CHEME -- A Program for Balancing Chemical Equations

by

Stephen N. Au

September 15, 1972

CHEME

I. Introduction

On October, 1968 George Copley[1] wrote an article on balancing chemical equations by using matrix algebra. This article aroused my interest in using a computer to solve systems of chemical equations since Copley provided a simpler, more complete, and more precise method to deal with the problem than the others[2]. Although modifications have been made to simplify the calculations, the basic mathmatical set-up of the solution followed closely the method described by Copley.

II. Mathmatical Background

An understanding and operation in reducing a matrix or solving a set of homogeneous equations is essential in balancing chemical equations in this program. A discussion of solving a set of homogeneous equations is discussed by Paul Shields[3].

For purpose of illustration, three chemical systems are used as examples. The subscript 'p' and 'r' stands for product, and reactant, respectively.

A. Consider a chemical system made up of Al_2O_3, H^+, H_2O and Al^{+3}. If we let a, b, c, and d to represent the stoichiometric factor of the species, we would have the following:

$$a\ Al_2O_3 + b\ H^+ + c\ H_2O + d\ Al^{+3} = 0$$

Beware that the above is only mathmatically correct and is chemically invalid. It would make sense in Chemistry, as shown later, if some of the variables are negative and all of the negative terms are being placed on the other side of the equation.

By the conservation of matter and charge in a chemical equation, the following is deduced:

$$Al_r = Al_p \quad \text{or} \quad Al_p - Al_r = 0$$
$$H_r = H_p \quad \text{or} \quad H_p - H_r = 0$$
$$O_r = O_p \quad \text{or} \quad O_p - O_r = 0$$

Relating the above to the overall chemical equation, we can set

$$b + (\text{as in } H^+) + 3d + (\text{as in } Al^{+3}) = 0$$
$$2a \text{ Al } (\text{as in } Al_2O_3) + d \text{ Al } (\text{as in } Al^{+3}) = 0$$
$$b \text{ H } (\text{as in } H^+) + 2c \text{ H } (\text{as in } H_2O) = 0$$
$$3a \text{ O } (\text{as in } Al_2O_3) + c \text{ O } (\text{as in } H_2O) = 0$$

The positive sign is used in here other than the negative sign because a, b, c, and d can take in negative values as well. Note that + denotes the positive charge. Expressing in matrix form gives:

$$\begin{array}{c} \\ Al \\ O \\ H \\ + \end{array} \begin{array}{c} \begin{array}{cccc} Al_2O_3 & H^+ & H_2O & Al^{+3} \end{array} \\ \begin{bmatrix} 2 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{bmatrix} \end{array} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

It can be seen that the vertical column refers to the particular compound and the horizontal row gives the number of the particular element present in the compound. For example, Al_2O_3 has 2 Al, 3 O, 0 H, and 0 + . If the matrix multiplication is carried out, it is easily seen that the matrix expression is just another way of stating the algebraic equations. Actually, if the user is not familiar with the matrix algebra, he can readily solve the equations by solving a system of simultaneous equations.

Reducing the matrix gives

$$\begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 3/2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 6 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In other words,

$$2a = -d$$
$$2b = 6d$$
$$2c = -3d$$

Letting d=2 and rearranging the negative terms to the other side of the original overall equation, a balanced chemical equation is thus obtained:

$$3\ H_2O + 2\ Al^{+3} = Al_2O_3 + 6\ H^+$$

B. As another example, examine the following system:

$a\ IO_4^- + b\ Mn^{+2} + c\ H_2O + d\ MnO_4 + e\ IO_3^- + f\ H^+ = 0$

At the first glance, it may be tempting to set up the matrix as shown below;

$$\begin{array}{c} \\ I \\ O \\ - \\ H \\ Mn \\ + \end{array} \begin{array}{c} \begin{array}{cccccc} IO_4^- & Mn^{+2} & H_2O & MnO_4 & IO_3^- & H^+ \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 1 & 4 & 3 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

However, a second thought would refute the above because of the fact that positive charge and negative charge may not be conserved in the reaction; although it is true that the net charge is conserved. To solve the problem, we have to use the net charge in the equation, namely that

$$(\text{net charge})_r = (\text{net charge})_p$$

$$(\text{positive charge} + \text{negative charge})_r = (\text{positive charge} + \text{negative charge})_p$$

or

$$((\text{positive charge})_p - (\text{positive charge})_r) + ((\text{negative charge})_p - (\text{negative charge})_r) = 0$$

and vice versa. In terms of the stoichiometric coefficient matrix, this means that

$$\text{row}_+ - \text{row}_- = 0$$

since both positive charge and negative charge are represented by positive numbers. Following the discussion, the system should be written as

$$\begin{array}{c} \\ I \\ O \\ \epsilon \\ H \\ Mn \end{array} \begin{array}{c} \begin{array}{cccccc} IO_4^- & Mn^{+2} & H_2O & MnO_4 & IO_3^- & H^+ \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 1 & 4 & 3 & 0 \\ -1 & 2 & 0 & 0 & -1 & 1 \\ 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where ϵ is the net charge of the compound. A more convenient way to interpret ϵ is to follow the convention that positive number represent positive charge while negative charge are represented by negative numbers. Reducing, we can have

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 5/6 \\ 0 & 1 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & -1/3 \\ 0 & 0 & 0 & 0 & 1 & -5/6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

or equivalently,

$$\begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 5 \\ 0 & 6 & 0 & 0 & 0 & 2 \\ 0 & 0 & 6 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 & 0 & -2 \\ 0 & 0 & 0 & 0 & 6 & -5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Setting f to be 6, we can read off the solution as

$$a = -5$$
$$b = -2$$
$$c = 3$$
$$d = 2$$
$$e = 5$$

Referring back to the overall equation, the following is obtained;

$$5\ IO_4^- + 2\ Mn^{+2} + 3\ H_2O = 5\ IO_3^- + 2\ MnO_4^- + 6\ H^+$$

C. The above examples can be think of as a one-variable--dependence system (vocality=1) because the equation can be solved if the numerical value of one of the variables is known. In this example, a multi-variable-dependence system (vocality greater than or equal to 2) is shown in which case the numerical values of more than one variable must be known for the solution of the equations. A system such as

$$a\ e^- + b\ H^+ + c\ O_2 + d\ H_2O + e\ O_3 = 0$$

gives the following set-up.

$$\begin{array}{c} \\ \epsilon \\ H \\ O \end{array} \begin{array}{c} \begin{array}{ccccc} e^- & H^+ & O_2 & H_2O & O_3 \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 2 & 1 & 3 \end{bmatrix} \end{array} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

After reducting, we have

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1/2 & 3/2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Converting to integers, we have

$$\begin{bmatrix} 2 & 0 & 0 & 4 & 0 \\ 0 & 2 & 0 & 4 & 0 \\ 0 & 0 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This means that

$$2a + 4d = 0$$
$$2b + 4d = 0$$
$$2c + d + 3e = 0$$

As can be seen, the equation can be solved only when the numerical value of two of the variables is known. In this case, it is easier to substitute for d and e. By letting d to be 1 and e to be 1, we have

$$2\,e^- + 2\,H^+ + 2O_2 = H_2O + O_3$$

since $a = -2$

$b = -2$

$c = -2$

Another possible substitution would be d=-1 and e = 1 which leads to

$a = 2$

$b = 2$

$c = -1$

The balanced equation for this set is

$$O_2 + H_2O = O_3 + 2\,e^- + 2\,H^+$$

By varying d and e, we can have entirely different chemical equations (and many others more !!!). With this kind of system, proper choice of substitution is essential to the solution of the problem.

III. Programming Aspect of CHEME

CHEME is a PL/I coded program. It was initiated in the summer of 1971 and got underway in January, 1972. Roughly speaking, CHEME can be divided into two parts. The first part interprets the input and the second part performs the necessary computations.

A. Input. CHEME reads in input cards. Only 72 columns out of the 80 columns are read by CHEME (i.e. col 1 to col. 72). The input is the empirical formula, written in a special notation, of all the involved chemical species. The special notation consists of

= denotes linkage between elements as they are in the same compound, and

* denotes subscript of the element.

Parentheses can also be used together with the chemical formulas. For example,

Chemistry Notation	Input Notation
H_2O	H*2=O
CO_2	C=O*2
H	H
H^+	H=+
OH^-	O=H=-
Al_2O_3	AL*2=O*3
CH_3NH_2	C=H*3=N=H*2
HNNN	H=N=N=N (or H=N*3)
$(NH_2)_2CO$	(N=H*2)*2=C=O
$(CH_3)_2CHCOOH$	(C=H*3)*2=C=H=C=O=O=H
$(Pt(NH_3)_6)Cl_4$	(PT=(N=H*3)*6)=CL*4

Blanks, comma, and semicolon can be used to seperate chemical formulas from one another. Different sets of chemical system can be solved in the same run by placing a dollar sign in column 72 between different sets and starting the next set in the following card (only 72 columns are used.) It is not necessary nor advisable to place the dollar sign in column 72 when the last set is reached or else the program attempts to read in

another set of data which would result in the raising of an error condition (i.e. end of file).
It is important to note that each formula can be written only once in the same set. As the program stands now, it can not differentiate between products and reactants while reading in the input, and such kind of information is not accepted by the program. The difference between products and reactants is determined by the user after the program runs because the program merely solve the mathmatical portion of the problem, in which case the reactants and the products are mathmatical equivalent in this program.

B. Output. There are two output formats for the balanced equation depending on the solution of the problem. Simulation of the computer outputs are shown below.

1. One-variable-dependence system. The chemical system in Example A is used and the following output is produced.

REACTANTS	PRODUCTS
2 AL=+*3	1 AL*2=O*3
3 H*2=O	6 H=+

For Example B, the output looks like:

REACTANTS	PRODUCTS
2 MN=O*4=-	5 I=O*4=-
5 I=O*3=-	2 MN=+*2
6 H=+	3 H*2=O

As stressed before, the term 'reactants' and 'products' are used loosely and may not have any real significance in the Chemistry sense. It is obvious that the balanced chemical equation can easily be deduced from the output.

2. Multi-variable-dependence system. The output for this system is not so obvious as that one before. Simple algebraic manipulation by the user is needed in order to obtain the desired balanced equation. A sample of the output for Example C is given below:

STOICHIOMETRIC FACTOR	CHEMICAL SPECIES
+ 4D	- 2A -
+ 4D	- 2B H=+
+ 1D + 3E	- 2C O*2
1D	1D * H*2=0
1E	1E O*3

The set {A, B, C, D, ..., W, X, Y, Z} stands for any algebraic variable and with proper choice of numerical substitution for the variables, one can set the specie either to be the reactant or the product. Reactants are represented by negative numbers while products are represented by positve numbers in this program. The user can reverse the interpretation as he sees fit.

From the output, an equivalence is implied between the STOICHIOMETRIC FACTOR and the coefficient of the CHEMICAL SPECIES, i.e.

$$+ 4D = - 2A$$
$$+ 4D = - 2B$$
$$+ 1D + 3E = - 2C$$
$$1D = 1D$$
$$1E = 1E$$

Substituting for D and E as in Example C lead to the solution of A, B, and C. After the numerical solution of the variables has been known, the user can easily deduce the chemical equation. Remember that the overall equation is written in the form of

$$A_1X_1 + A_2X_2 + A_3X_3 + \ldots A_nX_n = 0$$

where A is the stoichiometric factor,
X is the involved chemical species, and
N is the number of chemical species.

There are some cases in which the chemical equation can not be balanced. In that situation, a message

'***NO CHEMICAL EQUATION IS POSSIBLE***'

is printed out. Following the print out of the balanced equation, the molecular weights of the formulas are also available.

C. Structure. This section is prepared for people who are interested in the program structure of CHEME and may not be of interest to the general users.

Roughly speaking, CHEME can be divided into two parts. The first part interprets the input and the second part performs the necessary computations. A simplified overall flowchart of CHEME can be seen in Fig. 1.

There are eight subroutines in CHEME. Their functions are briefly described below.

1. REDUS REDUS reduces a given matrix to its reduced form. The subroutine helps in solving a set of homogeneous equations and constitutes the main backbone in computation.
2. OUTPUT This is a two-dimensional matrix output routine. It is used in conjunction with REDUS.
3. PROPOT. This block converts the stoichiometric coefficient matrix (in a one-dimensional linear vector form) to an integer set. The accuracy desired in this case is +/- .02 (e.g. 4.98 and 5.02 would both be interpreted as 5).

4. DATA This procedure reads in the input as character strings and determines compound formulas which are seperated from each other by blanks, a comma, or a semicolon.
5. SUBSIP SUBSIP determines the subscript of each element in a compound formula. This procedure contributes to the set-up of the stoichiometric coefficient matrix.
6. PAREN PAREN is a parenthesis scanner which picks out the locations of parentheses in the string. After the locations are found, matching is performed to pair up the left and right parentheses. It is used together with SUBSIP to determine parenthesized elements.
7. ELEMEN From a given compound formula, ELEMEN finds out all the elements, including the charge, that make up the compound.
8. LENG LENG is a 'length' function which determines the 'length' of a given string. Length in this program refers to the number of 'non-blank' characters, including imbedded blanks and excluding trailing blanks, assigned to a given string. It is used in SUBSIP to determine the length of the character string which represent the element.

D. Limitation. There are several major limitations within this program. They are listed below.
1. The program does not check for valid chemical formulas nor equations thus produced can be evaluated as chemically valid.
2. The program does not check for valid input which leaves the program 'unprotected' to mistakes made by users.

E. Conclusion. A program has been produced which can be used to balance chemical equations. It can be used to balance all types of equations including redox, acid-base, and many others as well. It is hoped that this program can give students in Chemistry a better understanding of the relations between Chemistry, Mathmatics, and Computer Science.

FOOTNOTES

[1]George Novello Copley, 'Linear Alegbra of Chemical Formulas and Equations', Chemistry, Vol 41, No 9, October, 1968.

[2]There are some other methods in solving the problem, like using minors. But they tend to be quite tedious and are not suitable for computer programming. These methods were released by Chemistry.

[3]Paul C. Shields, Elementary Linear Algebra, pp. 1-12, New York, Worth Publishers, inc., 1970.

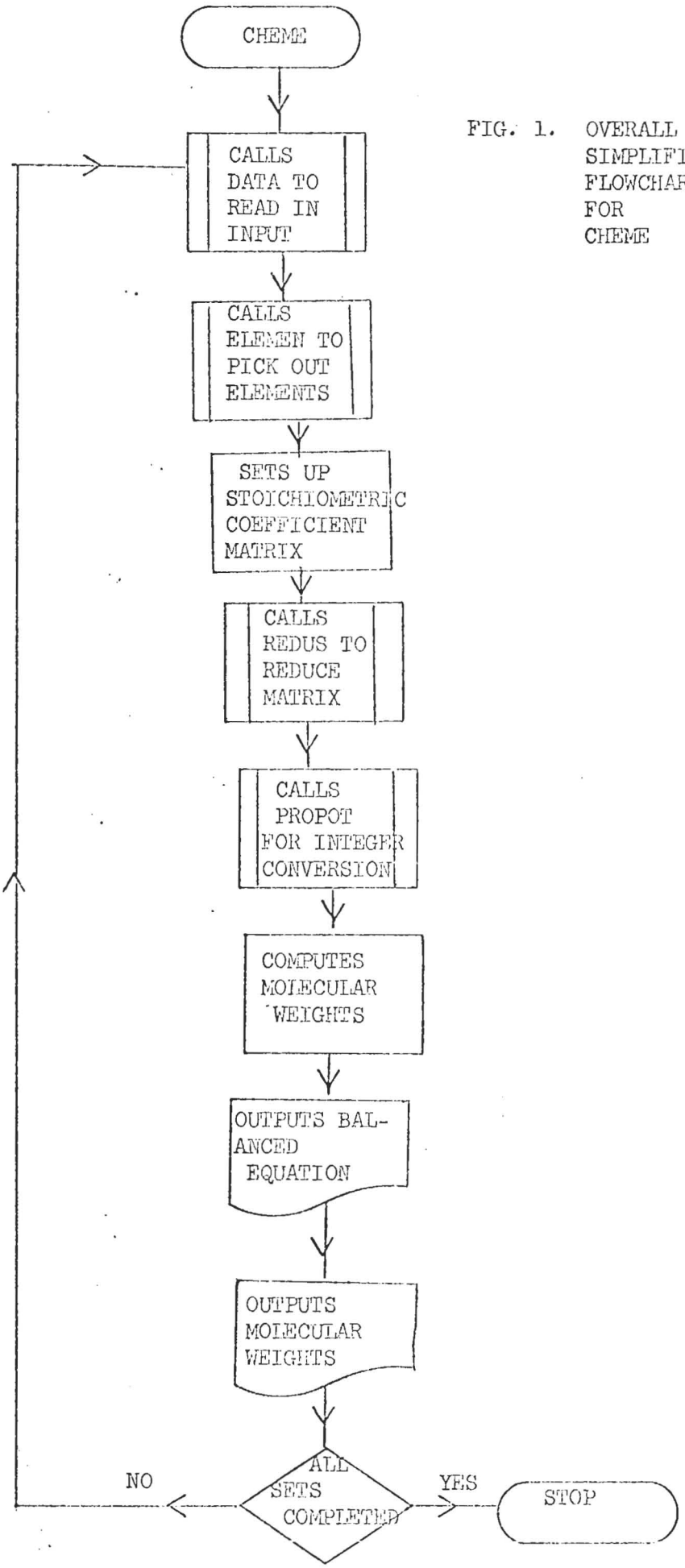

FIG. 1. OVERALL SIMPLIFIED FLOWCHART FOR CHEME