

# Synthesis of Reaction Pathways via Systematic Integration

Mingheng Li

*Department of Chemical Engineering, University of California, Los Angeles, CA 90095*

## Abstract:

The environmental pollution has become, during the recent years, a crucial problem of public concern, and the authorities are requested to realize “pollution prevention” with systematic methods to replace the conventional “end-of-pipe” treatment techniques. The specification of the most desirable reaction routes that connect the initial reactants and the desired products, or known as reaction path synthesis, represents a key step in arriving at environmentally benign process designs. One of the most famous examples is the Solvay Soda Ash Process, which converts sodium chloride (salt) and calcium carbonate (limestone) into sodium carbonate (soda ash) and calcium chloride by six reaction clusters that are equivalent to the infeasible overall target reaction from a mass balance point of view, while are all thermodynamically feasible under industrial conditions.

In this work, we will present the synthesis of thermodynamically feasible reaction pathways to achieve a thermodynamically infeasible overall target reaction, from a mathematical prospective. A notion of irreducible reaction (IRR) whose reactive system has one and only one degree of freedom, was initially introduced, and the particular properties of IRRs were elaborated. An efficient algorithm which is based on Singular Value Decomposition (SVD) to enumerate the finite number of IRRs from a large species database, with an arbitrarily pre-specified maximum number of substances allowed in each individual reaction, was developed and its completeness was strictly proved using linear algebra theories. Involved with only operations on matrices, this algorithm circumvents the conventional Mixed-Integer Linear Programming (MILP) methods and thus, the computational load can be significantly reduced, especially in high dimensional cases. Following the introduction of IRRs, the issue of synthesis of thermodynamically feasible reaction pathways to realize a thermodynamically infeasible overall reaction was formulated and solved as an optimization problem to satisfy overall mass balance, thermodynamics and other chosen performance criteria and constraints.

**Key words:** process design; reaction path synthesis; singular value decomposition; stoichiometry; algorithm.

## 1 Introduction

The environmental pollution has become, during the recent years, a crucial problem of public concern, and the authorities are requested to realize “pollution prevention” with systematic methods to replace the conventional “end-of-pipe” treatment techniques. The specification of the most desirable reaction routes that connect the initial reactants and the desired products, or known as reaction path synthesis, represents a key step in arriving at environmentally benign process designs.

Because of the inherent difficulties of this problem (e.g. infinite solutions, unpredictable kinetics), any attempt to take all of its engineering characteristics into consideration is impractical. During the past, main concerns on screening of potential reaction paths are concentrated on atom balance as well as feasibility in terms of thermodynamics, and several procedures based on stoichiometry were put forward to simplify the problem of reaction path synthesis, such as degrees of freedom of a reactive system [4, 5], maximum number of substances involved in a reaction [2, 3], and upper bound of the magnitude of each stoichiometric coefficient [6] etc. In our previous work [8, 7], we introduced a notion of Simple Stoichiometric Reaction (SSR), which owns a property that its reactive system has one and only one degree of freedom and has simple integer stoichiometric coefficients, to reduce the hunting zone of feasible reaction clusters. This type of reactions were also mentioned as Irreducible Reaction (namely, reaction that can not reduce to two or more reactions) by Tow and Rudd [11]. To avoid confusion, we followed this notion in the present paper.

There are several significant advantages brought about by the introduction of IRR, which include (1) IRRs in a reactive system are *finite*; (2) each IRR has very simple stoichiometry, which accords with industrial practice (Tow and Rudd [11] argued that almost all the industrial reactions are of this kind); and (3) each reaction candidate is generated and may be pre-evaluated in terms of kinetics and environmental impact before the reaction routes are synthesized, and consequently, reactions with undesirable properties in the final optimal solution can be avoided. Despite of the promising application of IRRs in the synthesis of reaction pathways, the unique characteristics of IRRs and the efficient procedure for the generation of IRRs, especially in high dimensional cases, remain unaddressed. At this point, it is important to note that although Tow and Rudd proposed a method to avoid almost all reducible reactions by simply not attempting to use more than  $(s+1)$  chemicals to achieve  $s$  element balances [11], we will prove later that this condition is either necessary or sufficient for the generation of IRRs.

In this paper, the particular properties of IRRs were elaborated. An efficient algorithm which is based on Singular Value Decomposition (SVD) to enumerate the finite number of IRRs with an arbitrarily pre-specified maximum number of substances allowed in each reaction formula from the atomic matrix of a reactive system, was also developed and strictly proved using linear algebra theories. Involved with only simple operations on matrices, this algorithm circumvents the conventional Mixed-Integer Linear Programming (MILP) methods [2] and thus, the computational load can be significantly reduced, especially in high dimensional cases. Following the introduction of IRRs, issues of reaction path synthesis such as identification of the most desirable routes for the production of a specified product [3], realization of an unfeasible reaction by several thermodynamically feasible reaction clusters [6, 12], and realization of Zero Avoidable Pollution (ZAP) by recycled use of undesirable components or joint production of other products, etc. may all be formulated as computationally tractable optimization problems. Examples will be studied to illustrate the proposed procedure.

## 2 Stoichiometry

Without lack of universality, let's consider a reactive system involving totally  $s$  substances and  $\varepsilon$  elements (or stable functional groups). Its atom balance equation can be written as [1]:

$$A_{\varepsilon \times s} x_{s \times 1} = 0 \quad (1)$$

where  $A$  is the  $\varepsilon$  by  $s$  atomic matrix with rank  $r$ , representing the number of each chemical element contained in each component (reactant or product), and  $x$  is the  $s$  by 1 column vector corresponding to the stoichiometric coefficient of each component, positive for products and negative for reactants. Usually,  $r \leq \varepsilon < s - 1$  and therefore Eq.1 can be simplified as

$$A_{r \times s} x_{s \times 1} = 0 \quad (2)$$

According to linear algebra theory, the possible solutions to Eq.2 are infinite. For example, for a reactive system involving  $C$ ,  $CO$ ,  $CO_2$ ,  $H_2$ ,  $O_2$ , and  $H_2O$  whose atom balance equation is

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 2 \\ 0 & 1 & 2 & 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = 0 \quad (3)$$

where  $x_1 \sim x_6$  are the coefficients for the aforementioned six substances, the reader may easily give several solutions to this equation, as shown in Table 1.

Table 1: Possible solutions to the atom balance equation.

No.	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
1	1	-2	1	0	0	0
2	2	-2	0	0	1	0
3	1	-1	0	-1	0	1
4	1	0	-1	0	1	0
5	0.5	0	-0.5	-1	0	1
6	0	0	0	-1	-0.5	1
7	0	2	-2	0	1	0
8	0	1	-1	-1	0	1
9	1	2	-2	0	1	0
10	0	0.22	-0.22	-1	-0.39	1
⋮						

From Table 1 it can be seen that the first reactions, no matter there are how many substances involved in each reaction, share a property that the reactive system constituted by substances showing up in the reaction formula (i.e. with nonzero stoichiometric coefficient) has one and only one degree of

freedom, and therefore, there is only one free variable in the coefficient vector. In the first reaction, for example, once the stoichiometric coefficient of any one substance (e.g.  $CO_2$ ) is specified to be 1, the coefficients for all the other substances ( $C$  and  $CO$ ) can be uniquely determined (1 and -2, respectively). However, when the stoichiometric coefficient of one substance in the 9th reaction (e.g.  $C$ ) is specified to be 1, the coefficients for the remaining substances ( $CO$ ,  $CO_2$ , and  $O_2$ ) cannot be uniquely determined (e.g. they can be -4, 3, and -1, respectively); this is because its reactive system has two degrees of freedom. Further observation shows that if a reactive system has only one degree of freedom, all the entries of the coefficient vector can be converted to simple integers if the latter is multiplied by an appropriate factor, which is guaranteed by the simple integer number of atoms contained in each substance. In the following text, we will call such kind of reactions Irreducible Reactions (IRRs), namely, reactions having very simple stoichiometric coefficients. Reactive systems with more than one degree of freedom, however, may lead to very complex coefficient vectors ( e.g. the one given in reaction 10), which should be avoided as much as possible in the synthesis of reaction clusters and reaction pathways. Many industrial reactions can have complex stoichiometric coefficients of many components in one reaction (for example, cracking of long chain hydrocarbons or many biochemical reactions) such as No. 10 in Table 1. These complex reactions can be expressed by the combination of two or more IRRs.

It can be proved that the total number of reactions sharing the above property in a reactive scheme is very limited (e.g. only 8 for the reactive system represented by Eq.3) instead of infinite if proportional coefficient vectors are considered to represent the same reaction (see the next section for details). Because main reactions used in industrial applications fall into this category, and any complex reaction that have two or more degrees of freedom can be expressed by the linear combination of a set of  $s - r$  linearly independent IRRs, the concept of IRR may play an important role in the synthesis of thermodynamically feasible reaction clusters and reaction pathways. In the following sections, issues such as the maximum number of IRRs in a reactive system, the algorithm for generating such reactions from the atomic matrix of the whole system, and its application in reaction path synthesis will be addressed.

### 3 IRR and its properties

Based on the above analysis, the definition of IRR is described as follows:

*A reaction is called an IRR if the reactive system constituted by substances showing up in the reaction formula has one and only one degree of freedom. Alternatively, once the stoichiometric coefficient of any one substance is specified to be 1, the coefficients of all the other substances in the reaction formula can be solely determined.*

Following the definition of IRR, several conclusions listed below can be drawn.

**Theorem:**

1. The atomic matrix associated with an IRR satisfies  $s' = r' + 1$ , where  $r'$  is its rank and  $s'$  is the

number of substances involved in the IRR. Furthermore, an IRR can not reduce to two distinct individual reactions without other substances showing up.

2. If a reactive system satisfies  $s' = r' + 1$ , it contains one and only one IRR. However, some substances may not show up in the resulting reaction formula (i.e. the coefficients of these substances are zeros). In addition, if  $s' > r' + 1$ , the reactive system has more than one IRR.
3. The number of substances involved in an IRR is less than or equal to  $r + 1$ , where  $r$  is the rank of the atomic matrix of the whole reactive system.
4. For an IRR in a reactive system, one can always find at least one set of  $r + 1$  substances that include all the components in this IRR, such that the atomic matrix composed by these  $r + 1$  substances has rank  $r$ . Furthermore, if there exists more than one set of such  $r + 1$  substances, they will ultimately reduce to the same IRR.
5. The total number of IRRs in a reactive system is less than or equal to  $C_s^{r+1}$ .

**Proof:**

1. Without lack of universality, let's assume the coefficient of the  $s'$ th substance in an IRR is specified to be 1, then the atom balance equation  $A'_{r' \times s'} x'_{s' \times 1} = 0$  can be converted into

$$\begin{bmatrix} A'_{r' \times (s'-1)} & A'_{r' \times 1} \end{bmatrix} \begin{bmatrix} x'_{(s'-1) \times 1} \\ 1 \end{bmatrix} = 0 \quad (4)$$

or

$$A'_{r' \times (s'-1)} x'_{(s'-1) \times 1} = -A'_{r' \times 1} \quad (5)$$

Because the above linear equation has a unique solution,  $A'_{r' \times (s'-1)}$  should be full rank and consequently  $s' = r' + 1$ . In addition, if an IRR can be decomposed into two distinct reactions without other substances, i.e.  $x_0 = x_1 + x_2$ , where  $x_0 \sim x_2$  are three solution vectors, of which all the entries are non-zeros, to the same atomic matrix. From Eq.5 one concludes that  $x_0 = x_1 = x_2$ , which yields  $x_0 = x_1 = x_2 = 0$ . Therefore, an IRR cannot reduce to two distinct reactions without other substances shown up. However, one should be reminded that the sum of two IRRs may lead to another IRR, in which case some substances must annihilate in the resulting IRR.

2. Apparently the condition  $s' = r' + 1$  doesn't guarantee that each element in vector  $x'$  in Eq.5 is nonzero. However, if the coefficient of a substance is zero, it doesn't show up in the resulting reaction formula. Note the rank of the atomic matrix is only 1 less than the number of substances involved in this subsystem, any atomic vector associated with a zero element in the coefficient vector  $x'$  should be linearly independent of any other  $s' - 1$  atomic vectors. Therefore, in the reduced subsystem which contain substances associated with non-zero elements in the coefficient vector, the total number of substances involved and the rank of the corresponding atomic matrix will be  $s'' = s' - z$  and  $r'' = r' - z$ , respectively, where  $z$  is the number of zero elements in the coefficient vector. Because it satisfy  $s'' = r'' + 1$ , it contains one and only one IRR.

Furthermore, if a reactive system satisfies  $s' > r' + 1$ , there are at least two substances whose atomic vectors can both be uniquely expressed by other  $r'$  linearly independent atomic vectors, i.e. the system has more than one IRR. Similarly, if a reaction satisfies  $s' = r' + 2$ , it can be decomposed to two IRRs, which share  $r'$  substances. For example,  $3C + 2O_2 = 2CO + CO_2$  is the sum of the following two IRRs  $2C + O_2 = 2CO$  and  $C + O_2 = CO_2$ .

3. Because any atomic matrix associated with a subsystem has rank less than or equal to  $r$ , which is the rank of the atomic matrix of the whole system, based on Theorem 1,  $s' = r' + 1 \leq r + 1$ .
4. Suppose the atomic matrix of an IRR is  $r'$  by  $r' + 1$ . Apparently, there are  $r'$  substances whose atomic vectors are linearly independent. Because the rank of the atomic matrix of the whole system is  $r$ , one can always find other  $r - r'$  substances such that these  $r$  atomic vectors are linearly independent. Although a scenario that more than one set of  $r - r'$  substances which satisfies the above condition can be easily assumed, they will ultimately reduce to the same IRR because Theorem 2 guarantees a system satisfying  $s' = r' + 1$  has only one IRR.
5. According to Theorem 4, one can set up a mapping from any one IRR to at least one set of  $r + 1$  substances, which satisfies (i) its atomic matrix has rank  $r$ ; and (ii) it will result in the aforementioned IRR only. Therefore, the total number of IRRs in a reactive system is less than or equal to the total possibilities of picking  $r + 1$  substances out of the whole system, i.e.  $C_s^{r+1}$ . In practice, the number of IRRs may be much less than this upper bound because (i) not all the enumerated atomic matrices composed by  $r + 1$  substances has rank  $r$ , and (ii) different sets of  $r + 1$  substances may reduce to the same IRR.

The above analysis provides an approach to generate all the IRRs in a reactive system through the enumeration of possible combinations of  $r + 1$  substances whose associated atomic matrix has a rank of  $r$ . Note any  $r + 1$  substances are automatically linearly dependent, the above problem is equivalent to the enumeration of  $r$  by  $r$  full rank matrices from the atomic matrix of the whole system.

**Remark 1:** To elaborate the properties of IRR, we used the reduced atomic matrix  $A_{r \times s}$  instead of the original atomic matrix  $A_{\epsilon \times s}$ . This is because the row vector in the latter may be linearly dependent.

For instance, the atomic matrix for an IRR  $C_2H_2 + Cl_2 = C_2H_2Cl_2$  is  $\begin{bmatrix} 2 & 0 & 2 \\ 2 & 0 & 2 \\ 0 & 2 & 2 \end{bmatrix}$ , whose rank is 2 while the number of chemical elements involved is 3. This example also shows that the IRRs generated by simply not attempting to use more than  $(s + 1)$  chemicals to achieve  $s$  element balances (see, for example, [11]) may be incomplete.

## 4 Algorithm for enumeration of IRRs in a reactive system

By multiplying a full-rank column translation matrix  $P_{s \times s}$  (to relocate the sequence of each column vector in a matrix) and its inverse in between of  $A_{r \times s}$  and  $x_{s \times 1}$ , Eq.2 becomes

$$[A_{r \times s} P_{s \times s}] [P_{s \times s}^{-1} x_{s \times 1}] = 0 \quad (6)$$

Dividing the new matrix and new vector both into two parts, i.e.  $[A_{r \times s} P_{s \times s}] = [A_{r \times r} \quad A_{r \times (s-r)}]$  and  $[P_{s \times s}^{-1} x_{s \times 1}] = \begin{bmatrix} x_{r \times 1} \\ x_{(s-r) \times 1} \end{bmatrix}$ , the above equation yields

$$\begin{bmatrix} A_{r \times r} & A_{r \times (s-r)} \end{bmatrix} \begin{bmatrix} x_{r \times 1} \\ x_{(s-r) \times 1} \end{bmatrix} = 0 \quad (7)$$

If  $A_{r \times r}$  is full rank, the above equation can be further simplified to

$$x_{r \times 1} = -A_{r \times r}^{-1} A_{r \times (s-r)} x_{(s-r) \times 1} \quad (8)$$

or,

$$\begin{bmatrix} x_{r \times 1} \\ x_{(s-r) \times 1} \end{bmatrix} = \begin{bmatrix} -A_{r \times r}^{-1} A_{r \times (s-r)} \\ I_{(s-r) \times (s-r)} \end{bmatrix} x_{(s-r) \times 1} \quad (9)$$

where  $I_{(s-r) \times (s-r)}$  is a unit matrix. Finally we obtain a general expression of the solution

$$x_{s \times 1} = P_{s \times s} \begin{bmatrix} -A_{r \times r}^{-1} A_{r \times (s-r)} \\ I_{(s-r) \times (s-r)} \end{bmatrix} x_{(s-r) \times 1} \quad (10)$$

to the atom balance equation, where  $x_{(s-r) \times 1}$  can be any vectors whose elements are all nonnegative.

It can be proved using Theorem 2 that any column vector in the  $s$  by  $s - r$  matrix

$$P_{s \times s} \begin{bmatrix} -A_{r \times r}^{-1} A_{r \times (s-r)} \\ I_{(s-r) \times (s-r)} \end{bmatrix} \quad (11)$$

is an IRR, which contains at most  $r + 1$  substances in the reaction formula, and satisfies  $r' = s' - 1$  because one can safely set  $r' = r$  and  $s' = r + 1$ . By enumerating all the possible full rank matrix  $A_{r \times r}$  (through the column translation matrix  $P_{s \times s}$ ), one can generate all the IRRs in the reactive system (guaranteed by Theorem 4). Note this approach tends to repeat an IRR candidate up to  $r + 1$  times, so that the total number of IRRs should be less than or equal to  $\frac{(s-r)C_s^r}{r+1} = C_s^{r+1}$ , which is the same as the one in Theorem 5. By the same token, because not all  $A_{r \times r}$  is full rank and different combinations of  $r + 1$  substances may reduce to the same IRR, the number of IRRs is usually less than its upper bound.

## 5 Improved algorithm based on singular value decomposition

Normally speaking, not all reactions with satisfied stoichiometry seem to occur in practice. Reactions with very complex kinetics, for example, involving 6 or more substances in a single step reaction don't

likely to take place practically. One method to fulfill this objective is to set the maximum number of nonzero elements allowed in each coefficient vector after it has been solved. However, for high dimensional cases in which the rank of the atomic matrix and the number of substances are both very large, it will be computationally untractable to calculate the inverse of a huge size matrix for a tremendous number of times. Under such a circumstance, an improved method which is based on singular value decomposition, can be used to modify the procedure proposed in Section 4.

Let's assume the maximum number of substances allowed in a reaction is  $M + 1$  ( $M \leq r$ ). Instead of picking out a square matrix from the atom matrix  $A_{r \times s}$ , we divide the latter into two matrices which contain  $M$  and  $(s - M)$  columns, respectively, i.e.

$$\begin{bmatrix} A_{r \times M} & A_{r \times (s-M)} \end{bmatrix} \begin{bmatrix} x_{M \times 1} \\ x_{(s-M) \times 1} \end{bmatrix} = 0 \quad (12)$$

By singular value decomposition on  $A_{r \times M}$ , i.e.  $A_{r \times M} = S_{r \times r} V_{r \times M} D_{M \times M}^T$ , where  $S_{r \times r}$  and  $D_{M \times M}$  are full rank matrices and  $V_{r \times M}$  is a matrix whose non-diagonal entries are zeros, the above equation yields,

$$\begin{bmatrix} S_{r \times r} V_{r \times M} D_{M \times M}^T & A_{r \times (s-M)} \end{bmatrix} \begin{bmatrix} x_{M \times 1} \\ x_{(s-M) \times 1} \end{bmatrix} = 0 \quad (13)$$

Note the solution structure of a linear equation  $Ax = 0$  will not change if it's multiplied on the left by a full rank matrix, Eq.13 is equivalent to

$$\begin{bmatrix} V_{r \times M} D_{M \times M}^T & S_{r \times r}^{-1} A_{r \times (s-M)} \end{bmatrix} \begin{bmatrix} x_{M \times 1} \\ x_{(s-M) \times 1} \end{bmatrix} = 0 \quad (14)$$

If the rank of  $V_{r \times M}$  is  $M$  (it will be called "the first condition" in the following text), i.e.  $V_{r \times M} = \begin{bmatrix} \tilde{V}_{M \times M} \\ 0_{(r-M) \times M} \end{bmatrix}$ , and  $\tilde{V}_{M \times M}$  is full rank, then the above equation yields

$$\begin{bmatrix} (\tilde{V} D^T)_{M \times M} & \vdots \\ 0_{(r-M) \times M} & \vdots \end{bmatrix} \begin{bmatrix} S_{r \times r}^{-1} A_{r \times (s-M)} \end{bmatrix} \begin{bmatrix} x_{M \times 1} \\ x_{(s-M) \times 1} \end{bmatrix} = 0 \quad (15)$$

Furthermore, if any column vector in  $S_{r \times r}^{-1} A_{r \times (s-M)}$ , say,  $\alpha_{r \times 1}$ , whose last  $r - M$  entries are all zeros (it will be called "the second condition" in the following text), from the above equation one can get

$$x_{M \times 1} = -(\tilde{V} D^T)_{M \times M}^{-1} \alpha_{M \times 1} x_{1 \times 1} \quad (16)$$

where  $\alpha_{M \times 1}$  represents the column vector composed by the first  $M$  entries of  $\alpha_{r \times 1}$ , and  $x_{1 \times 1}$  is its corresponding stoichiometric coefficient. Therefore, one can solve the coefficient vector of an IRR by

$$x_{(M+1) \times 1} = \begin{bmatrix} -(\tilde{V} D^T)_{M \times M}^{-1} \alpha_{M \times 1} \\ 1 \end{bmatrix} \quad (17)$$

which contains at most  $M + 1$  substances and their corresponding indices in the matrix can be determined simultaneously.



Following a similar procedure described in Section 4, one can prove that Eq.17 represents an IRR and all the IRRs which have less than or equal to  $M + 1$  substances in a reactive system can be generated by the enumeration of  $A_{r \times M}$  which satisfies the two aforementioned conditions. In fact, the algorithm based on SVD is more generalized than the one proposed in Section 4. When  $M$  is taken to be  $r$ , it collapses to the latter because, the second condition is satisfied automatically. Furthermore, from the development of this algorithm, it can also be concluded that  $A_{r \times s}$  can be safely replaced by  $A_{\varepsilon \times s}$  without any change in the final solution.

## 6 Computer simulation procedure

Based on the fact that each IRR is constituted by at most  $M + 1$  substances, one can use a solution index vector ( $SI$ ) and a solution vector ( $SV$ ), both of size  $M + 1$ , to adequately record the indices and the corresponding stoichiometric coefficients of all the substances in an applicable solution. Because the proposed algorithm tends to generate proportional solutions which should be considered as the same IRR, how to avoid unnecessary computation is of utmost importance in arriving at efficient, redundancy-proof IRR generation procedures. Before an  $SV$  is solved, a judgement should be made based on the current applicable  $SI$  regarding whether it will result in a redundant solution, thus automatically skipping the calculation of  $SV$  once it is predicted to lead to an IRR solved already. There are two kind of cases which may be considered to be “proportional solutions”. One is that the current  $SI$  has shown before, with only a different order of each entry, which will definitely lead to the same IRR. The other is that there exists a previous  $SI$ , whose associated substances with nonzero entries in its corresponding  $SV$  all show up in the system represented by the current  $SI$ . According to Theorem 4, the current  $SI$  will eventually lead to an  $SV$  whose corresponding IRR has shown previously.

In computer simulations, the enumeration of matrix  $A_{r \times M}$  is realized not by multiplication of column translation matrix  $P$ , but through an iterative procedure on the column index vector  $c$ , which represents the indices of the first  $M$  column vectors in the atomic matrix (or the first  $M$  entries in  $SI$ ). The original  $c^0$  is  $[ 1 \ 2 \ \cdots \ M ]$  and the final  $c^f$  is  $[ s - M \ s - M + 1 \ \cdots \ s - 1 ]$ . Based on the current column index vector  $c^k$ , an algorithm is applied to calculate  $c^{k+1}$ , whose entries are sorted automatically in an ascending order. For each  $c$ , the last entry of  $SI$  can be numbers from  $c(M) + 1$  to  $s$ , thus counting off all the possible  $C_s^{M+1}$  different  $SI$ s. By enumerating  $SI$  in a monotonic way, the first case of “proportional solutions” is avoided automatically. To avoid the second case of “proportional solutions”, each entry in an  $SI$  associated with zero entry in its corresponding  $SV$  is set to zero after the latter has been solved. Once the program detects that a previous  $SI$  whose nonzero entries all show up in the current  $SI$ , it will skip to the next applicable  $SI$ .

**Remark:** In the above procedure, solution index vector  $SI$  is enumerated from  $[ 1 \ 2 \ \cdots \ M \ | \ M + 1 ]$  to  $[ s - M \ s - M + 1 \ \cdots \ s - 1 \ | \ s ]$ . Because for each applicable  $SI$  that may lead to an IRR, the last entry of its corresponding  $SV$  has a nonzero value of 1, i.e. its associated substance will defi-

nitely show up in the resulting IRR, the above procedure will ignore any potential IRR(s) constituted ONLY by the first  $M$  substances. This issue can be solved by (i) purposely listing the atomic matrix such that its first  $M$  column vectors are linear independent; or (ii) adding any IRR in the subsystem constituted by the first  $M$  substances, if applicable, to the solution sets.

**Algorithm for generating  $c^{k+1}$  based on  $c^k$ :**

(a) *calculate number of elements in the first  $M$  entries of SI which need to be changed ( $nec$ )*

$nec = 1$

**for**  $i = 1 : M$

**if**  $c^k(i) \equiv n + i - M - 1$

$nec = nec + 1$

**end if**

(b) *change the sequence of the first  $M$  column vector(s), which are sorted in ascending order automatically*

**if**  $nec = 1$

$c^{k+1}(M) = c^k(M) + 1$

**else**

$c^{k+1}(M - nec + 1) = c^k(M - nec + 1) + 1$

**for**  $j = M - nec + 2 : M$

$c^{k+1}(j) = c^{k+1}(j - 1) + 1$

**end if**

## 7 Illustrative examples

### Example 1

The reactive system for the production of a kind of herbicide - 1-naphthyl-N-methyl carbamate [3], was studied based on the above procedure. The scale of this case is small and it is suitable for the demonstration of the method proposed previously. Instead of generating the reactions with herbicide - 1-naphthyl-N-methyl carbamate as the product, we enumerated all the IRRs from this database. Firstly, the atomic matrix of this reactive system was written as follows:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 2 & 0 & 1 & 1 & 2 & 0 & 2 \\ 0 & 2 & 1 & 0 & 5 & 2 & 5 & 0 & 3 & 1 & 4 \\ 2 & 0 & 0 & 2 & 1 & 1 & 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} C \\ H \\ O \\ N \\ Cl \\ X \end{matrix} \quad (18)$$

Note we used  $X$  to represent the stable functional group naphthyl. To demonstrate the generation method based on SVD, initially we chose  $M$  to be 2, in which case all the reactions generated should involve at most 3 substances. Note that the dimensions of  $SI$  and  $SV$  are both 3 and the column index vector  $c$  should be enumerated from  $[1\ 2]$  ( $c^0$ ) to  $[9\ 10]$  ( $c^f$ ). When  $c = [1\ 2]$ , the above matrix is first divided into two parts as

$$A_1 = \begin{bmatrix} 0 & 0 \\ 0 & 2 \\ 2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 1 & 2 & 0 & 2 \\ 1 & 0 & 5 & 2 & 5 & 0 & 3 & 1 & 4 \\ 0 & 2 & 1 & 1 & 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (19)$$

By taking out the zero rows in  $A_1$  and deleting the columns in  $A_2$  whose corresponding rows are not zeros and also taking out these zeros rows, one get the following compressed matrices:

$$A_{10} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}, \quad A_{20} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad (20)$$

The index of  $A_{10}$  and  $A_{20}$  are  $[1, 2]$  and  $[6]$ , respectively. By singular value decomposition ( $A_{10}$  is full-rank square matrix occasionally, but sometimes it may not be)

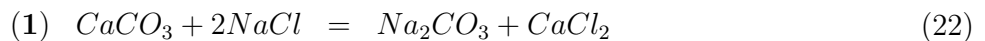
$$S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad D^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (21)$$

In this case  $\tilde{V} = V$  and thus one solution with  $SV$  generated by using Eq.17  $[-1, -0.5, 1]$  and corresponding  $SI = [1, 2, 6]$ .

The number of IRRs generated in this reactive system for  $M = 2, 3$  and  $4$  is  $3, 8,$  and  $16,$  respectively, as shown in table 2. All the simulations were ran using a Pentium III-1.0G laptop with a program code written by Matlab and the CPU time consumed for each case was less than 0.5 sec.

## Example 2

The famous Solvay Soda Ash Process converts sodium chloride (salt) and calcium carbonate (limestone) into sodium carbonate (soda ash) and calcium chloride, a thermodynamically infeasible reaction



by six reaction clusters that are equivalent to the infeasible overall target reaction while are all thermodynamically feasible under industrial conditions [9, 10]

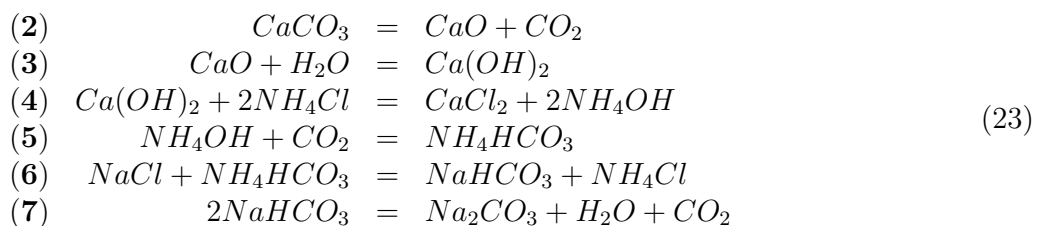


Table 2: IRRs in the 1-naphthyl-N-methyl carbamate production system<sup>a</sup>

M	SI	SV	CPU (sec)
2	1 2 6	-0.5 -1 1	0.06
	2 5 9	1 -1 1	
	9 10 11	-1 -1 1	
3	1 5 6 9	-0.5 -1 1 1	0.14
	2 5 10 11	1 -1 -1 1	
	3 4 7 11	1 -1 -1 1	
	3 4 8 10	-1 -1 1 1	
	3 7 8 9	2 -1 -1 1	
4	1 5 6 10 11	-0.5 -1 1 -1 1	0.33
	2 3 5 7 8	1 -2 -1 1 1	
	3 4 7 9 10	1 -1 -1 1 1	
	3 4 8 9 11	-1 -1 1 -1 1	
	3 7 8 10 11	2 -1 -1 -1 1	
	4 7 8 9 10	-1 -0.5 0.5 0.5 1	
	4 7 8 9 11	-1 -0.5 0.5 -0.5 1	
4 7 8 10 11	-2 -1 1 1 1		

<sup>a</sup> 1, oxygen; 2, hydrogen; 3, hydrogen chloride; 4,  $\alpha$ -naphthol chloroformate; 5, methyl formamide; 6, water; 7, methyl amine; 8, phosgene; 9, methyl isocyanate; 10,  $\alpha$ -naphthol isocyanate; 11, 1-naphthyl-N-methyl carbamate

In this work, we tested a system constituted by 27 components which includes all the substances shown above. The program was executed with  $M = 3$  and it screened 335 IRRs in 13.3 sec with all the aforementioned seven reactions included, which validated the efficiency of the proposed procedure.

Based on the proposed generation method of IRRs, the problem of thermodynamically feasible reaction pathways synthesis can be easily formulated and solved as an optimization problem based on the strategies proposed in [6], which will be covered in a full-length paper and will not be discussed here.

## References

- [1] R. Aris and R. H. S. Mah. Independence of Chemical Reactions. *Ind. Eng. Chem. Fund.*, 2:90–94, 1963.
- [2] A. Buxton, A. G. Livingston, and E. N. Pistikopoulos. Reaction Path Synthesis for Environmental Impact Minimization. *Computers Chem. Eng.*, 21:959–964, 1997.
- [3] E. W. Crabtree and M. M. El-Halwagi. Synthesis of Environmentally Acceptable Reactions. In *Pollution Prevention via Process and Product Modifications, AIChE Symp. Ser. 90*, pages 117–127, New York, 1995.

Table 3: Identified solvay reactions in the IRR sets

sequence in IRR sets	No.	SI				SV			
8	(3)	0	4	7	12	0	-1	1	1
10	(2)	0	6	7	10	0	-1	1	1
27	(5)	0	10	22	24	0	1	-1	1
83	(4)	4	5	21	24	-0.5	0.5	-1.0	1
89	(1)	5	6	17	18	-0.5	0.5	-0.5	1
111	(7)	10	12	17	19	-0.5	-0.5	-0.5	1
323	(6)	18	19	21	22	1	-1	-1	1

<sup>b</sup> 1, C; 2,  $Ca(HCO_3)_2$ ; 3,  $Ca(NO_3)_2$ ; 4,  $Ca(OH)_2$ ; 5,  $CaCl_2$ ; 6,  $CaCO_3$ ; 7,  $CaO$ ; 8,  $Cl_2$ ; 9,  $CO$ ; 10,  $CO_2$ ; 11,  $H_2$ ; 12,  $H_2O$ ; 13,  $HCl$ ; 14,  $HNO_3$ ; 15,  $N_2$ ; 16,  $N_2O_5$ ; 17,  $Na_2CO_3$ ; 18,  $NaCl$ ; 19,  $NaHCO_3$ ; 20,  $NH_3$ ; 21,  $NH_4Cl$ ; 22,  $NH_4HCO_3$ ; 23,  $NH_4NO_3$ ; 24,  $NH_4OH$ ; 25,  $NO$ ; 26,  $NO_2$ ; 27,  $O_2$ .

- [4] T. Fornari, E. Rotstein, and G. Stephanopoulos. Studies on the Synthesis of Chemical Reaction Paths -II. Reaction Schemes with Two Degrees of Freedom. *Chem. Eng. Sci.*, 44:1569–1579, 1989.
- [5] T. Fornari and G. Stephanopoulos. Synthesis of Chemical Reaction Path Synthesis: Economic and Specification Constraint. *Chem. Eng. Comm.*, 129:159–182, 1994.
- [6] K. Holiastos and V. Manousiouthakis. Automatic Synthesis of Thermodynamically Feasible Reaction Clusters. *AIChE J.*, 44:164–173, 1998.
- [7] M. Li, S. Hu, Y. Li, and J. Shen. A Hierarchical Optimization Method for Reaction Path Synthesis. *Ind. Eng. Chem. Res.*, 39:4315–4319, 2000.
- [8] M. Li, S. Hu, Y. Li, and J. Shen. Reaction Path Synthesis for a Mass Closed-cycle System. *Computers Chem. Eng.*, 24:1215–1221, 2000.
- [9] D. May and D. F. Rudd. Development of Solvay Clusters of Chemical Reactions. *Chem. Eng. Sci.*, 31:59–69, 1976.
- [10] D. F. Rudd. Accessible Designs in Solvay Clusters Synthesis. *Chem. Eng. Sci.*, 31:701–703, 1976.
- [11] D. S. Tow and D. F. Rudd. Pattern Recognition in Computer-Aided Synthesis of Industrial Reactions. *Ind. Eng. Chem. Res.*, 28:1619–1628, 1989.
- [12] R. H. Wentorf and R. E. Hanneman. Thermochemical Hydrogen Generation. *Science*, 185:311–319, 1974.