Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems

C. Wagner*

Institute of Pharmacology, University of Bern, Friedbuehlstr. 49, CH-3010 Bern, Switzerland

Received: February 28, 2003; In Final Form: December 8, 2003

The analysis of a chemical reaction network by elementary flux modes is a very elegant method to deal with the stationary states of the system. Each steady state of the network can be represented as a convex combination of these modes. They are elements of the nullspace of the stoichiometry matrix due to the imposed steady-state condition. We propose an approach, which first derives the basis vectors of the nullspace and then calculates the elementary modes by an apt linear combination of the basis vectors. The algorithm exploits the special representation of the nullspace matrix in the space of flows and the fact that elementary modes consist of a minimal set of flows. These two ingredients lead to construction rules, which diminish the combinatorial possibilities to design elementary modes and, hence, reduce the computational costs. Further, we show that the algorithm also accounts for reversible reactions. If a system includes reversible reactions, it can be transformed into an unidirectional network by considering the forward and the backward flow separately. We derive a projection operator, which reveals the interrelationship between the two representations.

Introduction

Metabolic pathways of microorganisms on a nutrition substrate are assumed to be in a steady state. In mathematical terms, this state is expressed by a zero first derivative with respect to time of the concentrations. If flows are considered instead of concentrations, then the steady-state condition reads

$$\mathbf{N}\nu = 0 \tag{1}$$

where **N** represents the stoichiometry matrix and ν the flow vector. The matrix **N** is of size $m \times n$ where m and n respectively denote the number of species and the number of flows of the network. In the stationary state, the flows across the network are constant (here $\neq 0$) and sum up to zero for each metabolite. The concept of elementary modes is a very elegant method to analyze steady states of a network, which was first introduced by Clarke.[1] Elementary modes are minimal sets of flows, which properly connect inputs to the outputs of a biochemical system so that no accumulation of species within the network occurs. In addition, they also account for the presence of internal cycles. In the approach promoted by Clarke, all reversible reactions of the network are represented by two different unidirectional flows. Then, the set of normalized elementary modes is unique for a given network structure and each steady state of the system can be uniquely expressed as a nonnegative linear combination of these elementary modes. The only information, which enters the analysis is the stoichiometry matrix of the network. All feasible states including the elementary modes are in the nullspace of this matrix due to the imposed steady-state condition. In geometric terms, elementary modes represent in this case extremal vectors of the nullspace which limit the set of allowed flux vectors. The former can be represented as vertexes of a polytope in flux space and the latter as interior points of the polytope. An introduction to the topic can be found in the review of Stucki.[2] Further, a concise description of the method and the extension to reversible flows is presented in the book of Heinrich and Schuster.[3]

Although the basic theory was developed in the 1970s, its application to larger systems was realized during the past few years.[4,5] Due to the completely sequenced genome of the bacterium *E. coli*, Edwards and Palsson[6] achieved to reconstruct the major part of the metabolic map of the microorganism. They first determined the space of feasible metabolic steady states and then used a target function to optimize growth, which can be solved by linear programming. In an accompanying experiment they indeed observed that the bacterium undergoes adaptive evolution to achieve the in silico predicted growth.[7] Moreover, Stelling and co-workers[8] used a reduced model of the central carbon metabolism to study the growth of wild type and mutated *E. coli*. The growth of phenotypes was computationally foretold by the elementary mode analysis, and the result was experimentally tested. In the overwhelming majority of cases, the flux mode analysis correctly predicted the experimental outcome.

If a system is represented by a stoichiometry matrix **N** of size $m \times n$ and of rank m, its extreme solutions ν determined by eq 1 contain maximally $m + 1$ coefficients unequal to zero.[1] The additional coefficient is due to loss of the information about the absolute values of flows, which allows for an arbitrary normalization condition. Therefore, Clarke[1,9] proposed to solve the $\binom{n}{m+1}$ possible linear equation systems in order to determine the elementary modes. As a consequence, if m is not close to 1 or n, the number of possible modes grows rapidly with increasing network size making the elementary mode analysis impossible. Interestingly, only a small part of the $\binom{n}{m+1}$ possibilities remain finally as extremal currents. A further algorithm was suggested by von Hohenbalken,[10] which is based on geometric considerations and which strongly reduced the computational expenditure. However, it requires a time-consuming least distance method to eliminate non extremal currents. During the past decade Schuster and co-workers[11−13] introduced the tableau method[14] to the field. The algorithm was improved in such a way that nonelementary modes are recognized and eliminated in an early stage of computation making the analysis of larger systems feasible.

The algorithm introduced by Schuster et al. works in a similar manner as the Gaussian elimination procedure to compute the

* To whom correspondence should be addressed. Tel: +41 31 632 9991. Fax: +41 31 632 4992. E-mail: clemens.wagner@pki.unibe.ch.

10.1021/jp034523f CCC: $27.50
Published on Web 01/27/2004

integer nullspace basis of a matrix. The transpose of the stoichiometry matrix is extended by the identity matrix to the right side yielding the 0th tableau

$$\mathbf{T}_0 = [\mathbf{N}^T|\mathbf{Id}] \tag{2}$$

The first tableau $\mathbf{T}_1$ is then obtained by combining rows of $\mathbf{T}_0$ in pairs so that the first column of $\mathbf{N}^T$ becomes the null vector. In contrast to the nullspace basis vector calculation, the determination of elementary modes requires the inclusion of all possible pairs of rows. However, not all possible combinations lead to an elementary mode. Therefore, only selected combinations are added to the tableau using the selection rules given in ref 4. After iterative application of the procedure the final tableau is reached when $\mathbf{N}^T$ is nullified.

Another natural way to find the elementary modes of a reaction network is to calculate the integer basis vectors of the nullspace and to determine the elementary modes by a linear combination of the basis vectors. The nullspace basis can be computed via standard techniques, which provides $n - m$ basis vectors. In 1982 Happel and Sellers[15] developed an approach, which uses the nullspace matrix (basis vectors arranged as matrix) to determine the elementary modes. It is based on the fact that elementary modes have maximally $m + 1$ nonzero entries and at least $n - (m + 1)$ zeros. They selected $n - (m + 1)$ columns from the nullspace matrix forming a sub-matrix. The coefficients of the linear combination are then determined by the condition that the coefficient vector is in the nullspace of this sub-matrix. However, the algorithm requires the same exhaustive search as the Clarke algorithm, because $\binom{n}{m+1} = \binom{n}{n-(m+1)}$.

Like Happel and Sellers, we also define elementary modes as linear combinations of nullspace basis vectors. However, in addition, we exploit the fact that the matrix **K** formed by the nullspace basis vectors has a well-defined representation

$$\mathbf{K} = \mathbf{D}[\mathbf{K}'|\mathbf{Id}]. \tag{3}$$

For a system with a $m \times n$ stoichiometry matrix **N** of rank m, the associated nullspace matrix **K** is of size $(n - m) \times n$, which can be split into $\mathbf{K}'$ of size $(n - m) \times m$ and an $n - m$ identity matrix. The matrix **D** is diagonal with coefficients so that **K** is integer valued. The basic idea behind our approach is that elementary modes have a minimal set of flows. Minimal means that there is no mode with a set of flows, which is a subset of the flows of the elementary mode. Therefore, the aim of the algorithm is to construct linear combinations of nullspace basis vectors so that the number of flows is minimized. As we will show, this can be done by combining rows of **K** so that some of the flows in $\mathbf{K}'$ get canceled. If we do it in a systematic way, we obtain all elementary modes of the system. The minimality condition also defines the selection rule for elementary modes. The Schuster algorithm and the present algorithm work in a very similar way (pairwise cancellation of flows). In both cases, a major determinant of the computational expenditure is the size of the initial matrix. Using the Schuster algorithm, the initial matrix is of size $n \times m$ ($\mathbf{N}^T$ in eq 2) whereas in our case the initial matrix shrinks to the size of $(n - m) \times m$ ($\mathbf{K}'$ in eq 3). This strongly reduces the combinatorial problem to determine the elementary modes and therefore the computational costs.

The present paper develops the nullspace approach in three steps. First we derive the algorithm for a system with exclusively unidirectional flows where reversible reactions are represented as forward and backward flows. Second, we consider networks with a combination of reversible and irreversible flows. Finally, we show that the relation between the results of the first two steps is expressed by a projection operator.

Definitions

Without limiting generality, we set in the following the diagonal matrix **D** in eq 3 to **Id**. The starting point to determine the nullspace basis vectors of the stoichiometry matrix **N** is a two-dimensional array consisting of $\mathbf{N}^T$ extended by an identity matrix of adapted size. It is equal to $\mathbf{T}_0$ (eq 2) and has the following shape when written in full length

$$\begin{pmatrix} & & & & \nu_1 & \nu_2 & \dots & \nu_n \\ \nu_1 & | & s_{11} & \dots & s_{1m} & 1 & 0 & \dots & 0 \\ \nu_2 & | & s_{21} & \dots & s_{2m} & 0 & 1 & \dots & 0 \\ \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu_n & | & s_{n1} & \dots & s_{nm} & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{4}$$

Here, s_{ij} denotes the entries of the stoichiometry matrix, and n is the number of reactions and m the number of species. The ν_i represents the flow of the ith reaction. By applying the Gaussian algorithm to the left part of eq 4 the basis vectors of the nullspace appear in the right part. If m is larger than the rank of **N**, then there are conserved species in the network like enzymes, which catalyze a reaction. These dependences of rows of the stoichiometry matrix are automatically eliminated while calculating the nullspace basis vectors. In the following, we assume that m corresponds to the rank of **N**. Then, the nullspace matrix formed by the nullspace basis vectors e_1 to e_{n-m} (row vectors) assumes the following shape (see eq 3):

$$\begin{pmatrix} & \nu_1 & \dots & \nu_m & \nu_{m+1} & \nu_{m+2} & \dots & \nu_n \\ e_1 & | & k_{11} & \dots & k_{1m} & 1 & 0 & \dots & 0 \\ e_2 & | & k_{21} & \dots & k_{2m} & 0 & 1 & \dots & 0 \\ \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e_{n-m} & | & k_{(n-m)1} & \dots & k_{(n-m)m} & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{5}$$

where k_{ij} denotes the jth entry of the ith basis vector. The first line in eq 5 shows that each entry of a basis vector is associated with a flow (reaction). Moreover, the last $n - m$ entries represent labels of basis vectors. Sometimes it might be necessary to permute the flows in order to obtain the form given in eq 5.

Because elementary modes are elements of the nullspace they are linear combinations of nullspace basis vectors

$$\text{elm} = \sum_j \alpha_j e_j \tag{6}$$

where elm and e_j denote the elementary mode and the nullspace basis vectors, respectively, and α_j the coefficients. Taking e_1 to e_{n-m} as the basis, the linear combination of a flux mode can be recognized in the last $m - n$ entries of the mode. From Clarke's work,[1] it is known that elementary modes have maximally $m + 1$ nonzero entries. Therefore, each flux mode consists of $l \leq \min\{m + 1, n - m\}$ basis vectors and has the following shape after suitable permutation of the last $m - n$ entries:

$$\text{elm} = (* * *...*, \alpha_1\ \alpha_2...\alpha_l\ 0...0). \tag{7}$$

The elementary mode can be divided into two parts. The right part (after the comma) shows the linear combination of basis vectors of which the mode is made of. Due to the increase of nonzero entries in this part, the number of zeros in the left part has to increase. At least $l - 1$ stars on the left-hand side in eq

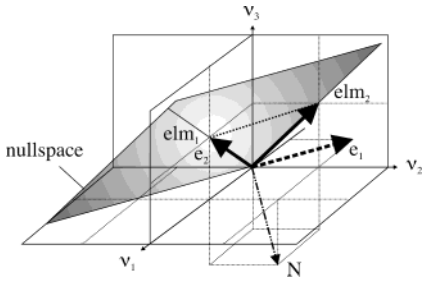

Figure 1. Flow space according to the reaction scheme given in eq 9. The stoichiometry matrix reads $N = (1\ 1\ -1)$, which determines the plane equation $\nu_1 + \nu_2 - \nu_3 = 0$ of the nullspace. The Gauss algorithm applied to the stoichiometry matrix yields the two basis vectors of the nullspace $e_1 = (-1\ 1\ 0)$ and $e_2 = (1\ 0\ 1)$. Applying the presented algorithm to e_1 and e_2 determines the two elementary modes $elm_1 = (1\ 0\ 1)$ and $elm_2 = (0\ 1\ 1)$.

7 must be zero by the combination. Using eq 6, this condition yields a set of equations for the coefficients $\alpha_1 ... \alpha_l$

$$\{\alpha_1 k_{1,i} + \alpha_2 k_{2,i} + ... + \alpha_l k_{l,i} = 0\}_{i=1...m} \tag{8}$$

From these m conditions, we can select $l - 1$ independent ones in order to determine $\alpha_1 ... \alpha_l$ up to multiples. Solutions, which fulfill $l - 1$ independent conditions of eq 8, are defined as elementary modes of the network. Obviously, the modes constructed in this way are flux modes as defined by Schuster et al.[3] Furthermore, they are not decomposable and they are therefore elementary flux modes. This can be proven in the following way. Consider an elementary mode elm constructed as described above, which should be decomposed into two modes m_1 and m_2. These two modes must have the following properties: 1. they must be elements of a proper subspace of the space S_{elm}, which is spanned by the basis vectors of elm; 2. the union of the space S_{m1} of m_1 and the space S_{m2} of m_2 must be equal to S_{elm}, $S_{m1} \cup S_{m2} = S_{elm}$. Compared to elm, the first condition gives rise to at least one additional zero in the right part of m_1 and m_2, which represents the decomposition into nullspace basis vectors. The second condition warrants that the basis vectors of a combination of m_1 and m_2 spans the same space as S_{elm}. Because the elementary mode elm is a linear combination of m_1 and m_2, the two modes have zeros at the same $l - 1$ positions in the left part as the elementary mode has. However, if such conditions for m_1 and m_2 exist, the rank of the matrix, which determines the coefficients $\alpha_1 ... \alpha_l$ (eq 8) is less than $l - 1$, which contradicts the construction rule of elementary modes.

These considerations lead to a new view of elementary modes. For a mode m, the set of zeros is defined as $Z(m) = \{i, m(i) = 0\}$ and the set of flows $F(m) = \{i, m(i) \neq 0\}$. Imagine that we have calculated all flux modes with a different set of flows. Now we sort the modes in chains so that the set of flows of the subsequent mode $F(m_{k+1})$ is a subset of the set of flows of the previous mode $F(m_k)$, i.e., $F(m_{k+1}) \subset F(m_k)$. All end points of these sequences are then elementary modes. As a consequence, elementary modes have a minimal set of flows and a maximal set of zeros. Just recently Schuster et al.[16] came to a similar conclusion.

The Algorithm

To derive the new algorithm, we first take only irreversible reactions into account and show subsequently the extension to reversible reactions. We start with a simple example, which can be represented in the 3d flow space (see Figure 1). Consider,

T_0	T_1	T_2	T_3	•••	T_m
ν_1	ν_2	ν_3			
	T_0 $\nu_1 = 0$	T_1 { T_0 $\nu_1 = 0$ $\nu_2 = 0$ $\nu_1, \nu_2 = 0$	T_2 { T_0 $\nu_1 = 0$ $\nu_2 = 0$ $\nu_1, \nu_2 = 0$ $\nu_3 = 0$ $\nu_1, \nu_3 = 0$ $\nu_2, \nu_3 = 0$ $\nu_1, \nu_2, \nu_3 = 0$		

Figure 2. Strategy to construct modes with all possible combinations of deleted flows in the first m entries. The first row shows the tableaux sequence starting at $[\mathbf{T}_0]$, the nullspace matrix, and ending at $[\mathbf{T}_m]$. The second row presents the flows (ν_i), which get canceled by pairwise combination of modes of the tableau in the same column ($\mathbf{T}_i$). As displayed in the third row, a new tableau ($\mathbf{T}_{i+1}$) is obtained by adding accepted combinations (modes with deleted (ν_i) but minimal) to the old tableau ($\mathbf{T}_i$).

we have a single species X with two input flows ν_1 and ν_2 and one output flow ν_3

$$\begin{array}{c} \nu_2 \\ \downarrow \\ \nu_1 \rightarrow X \rightarrow \nu_3 \end{array} \tag{9}$$

The stoichiometry matrix is then given by $\boldsymbol{N} = (1\ 1\ -1)$. Here we have $m = 1$ species and $n = 3$ reactions, so the nullspace has dimension $n - m = 2$. The two basis vectors read

$$e_1\ (-1\ 1\ 0)$$

$$e_2\ (1\ 0\ 1) \tag{10}$$

The strategy of the approach is based on the principle that an elementary mode contains a minimal the set of flows consistent with the nullspace (not all flows zero). Therefore we combine nullspace basis vectors so that at least one flow vanishes in the new mode. As can be recognized in eq 5, nullspace basis vectors cannot be combined such that one of the last $n - m$ flows are canceled (only one number in a column). Therefore, only the first m flows can be used to explore the nullspace for elementary modes. The algorithm systematically constructs modes with an increasing number of deleted flows in the first m entries. The calculated modes are then tested, and only modes with a minimalset of flows are retained. In a final step, all elementary modes, which satisfy the irreversibility condition, are collected.

To find the linear combinations of nullspace basis vectors with different sequences of deleted flows in the first m entries, we start with the nullspace matrix as the 0th tableau $\mathbf{T}_0$. The nullspace basis vectors are elementary modes of the system if they fulfill the sign restriction imposed by the irreversibility conditions. The first tableau $\mathbf{T}_1$ is obtained by processing the first column of $\mathbf{T}_0$. It means that all combinations in pairs of row vectors of $\mathbf{T}_0$ are calculated, which lead to a zero in the first entry. The new vectors are added at the bottom of $\mathbf{T}_0$ to produce $\mathbf{T}_1$. Thus, we computed all combinations, which delete the first flow ν_1. In the next step, we treat the second column of $\mathbf{T}_1$ in order to get $\mathbf{T}_2$. Thereby, all combinations in pairs are taken into account, which deletes ν_2. Because all modes with $\nu_1 = 0$ are already in the tableau, we obtain during the actual step also the modes with $\nu_1 = 0$ and $\nu_2 = 0$. By proceeding in

this way until the *m*th column, the method accounts for the deletion of all possible flow sequences up to length *m*. The strategy of the approach is visualized in Figure 2. Due to the irreversibility condition, the coefficients of the combinations are restricted to positive values.

In the example shown in Figure 1, $\mathbf{T}_0$ is given by e_1 and e_2. Because $m = 1$, only the first column has to be processed. To produce a zero in the first entry, e_1 and e_2 have to be added in order to furnish $\mathbf{T}_1$

$$\mathbf{T}_1 = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{11}$$

All row vectors of $\mathbf{T}_1$ are minimal in the sense that the set of flows of one row vector is not a subset of any other one. However, only the last two rows are elementary modes of the system because they satisfy the irreversibility conditions.

To apply the algorithm to large systems, it is crucial to eliminate nonelementary modes during computation. To do so, we use again the principle that elementary modes consists of a minimal set of flows. Consider the case where we already have calculated the tableau $\mathbf{T}_i$ and where the tableau $\mathbf{T}_{i+1}$ is partially filled. The set of flows of a new combination of rows of $\mathbf{T}_i$, which is a candidate for $\mathbf{T}_{i+1}$, is compared to the set of flows of all modes already in $\mathbf{T}_{i+1}$. Now, we can distinguish three cases. First, if there is a mode in $\mathbf{T}_{i+1}$, which contains the same set of flows, the two vectors are multiples of each other. Thus, the new combination must not be added and is ignored. Second, if we find a mode in $\mathbf{T}_{i+1}$, which flows from a subset of the flows of the new combination, then the new mode is not minimal and must also not be added to the tableau. Third, if the set of flows of the new mode is a subset of the flows of a mode already in $\mathbf{T}_{i+1}$, then the mode already in $\mathbf{T}_{i+1}$ must be replaced by the new mode. However, this situation rarely occurs and the fulfillment of the third rule is left to a final elimination round at the end of the program.

Extension to Reversible Reactions

The inclusion of reversible reactions or flows can be done in two different ways. Either the back reaction is represented by an additional irreversible flow or the sign of the flow in the elementary modes might become negative. The representation of a reversible reaction by two irreversible flows results in an increased stoichiometry matrix and therefore increases the computation time of elementary modes. Allowing for negative flows leaves the size of the stoichiometry matrix unchanged but the representation of modes as convex combination of elementary ones becomes ambiguous (see example in ref 16).

A reversible mode is defined as a mode that contains only reversible flows. In contrast to irreversible modes (which contain at least one irreversible flow), reversible modes can be added and subtracted to other modes because they run in both directions. As we have shown in eq 5, all basis vectors are labeled by a flow due to the identity matrix in the right part. As a consequence, if the flows associated with the identity matrix are all irreversible, then there exists no reversible elementary mode. If one would change the sign of one of the basis vectors, the sign restriction imposed on irreversible flows would be violated. If all basis vectors are irreversible, then only combinations with positive factors are allowed. In general, the basis vectors of the nullspace form a tableau, which can be divided in reversible and irreversible basis vectors. New vectors are constructed by mutual annihilation of an entry, using positive and negative factors for reversible (basis) vectors and only positive factors for irreversible (basis) vectors. In the final step, only those elementary modes are selected from the tableau, which have positive entries at the positions of irreversible flows.

To illustrate the algorithm, we use the reaction scheme of monosaccharide metabolism already studied by Schuster et al.[4] The model system uses $m = 5$ species, which are connected by $n = 9$ flows, whereof the first $r = 4$ flows are reversible. The reversible flows are separated from the irreversible flows by a vertical line. The stoichiometry matrix is given by

$$\mathbf{N} = \left(\begin{array}{cccc|ccccc} 0 & 0 & -1 & -2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 2 & 0 & 0 & -1 & & 0 \\ 0 & 2 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 \end{array}\right) \tag{12}$$

The tableau of $n - m = 4$ basis vectors contains $n = 9$ columns for n flows, and the first $r = 4$ flows are again reversible. Thus, it reads

$$\mathbf{T_0} = \left(\begin{array}{cccc|ccccc} -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 2 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 3 \end{array}\right) \tag{13}$$

To deal with integers, the diagonal of the matrix **D** reads (3 1 1 3). In this example, each basis vector reveals in addition to the first four reversible flows also at least one irreversible one. Therefore, all basis vectors are irreversible and may only be combined by positive factors. After treating the first column, we obtain 4 new vectors containing a zero in the first entry. The new tableau reads

$$\mathbf{T_1} = \left(\begin{array}{cccc|ccccc} -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 2 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 3 \\ 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & -2 & 2 & -1 & -5 & 0 & 0 & 2 & 3 \end{array}\right) \tag{14}$$

The treatment of columns 2−4 do not provide any new independent vector. Therefore, we jump directly to the fifth tableau

$$\mathbf{T'_5} = \left(\begin{array}{cccc|ccccc} -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 2 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 3 \\ 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & -2 & 2 & -1 & -5 & 0 & 0 & 2 & 3 \\ -5 & -1 & 2 & 2 & 0 & 6 & 0 & 1 & 0 \\ 5 & 1 & 4 & -2 & 0 & 0 & 1 & 0 & 6 \end{array}\right) \tag{15}$$

All further columns have only positive entries and cannot be combined. Thus, we can select from $\mathbf{T'_5}$ the elementary modes, which fulfill the sign restriction (the last $n - r = 5$ flows must be positive). The final tableau then reads

$$\mathbf{T}_5 = \left(\begin{array}{rrrr|rrrrr} -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -5 & -1 & 2 & 2 & 0 & 6 & 0 & 1 & 0 \\ 5 & 1 & 4 & -2 & 0 & 0 & 1 & 0 & 6 \end{array}\right) \quad (16)$$

A comparison of the computational effort of the Schuster and the present algorithm is given in Table 1. The first row of both algorithms displays the growth of the tableaux during computation. It shows that the size of the tableaux of the present algorithm is always less than the tableau size obtained by the Schuster method. Further, the reduction of the initial matrix from 5×9 (Schuster) to 4×5 (present) leads to a decreased number of new row (or elementary mode) calculations (totally: from 25 to 6) and to a diminished number of irrelevant mode checks (totally: from 23 to 4). The estimation omits the final irreversibility test of the present algorithm because it is a simple sign check.

It is important to note that no sign restriction is imposed on the elimination rule. Therefore, elementary modes are recognized by their minimal set of flows also in the presence of reversible reactions. As a consequence, the reverse of a reversible elementary mode is eliminated by the mode check and has to be added in a separate step at the end. In the Appendix, a pseudo code of the algorithm is given, which is divided in three parts. In the preprocessing step, the integer nullspace matrix is calculated and the Dulmage−Mendelsohn permutation applied. The latter admits to decompose the network into independent subsystems, which can be processed separately. The main part of the program calculates the sequence of the tableaux. It accounts only for systems with irreversible basis vectors. The extension to reversible basis vectors is straightforward. An additional column permutation (colperm) is applied at the beginning of the main program, which is very efficient to further reduce the computational load. Finally, in the postprocessing step, the modes are normalized and checked for minimality and irreversibility. Further, the inverted permutations have to be applied.

Link between Reversible and Irreversible Approach

If one decides to use for each back reaction an additional flow, all nullspace basis vectors are obviously irreversible and only positive combinations can be used to construct the elementary modes. To illustrate the connection between this approach and the previous one, we use the same example as above. The superscript irr denotes that all flows are irreversible. For each back reaction, a column is added to the stoichiometry matrix, which is a negative copy of the forward reaction. In our case with $r = 4$ reversible flows, the stoichiometry matrix becomes extended by a negative copy of the first 4 columns

$$\mathbf{N}^{\mathbf{irr}} = \left(\begin{array}{rrrrrrrrrrrrr} 0 & 0 & -1 & -2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & -2 \\ 0 & 2 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 1 \end{array}\right) \quad (17)$$

The number of flows rises to 13 and the number of nullspace basis vectors from 4 to 8

TABLE 1: Comparison of the Computational Costs for the Schuster and the Present Algorithm[a]

	T_0	T_1	T_2	T_3	T_4	T_5
size (Schuster)	9×14	9×13	9×12	11×11	11×10	7×9
new rows		3	3	6	8	5
irrelevant mode check		0	0	3	11	9
size (present algorithm)	4×9	8×9	8×9	8×9	8×9	7×9
new elementary modes		4	0	0	0	2
irrelevant mode check		0	1	0	1	2

[a] The first row of both methods shows the growth of the size of the tableaux from the first ($\mathbf{T}_0$) to the final tableau ($\mathbf{T}_5$). The second and the third lines display the number of new rows, which have to be calculated and the number of mode checks for early elimination of non elementary modes.

$$\mathbf{T}_0^{\mathrm{irr}} = \left(\begin{array}{rrrrrrrrrrrrr} -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & -1 & -1 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \quad (18)$$

The algorithm provides 11 independent solutions

$$\mathbf{T}_5^{\mathrm{irr}} = \left(\begin{array}{rrrrrrrrrrrrr} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 5 & 1 & 4 & 0 & 0 & 0 & 1 & 0 & 6 & 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 2 & 0 & 6 & 0 & 1 & 0 & 5 & 1 & 0 & 0 \end{array}\right) \quad (19)$$

The first four elementary modes denote artificial or spurious cycles because they represent reversible reactions. The calculated elementary modes can be projected to the $n = 9$ dimensional subspace with $r = 4$ reversible reactions. The projection operator leaves the first n entries of eq 19 unchanged, whereas the last r components are projected with a negative sign onto the first r coordinates. In general, the operator reads

$$\mathbf{P} = \left(\begin{array}{c} \mathbf{Id}_n \\ \hline -\mathbf{Id}_r \end{array}\right) \quad (20)$$

where $\mathbf{Id}_n$ and $\mathbf{Id}_r$ denote the identity matrix of size n and r, respectively. If it is applied to the extremal currents of eq 19 ($\mathbf{T}_5^{\mathrm{irr}} \times \mathbf{P}$), we obtain the same elementary modes as shown in eq 16. The spurious elementary modes are in the nullspace of the projection operator. All other modes, which contain the forward and the backward flow have to be composites of elementary modes and of spurious cycles and are thus eliminated by the algorithm. Furthermore, in eq 19, where all flows are irreversible, the non spurious elementary modes can only have an entry at either the forward or the backward flow. Therefore, if one neglects the spurious modes, the number of elementary modes and the number of extremal currents are the same.

Summary

We developed a new approach to determine the elementary modes of a chemical reaction system. The algorithm is based

PSEUDOCODE

All commands correspond to Matlab (Mathworks Inc.) commands.

N: stoichiometry matrix
K: nullspace basis matrix
D: diagonal matrix for integer basis
K_i: submatrix of K
m: number of rows of N
n: number of columns of N

```
1. Preprocessing
    % order N so that reversible flows are in the left part;
    K = null(N,'r');                    % calculates the rational basis of the nullspace
    % determine D;
    K = D*K;                            % integer basis of the nullspace
    K = dmperm(K);                      % Dulmage-Mendelsohn decomposition of K in K_i submatrices

2. Main Program
    % here we consider only irreversible basis vectors, which is the most prominent case. The extension to
    % reversible basis vectors is straight forward.
    For all submatrices K_i of K
        T_0 = colperm(K_i);             % column permutation of K_i
        L_1 = length(T_0(:,1)) +1;      % L_1: next open space for a new elementary mode
        For col = 1 : n_i               % n_i: number of columns of T_0
            T_col = T_col-1;            % starting tableau is equal to the previous tableau
            H = length(T_col-1(:,1));   % H: number of rows of T_col
            For h_1=1:(H-1)             % test all possible combinations in pairs
                For h_2=(h_1+1):H
                    if T_col-1(h1,col)*T_col-1(h2,col) < 0      % for irreversible vectors combinations are
                                                                % only possible if product is < 0
                        mode_1 = T_col-1(h1,:);
                        mode_2 = T_col-1(h2,:);
                        mode = abs(T_col-1(h2,col))*mode_1 + abs(T_col-1(h1,col))*mode_2;
                        notin = 1;
                        h_3 = 1;
                        while h_3 < L_1 & notin > 0             % check mode against modes in T_col
                            if Z(mode) ⊆ Z(T_col(h_3,:))        % if mode is not minimal compared to
                                notin=0;                        % modes in T_col then do not add mode
                            end;
                            h_3 = h_3 + 1;
                        end;
                        if (notin)                              % if notin = 1 add mode
                            T_col(L_1,:) = mode;
                            L_1 = L_1+1;
                        end;
                    end;
                end;
            end;
        end;

3. Post-Processing
        % normalize rows of T_col to get small integer values;
        % sort T_col by the inverse of colperm;
        % Collect all modes, which are in accordance with the irreversibility conditions;
        % Final test for minimality;
    end;
    % sort all T_col by the inverse of dmperm
```

Figure 3. Pseudocode of the algorithm.

on the special representation of the nullspace basis matrix in the flowspace. It can be arranged in such a way that it splits up in a $(n - m) \times m$ submatrix and an $(n - m)$ identity matrix. Only the former gives rise to combinations of nullspace basis vectors to compose the elementary modes of the network. The construction rule is based on the fact that elementary modes contain a minimal set of flows (not all zero). Therefore, we check out all combinations of basis vectors, which lead to additional zeros in the first m entries. The strategy to solve this combinatorial problem is shown in Figure 2. Thereby pairs of vectors are combined so that at least one flow gets deleted. During computation, only those modes are collected for further processing, which contain a minimal set of flows.

To illustrate the approach, we give a geometrical interpretation of the algorithm using the simple example defined in eq 9 and displayed in Figure 1. We divide the flow space in two different subspaces. The first subspace SF_1 is given by the first flow $m = 1$, whereas the second subspace SF_2 is determined by the last $n - m = 2$ flows. The projection to SF_2 reveals that basis vectors are orthogonal in this subspace and can therefore not be used to generate elementary modes. In Figure 1, SF_2 corresponds to the $\nu_2 - \nu_3$ plane. In contrast, if the basis vectors are projected to SF_1, the flows overlap and give rise to combinations, where at least one flow gets annihilated. In general, the algorithm scans all subspaces of SF_1 with all combinations of deleted flows and selects modes, which satisfy the minimality condition. In our example (see Figure 1), the subspace SF_1 is given by the ν_1 coordinate. Therefore, the single subspace of SF_1, which has to be considered, is the origin. Using the condition $\nu_1 = 0$, we obtain the mode (0 1 1). Because its

set of flows is not a subset of the set of flows of the other two modes (e_1 and e_2), the new mode is elementary.

Compared to Schuster's approach, the major achievement of the present algorithm is the size reduction of the initial matrix, which has to be processed. It shrinks from $n \times m$ ($\mathbf{N}^T$) to $(n - m) \times m$ ($\mathbf{K}'$). As a consequence, the computational expenditure to determine the elementary modes is markedly reduced (see Table 1). The advantage is 2-fold: first, the diminished size of the tableaux reduces the required memory space, and second, the decreased number of computed modes and mode checks reduces the CPU time. Further, the algorithm accounts for the presence of conserved species in the network because linear dependent rows of the stoichiometry matrix are eliminated by the nullspace calculation. Moreover, subnetworks, which subsist in equilibrium whenever the system is at steady state, appear as zero submatrixes in the nullspace matrix. If the system consists of independent subnetworks, the nullspace matrix can be block-diagonalized by the Dulmage−Mendelsohn algorithm[17] and the nullspace approach can be applied to the subspaces. It is important to note that the time for preprocessing the stoichiometry matrix is in general negligible compared to the computation time of the elementary modes. Therefore, the calculation of nullspace basis vectors is of no consequence. Further reduction of the computational costs can be achieved by sorting the columns of the nullspace matrix. A very efficient method is to permute the columns so that the number of entries in a column increases in the same direction as the matrix is processed. To appreciate the adaptive nature of the nullspace approach, consider a network with n flows and $m = n - 1$ species. The nullspace is given by a single basis vector, which represents the only elementary mode. So the nullspace method is already finished after the preprocessing step. In contrast, using for example the Schuster algorithm, one still would have to deal with a matrix of size $n \times (n - 1)$. Note that also in the opposite case, where the system consists of a single species $m = 1$ and n flows, the initial matrix $\mathbf{K}'$ of the nullspace approach is still smaller than the stoichiometry matrix $\mathbf{N}$.

However, the work that has to be done to determine the elementary modes does not only depend on the size of the initial matrix but it also depends on the number of zeros in this matrix. For an efficient computation of the modes, it is therefore crucial to keep this number as high as possible during the preprocessing step. Occasionally it can happen, that the number of entries in the first m-columns of the nullspace basis vectors is larger than the number of entries in the stoichiometry matrix, which partly reduces the benefits of the algorithm. The derivation of a procedure, which optimizes the number of zeros in the nullspace approach will be the topic of future work.

Acknowledgment. I thank Jörg Stucki for stimulating discussions and for helpful comments on the manuscript. The support of the Swiss National Science Foundation (Grant No. 31-49745.96 and Grant No. 3100A0-102269) is gratefully acknowledged.

References and Notes

(1) Clarke, B. *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley: New York, 1980.
(2) Stucki, J. *Prog. Biophys. Mol. Biol* **1978**, *33*, 99.
(3) Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman & Hall, New York, 1996.
(4) Schuster, S.; Fell, D.; Dandekar, T. *Nat. Biotechnol.* **2000**, *18*, 326.
(5) Edwards, J.; Ibarra, R.; Palsson, B. *Nat. Biotechnol.* **2001**, *19*, 125.
(6) Edwards, J.; Palsson, B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5528.
(7) Ibarra, R.; Edwards, J.; Palsson, B. *Nature* **2002**, *420*, 186.
(8) Stelling, J.; Klamt, S.; Bettenbrock, K.; Schuster, S.; Gilles, E. *Nature* **2002**, *420*, 190.
(9) Clarke, B. *J. Chem. Phys.* **1981**, *75*, 4970.
(10) von Hohenbalken, B. C. B.; Lewis, J. *J. Comput. Appl. Math.* **1987**, *19*, 231.
(11) Schuster, S.; Hoefer, T. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2561.
(12) Schuster, R.; Schuster, S. *Comput. Appl. Biosci.* **1993**, *9*, 79.
(13) Schuster, S.; Hilgetag, C. *J. Phys. Chem.* **1995**, *99*, 8017.
(14) Nozicka, F.; Guddat, J.; Hollatz, H.; Bank, B. *Theorie der Linearen Parametrischen Optimierung*; Akademie-Verlag: Berlin, 1974.
(15) Happel, J.; Sellers, P. *Ind. Eng. Chem. Fundam.* **1982**, *21*, 67.
(16) Schuster, S.; Hilgetag, C.; Woods, J.; Fell, D. *J. Math. Biol.* **2002**, *45*, 153.
(17) Schuster, S.; Schuster, R. *J. Math. Chem.* **1991**, *6*, 17.