

# A new general algorithmic method in reaction syntheses using linear algebra

István Szalkai

Department of Mathematics, University of Veszprém, P.O. Box 158, Veszprém, Hungary H-8201

Received 19 June 2000

We discuss here a new general linear algebraic method (both model and algorithm) for describing and generating (among others) minimal *reactions* and also minimal *mechanisms* in stoichiometry, or dimensionless groups in physics as well. (Further applications in *process network syntheses* will be discussed in [1].) With some minor modifications of the *input* this method can be extended for *several* related questions: for generating direct and overall reactions, direct (steady state) mechanisms, for finding the possible resulting (overall) reactions among all possible mechanisms, etc.

Computational results in section 4 show the speed of our algorithm.

We give also mathematical background and results in sections 3, 5 and 6. However, we do not restrict ourselves to mathematics only, we also talk on the language of chemistry, too.

The theoretical results in sections 3.2, 3.3, 5 and the computational examples in section 4 are completely new, further theoretical results will appear in [1,2] and in [3].

**KEY WORDS:** linear algebraic method, minimal reaction, direct mechanism, simplexes

## 1. Introduction and history

Mathematical formulation and even algorithmic approaches of finding minimal (direct) and overall reactions and mechanisms grew a great literature in the last decade. We refer here only to the works of Happel, Sellers and Otarod [4–6], of Pethő [7–9], of Bertók [10], Fishtik, Alexander and Datta [11,12] and of the present author [1,2,13–16] and [3].

The main purpose of the present paper is to highlight the versatility of our mathematical model and the algorithm based on it. Our theory is based on Prof. Pethő's work [7] and is applied for several examples in section 2. For fully understanding of the generality of our method and algorithm we devote section 2 for the mathematical approach and examples of different types (as *atoms*, *reactions*, *mechanisms*, *process network syntheses* or *dimensionless units in physics*). Though the mathematical notion ('simplex') we introduce has its own curiosity in mathematics, here we focus on chemical aspects of its applications only. The mathematical investigations of simplexes are published in [14–16] and are continued in [2] and in [3]. In section 3 we introduce our new general algorithm which idea was first published in [13] in 1991. Our algorithm

differs in its basis from Happel’s et al. in [6] and Bertók’s [10], it is fully automatic, and computer experiments show that it is considerably fast (see section 4).

Now in this present paper we apply this method for many further applications. Surprisingly enough only slight modifications were necessary of the *input set* for many different and effective further applications. These modifications are discussed in section 3.2.

Let us remark also for readers of nonmathematical interest that the results of [14] and [15] give sharp bounds for the *number* of simplexes in a given vector set. In other words, these formulas give immediately bounds for the *number* of chemical *minimal reactions/direct mechanisms* in a given set of *species* (groups of atoms or functional bonds)/*reactions* since the number of the involved elements/species is known and fixed. For the readers’ convenience we list these formulas in the last section, in the appendix.

Finally, in section 5 we deal with some other general mathematical questions. Special thanks are given to Prof. Árpád Pethő, Universität Hannover and Prof. Ferenc Friedler, University of Veszprém.

## 2. Mathematical formulation

Our mathematical model is standard and quite well-known, first described in [7] and [13], partly similar to Sellers and others [5,6]. Let us highlight in advance that our method is exactly the *same* both for *reactions*, *mechanisms*, and *dimensionless groups*. We can describe all these problems in the language of linear combinations of vectors, and in order to find the desired object (minimal reactions/mechanisms/dimensionless groups) we need to find, in all of the three above cases, minimal *linearly dependent* subsets of these vectors. We call these subsets of vectors *simplexes*. Since our algorithm finds *all* simplexes in *any* set of given vectors, it can be applied immediately for all of the three problems above.

### 2.1. Reactions

If the chemical species (or groups of atoms, i.e., functional bonds only)  $A_1, \dots, A_n$  consists of the atoms  $E_1, \dots, E_m$  as

$$A_j = \sum_{i=1}^m a_{i,j} \cdot E_i,$$

where  $a_{i,j} \in \mathbb{N}$  for  $j = 1, \dots, n$  and  $i = 1, \dots, m$ , and the sets  $\{A_1, \dots, A_n\}$  and  $\{E_1, \dots, E_m\}$  are fixed, then we can *assign* the species  $A_j$  to the vector

$$\mathbf{A}_j := [a_{1,j}, \dots, a_{m,j}]^T$$

for  $j = 1, \dots, n$ , which vectors<sup>1</sup> are elements of the  $m$ -dimensional linear space, where a (natural) set of base vectors is  $\{E_1, \dots, E_m\}$ . Now, a (possible) chemical **reaction**

<sup>1</sup> We will not emphasize the difference between the species  $A_j$  and the vectors  $\mathbf{A}_j$ .

among the species  $\{A_j: j \in S\}$  does exist for any  $S \subseteq \{1, \dots, n\}$  if and only if the system of homogen linear equations

$$\sum_{j \in S} x_j \cdot \mathbf{A}_j = \mathbf{0} \quad (0) \quad (1)$$

has nontrivial solution for  $x_j \in \mathbb{R}$  ( $j \in S$ ) by the law of mass balance (conservation of material). A solution  $\{x_j: j \in S\}$  easily determines uniquely a chemical reaction among the species  $\{A_j: j \in S\}$ .<sup>2,3</sup> (The structure of the solutions of homogeneous linear equations is discussed in [17].)

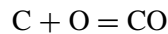
For example, consider the species  $A_1 = \text{C}$ ,  $A_2 = \text{O}$ ,  $A_3 = \text{CO}$  and  $A_4 = \text{CO}_2$ , that is we use the elements  $E_1 = \text{C}$  and  $E_2 = \text{O}$ . So  $m = 2$  (= the number of atoms = the dimension), and  $n = 4$  (= the number of species/vectors). The columns of the following table represent the vectors  $\mathbf{A}_1$  through  $\mathbf{A}_4$ :

	$A_1$	$A_2$	$A_3$	$A_4$	
C	1	0	1	1	(1)
O	0	1	1	2	
	C	O	CO	CO <sub>2</sub>	

Now, for example, taking  $S = \{A_1, A_2, A_3\}$  we get the solution<sup>4</sup>

$$X_1 = [x_1, x_2, x_3, x_4]^T = [1, 1, -1, 0]^T$$

which corresponds to the reaction



The assumption that the equation (0) has nontrivial solutions<sup>5</sup> is equivalent to that the vector-set  $\{A_j: j \in S\}$  is *linearly dependent*. The corresponding reaction is called **minimal** iff for no  $T \subset S$ ,  $T \neq S$  there might be any reaction among the species  $\{A_j: j \in T\}$ . That is, the vector-set  $\{A_j: j \in T\}$  is *linearly independent* for  $T \subset S$ ,  $T \neq S$ .

This motivates the following definition (see [8], [13] or [14]):

**Definition 2.1.** A set of vectors  $B = \{b_j: j \in S\} \subseteq \mathbb{R}^n$  is called a **simplex** iff  $B$  is linearly dependent but all of its proper subsets  $\{b_j: j \in T\}$  are linearly independent for any  $T \subset S$ ,  $T \neq S$ .

<sup>2</sup> Species (groups of atoms or functional bonds) with positive coefficients represent the right-hand side of the reaction equality while species with negative coefficients are moved to the left-hand side.

<sup>3</sup> All the solutions for the unknowns  $x_j$  for  $j \in S$  are rational since the components  $a_{i,j}$  of the vectors  $A_j$  – the coefficients of the homogeneous linear equation (0) – are all integers.

<sup>4</sup> The coordinates of  $X$ , corresponding to vectors  $A_j$  which do NOT belong to the simplex  $S$ , are always 0.

<sup>5</sup> Of course the reactions obtained in the described way are only possibilities, since, for example, the reaction  $2\text{Au} + 6\text{HCl} = 2\text{AuCl}_3 + 3\text{H}_2$  does not yield under normal conditions!

The only simplexes in the above example are (we write the sub-indexes of the vectors  $A_i$  only)  $S_1 = \{1, 2, 3\}$ ,  $S_2 = \{1, 2, 4\}$ ,  $S_3 = \{1, 3, 4\}$  and  $S_4 = \{2, 3, 4\}$ . After solving the corresponding equations (0) we get the following (complete) list of minimal reactions:

$$\begin{aligned} X_1 &= [1, 1, -1, 0]^T : \text{C} + \text{O} = \text{CO} \\ X_2 &= [1, 2, 0, -1]^T : \text{C} + \text{O}_2 = \text{CO}_2 \\ X_3 &= [1, 0, -2, 1]^T : \text{C} + \text{CO}_2 = 2\text{CO} \\ X_4 &= [0, 1, 1, -1]^T : \text{O} + \text{CO} = \text{CO}_2 \end{aligned} \quad (2)$$

(We continue *this* example in the next section.)

## 2.2. Mechanisms

We can build up **mechanisms** (and also *minimal* ones) from any set of reactions just on the same way as we built up reactions from species in the previous subsection. (Mechanisms are the main topic of our present research, we investigate them in more detail in sections 3.2 and 4.)

Well, if we are given the reactions  $X_1, \dots, X_k$  which use the species (groups of atoms or functional bonds)  $A_1, \dots, A_n$  then we can correspond these reactions to the vectors  $\mathbf{X}_1, \dots, \mathbf{X}_k \in \mathbb{R}^n$  in the  $n$ -dimensional linear space  $\mathbb{R}^n$ , where the base vectors are  $\mathbf{A}_1, \dots, \mathbf{A}_n$  (= the standard base) representing each speci involved in the reactions. That is:

$$\mathbf{X}_j = \sum_{i=1}^n b_{i,j} \cdot \mathbf{A}_i,$$

where  $b_{i,j} \in \mathbb{Z}$  for  $j = 1, \dots, k$  and  $i = 1, \dots, n$ .<sup>6,7</sup> Then *any* linear combination

$$\mathcal{M} = \sum_{j=1}^k \lambda_j \cdot \mathbf{X}_j \quad (3)$$

of the vectors  $\mathbf{X}_j$  with integer (or rational<sup>8</sup>) coefficients  $\lambda_j \in \mathbb{Z}$  represents a (possible) *mechanism*  $\mathcal{M}$  in a natural way: during that mechanism the reaction  $X_j$  takes place  $\lambda_j$ -many times (see the example below). Negative coefficients of course, mean that the corresponding reactions take place in reversed order. This reaction uniquely can be described by the vector of the coefficients

$$\underline{\lambda} := [\lambda_1, \dots, \lambda_k]^T \in \mathbb{Z}^k.$$

<sup>6</sup> Moving the terms with negative coefficients  $b_{i,j} < 0$  to the left-hand side of the equality (initial materials of the reaction) and leaving the others in the right-hand side (resulting materials) we get the usual form  $\sum_{i \in X} b'_{i,j} A_i = \sum_{i \in Y} b'_{i,j} A_i$  the mechanism.

<sup>7</sup> By the law of the mass-balance we surely must have  $\mathbf{A}\mathbf{X} = \mathbf{0}$  for the matrix  $\mathbf{A} := [A_1, \dots, A_n]$  which “codes” the sum-formulae of the species  $A_1, \dots, A_n$ .

<sup>8</sup> After multiplying all the coefficients by the common denominator we get integer coefficients.

Let us emphasize here that the detailed flow of the mechanism can be detected from the above vector of the coefficients *but* not at all from the sum-vector of the linear combination in (3).

For further investigations we are advised to extract the notion of the **resulting (overall) reaction** denoted by  $\mathcal{R}(\mathcal{M})$  or  $\mathcal{R}(\underline{\lambda})$  corresponding to the mechanism  $\mathcal{M}$  in (3) as (see [5]):

$$\mathcal{R}(\underline{\lambda}) := \sum_{j=1}^k \lambda_j \mathbf{X}_j \in \mathbb{R}^n,$$

which is the final reaction, caused by the reactions  $\mathbf{X}_j$ ,  $\lambda_j$ -many times one after another,  $1 \leq j \leq k$ .

Let us remark that this model does not reveal the *order* of or any other connection among the reactions in a mechanism, B. Bertók recently introduced a graph-theoretical approach in [4] for that problem.

In the practice we distinguish *terminal species* (the starting and final ones that we are interested in) and *intermediate* ones (which occur only during the mechanism). These terms are from [6], the terms *chemical* and *active species* are also in use. According to this distinction of species (or just groups of atoms/functional bonds), mechanisms are called **steady state mechanisms** [4] if the corresponding reaction (called **overall reaction**) contains only terminal (chemical) species.

The mechanism is called **direct** or **minimal mechanism** [4] if the set of active reactions (vectors with nonzero coefficient)

$$S(\lambda) := \{j \leq k \mid \lambda_j \neq 0\}$$

cannot be decreased to yield the same resulting (overall) reaction  $\mathcal{R}(\underline{\lambda})$ . In other words: there is *no proper* subset  $S' \subsetneq S(\underline{\lambda})$  such that  $\mathcal{R}(\underline{\mu}) = \alpha \cdot \mathcal{R}(\underline{\lambda})$  and  $S(\underline{\mu}) = S' \subsetneq S(\underline{\lambda})$  for some coefficient-vector  $\underline{\mu} \in \mathbb{Z}^k$  and rational number  $\alpha \in \mathbb{Q}$ . In this case the corresponding reaction  $\mathcal{R}(\underline{\lambda})$  is called **simple** or **minimal reaction**.

Using our terminology, the solutions of the linear equations

$$\sum_{j \in S} y_j \mathbf{X}_j = 0 \tag{4}$$

are minimal mechanisms *iff* the set  $S \subseteq \{1, \dots, k\}$  is a *simplex*, again!

Milner uses the term **direct path** and Sellers [5] the **cycle-free mechanism** for minimal mechanisms.

With slight modifications of our pure linear algebraic algorithm (detailed in section 3.1) we can solve several problems: we can find all minimal mechanisms resulting to any given reaction, or we can find all overall reactions if the sets of terminal and intermediate (chemical and active) species are given. These and other applications are explained in detail in section 3.2 of our present paper. Happel, Sellers and Otarod in [6] and Bertók in [10] published algorithms for finding direct steady-state mechanisms. Pethő

and Kumar in [19] presented an output for minimal reactions only. Our computational results are listed and the related ones are compared to the above ones in section 4.

Continuing our example from the previous subsection, the coordinates of the vectors  $\mathbf{X}_1, \dots, \mathbf{X}_4$  (= [minimal] reactions, see (1) and (2)) are

	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$	
$A_1 = \text{C}$	1	1	0	1	(5)
$A_2 = \text{O}$	1	2	1	0	
$A_3 = \text{CO}$	-1	0	1	-2	
$A_4 = \text{CO}_2$	0	-1	-1	1	

Funny enough in our example we have again *four* simplexes. Solving the corresponding equations, we get the following list of minimal mechanisms (cycles):

$$\begin{aligned}
 Y_1 = [-1, 0, 1, 1]^T : & \text{C} + \text{O} \leftarrow \text{CO} \\
 & \text{O} + \text{CO} \rightarrow \text{CO}_2 \\
 & \text{C} + \text{CO}_2 \rightarrow 2\text{CO} \\
 Y_2 = [-1, 1, -1, 0]^T : & \text{C} + \text{O} \leftarrow \text{CO} \\
 & \text{C} + 2\text{CO} \rightarrow \text{CO}_2 \\
 & \text{O} + \text{CO} \leftarrow \text{CO}_2 \\
 Y_3 = [2, -1, 0, -1]^T : & 2(\text{C} + \text{O} \rightarrow \text{CO}) \\
 & \text{C} + 2\text{O} \leftarrow \text{CO}_2 \\
 & \text{C} + \text{CO}_2 \leftarrow 2\text{CO} \\
 Y_4 = [0, -1, 2, 1]^T : & \text{C} + 2\text{O} \leftarrow \text{CO}_2 \\
 & 2(\text{O} + \text{CO} \rightarrow \text{CO}_2) \\
 & \text{C} + \text{CO}_2 \rightarrow 2\text{CO}
 \end{aligned}
 \tag{6}$$

For all these mechanisms  $Y_i$  the resulting (overall) reaction  $\mathcal{R}(Y_i) = 0$  (the void reaction),  $i = 1, 2, 3, 4$ .

### 2.3. Dimensionless groups

This application was first mentioned in [7] and is not from stoichiometry, but reveals the structure of composite measure units in physics: we can easily (fully automatically) find minimal identities among physical quantities by finding (all) simplexes again in a certain set of vectors.

Let there be given  $m$  elementary quantities (mass, length, time, etc.) denoted by  $E_1, \dots, E_m$  and  $n$  quantities  $A_1, \dots, A_n$  ( $n \in \mathbb{N}$  is any number) as

$$A_j = \prod_{i=1}^m E_i^{a_{i,j}},$$

where  $a_{i,j} \in \mathbb{Z}$  for  $j = 1, \dots, n$  and  $i = 1, \dots, m$ . Clearly we can assign the quantities  $A_j$  to the vectors

$$\mathbf{A}_j := [a_{1,j}, \dots, a_{m,j}]^T \in \mathbb{R}^m$$

for  $j = 1, \dots, n$ , which vectors are in the  $m$ -dimensional linear space  $\mathbb{R}^m$ . Now, a (possible) **dimensionless group (real number)** of the quantities  $\{A_j: j \in S\}$  does exist for any  $S \subseteq \{1, \dots, n\}$  iff the equality

$$\prod_{j \in S} A_j^{x_j} = 1 \quad (7)$$

holds, or equivalently (considering the exponents) the homogeneous system of linear equations

$$\sum_{j \in S} x_j \cdot \mathbf{A}_j = \mathbf{0}$$

has nontrivial solution for  $x_j \in \mathbb{R}$  ( $j \in S$ ). That is, we again reached to the system of linear homogen equations (0) and to the problem of finding simplexes!

For example, consider the flow of a fluid through a heated tube and the heat transfer between the pipe wall and the fluid. Consider now the following quantities:

$$\begin{aligned} A_1 = \text{tube diameter} &= d(\ell)(\text{length, basic quantity}), \\ A_2 = \text{linear velocity} &= v(s/t), \\ A_3 = \text{fluid density} &= \rho(m/\ell^3), \\ A_4 = \text{viscosity} &= \nu(m/\ell t), \\ A_5 = \text{heat capacity} &= \kappa(A/t^2 T), \\ A_6 = \text{heat transfer coeff.} &= \lambda(m/t^3 T), \\ A_7 = \text{thermal conductivity} &= \mu(m\ell/t^3 T). \end{aligned}$$

In matrix form:

	$m$	$\ell$	$t$	$T$	
$A_1 = d$	0	1	0	0	
$A_2 = v$	0	1	-1	0	
$A_3 = \rho$	1	-3	0	0	
$A_4 = \nu$	1	-1	-1	0	
$A_5 = \kappa$	0	2	-2	-1	
$A_6 = \lambda$	1	0	-3	-1	
$A_7 = \mu$	1	1	-3	-1	(8)

For example, one minimal dimensionless group  $X_1$  of the seven ones is the following:

$$X_1 = [0, 0, 0, 1, 1, 0, -1]^T,$$

which corresponds to the equality

$$\nu \cdot \kappa = \mu \cdot c$$

for some constant  $c \in \mathbb{R}$ .

#### 2.4. General remarks

As we have seen above in three different example, the main mathematical problem is the same:

$$\text{“In a given vector-set } H = \{v_1, v_2, \dots, v_M\} \in \mathbb{R}^N \text{ we have to find all simplexes } S \subseteq H \text{ (i.e., minimal dependent subsets of } H\text{)”}. \quad (9)$$

Our algorithm, introduced in [13] (and briefly sketched in the next section) solves directly the above computational problem, so can be applied directly for any of the three above practical problems. With some minor modifications we apply it for some other special questions, mainly in study of reaction mechanisms in section 3.2. See also [1] and [2].

Let us remark here that the *size* of the simplexes must *not* be the same, moreover they do not have any connection with the bases of the linear space (of all possible reactions/mechanisms/dimensionless groups). However, the *number* of simplexes in a given set of vectors (in other words, the number of the possible minimal mechanisms/reactions) is only partially solved question, hard in full details. Our recent papers [14] and [15] give almost full answers, the main results and open questions are listed also in the appendix of the present paper, for the readers’ convenience. (The paper [16] generalizes the mathematical notion of simplexes and counts their number not only in linear spaces but in more general structures called *matroids*.) Examples in section 4 show that large sets of vectors in high dimension might have few simplexes only, and small sets in low dimension might have many simplexes. This means, it is very hard to give the *exact* number of simplexes in a given vector-set!

As Prof. Árpád Pethő in [17] shows, *all* solutions (not only minimals) of systems of homogeneous linear equations (0) can be calculated from minimal ones, this was the first case he introduced the notion of simplexes.

Let us mention further that solving the system of linear equations (0) the solution vector can be assumed to have integer coordinates (components) only since (0) has integer coefficients, and we can multiply the rational components of the solution vector by the common denominator.

A general linear algebraic notion (linear functional) is applied to chemical reactions and mechanisms in [9] where it is called the *valuation operator*, this question is investigated in [2].

As sections 2.1 and 2.2 above indicate together, we could build an infinite sequence of **hierarchies**: (1) vectors (species/groups of atoms/functional bonds); (2) solution vectors corresponding to simplexes (i.e., reactions); (3) solution vectors of simplexes of the above vector-set (i.e., mechanisms in our example), **and so on**...! Details and further properties of this *infinite* hierarchy will be investigated in [7] and [3].



### 3. The algorithm and extensions

As we have indicated before, our algorithm, published in [13] solves the mathematical problem (9) in general, so one can apply it directly for finding minimal either reactions or mechanisms or dimensionless groups. After a short description of the algorithm in the next section, we discuss a couple of simple modifications of it for solving other related questions (mainly for studying mechanisms). Computer experiments are discussed in section 4.

#### 3.1. The algorithm

So we are given a set of vectors  $H \subseteq \mathbb{R}^n$  (input) and we have to find all simplexes  $S \subseteq H$ , i.e., minimal dependent subsets of  $H$ .

The only interesting part of the algorithm might be how to generate (= book-keeping + modifying) all *simplexes* of a given set of vectors. (Checking whether a chosen set of vectors is simplex or not is standard.)

We chose the lexicographic enumeration of all (possible) simplexes (subsets) and the “back-and-forth” method for modifying. To store the elements of a subset  $S \subseteq H$  of vectors  $H$  (or equivalently the *indexes* of the vectors) we chose a string (called **simplex**[ ] in our program) with an information-character on the last byte of it. (We put the info character to the last position only for convenience. Also for convenience we labeled the vectors with the characters A, B, . . .)

The last info character of this string (representing the subset  $S \subseteq H$  we just examine for being simplex) is one of the following:

- ‘ ’ (space) –  $S$  is untested,
- ‘i’ – the whole  $S$  is independent,
- ‘d’ – one of the proper subsets of  $S$  is dependent,
- ‘s’ –  $S$  is a simplex.

The procedure which modifies the vector-subset (*procedure modify*) always puts a space to the last info character, is enclosed at the end of the paper [13]. The main program keeps testing whether the new vector-set is a simplex or not and fills out the last info character according to this. After this calls the main program *procedure modify* for modifying the vectors-subset, and solves (0) where the columns of the coefficient matrix are exactly the vectors of the actual simplex. One of the most crucial part of the algorithm’s speed is: how many times do we need to check the linear dependency of a subset  $T \subseteq H$  of the input vector-set  $H$  for finding all the simplexes  $S \subseteq H$ . It is polynomial in the number of the input vectors, as we count it at the beginning of section 4. Moreover, the number of these checks is displayed in each of our examples in section 4, always in the last line of the tables. Perhaps a smart data-handling could avoid repeated checkings of subsets  $T \subseteq H$  but we do not think it would fasten considerably our algorithm.

Now let us see our *procedure modify* in more details. (See also the enclosed program list at the end of [13].)  $N$  denotes the number of the given vectors, while  $S$  will

denote the set of the vectors *we just examine* (or shortly the indices of them only) with the last information character, that is

$$S \subset \{1, 2, \dots, N, ' ', 'i', 'd', 's'\}.$$

Further,  $c$  always denotes an arbitrary character,  $k, t \leq N$  and  $T$  is any subset of  $\{1, 2, \dots, t-1\}$  if  $t$  is fixed. Now, the pseudocode of the main point of our algorithm is the following:

```

S := {1};
while not end do begin
  if S = {k, k + 1, ..., N; c} and c ≠ d then END;
  if S = {k, k + 1, ..., N; 'd'} then S := {k, k + 1, ..., N - 1, N; ' '};
  if S = {T, t, N; c} then S := {T, t + 1; ' '};
  if S = {T, t; 'i'} then S := {T, t, t + 1; ' '};
  if S = {T, t; 'd'} then S := {T, t + 1; ' '};
  if S = {T, t; 's'} then S := {T, t + 1; ' '};
end ;

```

The program does not miss any simplex because, roughly speaking, it checks the (candidate) subsets of the given vector-set in *lexicographical* order.

A working routine in Pascal language is shown at the end of the paper [13], some computational results are listed in section 4 of the present paper.

### 3.2. The extensions

Surprisingly enough a couple of related questions can also be solved by slight modifications *not* of the above algorithm *but*, however, of the input set of vectors only!

We explain here our ideas mainly refereeing for mechanisms, though the algorithm itself and the following ideas can be used in general in any other problem, in *any* linear space.

#### 3.2.0. Reducing the dimension

Using some easy observations we can reduce the *size* of the input and so the running time of our computer.

(a) Clearly any vector (reaction) which is linearly independent of the others must be omitted since no simplexes could contain it. Though any systematic search for *all* these kind of vectors requires considerable time, for huge datasets (about 30 vectors in 30 dimension, i.e., hours of CPU time) this would offer a remarkable time saving, using some minute elementary algebraic computation.

For an outstanding example let us highlight here the case of Methane–Methanol example of [6] (which is also our example 4.6 below) where especially the reactions  $S_9$  and  $S_{11}$  both contain exactly one extra speci (a *single* coordinate), namely  $C_2H_6$  and  $CH_3OCH_3$  respectively, which do *not* occur in any other reactions. This clearly means

that they must be linearly independent of the others. And in fact,  $S_9$  and  $S_{11}$  are not listed at all in table X in [6], which also can be seen in subsection our 4.6 below: table 4.6(a) shows the original example while in table 4.6(b) we can see the effect of omitting  $S_9$  and  $S_{11}$ .

(b) If we have a vector with exactly *two* nonzero coordinates it may also be omitted and further the dimension of all the remaining vectors may be decreased by one.

In the language of *chemistry*: such a vector stands for a reaction of type  $A = \lambda B$  for some (positive) number  $\lambda \in \mathbb{R}$ . This clearly implies that the species (groups of atoms/functional bonds)  $A$  and  $B$  are equivalent. That is, the speci  $A$  can be replaced to  $\lambda B$ , for example, in all the other reactions while the speci  $A$  and the reaction  $A = \lambda B$  can be omitted. This decreases both the number of species (dimension) and of reactions (vectors). Of course to each mechanism  $\mathcal{M}^-$  in this reduced lower dimensional space we might add the reaction  $A = \lambda B$  if the speci  $B$  occurs in  $\mathcal{M}^-$  to form a real mechanism  $\mathcal{M}$  in the original problem. (Of course  $\mathcal{M}$  must *not* be minimal, so if we search for *minimal* mechanisms we might have also to examine its minimality after.)

Let us examine this construction in the language of *mathematics*, too.

Let the vector-set  $H \subseteq \mathbb{R}^N$  be given and suppose that the vector  $\mathbf{X} \in H$  has *exactly two* nonzero coordinates  $x_u = \lambda \cdot x_v$  where  $\lambda \in \mathbb{R} \setminus \{0\}$  is arbitrary nonzero real number and  $\mathbf{X} = [x_1, \dots, x_N]^T$ . For each vector  $\mathbf{Y} \in H$  let us now subtract the  $v$ th coordinate  $\lambda$ -times from the  $u$ th one and delete the  $v$ th coordinate from  $\mathbf{Y}$  to get the vector  $\mathbf{Y}^- \in \mathbb{R}^{N-1}$ . In formula: from the vector  $\mathbf{Y} = [y_1, \dots, y_N]^T \in \mathbb{R}^N$  we construct the vector

$$\mathbf{Y}^- := [y_1, \dots, y_u - \lambda \cdot y_v, \dots, y_{v-1}, y_{v+1}, \dots, y_N]^T \in \mathbb{R}^{N-1}$$

if we suppose  $u < v$ . Clearly  $\mathbf{X}^- = \mathbf{0}$ . Let now

$$H^- := \{\mathbf{Y}^- \mid \mathbf{Y} \in H, \mathbf{Y} \neq \mathbf{X}\} \subseteq \mathbb{R}^{N-1}.$$

**Lemma 3.0.** For any  $S^- \subseteq H^-$  we have that  $S^-$  is linearly independent iff the set  $S \cup \{\mathbf{X}\}$  is linearly independent, where  $S := \{\mathbf{Y} \mid \mathbf{Y}^- \in S^-\}$ .

*Proof.* Let  $S := \{Y_i \mid i \leq t\}$ . Since only the  $u$ th coordinates of the vectors  $\mathbf{Y} \in H$  were changed, and further all other coordinates of  $\mathbf{X}$  were  $\mathbf{0}$  which were not deleted, we have to focus on the  $u$ th coordinates  $y_u^i$  of  $\mathbf{Y}_i$  only. (Recall that the  $u$ th coordinate of  $\mathbf{Y}_i^-$  is  $y_u^i - \lambda \cdot y_v^i$ .) This means, that for any set of coefficients  $\mu_1, \dots, \mu_t \in \mathbb{R}$  we have

$$\begin{aligned} \sum_{i=1}^t \mu_i \mathbf{Y}_i^- = \mathbf{0} & \quad \text{iff} \quad \sum_{i=1}^t \mu_i (y_u^i - \lambda \cdot y_v^i) = 0 & \quad \text{iff} \\ \sum_{i=1}^t \mu_i y_u^i = \lambda \cdot \sum_{i=1}^t \mu_i y_v^i =: c & \quad \text{iff} \quad \sum_{i=1}^t \mu_i \mathbf{Y}_i = \frac{c}{x_u} \cdot \mathbf{X}. & \quad \square \end{aligned}$$

Now let us consider any subset  $S^- \subseteq H^-$ . Using the previous lemma we have that  $S^-$  is simplex iff  $S \cup \{\mathbf{X}\}$  is dependent and  $S \setminus \{\mathbf{Y}\} \cup \{\mathbf{X}\}$  is independent for all  $\mathbf{Y} \in S$ ,  $\mathbf{Y} \neq \mathbf{X}$  iff either  $S \cup \{\mathbf{X}\}$  is a simplex or  $S$  is dependent but in the latter case we must have a simplex  $T \subseteq S$  not containing  $\mathbf{X}$ . Using this observation we can reduce the search for simplexes in  $H$ : we have to search for them among the set  $S \cup \{\mathbf{X}\}$  and all the subsets of  $S$  for every simplex  $S^- \subseteq H^-$ .

Let us emphasize that we did not mention such difficulties above when we spoke in the language of chemistry. There we did not bother of finding *minimal original reactions* in the previous paragraph what caused this difference.

This reduction is illustrated in our example 4.7 which is taken from [6], the example of conversion of Glucose to Pyruvate. The first part of table 4.7 show the original problem<sup>9</sup>, for the second part we applied the above reduction for all the *eight* reaction of type  $A = \lambda B$  of the original problem. For the last part we applied one further reduction for a resulted new reaction of this type. The CPU time decreased from 93.00 to 0.00 sec. (Well, we did not make any computations for decoding – finding simplexes in the original problem from the reduced ones, so far.)

### 3.2.1. Searching for (direct) overall reactions

In this section we consider the case when the resulting (overall) reaction  $\mathbf{X}_{k+1} = \mathcal{R}(\underline{\lambda})$  is *not* known but the lists of the *terminal (chemical) species* (what we want to synthesize and what are the initial ones), and *intermediate (active) species* (which occur only during the mechanism) are given. (For terminology see [4] or our section 2.2 above.) Remark that all species are supposed to be either terminal or intermediate ones.

Our goal now is to find *all* mechanisms (called **steady-state** ones) of which corresponding reactions contain terminals (chemical) species only, these reactions are called **overall** ones. Again, our algorithm provides minimal mechanisms (i.e., **direct** or **cycle-free** ones) with their minimal (i.e., simple) reactions.

(i) For this purpose enlarge first the given set of vectors  $\{\mathbf{X}_1, \dots, \mathbf{X}_k\} \subseteq \mathbb{R}^N$  with new “ideal” vectors  $\mathbf{V}_t \in \mathbb{R}^N$  to separate terminate and intermediate species: one new vector  $V_t$  for each terminate species  $A_t$  where  $t \in T$ . In more detail, let all but the  $t$ th coordinates of  $V_t$  be equal to 0 while let the  $t$ th coordinate (representing  $A_t$ ) of  $V_t$  be equal to 1.

Now, from *any* mechanism

$$\sum_{j=1}^k \lambda_j \cdot X_j + \sum_{t \in T} \mu_t \cdot V_t = \mathbf{0}, \quad (10)$$

we can extract the overall reaction

$$R := - \sum_{t \in T} \mu_t \cdot V_t, \quad (11)$$

<sup>9</sup> Section 3.2.1 explains the differences among these three parts.

where, as usual, the initial species are with negative coefficients and the synthesised ones with positive coefficients. Clearly now the mechanism

$$\mathcal{M} := \sum_{j=1}^k \lambda_j \cdot X_j \quad (12)$$

has exactly the resulting (overall) reaction

$$\mathcal{R}(\mathcal{M}) = \mathbf{R}.$$

Of course we have to consider only the simplexes of the extended vector-set

$$S \subseteq \{\mathbf{X}_1, \dots, \mathbf{X}_k\} \cup \{V_t \mid t \in T\} \quad (13)$$

which contain at least one new vector  $V_t$ . Otherwise we would get mechanisms  $\mathcal{M}$  resulting the void reaction  $\mathcal{R}(\mathcal{M}) = \mathbf{0}$  (as without introducing the new vectors  $\{V_t \mid t \in T\}$ ).

We have to find *simplexes*  $S$  in the set (13) since we are interested in *minimal* mechanisms. This might be an essential restriction for finding *all* mechanisms. In (iii) below we discuss another, two-stage solution for finding *all* minimal (direct) overall reactions and minimal mechanisms resulting the corresponding overall reaction, though the method we just presented is a direct, *one-stage* one for the same problem. In section 3.3 we give an exact mathematical proof for that the present method and the one in (iii) give the *same* output. In other words, the above method based on (12) and (13) really gives *all* minimal reactions and their minimal mechanisms. So we can use our algorithm automatically in *one stage* as we discussed above in (12) and (13), which means a fast solution to our problem.

After we have extended our vector-set  $\{\mathbf{X}_i: i \leq k\}$  with the new vectors  $\{\mathbf{V}_j: j \in T\}$  to the set

$$H := \{\mathbf{X}_i: i \leq k\} \cup \{\mathbf{V}_j: j \in T\}$$

we may run our algorithm either for finding *all* simplexes in  $H$  (**VarAll**) or *only* those which contain at least one of the new vectors  $\{\mathbf{V}_j: j \in T\}$  (**VarOnly**) and we can compare these runs to when we searched for all simplexes in the original vector-set  $\{\mathbf{X}_i: i \leq k\}$  (**VarOrig**). Clearly for all computational quantities  $\nu$  (time, number of simplexes, checked subsets, etc.) we must have

$$\nu(\text{VarAll}) = \nu(\text{VarOnly}) + \nu(\text{VarOrig}). \quad (14)$$

This easy fact can be also seen in each of in tables 4.5 through 4.7 in section 4.

(ii) We must not forget the *law of mass balance (remaining of the material)* in the mechanism  $\mathcal{M}$ , which results the reaction  $\mathbf{R} = \mathcal{R}(\mathcal{M}) = -\sum_{t \in T} \mu_t \cdot V_t$  by (11).

This is not a problem from *chemical* point of view: since all the given reactions (input vectors) satisfy this law, their linear combination satisfies, too, since by (10) we have  $\mathbf{R} = \sum_{j=1}^k \lambda_j \cdot X_j$ . (In other words, this law holds automatically for the output mechanism assuming it hold for all of the input reactions.)

From *mathematical* point of view: this law is equivalent to that the vector  $\mathbf{R} \in \mathbb{R}^N$  must satisfy the requirement

$$\mathbf{B} \cdot \mathbf{R} = \mathbf{0} \quad (15)$$

for the matrix  $\mathbf{B} \in \mathbb{R}^{m \times N}$  which codes the sum-formulae of *all* the species (groups of atoms/functional bonds) involved in our problems: in any of the input vectors  $\{\mathbf{X}_i: i \leq k\}$  did it, as we described in section 2.1. The equality (15) might be curious if we used only (11) as  $\mathbf{R} = -\sum_{t \in T} \mu_t \cdot V_t$  but using

$$\mathbf{R} = \sum_{j=1}^k \lambda_j \cdot X_j$$

from (10) and the assumption

$$\mathbf{B} \cdot X_j = \mathbf{0} \quad (\forall j \leq k)$$

we clearly have

$$\mathbf{B} \cdot \mathbf{R} = \sum_{j=1}^k \lambda_j \mathbf{B} X_j = \mathbf{0}.$$

Not only for still unconvinced readers but for an additional explanation of the proof in section 3.3, let us mention the next variant for ensuring the law of the mass balance. Let us first enlarge the dimension of all the vectors  $\{\mathbf{X}_i: i \leq k\}$  and  $\{V_j: j \leq T\}$  with as many new coordinates as many atoms the vectors  $\{V_j: j \leq T\}$  (i.e., the *terminate* species) are built up. Next, let us code the sum-formula of these new vectors  $V_j$  in their own new coordinates as we described in section 3.2.1, while we let all the new coordinates of the old vectors  $\{\mathbf{X}_1, \dots, \mathbf{X}_k\}$  to be equal to 0. We do *not* think it would be interesting to present such computational runs which (of course) gave the same results but in some more time!

Let us emphasize again that running our algorithm with the modified dataset (as in (i)) we immediately get (in one run) *also* the steady-state *mechanism*

$$\mathcal{M} := \sum_{j=1}^k \lambda_j \cdot X_j$$

with  $\mathcal{R}(\mathcal{M}) = R$ , that is the mechanisms which yield exactly the reaction  $R$ . Since our algorithm lists *all* simplexes of the vector-set  $\{\mathbf{X}_i: i \leq k\} \cup \{\mathbf{V}_j: j \in T\}$  in its full run, we have only to separate all mechanisms corresponding to several reactions, so we get *all* (direct) steady-state mechanisms with *all* their (simple) overall reactions in a single run. This means that in the present case there is *no* need for further running the variant described in section 3.2.2.

(iii) *Another variant* to the problem of finding all minimal overall reactions and minimal (direct) mechanisms leading to them would be the following. Let us first search *all* the possible minimal reactions among the terminate species as it is described in section 3.2.1. Then for each minimal reaction let us search separately the minimal (di-

rect/steady state) mechanisms among the original given set of reactions, which mechanisms belong to the actual minimal reaction as described in the next section 3.2.2.

This idea requires *several* lower dimensional runs of our algorithm (and transferring data among them) instead of a single higher dimensional one as we suggested in (i). Let us warn however our readers that in the preliminary search for all the possible minimal reactions among the terminate species we may get also reactions which can *not* be built up from the original given set of reactions and so imply void runs in the next step!

Computational examples for these comparisons are presented in sections 4.5 through 4.7. In section 4.5 we explain the most details: the first column of table 4.5 examines the possible reactions among the terminate species (mentioned in (i) above) but we did not any use of this computation for the other columns.<sup>10</sup> Column 2 investigates all possible mechanisms among the original reactions. Columns 3 and 4 treat with introducing the new vectors  $\{V_t: t \in T\}$  only (both without and with halting the programme after leaving all new vectors).

### 3.2.2. (Direct) steady state mechanisms

In this subsection we deal with the case when we *do* know already one or more resulting (overall) reactions (determined either by chemical or other mathematical method) to which we want to determine all the (minimal) mechanisms leading to these given reactions.

Let us deal first with the case when we are given a single resulting reaction  $R$ . For our purpose let us first extend the given vector-set

$$H := \{\mathbf{X}_1, \dots, \mathbf{X}_k\}$$

with the new vector  $\mathbf{X}_{k+1} := R$  and then let make our algorithm list *only* the simplexes containing  $R$ .

Since any mechanism (3) which results the reaction  $R$  can be trivially transformed to the vanishing linear combination

$$\sum_{j=1}^{k+1} \lambda_j \cdot X_j = \mathbf{0}$$

(which is called a *cycle* in [5]), we should only subtract  $\mathbf{X}_{k+1}$  from the above equality and the remaining mechanism

$$\mathcal{M} := -\frac{1}{\lambda_{k+1}} \sum_{j=1}^k \lambda_j \cdot X_j$$

would certainly result the reaction  $R$ , that is  $\mathcal{R}(\mathcal{M}) = R$ .  $\mathcal{M}$  is minimal of course.

<sup>10</sup> So far we did not made any computer experiments for the idea presented in (iii) since all our computations concerning mechanisms were below 25 minutes, the long-running example in section 4.4 is for all the minimal mechanism where the resulting (overall) reaction is already given.

Since we are looking only for simplexes which do contain the fixed vector  $R = \mathbf{X}_{k+1}$ , we have to apply our bounding formulas (see appendix) for  $k$  many vectors only! In other words, the computing time decreases by a factor of approximately  $(k - N)/(k + 1) = 1 - (N + 1)/(k + 1)$  when we search for simplexes containing  $\mathbf{R}$  (as a fixed element) instead of containing all possible subsets of  $H \cup \{R\}$ , which is a  $(k + 1)$ -element subset of  $\mathbb{R}^N$ .

The above trick can be extended easily to the case when we are given more than one fixed resulting (overall) reactions  $\mathbf{R}_1, \dots, \mathbf{R}_t$  at the same time. We simply have to add these vectors to  $H$  and we have to consider only the simplexes

$$S \subseteq H \cup \{\mathbf{R}_1, \dots, \mathbf{R}_t\}$$

which contain *exactly one* of the new vectors  $\{\mathbf{R}_1, \dots, \mathbf{R}_t\}$ . Clearly this parallel computing is advised for small  $t$  only since we are interested in simplexes  $S$  with this property and checking that  $|S \cap \{\mathbf{R}_1, \dots, \mathbf{R}_t\}| = 1$  would require some time for large  $t$ . For large  $t$ , which has magnitude of  $k$ , we recommend to run this modified algorithm separately for each set  $H \cup \{\mathbf{R}_i\}$  for all  $i \leq t$ .

### 3.2.3. Neither reactions nor terminate species are known

We can also handle the case when *no* terminal (chemical) species are selected at all at the beginning but we want to find all overall reactions. By our algorithm we search for all the possible simplexes and compute (all) the corresponding resulting (overall) reactions which give the answer. Observe however that in the meantime we have already listed the corresponding overall mechanisms, too. This means that there is no need for any further computing.

### 3.3. Proof of the equivalence

In this subsection we give a (mathematical) proof for the equivalence of the variations of our algorithm given in (i) and in (iii) in section 3.2.1. Namely, we justify that the same set of *all* direct overall reactions and their minimal mechanisms can be obtained either with the one-stage method described in (i) or the two-stage variant introduced in (iii) in section 3.2.1.

So, we are given the (arbitrary) vector-set

$$\{\mathbf{X}_1, \dots, \mathbf{X}_k\} \subseteq \mathbb{R}^N.$$

Let further

$$\{\mathbf{V}_1, \dots, \mathbf{V}_t\} \subseteq \mathbb{R}^N$$

be *any* set of linearly independent set of vectors where  $t \leq N$ . Let us denote the set of all these vectors by  $H$ , i.e., let

$$H := \{\mathbf{X}_j: j \leq k\} \cup \{\mathbf{V}_i: i \leq t\}.$$

We have to prove the equivalence of the simplexes of the vector-set  $H$  and those of  $\{\mathbf{X}_j: j \leq k\} \cup \{\mathbf{S}_R\}$  for *certain* vector  $\mathbf{S}_R \in \mathbb{R}^N$ .



Let us start with an easy but general fact about simplexes.

**Lemma 3.1.** Any set of vectors  $U = \{u_1, \dots, u_m, v\}$  is a simplex if and only if the set  $\{u_1, \dots, u_m\}$  is linearly independent and all the coefficients  $\alpha_i$  in the existing equality

$$\mathbf{v} = \sum_{i=1}^m \alpha_i u_i \quad (16)$$

are different from zero.

*Proof.* If the set  $U \setminus \{u_{i_0}\}$  would be dependent for some  $i_0 \leq m$  then we would have

$$\beta_v \cdot \mathbf{v} + \sum_{i \neq i_0} \beta_i u_i = \mathbf{0}$$

and  $\beta_v \neq 0$  since the set  $\{u_1, \dots, u_m\}$  was assumed to be linearly independent. This implies

$$\mathbf{v} = \sum_{i \neq i_0} \frac{\beta_i}{\beta_v} u_i$$

which contradicts to (16) since the coefficient of  $u_{i_0}$  in the last linear combination is 0 while all the coefficients of  $u_i$  for each  $i \leq m$  constructing  $v$  are unique since the set  $\{u_1, \dots, u_m\}$  is assumed to be linearly independent.  $\square$

**Theorem 3.2.** For any simplex  $S \subseteq H$ ,  $S = \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{V}_i: i \in T\}$  where  $K \subseteq \{1, \dots, k\}$  and  $T \subseteq \{1, \dots, t\}$  are nonempty subsets, and for the vector

$$\mathbf{S}_R := - \sum_{j \in K} \mu_j \cdot \mathbf{X}_j = \sum_{i \in T} \lambda_i \cdot \mathbf{V}_i \quad (17)$$

(for the suitable coefficients  $\mu_j, \lambda_i \in \mathbb{R}$ ) the set  $S' := \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{S}_R\}$  is also a simplex.

*Proof.* Since  $S$  is minimal dependent, all the coefficients in (17) must be different from 0. Then lemma 3.1 completes the proof.  $\square$

The above theorem ensures that the method we described in (i) of 3.2.1 finds all simplexes (solutions) what the other variation in (iii) has found.

We will prove the other direction in theorem 3.3, what is not true for any set  $H$  of vectors and simplex  $S' = \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{S}_R\}$ , we also need the additional assumptions (first) through (fifth) below.

First we need that all the vectors  $\mathbf{X}_j$  ( $j \leq k$ ) satisfy

$$\mathbf{B} \cdot \mathbf{X}_j = \mathbf{0} \quad (\forall j \leq k) \quad (18)$$

for a given (arbitrary) matrix<sup>11</sup>  $\mathbf{B} \in \mathbb{R}^{m \times N}$ , though this is not required for the vectors  $\mathbf{V}_i$  ( $i \leq t$ )! Denote now

$$\{\mathbf{b}_i: i \leq N\} \subseteq \mathbb{R}^m$$

the set of column-vectors of  $\mathbf{B}$ .

*Second* we need that the vectors  $\mathbf{V}_i$  ( $i \leq t$ ) are the first  $t$  standard base vectors of  $\mathbb{R}^N$ , that is all but the  $i$ th coordinates of  $\mathbf{V}_i$  are equal to 0 while the  $i$ th coordinate equals to 1.

*Third* we have to observe the connection

$$\mathbf{B} \cdot \mathbf{V}_i = \mathbf{b}_i \quad (\forall i \leq t)$$

which is a trivial consequence of second.

Recall now that we want to prove that each solution (minimal overall reaction and direct mechanisms) received from the variant (iii) of our algorithm (in section 3.2.1) will also be provided by the variant in (i) *ibid.* This implies the following assumptions we may use:

*Fourth* let the vector  $\mathbf{S}_R$  in (17) and let the set  $S' = \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{S}_R\}$  be a simplex. By (iii) of section 3.2.1 we also have that the set

$$\{\mathbf{b}_i: i \in T\} \subseteq \mathbb{R}^m$$

is also a simplex!

Let us mention that by (17) we have

$$\mathbf{0} = \mathbf{B} \cdot \sum_{j \in K} \mu_j \mathbf{X}_j = \sum_{i \in T} \lambda_i \mathbf{B} \cdot \mathbf{V}_i = \sum_{i \in T} \lambda_i \mathbf{b}_i \quad (19)$$

which implies that  $\{\mathbf{b}_i: i \in T\}$  is dependent for *any* vector  $\mathbf{S}_R$  satisfying (17). The above result (fourth) implies the below one which will be useful in our proof:

*Fifth*, the above set of indices  $T$  satisfies the following minimality property: “ $T \subseteq \{1, \dots, t\}$  is minimal in the sense that there is no proper subset  $T' \subsetneq T$  for which

$$\sum_{j \in L} \mu'_j \mathbf{X}_j = \sum_{i \in T} \lambda'_i \mathbf{V}_i \quad (*)$$

would hold for some  $L \subseteq \{1, \dots, k\}$ ”.

Now we are ready to prove the converse of theorem 3.2. □

**Theorem 3.3.** Let  $S' = \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{S}_R\} \subseteq \mathbb{R}^N$  be any simplex where  $\mathbf{S}_R$  satisfies (17), further the above assumptions (first) through (fifth) hold. Then the set

$$S := \{\mathbf{X}_j: j \in K\} \cup \{\mathbf{V}_i: i \in T\}$$

is also a simplex.

<sup>11</sup> For chemists:  $\mathbf{B}$  codes the sum-formulae of *all* the species involved in the input reactions  $\mathbf{X}_j$  ( $j \leq k$ ) as we described in section 2.1.

*Proof.*  $S$  is clearly dependent by (17). The sets  $S \setminus \{\mathbf{V}_{i_0}\}$  for each  $i_{i_0} \in T$  are independent using the minimality property (\*) for  $T$ , since (fourth) implies that  $T$  is nonempty.

Let now  $j_0 \in K$  be arbitrary and suppose by contradiction that the set  $S \setminus \{\mathbf{X}_{j_0}\}$  is dependent. Let

$$\mathbf{S}_Q := - \sum_{j \in K} \mu'_j \mathbf{X}_j = \sum_{i \in T} \lambda'_i \mathbf{V}_i \quad (20)$$

for some  $\mu'_j, \lambda'_i \in \mathbb{R}$  where  $\mu'_{j_0} = 0$ . Arguing now as in (19) we get

$$\mathbf{0} = -\mathbf{B} \cdot \sum_{j \in K} \mu'_j \mathbf{X}_j = \sum_{i \in T} \lambda'_i \mathbf{B} \cdot \mathbf{V}_i = \sum_{i \in T} \lambda'_i \mathbf{b}_i. \quad (21)$$

Since the set  $\{\mathbf{b}_i: i \in T\}$  is simplex, all solutions of the homogeneous equality

$$\mathbf{0} = \sum_{i \in T} \gamma_i \mathbf{b}_i$$

are parallel, this is another easy characterization of simplexes. For our equalities in (19) and in (21) this implies that the coefficient vectors  $[\lambda_i: i \in T]^T$  and  $[\lambda'_i: i \in T]^T$  are parallel. This would imply by (17) and (20) that the vectors  $S_Q$  were  $S_R$  also parallel, i.e.,  $S_Q = \tau \cdot S_R$  for some  $\tau \in \mathbb{R}$ . Using (17) and (20) again this leads to a contradiction since the vectors  $\{\mathbf{X}_j: j \in K\}$  were supposed to be independent, moreover  $\mu_{j_0} = 0$  but also *none* of the coefficients  $\mu_j$  was 0, finally  $S_Q$  may not be the zero vector.  $\square$

The above theorem clearly justifies that each solution (minimal overall reaction and direct mechanisms) produced by the variant (iii) of section 3.2.1 is also given by the other one in (i) of the same subsection.

#### 4. Computer experiments

An easy analysis of the algorithm shows that our algorithm runs for  $\mathcal{O}(M^{n+1})$  time<sup>12</sup> (in the worst case) since all examined subsets of the  $M$  vectors have size at most  $n + 1$ , the *spanned* dimension (= range) + 1 of the input vector-set.

This is *polynomial* time in  $M$ , the number of vectors. The results in the appendix show that the maximal (possible) number of simplexes (= the size of the output) is really of this magnitude. According to this estimation everyday size inputs (some dozen of vectors in 10–20-dimensional spaces) require some seconds only on modern computers. Concrete computational experiments on (also large) datasets are shown below. We used a Packard–Bell PC with Pentium II processor of 360 MHz, in Turbo Pascal 6.0 in DOS mode.

The computing results of Happel, Otarod and Sellers [6] and Bertók [10] require also about this running time but their results are not better at all than of ours.

<sup>12</sup>For any functions  $f, g: \mathbb{N} \rightarrow \mathbb{N}$  we say that  $f = \mathcal{O}(g)$  (“big oh g”) iff for some positive constants  $c_1, c_2 \in \mathbb{R}$  we have  $c_1 < f(n)/g(n) < c_2$ , that is  $c_1 \cdot g(n) < f(n) < c_2 \cdot g(n)$  for all  $n \in \mathbb{N}$ .

When a so-called “resulting” (overall) reaction  $S_R$  is given (which we have build from a linear combination of the other vectors) we made run our program in two ways: we computed *all* simplexes (not regarding the extra role of  $S_R$ ) and we also computed the mechanisms resulting the (overall) reaction  $S_R$  (i.e., the simplexes *only* which contain  $S_R$ , see the section 3.2.2). Checking this latter requirement in each step slowed down slightly our computer but our formula

$$v(\text{VarAll}) = v(\text{VarOnly}) + v(\text{VarOrig})$$

from (14) in section 3.2.1(i) is justified now.

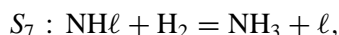
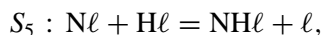
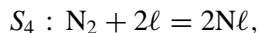
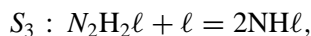
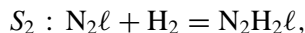
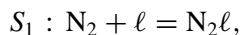
**4.1.** Our first example is taken from Amundson [20] and also can be found in [7]. We have given the groups of atoms (functional bonds) CO, CO<sub>2</sub>, O<sub>2</sub>, H<sub>2</sub>, CH<sub>2</sub>O, CH<sub>3</sub>OH, C<sub>2</sub>H<sub>5</sub>OH, (CH<sub>3</sub>)<sub>2</sub>CO, CH<sub>4</sub>, CH<sub>3</sub>CHO, H<sub>2</sub>O (of course there is *no* resulting (overall) reaction now). The 213 minimal reactions (simplexes) we also get by our computation, are listed in detail, e.g., in [7]. Since we use three atoms C, O, H (i.e., the input consists of 3-dimensional vectors) and there are no parallel vectors among the species we can use the sharper lower bound from corollary A.4.

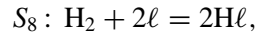
Table 4.1  
“Amundson”

$N$ (dimension of the vectorspace)	3
$n$ (dimension of what $H$ spans)	3
$M$ (number of input vectors: $ H $ )	11
$\text{simp}(H)$ (number of simplexes)	213
$1 + \binom{M-2}{3} + \binom{M-3}{2}$ (lower bound <sup>a</sup> )	$113 \leq$
$\binom{M}{n+1}$ (upper bound)	$\leq 330$
$t$ (computational time [sec])	0.22 s
number of checked subsets of $H$	502

<sup>a</sup> See appendix of the present paper.

**4.2.** Our next example is called “Ammonia” in [6] and it is the 4th example in [10]. The resulting (overall) and the examined (“*possible elementary*”) reactions are:





where  $\ell$  denotes the catalysator's surface.

The minimal mechanisms are

$$1) 3S_1 + 3S_2 + 3S_3 - 2S_4 - 4S_5 + 2S_6 + 2S_9 = S_R,$$

$$2) S_1 + S_2 + S_3 + 2S_6 + 2S_8 + 2S_9 = S_R,$$

$$3) S_1 + S_2 + S_3 + 2S_7 = S_R,$$

$$4) S_4 + 2S_5 - S_6 + 3S_7 - S_9 = S_R,$$

$$5) S_4 + 2S_5 + 2S_6 + 3S_8 + 2S_9 = S_R,$$

$$6) S_4 + 2S_5 + 2S_7 + S_8 = S_R,$$

$$7) -S_1 - S_2 - S_3 + S_4 + 2S_5 - S_6 + S_7 - S_9 = 0,$$

$$8) S_1 + S_2 + S_3 - S_4 - 2S_5 - S_8 = 0,$$

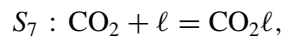
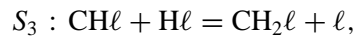
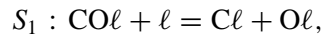
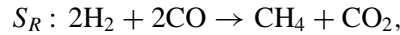
$$9) -S_6 + S_7 - S_8 - S_9 = 0.$$

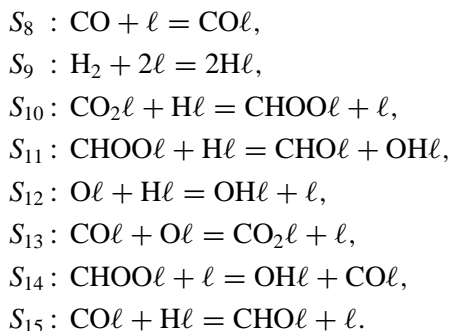
(The latter three mechanisms yield not  $S_R$  but the zero vector – a cycle.)

Table 4.2  
“Ammonia”

	Total	Containing $S_R$ only
$N$ (dimension of the vectorspace)	10	10
$n$ (dimension of what $H$ spans)	7	7
$M$ (number of input vectors: $ H $ )	10	10
$\text{simp}(H)$ (number of simplexes)	9	6
$b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$ (lower bound)	$3 \leq$	$1 \leq$
$\binom{M}{n+1}$ (upper bound)	$\leq 45$	$\leq 36$
$t$ (computational time [sec])	0.44 s	0.28 s
number of checked subsets of $H$	969	473

**4.3.** Our next example is Bertók's 5th example from [10] which is taken from [4]. The resulting (overall) and the examined (*possible elementary*) reactions are:





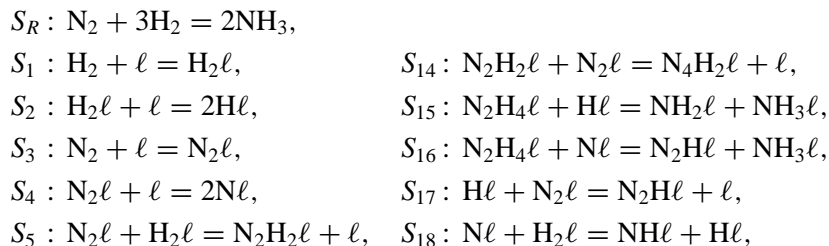
Now all the minimal mechanisms (the output) are:

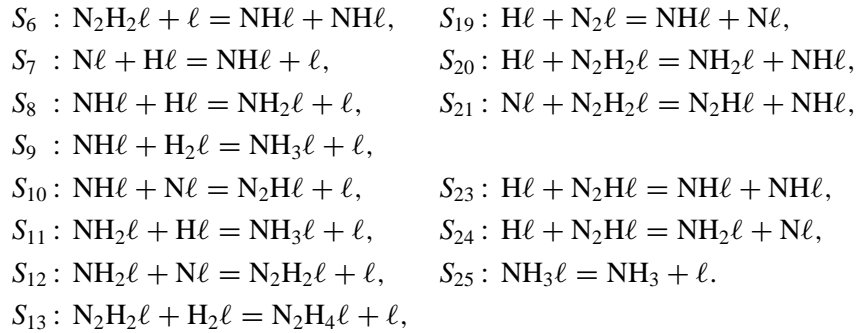
- 1)  $S_1 + S_2 + S_3 + S_4 + S_5 - S_7 + 2S_8 + 2S_9 - S_{10} - S_{11} + S_{12} + S_{15} = S_R$ ,
- 2)  $S_1 + S_2 + S_3 + S_4 + S_5 - S_7 + 2S_8 + 2S_9 - S_{10} + S_{12} - S_{14} = S_R$ ,
- 3)  $S_1 + S_2 + S_3 + S_4 + S_5 - S_7 + 2S_8 + 2S_9 + S_{13} = S_R$ ,
- 4)  $S_{10} + S_{11} - S_{12} + S_{13} - S_{15} = 0$ ,
- 5)  $S_{10} - S_{12} + S_{13} + S_{14} = 0$ ,
- 6)  $S_{11} - S_{14} - S_{15} = 0$ .

Table 4.3  
"Bertók\_5"

	Total	Containing $S_R$ only
$N$ (dimension of the vectorspace)	17	17
$n$ (dimension of what $H$ spans)	13	13
$M$ (number of input vectors: $ H $ )	16	16
$\text{simp}(H)$ (number of simplexes)	6	3
$b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$ (lower bound)	$4 \leq$	$1 \leq$
$\binom{M}{n+1}$ (upper bound)	$\leq 120$	$\leq 105$
$t$ (computational time [sec])	78.60 s	43.28 s
number of checked subsets of $H$	63,429	31,697

**4.4.** This example was provided for us by Mr. Bertók personally as `_rn.in` and was presented in [18]. The resulting (overall) and the examined (*possible elementary*) reactions now are ( $S_{22}$  is omitted by technical purposes):





Our calculations are summarized in the below table.

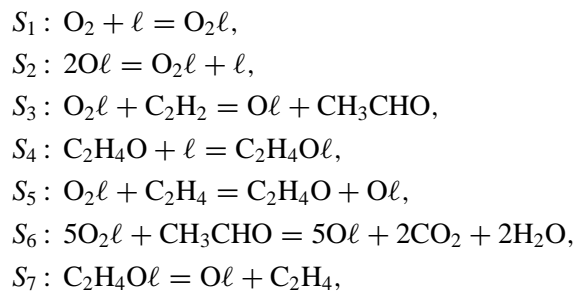
Table 4.4  
"Bertók\_rn.in"

	Total	Containing $S_R$ only
$N$ (dimension of the vectorspace)	15	15
$n$ (dimension of what $H$ spans)	14	14
$M$ (number of input vectors: $ H $ )	25	25
$\text{simp}(H)$ (number of simplexes)	5,609	3,585
$b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$ (lower bound)	$11 \leq$	$1 \leq$
$\binom{M}{n+1}$ (upper bound)	$\leq 3,268,760$	$\leq 1,961,256$
$t$ (computational time [sec])	$2.1 \cdot 10^4$ s	$1.2 \cdot 10^4$ s
	$\approx 5$ h 50 min	$\approx 3$ h 21 min
number of checked subsets of $H$	10,664,430	2,846,629

Let us mention that Bertók achieved the same list of simplexes as ours but in 13 hours computer run comparing to our 3 hours 21 minutes.

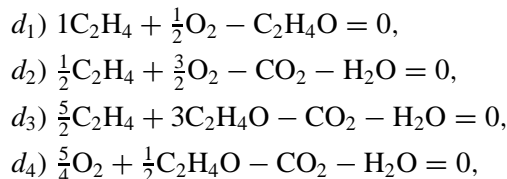
The following three examples illustrates the modifications described in section 3.2.

**4.5.** This example is introduced [6] as **Ethylene Oxide Synthesis**. We are given the following reactions



where the terminal (chemical) species are  $C_2H_4O$ ,  $C_2H_4$ ,  $O_2$ ,  $CO_2$  and  $H_2O$ , all the others are intermediate (active) ones.

As we indicated at the end of section 3.2.1 we can find all (possible) direct overall reactions among the above set of terminal species by our algorithm (as detailed in section 2.1). We got the following (complete) list of minimal reactions



in 0.00 sec which corresponds (in order) to  $d_1, d_3, d_4, d_2$ , of [6], respectively. Computational data concerning this run are shown in the first column of table 4.5.

The second column of this table shows searching for mechanisms (resulting the zero vector, i.e., the void reaction) among the original vectors (reactions  $S_1$  through  $S_7$ ). The single mechanisms we found shows that the reactions  $S_1, \dots, S_7$  are *not* linearly independent.

Columns 3 and 4 show computation when we introduced the new vectors  $V_1, \dots, V_5$  representing one-to-one the terminal (chemical) species as we suggested in the first part of section 3.2.1. For comparison of CPU time we run the algorithm first for *all* simplexes then for those which contain at least one new vector  $V_i$  only.

We have the following list of minimal mechanisms:

$$\begin{aligned} m_1 : & \frac{7}{6}\text{C}_2\text{H}_4 + \text{O}_2 - \text{C}_2\text{H}_4\text{O} - \frac{1}{3}\text{CO}_2 - \frac{1}{3}\text{H}_2\text{O} - S_1 + \frac{1}{6}S_3 - S_4 \\ & + \frac{1}{6}S_6 - S_7 = 0, \\ m_2 : & \text{C}_2\text{H}_4 + \frac{1}{2}\text{O}_2 - \text{C}_2\text{H}_4\text{O} + \frac{1}{2}S_1 - \frac{1}{2}S_2 - S_4 - S_7 = 0, \\ m_3 : & -\text{C}_2\text{H}_4 - \frac{1}{2}\text{O}_2 + \text{C}_2\text{H}_4\text{O} - \frac{1}{2}S_1 - \frac{1}{2}S_2 - S_5 = 0, \\ m_4 : & 2\text{C}_2\text{H}_4 + \text{O}_2 - 2\text{C}_2\text{H}_4\text{O} + S_1 - S_4 + S_5 - S_7 = 0, \\ m_5 : & -\text{C}_2\text{H}_4 - 3\text{O}_2 + 2\text{CO}_2 + 2\text{H}_2\text{O} - 3S_1 - 3S_2 - S_3 - S_6 = 0, \\ m_6 : & \frac{1}{3}\text{C}_2\text{H}_4 + \text{O}_2 - \frac{2}{3}\text{CO}_2 - \frac{2}{3}\text{H}_2\text{O} + S_1 + \frac{1}{3}S_3 - S_4 - S_5 + \frac{1}{3}S_6 - S_7 = 0, \\ m_7 : & \frac{5}{6}\text{C}_2\text{H}_4 - \text{C}_2\text{H}_4\text{O} + \frac{1}{3}\text{CO}_2 + \frac{1}{3}\text{H}_2\text{O} - S_2 - \frac{1}{6}S_3 - S_4 - \frac{1}{6}S_6 - S_7 = 0, \\ m_8 : & 5\text{C}_2\text{H}_4 - 6\text{C}_2\text{H}_4\text{O} + 2\text{CO}_2 + 2\text{H}_2\text{O} - S_3 + 6S_5 - S_6 = 0, \\ m_9 : & -\frac{5}{2}\text{O}_2 - \text{C}_2\text{H}_4\text{O} + 2\text{CO}_2 + 2\text{H}_2\text{O} - \frac{5}{2}S_1 - \frac{7}{2}S_2 - S_3 - S_4 - S_6 - S_7 = 0, \\ m_{10} : & -\frac{5}{2}\text{O}_2 - \text{C}_2\text{H}_4\text{O} + 2\text{CO}_2 + 2\text{H}_2\text{O} - \frac{5}{2}S_1 - \frac{5}{2}S_2 - S_3 + S_5 - S_6 = 0, \\ m_{11} : & \text{O}_2 + \frac{2}{5}\text{C}_2\text{H}_4\text{O} - \frac{4}{5}\text{CO}_2 - \frac{4}{5}\text{H}_2\text{O} + S_1 + \frac{2}{5}S_3 - S_4 - \frac{7}{5}S_5 \\ & + \frac{2}{5}S_6 - S_7 = 0, \\ m_{12} : & -S_2 - S_4 - S_5 - S_7 = 0. \end{aligned}$$

For comparing our above data to table VII of [6], let us remark that that table contains mechanisms *only* for the minimal (direct) reactions  $d_1$  and  $d_3$  and, moreover, its rows  $(m_2, d_3)$  and  $(m_3, d_3)$  are identical.



Table 4.5  
"Ethylene Oxid"

	Terminal species	Reactions only	With fictive vectors $V_i$	
			ALL simplexes	WITH $V_i$ only
$N$	3	10	10	10
$n$	3	6	9	9
$M$	5	7	12	12
$\text{simp}(H)$	4	1	12	11
$LB$	$2 \leq$	$1 \leq$	$3 \leq$	
$UB$	$\leq 5$	$\leq 1$	$\leq 66$	
$t$	0.00 s	0.06 s	1.87 s	1.80 s
$chk$	18	102	4,000	3,898

$N$  = dimension of the vectorspace,

$M$  = number of input vectors =  $|H|$ ,

$LB = b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$  (lower bound),

$t$  = computational time [sec],

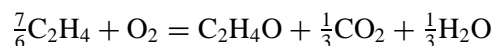
$n$  = dimension of what  $H$  spans,

$\text{simp}(H)$  = number of simplexes,

$UB = \binom{M}{n+1}$  (upper bound, if applicable),

$chk$  = number of checked subsets of  $H$ .

Further, the reaction

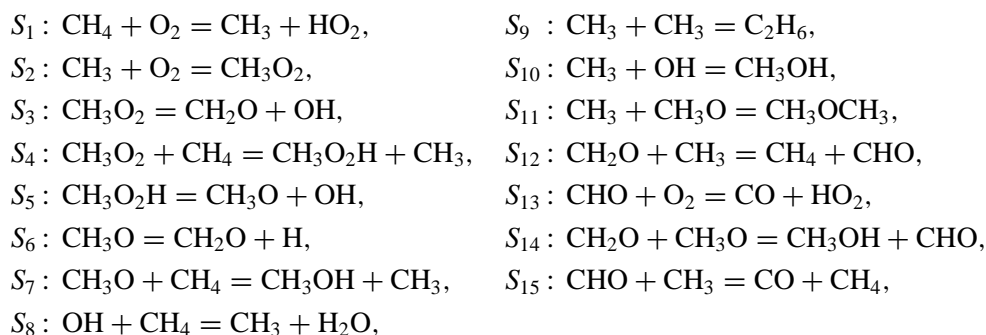


which can be extracted from our mechanism  $m_1$  above is *not* minimal, and in fact it is not listed among the minimal (direct) reactions  $d_1, d_2, d_3, d_4$  above we obtained by running our algorithm with the sum formulas of the terminal (chemical) species. The explanation is, that the vectors  $V_1, \dots, V_5$  are linearly independent which makes the vectors occurring in mechanism  $V_1$  a *simplex*!

We computed in the table 4.5 the lower and upper bounds ( $LB, UB$ ) according to the formulas of the appendix of the present paper but only in the case it is applicable (i.e., except from column 4).

Again, our formula  $\nu(\text{VarAll}) = \nu(\text{VarOnly}) + \nu(\text{VarOrig})$  from (14) in section 3.2.1(i) can be checked in table 4.5.

**4.6.** Our next example is introduced in [6] as **Methane to Methanol Conversion**. We are now given the reactions



where the terminal (chemical) species are  $\text{CH}_4, \text{O}_2, \text{CH}_3\text{OH}, \text{CO}$  and  $\text{H}_2\text{O}$ .

Again the first column of the table below compute all the direct overall reactions among terminal species while the second column deals with the (void) mechanisms among the given reactions  $S_1$  through  $S_{15}$ .

As we indicated in section 3.2.0, since only the reactions  $S_9$  and  $S_{11}$  contain the extra species  $C_2H_6$  and  $CH_3OCH$ , the vectors representing them are linearly independent from the others, so we can omit them and so we can reduce the number and the dimension of the problem by two. For comparison we made computer runs both with the original and with the reduced vectors: these are the two parts of the table below. This reduction helped us to decrease the CPU time from 27 min to 5 min!

The columns in both parts of the table show the various computations as in our previous example 4.5 we explained (the first columns in the two parts are the same).

Table 4.6  
"Methanol"

Before the reduction:				
	Terminal species	Reactions only	With fictive vectors $V_i$	
			ALL simplexes	WITH $V_i$ only
$N$	3	16	16	16
$n$	3	13	16	16
$M$	5	15	20	20
$simp(H)$	4	2	24	22
$LB$	$2 \leq$	$2 \leq$	$4 \leq$	
$UB$	$\leq 5$	$\leq 15$	$\leq 1140$	
$t$	0.00 s	30.38 s	1353 s $\approx$ 22 m	1323 s $\approx$ 22 m
$chk$	18	30,473	978,297	947,824

After the reduction:			
	Reactions only	With fictive vectors	
		ALL simplexes	WITH $V_i$ only
$N$	14	14	14
$n$	11	14	14
$M$	13	18	18
$simp(H)$	2	24	22
$LB$	$2 \leq$	$4 \leq$	
$UB$	$\leq 13$	$\leq 816$	
$t$	5.49 s	263 s $\approx$ 4 m	257 s $\approx$ 4 m
$chk$	7,623	244,611	236,988

$N$  = dimension of the vectorspace,

$M$  = number of input vectors =  $|H|$ ,

$LB = b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$  (lower bound),

$t$  computational time [sec],

$n$  = dimension of what  $H$  spans,

$simp(H)$  = number of simplexes,

$UB = \binom{M}{n+1}$  (upper bound, if applicable),

$chk$  = number of checked subsets of  $H$ .

The output set of mechanisms of the last two columns (i.e., when the input is the reduced set of reaction vectors  $S_i$  with  $V_1, \dots, V_5$ ) is the following:

$$\begin{aligned}
m_1 : & -2\text{CH}_4 - 2\text{O}_2 + \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - 2S_2 - S_3 - S_4 - S_5 - S_7 \\
& -2S_8 - S_{12} - S_{13} = 0, \\
m_2 : & -2\text{CH}_4 - 2\text{O}_2 + \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - 2S_2 - S_3 - S_4 - S_5 \\
& -2S_8 - S_{13} - S_{14} = 0, \\
m_3 : & -2\text{CH}_4 - 2\text{O}_2 + \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - 2S_2 - S_3 - S_4 - S_5 - S_7 \\
& -2S_8 - S_{12} - S_{15} = 0, \\
m_4 : & -2\text{CH}_4 - 2\text{O}_2 + \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - 2S_2 - S_3 - S_4 - S_5 - S_8 \\
& -S_{14} - S_{15} = 0, \\
m_5 : & -2\text{CH}_4 - \text{O}_2 + 2\text{CH}_3\text{OH} - S_2 - S_4 - S_5 - S_7 - S_{10} = 0, \\
m_6 : & -2\text{CH}_4 - \text{O}_2 + 2\text{CH}_3\text{OH} - S_2 - S_4 - S_5 - S_{10} + S_{12} - S_{14} = 0, \\
m_7 : & -\text{CH}_4 - \frac{6}{4}\text{O}_2 + \text{CO} + 2\text{H}_2\text{O} + S_1 - \frac{6}{4}S_2 - S_3 - \frac{2}{4}S_4 - \frac{2}{4}S_5 \\
& -\frac{2}{4}S_7 - 2S_8 + \frac{2}{4}S_{10} - S_{12} - S_{13} = 0, \\
m_8 : & -\text{CH}_4 - \frac{6}{4}\text{O}_2 + \text{CO} + 2\text{H}_2\text{O} + S_1 - \frac{6}{4}S_2 - S_3 - \frac{2}{4}S_4 - \frac{2}{4}S_5 \\
& + \frac{2}{4}S_7 - 2S_8 + \frac{2}{4}S_{10} - S_{13} - S_{14} = 0, \\
m_9 : & -2\text{CH}_4 - 3\text{O}_2 + 2\text{CO} + 4\text{H}_2\text{O} + 2S_1 - 3S_2 - 2S_3 - S_4 - S_5 - 4S_8 \\
& + S_{10} - S_{12} - 2S_{13} - S_{14} = 0, \\
m_{10} : & -\text{CH}_4 - \frac{6}{4}\text{O}_2 + \text{CO} + 2\text{H}_2\text{O} - \frac{6}{4}S_2 - S_3 - \frac{2}{4}S_4 - \frac{2}{4}S_5 - \frac{2}{4}S_7 \\
& -2S_8 + \frac{2}{4}S_{10} - S_{12} - S_{15} = 0, \\
m_{11} : & -\text{CH}_4 - \frac{6}{4}\text{O}_2 + \text{CO} + 2\text{H}_2\text{O} - \frac{6}{4}S_2 - S_3 - \frac{2}{4}S_4 - \frac{2}{4}S_5 + \frac{2}{4}S_7 \\
& -2S_8 + \frac{2}{4}S_{10} - S_{14} - S_{15} = 0, \\
m_{12} : & -\text{CH}_4 - \frac{3}{2}\text{O}_2 + \text{CO} + 2\text{H}_2\text{O} - \frac{3}{2}S_2 - S_3 - \frac{1}{2}S_4 - \frac{1}{2}S_5 - 2S_8 \\
& + \frac{1}{2}S_{10} - \frac{1}{2}S_{12} - \frac{1}{2}S_{14} - S_{15} = 0, \\
m_{13} : & 2\text{CH}_4 - 3\text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - S_3 + S_4 + S_5 + S_7 - 2S_8 \\
& + 2S_{10} - S_{12} - S_{13} = 0, \\
m_{14} : & 2\text{CH}_4 - 3\text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - S_3 + S_4 + S_5 + 2S_7 - 2S_8 \\
& + 2S_{10} - S_{13} - S_{14} = 0, \\
m_{15} : & -2\text{CH}_4 + 3\text{CH}_3\text{OH} - \text{CO} - 2\text{H}_2\text{O} - S_1 + S_3 - S_4 - S_5 + 2S_8 - 2S_{10} \\
& + 2S_{12} + S_{13} - S_{14} = 0, \\
m_{16} : & 2\text{CH}_4 - 3\text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - S_3 + S_4 + S_5 + S_7 - 2S_8 + 2S_{10} \\
& -S_{12} - S_{15} = 0, \\
m_{17} : & 2\text{CH}_4 - 3\text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - S_3 + S_4 + S_5 + 2S_7 - 2S_8 + 2S_{10} \\
& -S_{14} - S_{15} = 0, \\
m_{18} : & 2\text{CH}_4 - 3\text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - S_3 + S_4 + S_5 - 2S_8 + 2S_{10} - 2S_{12} \\
& + S_{14} - S_{15} = 0,
\end{aligned}$$

$$\begin{aligned}
m_{19} : & -\text{O}_2 - \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - S_2 - S_3 + S_7 - 2S_8 + S_{10} \\
& -S_{13} - S_{14} = 0, \\
m_{20} : & -\text{O}_2 - \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} + S_1 - S_2 - S_3 - 2S_8 + S_{10} - S_{12} - S_{13} = 0, \\
m_{21} : & -\text{O}_2 - \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - S_2 - S_3 + S_7 - 2S_8 + S_{10} - S_{14} - S_{15} = 0, \\
m_{22} : & -\text{O}_2 - \text{CH}_3\text{OH} + \text{CO} + 2\text{H}_2\text{O} - S_2 - S_3 - 2S_8 + S_{10} - S_{12} - S_{15} = 0, \\
m_{23} : & -S_1 + S_{13} - S_{15} = 0, \\
m_{24} : & S_7 + S_{12} - S_{14} = 0.
\end{aligned}$$

The computer running which served the above results is summarized in table 4.6.

**4.7.** Our last example is also taken from [6] as **Conversion of Glucose to Pyruvate**. Here the chemical species are abbreviated as

$$\begin{array}{ll}
C = \text{carbon dioxide}, & N = 6 - P \text{ gluconate}, \\
D = \text{dihydroxyacetone } P, & P = \text{pyruvate}, \\
E = \text{erythrose } 4 - P, & R = \text{ribose } 5 - P, \\
F = \text{fructose } 6 - P & S = \text{sedoheptulose } 7 - P, \\
G = \text{glucose } 6 - P, & X = \text{xylulose } 5 - P, \\
K = 2\text{-keto-3-deoxy } 6 - P \text{ gluconate}, & Y = \text{glyceraldehyde } 3 - P, \\
L = \text{ribulose } 5 - P, &
\end{array}$$

where the terminal (chemical) species are  $G$ ,  $P$  and  $C$ .

The original set of reactions is

$$\begin{array}{ll}
S_1 : R + X = S + Y, & S_8 : N = K, \\
S_2 : L = R, & S_9 : L = X, \\
S_3 : N = L + C, & S_{10} : E + X = Y + F, \\
S_4 : G = N, & S_{11} : Y = P, \\
S_5 : F = D + Y, & S_{12} : D = P, \\
S_6 : G = F, & S_{13} : K = Y + P, \\
S_7 : D = Y, & S_{14} : S + Y = E + F.
\end{array}$$

As we have discussed in section 3.2.0 all the six reactions of type  $A = \lambda B$  can be omitted with a suitable modification of the remaining others to reduce both dimension and the number of vectors. Let us emphasis here that we have to add the fictive vectors  $V_1, V_2, V_3$  before this reduction since their coordinates will also be modified. After this modification we get the following list of vectors

$V_1^-$	$V_2^-$	$V_3^-$	$S_1^-$	$S_3^-$	$S_5^-$	$S_{10}^-$	$S_{13}^-$	$S_{14}^-$
0	0	0	1	0	2	1	2	-1
0	1	0	0	1	0	0	0	0
0	0	1	-2	1	0	-1	0	0
1	0	0	0	-1	-1	1	-1	1
0	0	0	1	0	0	0	0	-1
0	0	0	0	0	0	-1	0	1

in which the rows correspond to the species  $P, C, X, K, S, E$ , respectively. This transformation reduced the CPU time from 93 sec to 0.10 sec!

For the reader's convenience we list here all the *three* sets of output mechanisms: the original one, after the first and after the second reduction (see the 2nd, 5th and the last but one columns of table 4.7 below).

Let us recall that the new vectors  $V_1, V_2, V_3$  originally stood for the species  $G, P, C$ , respectively. However the reduction steps eliminated the speci (row)  $G$  while all the remaining vectors were transformed to the vectors  $V_1^-, \dots, S_{14}^-$ .

The original set of (output) mechanisms:

$$m_1 : -\frac{1}{2}G + P + C - S_3 - S_4 + \frac{1}{2}S_5 + \frac{1}{2}S_6 + \frac{1}{2}S_7 - S_9 - S_{13} = 0,$$

$$m_2 : -G + 2P + C - S_3 - S_4 + S_7 - S_9 - S_{12} - S_{13} = 0,$$

$$m_3 : -G + 2P + C - S_3 - S_4 - S_9 - S_{11} - S_{13} = 0,$$

$$m_4 : -\frac{1}{2}G + P - S_4 + \frac{1}{2}S_5 + \frac{1}{2}S_6 + \frac{1}{2}S_7 - S_8 - S_{13} = 0,$$

$$m_5 : -G + 2P - S_4 + S_7 - S_8 - S_{12} - S_{13} = 0,$$

$$m_6 : -G + 2P - S_4 - S_8 - S_{11} - S_{13} = 0,$$

$$m_7 : -\frac{1}{2}G + P - \frac{1}{2}S_5 - \frac{1}{2}S_6 - \frac{1}{2}S_7 - S_{11} = 0,$$

$$m_8 : -\frac{1}{2}G + P - \frac{1}{2}S_5 - \frac{1}{2}S_6 + \frac{1}{2}S_7 - S_{12} = 0,$$

$$m_9 : -G + 2P - S_5 - S_6 - S_{11} - S_{12} = 0,$$

$$m_{10} : C - S_3 - S_4 + S_5 + S_6 + S_7 - S_9 + S_{11} - S_{13} = 0,$$

$$m_{11} : C - S_3 - S_4 + S_5 + S_6 - S_9 + S_{12} - S_{13} = 0,$$

$$m_{12} : C - S_3 + S_8 - S_9 = 0,$$

$$m_{13} : -S_4 + S_5 + S_6 + S_7 - S_8 + S_{11} - S_{13} = 0,$$

$$m_{14} : -S_4 + S_5 + S_6 - S_8 + S_{12} - S_{13} = 0,$$

$$m_{15} : S_7 + S_{11} - S_{12} = 0.$$

After the first reduction:

$$m_1^- : -V_1^- + V_2^- + V_3^- - S_3^- = 0,$$

$$m_2^- : -\frac{1}{2}V_1^- + 3V_2^- - S_1^- - 3S_3^- + \frac{1}{2}S_5^- - S_{10}^- - S_{14}^- = 0,$$

$$m_3^- : -\frac{1}{2}V_1^- + 3V_2^- - S_1^- - 3S_3^- - S_{10}^- + \frac{1}{2}S_{13}^- - S_{14}^- = 0,$$

$$m_4^- : \frac{5}{2}V_1^- - 3V_3^- - S_1^- + \frac{1}{2}S_5^- - S_{10}^- - S_{14}^- = 0,$$

$$m_5^- : \frac{5}{2}V_1^- - 3V_3^- - S_1^- - S_{10}^- + \frac{1}{2}S_{13}^- - S_{14}^- = 0,$$

Table 4.7.  
"Glucose"

Original reactions:

	Reactions only	With fictive vectors $V_i$	
		ALL simplexes	WITH $V_i$ only
$N$	13	13	13
$n$	12	13	13
$M$	14	17	17
$\text{simp}(H)$	3	15	12
$LB$	$2 \leq$	$4 \leq$	
$UB$	$\leq 14$	$\leq 680$	
$t$	8.00 s	93.00 s	85.00 s
$chk$	14,600	107,368	92,768

After the first reduction:

	Reactions only	With fictive vectors $V_i$	
		ALL simplexes	WITH $V_i$ only
$N$	6	6	6
$n$	5	6	6
$M$	6	9	9
$\text{simp}(H)$	1	8	7
$LB$	$1 \leq$	$3 \leq$	
$UB$	$\leq 1$	$\leq 36$	
$t$	0.00 s	0.10 s	0.10 s
$chk$	52	418	366

After the second reduction:

	Reactions only	With fictive vectors $V_i$	
		ALL simplexes	WITH $V_i$ only
$N$	5	5	5
$n$	4	5	5
$M$	4	7	7
$\text{simp}(H)$	0	4	4
$LB$	$0 \leq$	$2 \leq$	
$UB$	$\leq 0$	$\leq 7$	
$t$	0.00 s	0.00 s	0.00 s
$chk$	5	65	60

 $N$  = dimension of the vectorspace, $M$  = number of input vectors =  $|H|$ , $LB = b \cdot \binom{a+1}{2} + (n-b) \cdot \binom{a}{2}$  (lower bound), $t$  = computational time [sec], $n$  = dimension of what  $H$  spans, $\text{simp}(H)$  = number of simplexes, $UB = \binom{M}{n+1}$  (upper bound, if applicable), $chk$  = number of checked subsets of  $H$ .

$$\begin{aligned}
m_6^- &: \frac{5}{2}V_2^- - \frac{1}{2}V_3^- - S_1^- - \frac{5}{2}S_3^- + \frac{1}{2}S_5^- - S_{10}^- - S_{14}^- = 0, \\
m_7^- &: \frac{5}{2}V_2^- - \frac{1}{2}V_3^- - S_1^- - \frac{5}{2}S_3^- - S_{10}^- + \frac{1}{2}S_{13}^- - S_{14}^- = 0, \\
m_8^- &: S_5^- - S_{13}^- = 0.
\end{aligned}$$

Since two parallel vectors arised after this reduction:  $S_5^- \parallel S_{13}^-$  (moreover both of them are of form  $A = \lambda B$ ) we could make a second reduction, this can be seen in the last three columns of table 4.7. The output set of mechanisms is the following:

$$\begin{aligned}
m_1^- &: -V_1^- + V_2^- + V_3^- - S_3^- = 0, \\
m_2^- &: \frac{1}{2}V_1^- + 3V_2^- - S_1^- - 3S_3^- - S_{10}^- - S_{14}^- = 0, \\
m_3^- &: \frac{5}{2}V_1^- - 3V_3^- - S_1^- - S_{10}^- - S_{14}^- = 0, \\
m_4^- &: \frac{5}{2}V_2^- - \frac{1}{2}V_3^- - S_1^- - \frac{5}{2}S_3^- - S_{10}^- - S_{14}^- = 0.
\end{aligned}$$

This time we did not make any preliminary computation with terminal (chemical) species only.

## 5. Other mathematical questions

During our thorough theoretical investigations (mainly mathematical we mean) a dozen of other questions arose. We also have several results we plan to publish elsewhere, first in [1,2,16] and in [3].

We have already mentioned the general linear algebraic study of hierarchies among atoms, species, reactions, mechanisms, etc., this will be discussed in [3].

We also have mentioned the question of minimal number of simplexes if *no parallel* vectors are allowed among the vectors. Similar questions (with similar methods) can be studied in matroids, see [16]. In more general, one could ask what if assuming the minimal size of dependent subsets ('circles' in matroids) is *at least*  $k$ ? for some fixed  $k \in \mathbb{N}$ ? With Prof. Oxley together we have conjectures concerning this question we are working on. Other variant on the number of simplexes is when we have to count simplexes containing one fixed reaction  $S_R$  of the given ones, or the simplexes which contain *at least one* vector from the set  $\{V_1, \dots, V_t\}$ . Researches are in progress on these questions, too.

In [1] we plan to discuss in full detail the effect of extending the dimension of the vectors because of the law of mass balance (conservation of material) explained in section 3.2.1.

After a clear linear algebraic reformulation we deal with the question "*is there a (finite) set of mechanism/reactions which linear combinations would give the set of all mechanism/reactions*" also in [1]. (Convex linear combinations clearly are not sufficient as it is wellknown from linear algebra.)

Finally, the valuation operator (introduced in [9]) has also linear algebraic connections which is discussed in more generality and detail in [2].

## Appendix

Here we give the bounds for the number of simplexes (= minimal reactions/mechanisms/dimensionless groups) contained in a given set of vectors (= groups of atoms/reactions/physical quantities) we proved in [14] and [15]. The first theorems below give an exact characterization of the unique *structure* of the vector-sets when the number of the contained simplexes is maximal or minimal, which easily can be transformed into stoichiometrical results. Using them we give the numerical bounds for the number of simplexes.

Recall, that a set of vectors  $S \subset \mathbb{R}^N$  is called a **simplex** iff  $S$  is minimal linearly dependent, that is  $S$  itself is linearly dependent but all its subsets are independent.

For any set of vectors  $H \subset \mathbb{R}^N$ ,  $|H|$  denotes the *size* of  $H$ , while  $[H]$  is the subspace of  $\mathbb{R}^N$  spanned by  $H$ , and finally let  $\text{simp}(H)$  denote the number of simplexes contained in  $H$ . Now, our main results can be formulated as:

**Theorem A.1** ([14]). For any  $H \subset \mathbb{R}^N$  and  $\dim[H] = n$ ,  $\text{simp}(H)$  is **maximal** iff any  $n$  vector of  $H$  are linearly independent.

**Theorem A.2** [14]. For any  $H \subset \mathbb{R}^N$  and  $\dim[H] = n$ ,  $\text{simp}(H)$  is **minimal** if  $H$  contains of  $n$  linearly independent equivalence classes of almost the same size (i.e., the size difference is at most 1) where each equivalence class is a set of parallel vectors.

**Corollary** [14]. If  $H \subset \mathbb{R}^N$   $\dim[H] = n$  and  $|H| = M$  where  $M$  is a multiple of  $n$ , then

$$n \binom{\frac{M}{n}}{2} \leq \text{simp}(H) \leq \binom{M}{n+1}$$

and in the case  $n$  does *not* divide  $M$ , the *lower* bound is

$$b \binom{a+1}{2} + (n-b) \binom{a}{2},$$

where  $M = an + b$  and  $0 \leq b < n$ .

Let us recall that

$$\binom{u}{v} := \frac{u(u-1) \cdot (u-v+1)}{1 \cdot 2 \cdot \dots \cdot v}$$

is the binomial coefficient for any natural numbers  $u, v \in \mathbb{N}$ .

In the language of mechanisms, theorem 6.1 says that the number of direct (= minimal) mechanisms is *maximal* iff there are *no* small mechanisms. In the case when  $H$  spans the whole ground space  $\mathbb{R}^N$  this means that all minimal mechanisms must use *all* species (since the reactions involved in this mechanism must span the whole space). For fixed  $n$  (= dimension, i.e., number of species involved) the upper bound  $\binom{M}{n+1}$  is approximately  $\mathcal{O}(M^{n+1})$  (for definition of  $\mathcal{O}$  see footnote<sup>12</sup> in section 4).



If we want to have (or we expect) as *few* mechanisms as possible, by theorem 6.2 we must have many parallel species/reactions, more precisely vectors representing isomer species or the *same* reaction multiplied by a constant (with higher speed). Moreover the sizes of classes of pairwise parallel vectors (species/reactions) must be as equal as possible.

*The question for the minimum value of  $\text{simp}(H)$  is **open** if no parallel vectors are allowed in  $H$ .* Parallel vectors as species (groups of atoms) represent isomer species since their sumformulas are just the same, a multiple (by a real number) of each other, or simply a higher dose of the speci in question. If reactions are represented by vectors, parallel vectors represent the same reaction only but with different emphasis or speed of it. So, in the contrary of its mathematical difficulty, the problem on the number of simplexes when parallel vectors are *allowed* is also of importance for estimating the number of possible reactions/mechanisms. Though these interpretations of simplexes allow and make sense and importance of parallel vectors, to avoid counting the same reaction several times with different constant we must investigate the case when *no* parallel vectors are allowed or more sophisticated questions. Unfortunately so far we only have the below result from [15] for 3-dimensional vectors without parallel ones (e.g., when we our groups of atoms are build up from 3 atoms):<sup>13</sup>

**Theorem A.3** [15]. For any  $H \subset \mathbb{R}^3$  of fixed size but not equal to 3, 4 or 7 such that  $H$  spans  $\mathbb{R}^3$  and no parallel vectors are in  $H$ ,  $\text{simp}(H)$  is minimal if and only if  $H$  is contained in two intersecting planes, one of which is of size 3. In other words, when  $H$  contains three linearly independent vectors  $\{u_1, u_2, u_3\}$ , another vector  $v$  coplanar with  $u_1$  and  $u_2$ , and the rest  $H \setminus \{u_1, u_2, u_3, v\}$  is coplanar with  $u_2$  and  $u_3$ .

**Corollary A.4** [15]. If  $H \subset \mathbb{R}^N$  no parallel vectors are in  $H$ ,  $H$  spans  $\mathbb{R}^3$  and  $|H| = M$  where  $M$  differs from 3, 4 and 7, then

$$1 + \binom{M-2}{3} + \binom{M-3}{2} \leq \text{simp}(H).$$

This gives us a lower bound of magnitude  $\mathcal{O}(M^3)$ .

Several other conjectures were mentioned in section 5.

## References

- [1] I. Szalkai, Reaction and process network syntheses via linear algebraic methods, in preparation.
- [2] I. Szalkai, On valuation operators in stoichiometry and in reaction syntheses, J. Math. Chemistry, in press.
- [3] I. Szalkai and Á. Pethő, On generalized stoichiometry, in preparation.

<sup>13</sup> Moreover we also showed that this bound is sharp: we gave the unique construction with minimal number of simplexes for each size of  $H$ .

- [4] J. Happel and P.H. Sellers, Analysis of the possible mechanisms for a catalytic reaction system, *Adv. in Catal.* 32 (1983) 273–323.
- [5] P.H. Sellers, Combinatorial classification of chemical mechanisms, *SIAM J. Appl. Math.* 44 (1984) 784–792.
- [6] J. Happel, M. Otarod and P.H. Sellers, Mechanistic study of chemical reaction systems, *Ind. Eng. Chem. Res.* 29 (1990) 1057–1067.
- [7] Á. Pethő, The linear relationship between stoichiometry and dimensional analysis, *Chem. Eng. Technol.* 13 (1990) 328–332.
- [8] Á. Pethő, Mathematical discussion of the application of Hess’s law, *Hung. J. Ind. Chem.* 21 (1993) 35–38.
- [9] Á. Pethő, Further remarks on the analogy between stoichiometry and dimensional analysis: The valuation operation, *Hung. J. Ind. Chem.* 23 (1995) 229–231.
- [10] B. Bertók, A combinatorial algorithm for generating possible systems of elementary reactions, Thesis (1999) in Hungarian.
- [11] I. Fishtik, A. Alexander and R. Datta, Enumeration and discrimination of mechanisms in heterogeneous catalysis based on response reactions and unity bond index-quadratic exponential potential (UBI-QEP) method, *Surface Science* (1999), in press.
- [12] I. Fishtik and R. Datta, On the use of response reactions in the kinetic model of complex heterogeneous catalytic reactions, in: *React. Kin. and the Dev. of Catal. Proc.*, eds. G.F. Gromment and K.C. Waught (Elsevier, 1999).
- [13] I. Szalkai, Generating minimal reactions in stoichiometry using linear algebra, *Hung. J. Ind. Chem.* 19 (1991) 289–292, ref.: *Chemical Abstracts* 116:32761b (1992) vol. 114, No. 4, p. 160.
- [14] I. Szalkai and C. Laflamme, Counting simplexes in  $\mathbb{R}^n$ , *Hung. J. Ind. Chem.* 23 (1995) 237–240, ref.: *Chemical Abstracts* 123:290982:u (1995), vol. 123, No. 22, p. 310.
- [15] I. Szalkai, Counting simplexes in  $\mathbb{R}^n$ , *Electr. J. of Combinatorics*, Vol. 5 (1998) No. 1, Res. Paper No. 40, 11 pp, <http://www.combinatorics.org>, ref.: *Current Math. Publ.* (1998) No. 16.
- [16] I. Szalkai, Gy. Dósa, C. Laflamme and M. Hujter, On the maximal and minimal number of bases and simple circuits in matroids and the extremal constructions, to submit to *Electr. J. Combin.*
- [17] Á. Pethő, On a class of solutions of algebraic homogeneous linear equations, *Acta Math. Hungaricae* 18 (1967) 19–23.
- [18] L.T. Fan, B. Bertók and F. Friedler, Combinatorial framework for the systematic generation of reaction pathways, presented at the AIChE Annual Meeting, Dallas, TX, USA, Oct. 31–Nov. 5, 1999.
- [19] S. Kumar and S. Pethő, Note on a combinatorial problem for the stoichiometry of chemical reactions, *Intern. Chem. Eng.* 25 (1985) 767–769.
- [20] N.R. Amundson, Mathematical methods in chemical engineering, in: *Matrices and Their Applications* (Prentice Hall, Englewood Cliffs, 1966) pp. 53–54.