Theor Chim Acta (1989) 76: 47–51

Theoretica
Chimica Acta

Series analysis methods in enumeration of chemical isomers

Jerzy Cioslowski
Los Alamos National Laboratory, Theoretical Division, Group T12, MS J569,
Los Alamos, NM 87545, USA

(Received June 14, revised November 9, 1988/Accepted January 31, 1989)

Summary. Methods of series analysis are used to interpret the results of a computer enumeration of chemical isomers. It is demonstrated how a typical asymptotic behavior that controls the number of isomers arises from singular points of the generating function. The approach is tested on the examples of the isomers of alkanes, alkenes, alkyl radicals and polyenes.

Key words: Enumeration of chemical isomers — Generating function — Series analysis

1. Introduction

Even if the problem of enumeration of chemical isomers has been solved for particular structures (such as alkanes) a long time ago [1], the analytical formulas for the number of chemical isomers are still unknown in general. This means that we have to resort to computer simulations in order to count chemical structures. With the advent of fast computers, those simulations became increasingly popular [2, 3]. However, the methods of interpretation of the resulting data have remained to be elaborated. A successful interpretation would yield asymptotic formulas for the number of structures and, if one is fortunate enough, exact analytical expressions.

The chemical literature seems to be lacking such interpretations, the investigation by Gutman [4] being the only exception. In the present paper we show how the data on isomer enumeration can be analyzed by methods that are widely used by physicists in the lattice statistics.

2. Theory

Let n be an index associated with the chemical structures under investigation. It can be the number of carbon atoms or benzene rings, the length of a perimeter of a benzenoid hydrocarbon, etc. We will denote the number of chemical isomers pertinent to the index n by $M(n)$. The generating function is given by a formal series:

$$N(x) = \sum_{k=1}^{\infty} M(k)x^k. \tag{1}$$

We note that usually $M(k+1) \geqslant M(k)$ and therefore the series (1) has a finite radius of convergence. This means that the generating function possesses at least one singular point.

The simplest case of the singularity is a regular singular point with the asymptotics:

$$N(x) \rightarrow (x - x_c)^{\zeta_c} \quad \text{as } x \rightarrow x_c. \tag{2}$$

We call x_c the critical point of $N(x)$ and ζ_c the critical exponent. The singular point (2) gives rise to the asymptotic behavior of the coefficients $M(n)$:

$$M(n) \rightarrow (1/x_c)^n n^{-(\zeta_c + 1)} \quad \text{as } n \rightarrow \infty. \tag{3}$$

This kind of asymptotics, which is dominated by the critical point that has the smallest absolute value, is very often observed in practice [1]. The critical exponent is usually a rational number.

To illustrate the above points, we quote here the results by Harary and Read [5]. The number of catacondensed benzenoid hydrocarbons (both planar and non-planar), where the index n is the number of benzene rings, has the associated generating function that reads:

$$\begin{aligned} N_h(x) = {} & (24x^2)^{-1}[12 + 24x - 48x^2 - 24x^3 + (1-x)^{3/2}(1-5x)^{3/2} \\ & - 3(3+5x)(1-x^2)^{1/2}(1-5x^2)^{1/2} - 4(1-x^3)^{1/2}(1-5x^3)^{1/2}]. \end{aligned} \tag{4}$$

The critical point with the smallest absolute value is $x_c = 0.2$ and the corresponding critical exponent is $\zeta_c = 3/2$. Thus, we can expect the asymptotics:

$$M_h(n) \rightarrow 5^n n^{-5/2}. \tag{5}$$

The same result has been obtained previously after somehow lengthy algebraic manipulations [4, 5].

Our goal is to extract the unbiased values of x_c and ζ_c from a limited number of known coefficients, $M(n)$. The mathematical machinery required for this task is well known and widely used in lattice statistics [6]. One particular method is known under the names of the recurrence relations approach [7], the differential Padé approximants [8] or the Gaunt–Gammel approximants [9]. The formalism seems to be largely unknown to the chemical community. Therefore we recollect its basic points here, following the paper by Guttmann and Joyce [7].

We assume that we know $k+1$ consecutive coefficients of the series (1): $\{M(n_0), \ldots, M(n_0+k)\}$. We will use the abbreviation:

$$c_i = M(n_0+i), \qquad i = 0, \ldots, k. \tag{6}$$

First, we fit these coefficients to an $(m+1)$ term recurrence relation of the form:

$$F_m(x_j) = \sum_{i=0}^{m} [A_{i,2}(j-i)^2 + A_{i,1}(j-i) + A_{i,0}]c_{j-1} = 0, \tag{7}$$

with $A_{0,2} = 1$, $A_{0,0} = 0$ and $c_j = 0$ for $j < 0$. The unknown coefficients $A_{0,1}, A_{1,0}, A_{1,1}, A_{1,2}, \ldots, A_{m,0}, A_{m,1}, A_{m,2}$ are determined from the set of relations (6) for $j = 1, \ldots, 3m+1$. This is simply done by solving a system of $3m+1$ linear equations. The next step is to form two auxiliary polynomials:

$$Q(x) = x \sum_{i=0}^{m} A_{i,2}x^i, \tag{8}$$

and

$$R(x) = \sum_{i=0}^{m} (A_{i,2} + A_{i,1})x^i. \tag{9}$$

Eventually, the critical points are found as the zeros of $Q(x)$ and the corresponding critical exponents are given by

$$\zeta_c = 1 - R(x_c)/Q'(x_c). \tag{10}$$

Equation (10) follows from Eq. (6) in [7] by virtue of the l'Hôpital rule. It should be pointed out that the recurrence length, m, can assume all values from the range $1 \leqslant m \leqslant [(k-1)/3]$, where $[z]$ means the integer part of z. In general, the accuracy of x_c and ζ_c improves with increasing n_0 and m.

3. Examples and discussion

We tested the above formalism on the results of computer enumeration of alkanes, alkyl radicals, alkenes and polyenes. The computed critical points and critical exponents are given in Tables 1–4.

It appears that the differential Padé approximants are capable of yielding the values of x_c and ζ_c with at least five-digits accuracy. The convergence is faster for alkenes and polyenes than for alkanes and alkyl radicals. The asymptotics is controlled by the critical exponents of 3/2 for the structures built up from identical fragments (alkanes and polyenes) and 1/2 for the structures that have one distinct fragment (one double bond in alkenes and one unsaturated carbon atom in radicals). Interestingly enough, substitution does not change the value of x_c.

Table 1. Series analysis for alkanes (n is the number of carbon atoms)

m	$n_0 = 15$		$n_0 = 20$	
	x_c	ζ_c	x_c	ζ_c
8	0.35516	1.51021	0.35516	1.51090
9	0.35517	1.50803	0.35518	1.49979
10	0.35518	1.50137	0.35518	1.50067
11	0.35518	1.49824	0.35518	1.50003
12	0.35518	1.49927	0.35518	1.50024
13	0.35518	1.50018	0.35518	1.50000

Table 2. Series analysis for alkyl radicals (n is the number of carbon atoms)

m	$n_0 = 15$		$n_0 = 20$	
	x_c	ζ_c	x_c	ζ_c
3	0.35537	0.47019	0.35496	0.56315
4	0.35517	0.50466	0.35517	0.50358
5	0.35516	0.50620	0.35518	0.49862
6	0.35518	0.49981	0.35519	0.49731
7	0.35519	0.49770	0.35518	0.49990
8	0.35518	0.50001	0.35518	0.50008

Table 3. Series analysis for alkenes (n is the number of carbon atoms)

m	$n_0 = 10$		$n_0 = 15$	
	x_c	ζ_c	x_c	ζ_c
4	0.35533	0.46109	0.35512	0.52350
5	0.35516	0.50692	0.35518	0.49984
6	0.35517	0.50378	0.35518	0.50005
7	0.35517	0.50269	0.35518	0.49896
8	0.35518	0.49995		
9	0.35518	0.49841		

Table 4. Series analysis for polyenes (n is the number of carbon-carbon double bonds)

m	$n_0 = 10$		$n_0 = 15$	
	x_c	ζ_c	x_c	ζ_c
4	0.20938	1.40488	0.20915	1.50395
5	0.20912	1.51772	0.20928	1.41893
6	0.20909	1.54242	0.20916	1.49780
7	0.20915	1.50076	0.20915	1.49996
8	0.20915	1.49994		
9	0.20915	1.49999		

Once the value of critical exponent is deduced, one can obtain an extra accuracy in the critical point by using biased estimates for x_c [6]. If one assumes that x_c is a root of some simple function, this can be a first step in reconstruction of $N(n)$ from the computer-generated data. This can be very useful for further studies on the problem of enumeration of chemical structures.

Acknowledgement. This work has been performed under the auspices of the US DOE.

References

1. Polya G (1937) Acta Mach 68:145; Harary F, Palmer EP (1973) Graphical enumeration. Academic Press, New York
2. Knop JV, Müller WR, Szymanski K, Trinajstic N (1985) Computer generation of certain classes of molecules. Union of Chemists and Technologists of Croatia, Zagreb
3. Balaban AT, Brunvoll J, Cioslowski J, Cyvin BN, Cyvin SJ, Gutman I, He Wenchen, He Wenjie, Knop JV, Kovačevic M, Müller WR, Szymanski K, Tošic R, Trinajstic N (1987) Naturforsch Z 42a:863
4. Gutman I, (1987) Naturforsch Z 41a:1089
5. Harary F, Read RC (1970) Proc Edinburgh Math Soc 17:1
6. Hunter DL, Baker Jr GA (1973) Phys Rev B7:3346; B7:3376
7. Guttmann AJ, Joyce GS (1972) J Phys A5:L81
8. Hunter DL, Baker Jr GA (1979) Phys Rev B19:3808
9. Gammel JL (1973) Padé approximants and their applications. In: Graves-Morris PR (ed) Academic Press, London