolic Networks**. *BMC Bioinformatics* 2006, **7**(129).

40. Terzer M, Stelling J: **Large Scale Computation of Elementary Flux Modes with bit pattern trees**. *Bioinformatics* 2008.

41. Bernhard O Palsson NDP, Papin JA: **Development of network-based pathway definitions: the need to analyze real metabolic networks**. *Trends in Biotechnology* 2003, **21**(5):195–198.

Figures

Figure 1 - Metabolic map of E. coli central metabolism (reprinted with permission from Trinh, Unrean, Srienc [9] ©American Society for Microbiology).

Illustration of a sample metabolic network under study.

Figure 2 - Relationship of ethanol and biomass yields corresponding to anaerobic elementary modes (EFMs, shown in blue circles) and extreme pathways (ExPas, shown in red triangles).

This figure illustrates how the trade-off between ethanol production, biomass production (plus ease of implementation) might point to the use of non-extreme pathways.

Figure 3 - Elementary modes from a small network. Reprinted from Klamt & Stelling [25], ©Elsevier, used by permission

Figure showing elementary modes from example network, used in the example to illustrate the difference between elementary modes and extreme pathways, as suggested by Reviewer 1. Permission has been obtained through the ScienceDirect web site.

List of elementary modes shown in Fig. 3.

This figure lists all the elementary flux modes for the network shown in figure 3. This is part of Figure 6.

Stoichiometry matrix for Example shown in Fig. 3. Only the upper part corresponding to the internal metabolites is used here.

This figure shows the stoichiometry matrix for the example corresponding to Example 2 in the text and pictured in figure 3. This is part of Figure 6.

Figure 6 - Example illustrating how to use the rank test to distinguish extreme pathways from non-extreme elementary modes (see Example 2).

A boxed inset showing the steps to use the new rank test for 'extremeness'. The content of this figure (including figures 3, 5 , 3) should be re-formatted to suit the final Journal format and layout for a boxed inset.

Additional Files

Additional File 1 – Appendices

Appendices: A - Proofs of Theorem; B - Algorithm Overview; C - Examples.

File name: appendices.pdf

File Format: PDF

Description:

Appendix A contains proofs of theorems stated in the main paper.

Appendix B contains an overview of Canonical Basis Algorithm and Nullspace Algorithm.

Appendix C contains two computational examples illustrating the algorithms.

Appendices

Appendix A – Proofs of Theorems.

Appendix B – Algorithms.

Appendix C – Examples.

Appendix A – Proofs of Theorems

Proof (sketch) of Theorem 1. Assume WLOG that the rows of A are ordered so that

$$\mathbf{s} = \begin{pmatrix} \mathbf{s}_1 \\ 0 \end{pmatrix} = A\mathbf{x} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \mathbf{x},$$

where A_1, A_2 are $(d_1 \times q), (d_2 \times q)$, resp., and $\mathbf{s}_1 > 0$.

If the rank of A_2 is $q-1$, then its nullity is 1, and the vector $\mathbf{x}$ such that $A_2\mathbf{x} = 0$ is unique up to a scale factor. If $\mathbf{x}$ were a non-negative combination of two other valid flux vectors: $\mathbf{x} = \alpha\tilde{\mathbf{x}} + \beta\tilde{\tilde{\mathbf{x}}}$, with $\alpha, \beta \geq 0$, then $\mathbf{s} = \alpha\tilde{\mathbf{s}} + \beta\tilde{\tilde{\mathbf{s}}}$. Because $\mathbf{s}, \alpha, \tilde{\mathbf{s}}, \beta, \tilde{\tilde{\mathbf{s}}}$ are all non-negative, the parts $\tilde{\mathbf{s}}_2$, $\tilde{\tilde{\mathbf{s}}}_2$ must be zero. Because $\mathsf{nullity}(A_2) = 1$, that would mean $\tilde{\mathbf{x}}$, $\tilde{\tilde{\mathbf{x}}}$ are both multiples of $\mathbf{x}$. Hence, in order for $\mathbf{x}$ to be a non-negative combination of columns of a minimal R of a double description pair (A, R), all of which are valid flux vectors, $\mathbf{x}$ (or a multiple thereof) must appear as one of the columns in R.

Conversely, if (A, R) is a minimal double description pair, and $\mathbf{x}$ is a column of R (or a non-negative multiple thereof), then by definition of minimality, $\mathbf{x}$ cannot be a non-negative combination of any different valid flux vectors, which are themselves non-negative combinations of columns of R. If nullity of A_2 is bigger than one, then there is more than 1 vector satisfying the steady state condition. Suppose there were a different vector $\tilde{\mathbf{x}}$ not a multiple of $\mathbf{x}$ such that $A_2\tilde{\mathbf{x}} = 0$. We can assume $A_1\tilde{\mathbf{x}} > 0$, since we can add a sufficiently large positive multiple of $\mathbf{x}$ to $\tilde{\mathbf{x}}$ if necessary. Then one can find a combination $\tilde{\tilde{\mathbf{s}}} = A\mathbf{x} + \tilde{\alpha}A\tilde{\mathbf{x}}$ such that $\tilde{\tilde{\mathbf{s}}}_2 = 0$ and some element of $\tilde{\tilde{\mathbf{s}}}_1 \geq 0$ is zero. Since $A\mathbf{x} \geq 0$ and $A\tilde{\mathbf{x}} \geq 0$, it must be that $\tilde{\alpha} < 0$ Hence we have that $\mathbf{x}$ is a non-negative combination of $\tilde{\mathbf{x}}$ and $\tilde{\tilde{\mathbf{x}}} = \mathbf{x} + \tilde{\alpha}\tilde{\mathbf{x}}$. Thus $\mathbf{x}$ could not be present in a minimal R. Hence no such $\tilde{\mathbf{x}}$ can exist, and the nullity of A_2 is 1. □

Proof of Theorem 2. To simplify the notation, we locally use $\overline{\mathsf{Z}}$ as a shorthand for $\overline{\mathsf{Z}}(\mathbf{x})$.

Suppose that that $\mathbf{x}$ is not an elementary mode. Then $\mathbf{x}$ is a non-negative combination of two other different valid admissible pathways: $\mathbf{x} = \alpha_1\mathbf{x}^{(1)} + \alpha_2\mathbf{x}^{(2)}$ with $\alpha_1, \alpha_2 > 0$, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \geq 0$. The positions of the non-zero entries in $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ must be subsets of the positions $\overline{\mathsf{Z}}$. Hence $\mathbf{x}^{(1)}_{\overline{\mathsf{Z}}}$, $\mathbf{x}^{(2)}_{\overline{\mathsf{Z}}}$ must both be in the nullspace of $N_{*,\overline{\mathsf{Z}}}$ as well, and hence the right nullity of $N_{*,\overline{\mathsf{Z}}}$ must be at least 2.

On the other hand, if the nullity of $N_{*,\overline{\mathsf{Z}}}$ is at least 2, then $\mathbf{x}_{\overline{\mathsf{Z}}}$ must be the linear combination of two other vectors in the nullspace, and by adding suitable multiples of $\mathbf{x}_{\overline{\mathsf{Z}}}$ to these other vectors, we can ensure that these two other vectors are non-negative, and hence $\mathbf{x}_{\overline{\mathsf{Z}}}$ is not elementary.

Let $\overline{\mathsf{Z}} = \overline{\mathsf{Z}}(\mathbf{x})$ be the set of indices of the nonzero elements of an admissible flux vector $\mathbf{x}$, and let $x_{\overline{\mathsf{Z}}}$ the vector of those nonzero entries. Assume without loss of generality that all irreversible reactions are listed before any reversible reactions. Partition

$$x_{\overline{\mathsf{Z}}} = \begin{bmatrix} x_{\overline{\mathsf{Z}}_1} \\ x_{\overline{\mathsf{Z}}_2} \end{bmatrix}$$

where $\overline{\mathsf{Z}}_1$ denotes the indices of all the irreversible reactions. The cardinality of $\overline{\mathsf{Z}}$, denoted $|\overline{\mathsf{Z}}|$, is the number of non-zeros in $\mathbf{x}$, often called the *fill* of the vector. Let $N_{*,\overline{\mathsf{Z}}}$ denote the matrix of columns of N corresponding to the non-zeros in $\mathbf{x}$. This matrix is $m \times |\overline{\mathsf{Z}}|$.

Since $x_{\overline{\mathsf{Z}}}$ consists entirely of nonzero entries, we have that $x_{\overline{\mathsf{Z}}_1} > 0$ and that $N_{*,\overline{\mathsf{Z}}} \cdot \mathbf{x}_{\overline{\mathsf{Z}}} = 0$,

If $\mathsf{nullity}(N_{*,\overline{\mathsf{Z}}}) = 1$ then $\mathbf{x}_{\overline{\mathsf{Z}}}$ is the only possible vector such that $N_{*,\overline{\mathsf{Z}}} \cdot \mathbf{x}_{\overline{\mathsf{Z}}} = 0$, up to scalar multiples.

If $\mathsf{nullity}(N_{*,\overline{\mathsf{Z}}}) > 1$ then there is another non-zero $|\overline{\mathsf{Z}}|$-vector $\mathbf{v}$, not a multiple of $\mathbf{x}_{\overline{\mathsf{Z}}}$, such that $N_{*,\overline{\mathsf{Z}}} \cdot \mathbf{v} = 0$. Any linear combination $\mathbf{y} = \mathbf{x}_{\overline{\mathsf{Z}}} + c\mathbf{v}$ with $\mathbf{y}_1 = \mathbf{x}_{\overline{\mathsf{Z}}_1} + c\mathbf{v}_1 \geq 0$ also satisfies $N_{*,\overline{\mathsf{Z}}} \cdot \mathbf{y} = 0$ for any c. We choose $c \neq 0$ with smallest $|c|$ to introduce a zero element into the vector $\mathbf{y}$ while maintaining $\mathbf{y}_1 \geq 0$. The resulting $\mathbf{y}$ is an annihilating vector for $N_{*,\overline{\mathsf{Z}}}$ with fewer non-zeros than $\mathbf{x}_{\overline{\mathsf{Z}}}$. Hence $\mathbf{x}$ is not "elementary" according to Definition 1. □

Proof of Theorem 5. It is assumed that the stoichiometric matrix N has columns ordered so that first columns correspond to irreversible, and last columns correspond to reversible reactions. Idea for the proof of this proposition is to show that the processing of all rows corresponding to irreversible reactions corresponds to the execution of the double description method for the appropriately constructed matrix A. Suppose the

matrix N is decomposed as $N = (N_1,\, N_2)$ where N_2 is $m \times m$ non-singular matrix which includes all the reversible reactions. Then N may be further decomposed to account for reversible and irreversible reactions as $N = (N_1,\, N_{2,\mathrm{irr}},\, N_{2,\mathrm{rev}})$. Further, nullspace matrix of the Nullspace Algorithm is given as

$$X^{(1)} = \begin{pmatrix} I_{q-m} \\ -N_2^{-1}N_1 \end{pmatrix} = \begin{pmatrix} I \\ M_1 \\ M_2 \end{pmatrix}, \tag{11}$$

M_1 and M_2 being of size $(m - q_{\mathrm{rev}}) \times (q - m)$ and $q_{\mathrm{rev}} \times (q - m)$, respectively. We will assign the initial nullspace matrix of the Nullspace Algorithm to form the initial pair of the double description method whose A matrix is equal to:

$$A = \begin{pmatrix} N \\ -N \\ E \end{pmatrix} = \begin{matrix} \\ (m)\,\{ \\ (m)\,\{ \\ (n-q_{\mathrm{rev}})\,\{ \end{matrix} \begin{pmatrix} \overbrace{N_1}^{(n-m)} & \overbrace{N_{2,\mathrm{irr}}}^{(m-q_{\mathrm{rev}})} & \overbrace{N_{2,rev}}^{q_{\mathrm{rev}}} \\ -N_1 & -N_{2,\mathrm{irr}} & -N_{2,\mathrm{rev}} \\ I_{(\mathrm{n-m})} & 0 & 0 \\ 0 & I_{(\mathrm{m-q_{rev}})} & 0 \end{pmatrix} \tag{12}$$

The double description method will inspect which rows of the product $AX^{(1)}$ have negative values and make combinations of the columns with positive and negative elements in the processed row to form linear combinations and either adopt or reject them as candidates for extreme pathways. Since the aforementioned product $AX^{(1)}$ is equal to:

$$AX^{(1)} = \begin{pmatrix} 0 \\ 0 \\ I_{(\mathrm{n-m})} \\ M_1 \end{pmatrix}, \tag{13}$$

we may conclude that product $AX^{(1)}$ has to process submatrix M_1 identical to the submatrix which will be processed in the nullspace matrix of the nullspace algorithm, when only rows corresponding to irreversible reactions are treated. Both algorithms will form the same set of linear combinations of columns. □

Proof of Lemma 1. We show the proof as in [19] which demonstrates the existence of bijective mapping between the set of elementary modes obtained from the original matrix N with $q - q_{\mathrm{rev}}$ irreversible and q_{rev} reversible reactions and the reconfigured matrix N' with $q + q_{\mathrm{rev}}$ irreversible reactions obtained by splitting reversible reactions into two components. Let elementary mode corresponding to the original matrix have the form $x = \begin{pmatrix} x_{rev} \\ x_{irr} \end{pmatrix}$ and those of the reconfigured matrix the form $x' = \begin{pmatrix} x_{revp} \\ x_{irr} \\ x_{revn} \end{pmatrix}$. If a function $\phi : R^q \to R^{q+q_{\mathrm{rev}}}$ is defined as $x' = \phi(x) = \phi \begin{pmatrix} x_{rev} \\ x_{irr} \end{pmatrix} = \begin{pmatrix} pos(x_{rev}) \\ pos(x_{irr}) \\ pos(-x_{rev}) \end{pmatrix}$ and $pos(x) = \frac{x+|x|}{2}$ then elementary modes of the original matrix may serve to construct elementary mode of the reconfigured matrix. On the other hand, function $\psi : R^{q+q_{\mathrm{rev}}} \to R^q$ defined as $x = \psi \begin{pmatrix} x_{rev} \\ x_{irr} \end{pmatrix} = \begin{pmatrix} x_{revp} - x_{revn} \\ x_{irr} \end{pmatrix}$ maps every elementary mode corresponding to the reconfigured matrix into an elementary mode of the original matrix, unless the elementary mode of the reconfigured network is a spurious cycle.

□

Proof of Theorem 7. The reduced rank test depends on using the reduced row echelon form of the stoichiometry matrix, which has the form $N_{rref} = \begin{pmatrix} N_1 & I \end{pmatrix}$, and is obtained by suitable row operations on N. We assume without loss of generality that N_{rref} is $m \times q$ (by removing redundant rows in advance if necessary). The reduced row echelon, at the stage k of the Nullspace Algorithm, can be further decomposed to:

$$N_{\mathrm{rref}} = \begin{pmatrix} N_1 & I \end{pmatrix} = \begin{matrix} \\ (k)\,\{ \\ (m-k)\,\{ \end{matrix} \begin{pmatrix} \overbrace{P}^{(q-m)} & \overbrace{I_{\mathrm{k}}}^{k} & 0 \\ Q & 0 & I_{\mathrm{m-k}} \end{pmatrix}. \tag{14}$$

As mentioned in remark 2 we must select all the columns of the stoichiometric matrix whose indices correspond to nonzero elements among $x_1, \ldots, x_{q-m+k}$ at stage k and all the columns corresponding to indices

$q-m+k+1, ...q$. In the (14) this would correspond to selecting all last $m-k$ columns which contain in its lower part matrix I_{m-k}. To compute the rank of the submatrix obtained in this way we have:

$$\begin{aligned}
\mathsf{rank}\begin{pmatrix} P_{*,\overline{\mathtt{Z}}_\mathrm{a}} & I_{\mathrm{k},\overline{\mathtt{Z}}_\mathrm{b}} & 0 \\ Q_{*,\overline{\mathtt{Z}}_\mathrm{a}} & 0 & I_{\mathrm{m-k}} \end{pmatrix} &= \mathsf{rank}\begin{pmatrix} P_{*,\overline{\mathtt{Z}}_\mathrm{a}} & I_{\mathrm{k},\overline{\mathtt{Z}}_\mathrm{b}} & 0 \\ 0 & 0 & I_{\mathrm{m-k}} \end{pmatrix} \\
&= \mathsf{rank}\begin{pmatrix} P_{*,\overline{\mathtt{Z}}_\mathrm{a}} & I_{\mathrm{k},\overline{\mathtt{Z}}_\mathrm{b}} & 0 \\ 0 & 0 & I_{\mathrm{m-k}} \end{pmatrix} \\
&= \mathsf{rank}(P_{*,\overline{\mathtt{Z}}_\mathrm{a}},\, I_{\mathrm{k},\overline{\mathtt{Z}}_\mathrm{b}}) + \mathsf{rank}(I_{\mathrm{m-k}}) \\
&= \mathsf{rank}(P_{\mathtt{Z}_\mathrm{b},\overline{\mathtt{Z}}_\mathrm{a}}) + \mathsf{rank}(I_{\mathrm{k},\overline{\mathtt{Z}}_\mathrm{b}}) + \mathsf{rank}(I_{\mathrm{m-k}})
\end{aligned} \tag{15}$$

and x is an elementary mode if

$$\mathsf{rank}(P_{\mathtt{Z}_\mathrm{b},\overline{\mathtt{Z}}_\mathrm{a}}) = \left|\overline{\mathtt{Z}}_\mathrm{a}\right| - 1 \tag{16}$$

or expressed in terms of nullity of the matrix

$$\mathsf{nullity}(P_{\mathtt{Z}_\mathrm{b},\overline{\mathtt{Z}}_\mathrm{a}}) = 1. \tag{17}$$

The result from (17) shows that at k^{th} iteration it is sufficient to check the nullity of the submatrix obtained from matrix N_{rref} when columns corresponding to indices of non-zero elements in $a = [x_1, \ldots, x_{q-m}]$ and rows corresponding to indices of zero elements in vector $b = [x_{q-m+1}, \ldots, x_{q-m+k}]$ are selected among the first k rows. □

Appendix B – Algorithms

Having discussed the theory relating the stoichiometry problem to the general problem of double description pairs, we now turn to the algorithms for computing a double description pair and the related algorithms for computing all the elementary and/or extreme pathways for a stoichiometry problem.

General Double Description Method. The algorithm to compute an R forming a Double Description pair [DD pair] (1), starting with the matrix A, proceeds in a recursive manner. Let A_k denote the matrix consisting of the first k rows of A, where the ordering of the rows is arbitrary. Suppose A_k, R_k form a DD pair. The recursive process then proceeds to construct a DD pair A_{k+1}, R_{k+1}, where A_{k+1} is formed by appending the $k+1$-st row of A to A_k, and R_{k+1} is formed by taking all possible valid non-negative combinations of columns of R_k. The heart of the recursive algorithm consists of specifying the details of how R_{k+1} is constructed from R_k, proving that the result indeed forms a DD pair with A_{k+1}, and finding a proper way to initialize the algorithm.

Lemma 2 (DD Lemma [16])**.** *Let A_k denote the matrix consisting of the first k rows of the $d \times q$ matrix A. Any extreme ray with respect to A_{k+1} is a non-negative combination of at most two extreme ray with respect to A_k.*

Proof (sketch). Let $\mathbf{s}$ be an extreme ray *wrt* A_{k+1}. If it is already an extreme ray *wrt* A_k, we are trivially done. Let $\mathtt{Z}$ denote the indices of the zero entries in $A_{k+1} \cdot \mathbf{s}$, and let $A_{k+1,\,\mathtt{Z}}$, $A_{k,\,\mathtt{Z}}$ denote the rows of A_{k+1}, A_k corresponding to these zero entries, respectively. If $[A \cdot \mathbf{s}]_{k+1} \neq 0$ (i.e. the $k+1$-st entry of $A \cdot \mathbf{s}$ is not zero), then $A_{k+1,\,\mathtt{Z}} = A_{k,\,\mathtt{Z}}$, implying $\mathbf{s}$ is extreme ray *wrt* A_k and hence we would be done. So we are left to consider the case where $[A \cdot \mathbf{s}]_{k+1} = 0$, and $q - 1 = \mathsf{rank}(A_{k+1,\,\mathtt{Z}}) > \mathsf{rank}(A_{k,\,\mathtt{Z}}) = q - 2$. This also implies that the $k+1$-st row of A cannot be expressed as a linear combination of the rows of $A_{k,\,\mathtt{Z}}$.

Since $\mathbf{s}$ is a valid extreme ray *wrt* A_{k+1}, we have that $A_k \cdot \mathbf{s} \geq 0$, and $\mathbf{s}$ must be a non-negative combination of extreme ray *wrt* A_k: $\mathbf{s} = \alpha_1 \mathbf{x}_1 + \cdots \alpha_j \mathbf{x}_j$ where $\alpha_i > 0$, $\mathbf{x}_i$ extreme *wrt* A_k, for $i = 1, \ldots, j$.

It remains to show that $\mathbf{s}$ is the non-negative combination of exactly 2 rays that are extreme *wrt* A_k. Let $\mathtt{Z}_1$ denote the indices corresponding to zeros in $A_k \mathbf{x}_1$. Then select the rows of A_k indexed by $\mathtt{Z}_1$ to form $A_{k,\,\mathtt{Z}_1}$. This matrix has rank $q-1$ (since $\mathbf{x}_1$ is extreme *wrt* A_k).

Find the positive multiple β of $\mathbf{x}_1$ such that $A_k \cdot [\mathbf{s} - \beta \mathbf{x}_1]$ has at least one more zero entry (in position ν) compared to those of $A_k \cdot \mathbf{s}$, but maintaining $A_k \cdot [\mathbf{s} - \beta \mathbf{x}_1] \geq 0$. Let $\mathtt{Z}_\beta$ denote the indices of the zero entries in $A_k \cdot [\mathbf{s} - \beta \mathbf{x}_1]$. We have that $\mathtt{Z} \backslash \{k+1\} \subset \mathtt{Z}_\beta$, but the index ν is in $\mathtt{Z}_\beta$, but not in $\mathtt{Z}$. Since $[A \cdot \mathbf{s}]_\nu \neq 0$, but $[A_{k,\,\mathtt{Z}} \cdot \mathbf{s}] = 0$, the ν-th row of A cannot be a linear combination of the rows of $A_{k,\,\mathtt{Z}}$.

What is $\mathsf{rank}(A_{k,\mathsf{Z}_\beta})$? It must be no more than $q-1$ because its nullspace contains at least one non-null vector, namely $[\mathbf{s}-\beta\mathbf{x}_1]$. It must be at least 1 more than $\mathsf{rank}(A_{k,\mathsf{Z}})$ because A_{k,Z_β} includes all the rows in $\mathsf{rank}(A_{k,\mathsf{Z}})$ plus at least row ν not a linear combination of the rows of $A_{k,\mathsf{Z}}$. Since this latter matrix has rank $q-2$, the rank of A_{k,Z_β} must be exactly $q-1$, and hence $[\mathbf{s}-\beta\mathbf{x}_1]$ must be an extreme ray *wrt* A_k. So we have shown that $\mathbf{s}$ can be written as a non-negative combination of two extreme rays *wrt* A_k: $\mathbf{s} = \beta\mathbf{x}_1 + [\mathbf{s}-\beta\mathbf{x}_1]$.

We end with the remark that if $[A\cdot\mathbf{x}_1]_{k+1}$ were 0, then so would $[A\cdot[\mathbf{s}-\beta\mathbf{x}_1]]_{k+1}$, but then both $\mathbf{x}_1$ and $[\mathbf{s}-\beta\mathbf{x}_1]$ would be extreme *wrt* A_{k+1}. Hence these entries cannot be zero. □

Naive DD Algorithm. Suppose we have an initial DD pair (A_{k_0}, R_{k_0}), for some initial value of k_0, the DD Lemma gives a way to compute a DD pair for the entire matrix A.
For $k = k_0, \ldots, d-1$,

1. Form A_{k+1} appending the $k+1$-st row of A to A_k.
2. For every pair of columns $\mathbf{x}_1$, $\mathbf{x}_2$ of R_k (extreme rays *wrt* A_k), form a non-negative combination $\mathbf{s} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2$ such that $[A\cdot\mathbf{s}]_{k+1} = 0$. This is possible exactly when $[A\cdot\mathbf{x}_1]_{k+1}$ and $[A\cdot\mathbf{x}_2]_{k+1}$ are both nonzero and have opposite signs (see remark at the end of the proof of the DD Lemma).
3. Check that $\mathbf{s}$ is extreme *wrt* A_{k+1}, either by checking that $\mathsf{rank}(A_{k+1,\mathsf{Z}(\mathbf{s})}) = q-1$, or by checking that the set of indices $\mathsf{Z}(\mathbf{s})$ (set of indices of the zero entries in $A_{k+1}\cdot\mathbf{s}$) is not a subset of the corresponding set of zero indices for any existing column of R_k. Discard any vectors $\mathbf{s}$ failing this test.
4. Every column $\mathbf{x}$ of R_k such that $A_{k+1}\mathbf{x} \geq 0$ is already an extreme ray *wrt* A_{k+1}. So collect all columns $\mathbf{x}$ satisfying this condition, together with all vectors $\mathbf{s}$ found to be extreme *wrt* A_{k+1} in the previous step, to form R_{k+1}. The resulting R_{k+1} forms a DD pair with A_{k+1}. (Actually, if any q rows of A are linearly independent, then $A_{k+1}\mathbf{x} = 0$ cannot occur.)

At the end of this iteration, R_d will form a DD pair *wrt* A. It remains to figure out how to initialize the iteration. In the special case of (2), it is easy to construct an initial DD pair in two possible ways.

1. In the case the E block of (2) is a complete $q \times q$ identity matrix (i.e., all reactions are irreversible), then we can select the rows corresponding to the E in (2) by using the $R = I_{q\times q}$. The initial DD pair is $(E, I_{q\times q})$, and the recursive steps are used to enforce the remaining conditions $N\mathbf{x} \geq 0$, $-N\mathbf{x} \geq 0$. During the recursive algorithm, we add each row of N and the same row from $-N$ together to enforce the condition $N\mathbf{x} = 0$ directly. The resulting algorithm is the *Canonical Basis Algorithm* [12].
2. We can let the initial R be a basis for the nullspace of N, such that the first $q-m$ rows of R form an identity matrix. This is equivalent to setting the initial A_{k_0} to consist of all of the N and $-N$ parts of (2) plus the first $q-m$ rows of the E part. The recursive steps are used to enforce the non-negativity conditions *wrt* the remaining rows of E. This initialization works even when E is not a complete identity matrix, as long as E contains at least $q-m$ rows from the complete identity matrix, and the initial rank condition is still satisfied:
$$\mathsf{rank}(A_{k_0}) = \mathsf{rank}\begin{pmatrix} N \\ -N \\ I_{q-m\times q-m},\ 0 \end{pmatrix} = \mathsf{rank}\begin{pmatrix} N_1 & N_2 \\ -N_1 & -N_2 \\ I_{q-m\times q-m} & 0 \end{pmatrix} = q.$$
This is equivalent to the condition that the $m\times m$ submatrix N_2 is non-singular. The reactions (columns of N) may need to be permuted to meet this conditions. In terms of the stoichiometry, this works even if there are some reversible reactions, as long as there are at most m reversible reactions, and one can find m reactions, which must include all the reversible reactions, such that the stoichiometry *wrt* those reactions is "linearly independent." This leads to the *Nullspace Algorithm* [18, 19].

An overview of these two algorithms are given below.

Canonical Basis Algorithm [7]. Let the algorithm be illustrated on a simple one-metabolite example depicted in Figure 7 with a stoichiometry matrix $N = (1, -1, -1)$, representing a single metabolite M_1 and three reactions, R_1, R_2, R_3. Reaction R_1 produces the metabolite, while the other two consume it. We assume only R_3 is reversible.

In this algorithm, we start by enforcing the sign constraints $\mathbf{x} \geq 0$ at stages which correspond to irreversible reactions, and impose the constraints $N\mathbf{x} = 0$ one by one.

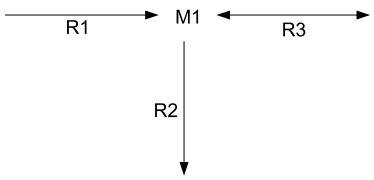

Figure 7: One-metabolite example.

1. Construct an initial basis guaranteed to encompass the answer, namely the columns of I. Call the initial basis $X^{(1)} = I$, and form the initial tableau

$$\mathbf{T}^{(1)} = \begin{pmatrix} NX^{(1)} \\ X^{(1)} \end{pmatrix}.$$

   As the columns of X are combined, the columns in top part of this tableau are combined the same way, and hence will always show NX. The goal is to find an X such that $NX = 0$ while maintaining $X \geq 0$.

   Initialize the iteration counter $k = 1$. Columns of N are sorted so that columns corresponding to irreversible reactions precede those columns corresponding to reversible reactions [12]:

$$T^{(1)} = \left(N^{(1)}_{IRREV}, N^{(1)}_{REV} \right).$$

   In the case of the one-metabolite example, the initial tableau is

$$\mathbf{T}^{(1)} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2. Columns with zero elements in k^{th} row of the matrix $\mathbf{T}^{(k)}$, are added to the empty matrix $\mathbf{T}^{(k+1)}$ In each iteration k, among the remaining columns in $\mathbf{T}^{(k)}$, three cases are considered to form linear combinations. First, all pairs of columns from column group $N^{(k)}_{REV}$ are combined between each other to form a linear combination which has a zero value in k^{th} row, and such are added to $N^{(k+1)}_{REV}$. Second, pair of columns from $N^{(k)}_{IRREV}$ with elements of opposite signs in the k^{th} row are combined to give the column with a zero element in k^{th} row of $\mathbf{T}^{(k)}$ and such are added to the block of columns in $N^{(k+1)}_{IRREV}$. Third, one column from $N^{(k)}_{REV}$ is combined with one column from $N^{(k)}_{IRREV}$ to form a linear combination with a zero element in k^{th} row, and the same is added to the $N^{(k+1)}_{IRREV}$ part of the matrix $T^{(k+1)}$. If columns i, j form one such pair, the new flux vector (lower part of tableau) will be

$$\tilde{\mathbf{x}} = \alpha \mathbf{x}^k_i + \beta \mathbf{x}^k_j,$$

   and the upper part of the tableau will be

$$N\tilde{\mathbf{x}} = \alpha N\mathbf{x}^k_i + \beta N\mathbf{x}^k_j,$$

   where $\alpha, \beta \geq 0$ are chosen so that the k-th entry in $N\tilde{\mathbf{x}}$ is zero.

   In the one-metabolite example, in the first and only round, the possible combinations yielding zeros in the first entry are (where in all cases $\alpha = \beta = 1$)

$$\begin{array}{rcccc} (i,j) & : & (1,2) & (1,3) & (2,3) \\ N\tilde{\mathbf{x}} & : & 0 & 0 & 0 \\ \tilde{\mathbf{x}} & : & \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \end{array}$$

3. Check that each prospective flux vector $\tilde{\mathbf{x}}$ generated in the previous step is "elementary," either by checking that no other flux vector has non-zero entries only in positions corresponding to the nonzeros in $\tilde{\mathbf{x}}$, or by checking the nullity condition. Eliminate all vectors that fail this test. Collect all the remaining new flux vectors into $X^{(k+1)}$.

   The new tableau has the final form of $\mathbf{T}^{(k+1)} = \begin{pmatrix} NX^{(k+1)} \\ X^{(k+1)} \end{pmatrix}$. Note that the first k rows of $NX^{(k+1)}$ are now entirely zero.

   In the simple one-metabolite example no vector is eliminated this way.

4. If $k < m$, increment $k \leftarrow k + 1$ and repeat from step 2.

5. After m steps, the set of columns in the lower part of the matrix, occupying last q rows represents the set of elementary modes.

   Alternatively, the algorithm could have been executed if we split every reversible reaction into two irreversible, expanding the size of the matrix $T^{(k)}$. Upon computing the set of elementary modes according to the reconfigured approach, it would additionally require the mapping to the original set of elementary modes using the Lemma 1.

Nullspace Algorithm [18]. Here we start with all possible vectors satisfying $N\mathbf{x} = 0$ and enforce one by one the non-negativity constraints in iterations which correspond to irreversible reactions.

1. Start with an initial set of rate vectors which form a basis for the nullspace of N, such that as many entries are non-negative as possible. Specifically, assume WLOG that the last m columns of N are linearly independent (permute the columns if necessary). These last m columns must include all the reversible reactions (if any), since we are in effect enforcing non-negativity conditions on the rates for the first $q - m$ reactions. Split $N = (N_1, N_2)$, where N_2 is $m \times m$ non-singular. Set

$$X^{(1)} = \begin{pmatrix} I_{q-m} \\ -N_2^{-1}N_1 \end{pmatrix},$$

   where I_{q-m} denotes a $(q-m) \times (q-m)$ identity matrix and N_2^{-1} is the matrix inverse of N_2.

   Initialize the iteration counter $k = 1$.

   In the simple one-metabolite example, $q = 3$, $m = 1$, $N_1 = [1 - 1]$, $N_2 = [-1]$, $N_2^{-1} = [-1]$ and the initial set of vectors is

$$X^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}.$$

   The goal is now to form pair-wise non-negative combinations of the columns to eliminate the negative entries in $-N_2^{-1}N_1$, the lower part of $X^{(1)}$. We do this row by row starting with the first row of $-N_2^{-1}N_1$, which is the $q - m + 1$-th row of $X^{(1)}$.

2. Form non-negative combinations of all possible pairs of columns of $X^{(k)}$ which yield zeros in the $q - m + k$-th row of $X^{(k)}$. Eliminate any that fail the test of being elementary. This test may require a scan of all the other columns to check the locations of the non-zeros, in a manner similar to the Canonical Basis Algorithm.

   In the simple one-metabolite example, the only pair of columns that can be combined to introduce zeros in position $q - m + 1 = 3$ is $(1, 2)$, yielding

$$\widetilde{X} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

   This mode satisfies the "elementariness" test.

3. If $q-m+k$-th reaction is irreversible reject those columns which have negative element, otherwise add Collect all the old columns in $X^{(k)}$ whose $q-m+k$-th entry is already non-negative, together with all the new columns that pass the test in the previous step. These columns form $X^{(k+1)}$.

   In the simple one-metabolite example, the first and second column from the original $X^{(1)}$ are carried over, together with the new column just created. The result is

$$X^{(2)} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

   This set of three elementary modes is the final result for this simple example.

4. If $k < m$, increment $k \leftarrow k+1$ and repeat from step 2.

5. After m iterations we obtain complete set of elementary modes.

Similar to the Canonical Basis Algorithm, the reversible reactions could be split into two irreversibles expanding the initial nullspace matrix by q_{rev} more columns and rows. Upon completion of the computation of the elementary modes, the same will have to be reverted back into the original version, using the mapping from the Lemma 1.

So we can think of this algorithm as starting with a set of valid elementary flux modes pretending that only $\rho_1, \ldots, \rho_{q-m}$ are irreversible, and imposing irreversibility on the remaining reactions $\rho_{q-m+1}, \ldots, \rho_q$ one at a time. If the reactions $\rho_{q-q_{\text{rev}}+1}, \ldots, \rho_q$ really are reversible, while all the earlier reactions are irreversible, we may stop the Nullspace Algorithm at stage $k = m - q_{\text{rev}}$, and we will be left with all the extreme pathways for this Stoichiometry matrix N. Algorithms for computation of elementary modes and extreme pathways are applied on the original matrix, since its reconfiguration would result in bigger matrix and more processing time. However, matrix reconfiguration is used to prove that algorithm finds all the elementary modes and extreme pathways of the metabolic network [19].

Appendix C – Examples

Example 4 (A One-Metabolite Example)**.** Refering to the one-metabolite example used in the description of Canonical Basis and Nullspace Algorithms the set of computed elementary modes is

$$X = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}. \tag{18}$$

Notice that the last column is the sum of the first two, so it is not an extreme pathway, but it is still elementary according to both Definition 1 and Theorem 2.

A corresponding double description pair consists of

$$A = \begin{pmatrix} N \\ -N \\ E \end{pmatrix} = \left(\begin{array}{ccc} 1 & -1 & -1 \\ \hline -1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right), \quad R = X.$$

The matrix X does not form a *minimal* double description pair. We find that

$$S = AX = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Applying the Extreme Ray Theorem 1 and taking each column in turn, we extract the rows of A corresponding the zero entries in that column:

$$A^{(1)} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}, A^{(2)} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, A^{(3)} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \end{pmatrix}.$$

We notice that the ranks of these three matrices are 2, 2, 1, and the nullities are 1, 1, 2, respectively. Hence only the first two columns satisfy the conditions to be extreme pathways and hence form a *minimal* double description pair. This connection between the stoichiometry problem and the general theory of double description pairs leads to a simple test to distinguish an extreme pathway from an elementary mode.

Example 5. Let N be the stoichiometric matrix representing a metabolic network with 3 metabolites m_1, m_2, m_3 and 6 reactions R_i, $1 \leq i \leq 6$. The network has four irreversible and two reversible reactions. The irreversible reactions are R_1, R_2, R_3, R_4, while the reversible ones are R_5, R_6. Initially, the columns are ordered so that irreversible reactions precede reversible reactions in the following manner:

$$N = \begin{pmatrix} 0 & 0 & 0 & -1 & 1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \end{pmatrix}. \tag{19}$$

The matrix columns are first ordered so that columns with more zero elements precede columns with fewer zero elements, maintaining the precedence of irreversible to reversible reactions. This heuristic is part of the Nullspace algorithm, and is intended to reduce the amount of elementary modes generated at intermediate steps. After this permutation, the matrix has the form:

$$N' = \begin{pmatrix} 0 & 0 & -1 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 1 & 0 & 0 \end{pmatrix} \tag{20}$$

and its reduced row echelon form is

$$N'_{\text{rref}} = \left(\begin{array}{ccc|ccc} 0 & -1 & 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 \end{array}\right) \tag{21}$$

To allow easier notation we will use $N'' = N'_{\text{rref}}$ in continuation of this example. The initial right nullspace matrix for N'' is

$$X^{(0)} = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 1 & \boxed{-1} \\ -1 & 1 & 0 \\ -1 & 1 & -1 \end{array}\right). \tag{22}$$

The nullspace algorithm finds all elementary modes in three steps, starting from the 4^{th} row of $X^{(0)}$. In the following, we illustrate the reduced rank test as applied to the Nullspace algorithm.

1. $k = 4$. The last two columns of $X^{(0)}$ are combined to form a new candidate column $\begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}^T$.

   The new matrix $X^{(1)}$ consists of the first two columns from $X^{(0)}$ together with the new candidate column:

   $$X^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix} \tag{23}$$

   The third column in the the initial matrix $X^{(0)}$ was eliminated because it does not satisfy the irreversibility requirement (the boxed coefficient was negative). To check that the candidate column (mode) is elementary, we must check the rank of submatrix of N'' consisting of columns 2, 3, 5, 6,

following the prescription in Theorem 2, where we include all the nonzeros among entries 1, 2, 3, 4, and all entries among those not yet processed (entries 5, 6). Equivalently, we use extract the same columns from the reduced row echelon form (21):

$$N''_{*,\overline{\mathsf{Z}}} = \left(\begin{array}{cc|cc} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{array}\right) \tag{24}$$

Instead of computing the nullity of this directly, we follow Theorem 7 to form a smaller matrix as follows. Subsets of indices which specify the submatrix are Z_b=[1] and $\overline{\mathsf{Z}}_\mathrm{a}$=[2,3], having that q=6, m=3, $a = [0, 1, 1]$ and $b = [0]$. The submatrix is found in the intersection of first row, and second and third columns of the matrix N''. The last two columns of $N_{*,\overline{\mathsf{Z}}}$ are derived from the "identity" block in (21), so we add multiples of these columns to the rest of the matrix to annihilate the corresponding row(s):

$$\mathsf{nullity}(N_{\mathsf{Z}_\mathrm{b},\overline{\mathsf{Z}}_\mathrm{a}}) = \mathsf{nullity}\begin{pmatrix} -1 & 1 \end{pmatrix} = 1, \tag{25}$$

This last reduced matrix consists of the columns of N'_{rref} corresponding to the nonzeros among the first $q - m$ entries and the rows corresponding to the zero entries among the entries $q - m + 1$ through $q - m + k$ of the candidate mode, according to Theorem 7.

We remark that at this stage we have processed all the rows of X corresponding to irreversible reactions, and hence according to Theorem 5 the matrix $X^{(1)}$ contains all the extreme pathways. The remaining steps compute elementary modes which are not extreme pathways.

2. $k = 5$. We combine columns of $X^{(1)}$ to annihilate coefficients in position $k = 5$ to form two new candidate columns, shown as the last two columns in $X^{(2)}$:

$$X^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 \end{pmatrix}. \tag{26}$$

The first three columns of $X^{(2)}$ are the columns of $X^{(1)}$, all of which are acceptable because the corresponding reaction (R_5) is reversible.

The reduced submatrix corresponding to the last column of $X^{(2)}$ is constructed as in (25):

$$\mathsf{nullity}\left(\begin{array}{ccc|c} 0 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & 1 \end{array}\right) = \mathsf{nullity}\left(\begin{array}{ccc|c} 0 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) = \mathsf{nullity}\begin{pmatrix} 0 & -1 & 1 \\ 1 & -1 & 0 \end{pmatrix} = 1, \tag{27}$$

We could also directly apply the result of Theorem 7 having that Z_b=[1,2] and $\overline{\mathsf{Z}}_\mathrm{a}$=[1,2,3], selecting first two rows and first three columns of the matrix N", to obtain the same final submatrix as above and compute its nullity. Similar reduced matrix is formed for the next to last column of $X^{(2)}$. In both cases, the nullities are equal to 1, so both columns are accepted.

3. $k = 6$. New candidate elementary modes are generated by combining pairs of columns (1,2) and (2,5) in the matrix $X^{(2)}$. These pairs of columns give the candidate modes $(1, 1, 0, 1, 0, 0)^T$ and $(1, 2, 1, 1, 1, 0)^T$. The first candidate is rejected because it is a duplicate of the 4^{th} mode in $X^{(2)}$. The second column is discarded because the nullity of the corresponding submatrix is not 1:

$$\mathsf{nullity}\left(\begin{array}{ccc|cc} 0 & -1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 \end{array}\right) = \mathsf{nullity}\left(\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 \end{array}\right) = \mathsf{nullity}\begin{pmatrix} 1 & -1 & 1 \end{pmatrix} = 2. \tag{28}$$

Actually, this column can be discarded without carrying out the rank test because it has too many non-zero elements so that the matrix (28) has too many columns, as discussed in 1.

The final set of obtained elementary modes is

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 \end{pmatrix} \tag{29}$$

of which only the first three (carried over from (23)) are extreme pathways.