match no. 13 pp. 7-15 1982

COMBINATORIAL CONSTRUCTION OF CHEMICAL REACTION EQUATIONS

Tetsuo Morikawa

School of Pharmaceutical Science, Toho University, Miyama, Funabashi-shi, Japan

(Received: October 25, 1982)

(Abstract)

A generating function method is developed for forming from a given set of molecules all types of chemical reaction equations that satisfy the law of conservation of chemical elements. Algorithms for computer calculation of the equations are presented.

Introduction. Here are simple examples of chemical reaction:

$$H_2 + O_2 \rightleftharpoons H_2 + O_2,\ 2H_2 + O_2 \rightleftharpoons 2H_2O,$$

$$3H_2 + O_2 \rightleftharpoons H_2 + 2H_2O,\ \ldots$$

Chemists derive such reaction equations from a set of molecules $\{H_2, O_2, H_2O\}$ by intuition and/or by trial and error, so as to balance the number of atoms (or chemical elements) in both sides of the equations. In the second equation the number of hydrogen (oxygen) atoms is equal to

4 (2). However, when the number of molecules in a given set increases, it would become more difficult to construct all of reaction equations by intuition. The present note will describe a general method that can form reaction equations in full by means of the so-called generating function.

Illustration. An infinite series $A(t) = \sum_{n=0}^{\infty} a_n t^n$ is called generating function associated with a sequence $\{a_n; n=0,1,2,\ldots\}$, t being any parameter(hereafter abbreviated as g.f.). We write a g.f. for choosing hydrogen molecules from the set $\{H_2, O_2, H_2O\}$ as

$$A(t) = 1 + at^2 + a^2t^4 + a^3t^6 + \ldots = (1 - at^2)^{-1}$$

where the indices (or exponents) on a refer to the number of hydrogen molecules, and those on t to that of hydrogen atoms. For oxygen and water molecules g.f.'s are expressed in a similar manner above as

$$B(u) = (1 - bu^2)^{-1} \quad \text{and} \quad C(t,u) = (1 - ct^2u)^{-1}$$

where the index on u represents the number of oxygen atoms. The product of A(t), B(u) and C(t,u)

$$R(t,u) = A(t)B(u)C(t,u) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} r(n,m)t^n u^m$$

becomes a new g.f., which obviously yields summands of molecules with the numbers (n and m) of the atoms. These subsets $r(n,m)$ suggest all possible expressions for separate parts of chemical reaction equations (the left or right sides of equations). In $r(4,2) = a^2b + c^2$ the first and second terms can be considered to be the meaning of $2H_2 + O_2$ and $2H_2O$, respectively.

<u>Definition</u>. Let $M = \{M^1, M^2, \ldots, M^g\}$ and $a = \{a_1, a_2, \ldots, a_g\}$ be the sets of g different molecular species and g corresponding parameters, respectively. Let $E = \{E^1, E^2, \ldots, E^f\}$ and $t = \{t_1, t_2, \ldots, t_f\}$ be the sets of f different elemental species and f corresponding parameters, respectively. Assume that the i-th molecule M^i consists of f elemental species whose numbers are $s_{i1}, s_{i2}, \ldots, s_{if}$, s_{ik} being natural numbers (positive integers or zeroes). Each row (and each column) of the matrix (s_{ik}) has at least one nonzero entry. We define a g.f. associated with $r(n_1,n_2,\ldots,n_f)$ by

$$R(t_1,t_2,\ldots,t_f) = A_1(t_1,t_2,\ldots,t_f)A_2(t_1,t_2,\ldots,t_f)\ldots$$
$$\ldots A_g(t_1,t_2,\ldots,t_f)$$
$$= \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\ldots\sum_{n_f=0}^{\infty} r(n_1,n_2,\ldots,n_f)t_1^{n_1}t_2^{n_2}\ldots t_f^{n_f}$$

in which $$A_i(t_1,t_2,\ldots,t_f) = (1 - (a_i t_1^{s_{i1}} t_2^{s_{i2}} \ldots t_f^{s_{if}}))^{-1}$$

Note that $a_i t_1^{s_{i1}} t_2^{s_{i2}} \ldots t_f^{s_{if}}$ is used to mean one of the i-th molecules, $(a_i t_1^{s_{i1}} t_2^{s_{i2}} \ldots t_f^{s_{if}})^2 = a_1^2 t_1^{2s_{i1}} t_2^{2s_{i2}} \ldots t_f^{2s_{if}}$ to two of those molecules, and so on.

Solution. (Step I) In order to determine $r(n_1, n_2, \ldots, n_f)$ explicitly, we consider a g.f. below

$$R_i(t_i) = A_1(t_i) A_2(t_i) \ldots A_g(t_i) = \sum_{n_i=0}^{\infty} r_i(n_i) t_i^{n_i}$$

where $A_k(t_i) = (1 - a_k t_i^{s_{ki}})^{-1}$. $A_k(t_i)$ is a g.f., which chooses fragments of the molecule M^k with respect to the element E^i. Indicating $r_i(n_i)$ for $a_1 = a_2 = \ldots = 1$ by $p_i(n_i)$ we get the g.f.

$$P_i(t_i) = (1 - t_i^{s_{1i}})^{-1}(1 - t_i^{s_{2i}})^{-1} \ldots (1 - t_i^{s_{gi}})^{-1}$$
$$= \sum_{n_i=0}^{\infty} p_i(n_i) t_i^{n_i}$$

This suggests that $p_i(n_i)$ is the number of ways of choosing different sets of solutions (namely, the number of partitions) for an algebraic equation (Diophantine equation in number theory, Reference 1)

$$n_i = s_{1i} x_1 + s_{2i} x_2 + \ldots + s_{gi} x_g$$

whose solutions $\{x_1, x_2, \ldots, x_g\}$ are sought in the set of natural numbers. (Note: $p_i(0) = 1$) We can then construct $r_i(n_i)$ using the $p_i(n_i)$ sets of solutions:

$$r_i(n_i) = \langle x_1(1), x_2(1), \ldots, x_g(1)\rangle + \langle x_1(2), x_2(2), \ldots, x_g(2)\rangle$$

$$+ \ldots + \langle x_1(p_i(n_i)), x_2(p_i(n_i)), \ldots, x_g(p_i(n_i))\rangle$$

The symbol $\langle x_1, x_2, \ldots, x_g\rangle$ is used to mean $a_1^{x_1} a_2^{x_2} \ldots a_g^{x_g}$, and is interpreted as $x_1M^1 + x_2M^2 + \ldots + x_gM^g$; $\{x_1(m), x_2(m), \ldots, x_g(m)\}$ is the m-th set of solutions. Note that if s_{ki} is equal to zero, x_k is indeterminate; then the index on a_k is denoted by the star * as $a_k{}^*$ for convenience of calculation below.

Multiplying both sides of $R_i(t_i)$ by $(A_1(t_i)A_2(t_i)\ldots A_g(t_i))^{-1}$ we have

$$1 = \sum_{n_i=0}^{\infty} (1 - a_1 t_i^{s_{1i}})(1 - a_2 t_i^{s_{2i}})\ldots$$

$$\ldots(1 - a_g t_i^{s_{gi}}) r_i(n_i) t_i^{n_i}$$

By comparison of the powers of t_i in both sides of this equation the recursive formulas are obtained, and make it possible to carry out the computer calculation of $r_i(n_i)$. (Note: $r_i(0) = 1$)

Example. For hydrogen,

$$1 = \sum_{n=0}^{\infty} (1 - at^2)(1 - ct^2) r_H(n) t^n$$

$$= \sum_{n=0}^{\infty} (r_H(n)t^n - (a + c)r_H(n)t^{n+2} + ac r_H(n)t^{n+4})$$

$$
\begin{array}{ll}
n=0 & 1 = r_H(0) \\
=1 & 0 = r_H(1) \\
=2 & 0 = r_H(2) - (a + c)r_H(0) \\
=3 & 0 = r_H(3) - (a + c)r_H(1) \\
=4 & 0 = r_H(4) - (a + c)r_H(2) + acr_H(0) \\
=5 & 0 = r_H(5) - (a + c)r_H(3) + acr_H(1) \\
\cdots & \cdots\cdots\cdots\cdots\cdots
\end{array}
$$

These give the general form

$$r_H(n) = (a + c)r_H(n-2) - acr_H(n-4), \quad r_H(0) = 1,$$
$$r_H(1) = 0, \quad r_H(2) = a + c, \quad r_H(3) = 0.$$

Inserting b* in $r_H(n)$ we have

$$R_H(t) = b^*(1 + (a + c)t^2 + (a^2 + ac + c^2)t^4 + (a^3 + a^2c + ac^2 + c^3)t^6 + \dots)$$

For oxygen,

$$R_O(u) = a^*(1 + cu + (b + c^2)u^2 + (bc + c^3)u^3 + (b^2 + bc^2 + c^4)u^4 + (b^2c + bc^3 + c^5)u^5 + \dots)$$

(Step II) We now return to $R(t_1,t_2,\dots,t_f)$. In order to obtain $r(n_1,n_2,\dots,n_f)$ we must solve a system of f simultaneous equations:

$$n_i = s_{1i}x_1 + s_{2i}x_2 + \dots + s_{gi}x_g \quad (i=1,2,\dots,f)$$

We introduce the symbol $\wedge$ that can conveniently display the solutions of the system; $\wedge$ acts only on the powers of a_i's according to

$$\langle x_1,x_2,\ldots,x_g\rangle \wedge \langle y_1,y_2,\ldots,y_g\rangle = (x_1:y_1)(x_2:y_2)\ldots(x_g:y_g)$$

where $(x_i:y_i) = a_i^{x_i}$ if $x_i = y_i$, and $(x_i:y_i) = 0$ otherwise. For instance,

$a_1^3a_2 \wedge a_2 = (3:0)(1:1) = 0\ a_2 = 0.$ Notice that $a_i^* \wedge a_i^{x_i} = a_i^{x_i}$, and $a_i^* \wedge a_j^* = 1\ (i \neq j)$. (Note: $(\langle x\rangle + \langle y\rangle) \wedge \langle z\rangle = (\langle x\rangle \wedge \langle z\rangle) + (\langle y\rangle \wedge \langle z\rangle)$)

The calculation of the $\wedge$ products of $R_i(t_i)$'s gives

$$R_1(t_1) \wedge R_2(t_2) \wedge \ldots \wedge R_f(t_f) = \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\ldots\sum_{n_f=0}^{\infty} r_1(n_1) \wedge r_2(n_2) \wedge \ldots \wedge r_f(n_f)t_1^{n_1}t_2^{n_2}\ldots t_f^{n_f}$$

of which the left hand side can be shown to be equal to the left hand side of $R(t_1,t_2,\ldots,t_f)$, after some calculations, for example,

$$(A_1(t_1)A_2(t_1)) \wedge (A_1(t_2)A_2(t_2)) = A_1(t_1,t_2)A_2(t_1,t_2)$$

where $A_k(t_1, t_2) = (1 - a_k t_1^{s_{k1}} t_2^{s_{k2}})^{-1}$. Thus $r(n_1, n_2, \ldots, n_f)$ is determined uniquely:

$$r(n_1, n_2, \ldots, n_f) = r_1(n_1) \wedge r_2(n_2) \wedge \ldots \wedge r_f(n_f)$$

Example. $r(n,m) = r_H(n) \wedge r_O(m)$

n ↓ \ m →	0	1	2	3	4	5	...
0	1	0	b	0	b^2	0	...
1	0	0	0	0	0	0	...
2	a	c	ab	bc	ab^2	b^2c	...
3	0	0	0	0	0	0	...
4	a^2	ac	a^2b+c^2	abc	$a^2b^2+bc^2$	ab^2c	...
5	0	0	0	0	0	0	...
6	a^3	a^2c	a^3b+ac^2	a^2bc+c^3	$a^3b^2+abc^2$	$a^2b^2c+bc^3$	...
⋮	..						

<u>Discussion</u>. We express $r(n_1, n_2, \ldots, n_f)$ for $a_1 = a_2 = \ldots = a_g = 1$ as $p(n_1, n_2, \ldots, n_f)$, which refer to the number of terms $\langle x_1, x_2, \ldots, x_g \rangle$'s in $r(n_1, n_2, \ldots, n_f)$. Since any term $\langle x_1, x_2, \ldots, x_g \rangle$ is interpreted as a representation for the left or right side of a reaction equation with the numbers $(n_1, n_2, \ldots, n_f)$ of atoms, we can conclude the following. (a) If $p(n_1, n_2, \ldots, n_f) = 0$, then no reaction occurs. (b) If $p(n_1, n_2, \ldots, n_f) = 1$ (except for $p(0,0,\ldots,0) = 1$), then it is possible to write the only type of reaction; in the equation the left side is identical with the right side. (c) The case: $p(n_1, n_2, \ldots, n_f) \geqq 2$.

(i) If a kind of term is chosen from $r(n_1, n_2, \ldots, n_f)$, and if it constitutes both sides of an equation, then the reaction is trivial as in (b). (ii) If any two terms are used to constitute an equation, then both sides of the reaction are different. The number of these types of reaction is $\binom{p(n_1, n_2, \ldots, n_f)}{2}$, (the combinations of $p(n_1, n_2, \ldots, n_f)$ things taken 2 at a time). (iii) To form a reaction equation with numbers n_1, n_2, ..., n_f, three (and four ...) terms cannot be chosen from $r(n_1, n_2, \ldots, n_f)$.

Remarks. In a substitution reaction

$$CH_3CH_2CH_2Cl + OH^- \rightleftharpoons CH_3CH_2CH_2OH + Cl^-$$

the elemental species can be taken as $CH_3CH_2CH_2^+$, Cl^-, and OH^-. In an electrochemical mechanism

$$H^+ + e^- \rightleftharpoons H$$

the electron is an elemental species. In

$$Cl + Cl_2 \rightleftharpoons Cl_2 + Cl$$

we can distinguish the left from the right side, because we are able to rewrite Cl and Cl_2 as, for example, ^{37}Cl and $^{35}Cl_2$.

Reference. 1. "Encyclopedic Dictionary of Mathematics", vol. 1, p.232,p.416, edited by S. Iyanaga and Y. Kawada, The MIT Press(1977).