

Approximation of any physicochemical constants of homologues with the use of recurrent functions

Igor G. Zenkevich

Received: 5 October 2006 / Accepted: 30 September 2008 / Published online: 14 July 2009
© Springer Science+Business Media, LLC 2009

Abstract Dependencies of various physicochemical constants of organic compounds (A) versus number of carbon atoms in the molecule within different homologous series [$A = f(n_C)$] usually are non-linear. The simplest recurrent equation $A(n + 1) = aA(n) + b$, connecting A -values for homologues ($n + 1$ carbon atoms) with the values of the same constants for previous members of series (n carbon atoms), indicates practically “ideal” linear character for most properties of organic compounds. It is the reasonable basis for approximation (or extrapolation) any physicochemical constants within any homologous series using the standard approach without special selection of appropriate algebraic functions. Principal mathematical properties of the function $A(n + 1) = aA(n) + b$ and some of its chemical applications are considered.

Keywords Organic compounds · Homologous series · Physicochemical constants · Recurrent function · Mathematical properties

1 Introduction

Various physicochemical constants of organic compounds (at first boiling points, T_b ; melting points, T_m ; refractive indices, n_D^{20} ; relative densities, d_4^{20} , etc.) are widely used in the characterization and identification of substances up to present [1]. However, the total multitude of organic compounds cannot be exhaustively characterized by all constants. Many compounds have been synthesized only once many years ago [2] and, as a result, such parameters as chromatographic retention indices (RI), ionization potentials, standard partition coefficients ($\log P$) and others cannot be determined for these

I. G. Zenkevich (✉)

Department of Gas Chromatographic Laboratory, Chemical Research Institute,
St. Petersburg State University, Universitetsky pr., 26, 198504 St. Petersburg, Russia
e-mail: izenkevich@mail15.com

“obsolete” objects. The number of isomers for higher members of homologous series rises up very quickly [3] and, hence, most of them never been or will be synthesized. At the same time, the properties of the simplest homologues of practically the all series are well-characterized. Mentioned objectives explain us the significance of the following problem: to evaluate physicochemical constants of higher homologues with the use of available data for the first members of series.

Numerous methods of the prediction the different physicochemical constants (A) of organic compounds elaborated until present [4] can be classified onto five principal groups:

- (i) Evaluation of the values of any properties using other constants of the same compounds [4,5];
- (ii) Comparison of the same constants for structural analogues of various series [6–9];
- (iii) Revealing of mathematical dependencies connecting the values of any properties versus the position of compounds within corresponding homologous series, i.e., by functions $A = f(n_C)$. In the simplest cases (e.g., for normal linear homologues) this position can be characterized unambiguously by the number of carbon atoms in the molecule (n_C) [4,10–14];
- (iv) Estimation of empirical dependencies of various properties versus molecular topological parameters (chemometrics approach) [15–19];
- (v) Calculations using additive schemes [4,20–24].

Methods of group (i) are based on the application of various physicochemical relationships connecting different constants of the same compounds one with others [4]. The typical example is the calculations using molar refractions (MR_D), which can be estimated using additive schemes [5]:

$$MR_D = \left(M/d_4^{20} \right) \left[\left(n_D^{20} \right)^2 - 1 \right] / \left[\left(n_D^{20} \right)^2 + 2 \right] \quad (1)$$

where M are molar masses.

The Eq. 1 allows us to evaluate density of any compound with the use of the refractive index value and *vice versa*.

Methods of type (ii) are based on the principle of structural analogy [6], or, by other words, on the correlations connecting the constants of compounds of any taxonomic groups (homologues, congeners) with those of members of other groups at the equivalency of their positions within them. Most often these relationships are used for boiling points, owing to the maximal precision of results comparing with those in other methods. For instance, $T_b(K)$ values for homologues of different series can be correlated using the following three-parameter linear-logarithmic equation [7,8]:

$$\log T_b(2) = a \log T_b(1) + bY + c \quad (2)$$

where Y parameter providing the equivalent positions of structural analogues within corresponding series ($Y = MW, n_C, MR_D$, number of heteroatoms, etc.), a, b, c coefficients calculated by least squares method (LSM).

Similar comparative methods are known for inorganic compounds. For example, the linear dependence of boiling points of tetrahalogenated silanes (SiF_4 , SiCl_4 , SiBr_4 , SiI_4) versus boiling points of analogous tetrahalogenated methanes (CF_4 , CCl_4 , CBr_4 , CI_4) permits us to estimate T_b values for any compound from every of these series using the data for corresponding analogue from another series [9]. However, this example has no practical significance, because reference T_b values for all mentioned compounds are known.

The use of non-linear functions $A = f(n_C)$ (group iii) seems most convenient in practice, but it is restricted by objective difficulties of their selection. Large part of these equations belongs to the type of empirical relationships that means they have no theoretical grounds at all. Namely, the most precise empirical equation for the approximation of boiling points of n -alkanes in the range C_1 – C_{100} is that of Kreglewsky and Zwolinsky [10]:

$$\log(a - T_b) = b - cn_C^k \quad (3)$$

where $a = 1078$, $b = 3.0319$, $c = 4.999 \times 10^{-3}$, and $k = 2/3$.

For approximation of refractive indices and relative densities the empirical approximating function $A = an_C^k + b$ was proposed [11]. This function is applicable only to those parameters of organic compounds, which tend not to infinity at the hypothetically infinite increase the number of carbon atoms in the molecule, but to non-infinite limits (at the condition $k < 1$). The values $\lim A(n_C \rightarrow \infty) = b$ were evaluated for refractive index (1.475 ± 0.003) and relative densities (0.857 ± 0.008). Four- [12, 13] and three-parameter dependencies [14] proposed for calculation of GC RIs with the use of boiling points of organic compounds can be classified as other examples of this approach.

The special approach has been proposed for optimization of the choice of non-linear functions for approximation of various properties of organic compounds versus number of carbon atoms in the molecule. The search of target function for initial data set $A(n_C)$ in many cases can be simplified if their second numerical derivatives A'' are taken into consideration, followed by their integrating twice. The application of the differences of high orders is not recommended owing to their small values. If the general regression function $g''(n_C) = A(n_C - B)^C$ can be fitted onto the data representing the second derivatives of reciprocal boiling points ($1/T_b(K)$), or absolute values of refractive indices or relative densities, then after integrating this function twice, we get the following relationship [15]:

$$g(n_C) = A(n_C - B)^{(C+2)} / [(C + 1)(C + 2)] + Dn_C + E, \quad (4)$$

where A , B , C , D , E constants.

Chemometrics approaches (group iv) imply application of empirical dependencies $A = f(x_1, x_2, \dots, x_n)$, where x_1, x_2, \dots, x_n are different molecular descriptors, like Wiener (W) [16, 17] and Hosoya (Z) [18, 19] indices.

The last group of methods (v) of the evaluation of physicochemical constants is based on long-time known additive schemes. As the illustration, so-called boiling numbers N [20, 21] should be mentioned as one of approaches in evaluation of boiling

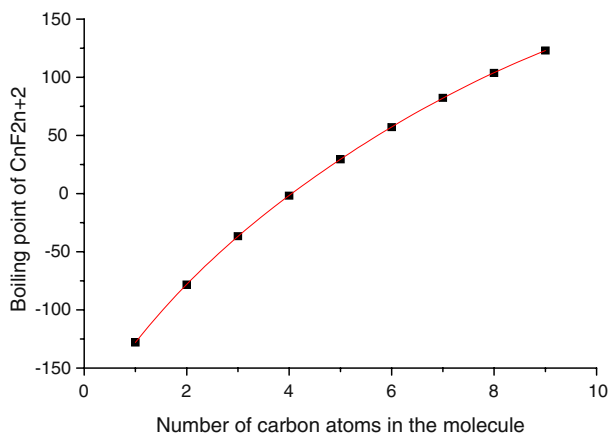


Fig. 1 Typical non-linear dependence of boiling points (T_b , °C) of perfluoro- n -alkanes versus number of carbon atoms in the molecule (no limit at $n_C \rightarrow \infty$)

points. The resulting N values for target compounds can be summarized using the set of atomic and group increments, $N = \sum N_i [N_i = 3.15(\text{CH}_2), 3.65(\text{CH}_3), 0.65(\text{F}), 3.40(\text{Cl}), 5.88(\text{Br}), 8.35(\text{I}), 4.20(\text{Si}), \text{etc.}]$, followed by application of empirical equation $T_b \approx 230 N^{1/3} - 270$. The principal disadvantage of this relationship is the presentation of numerical coefficients only with 1–2 valuable digits (230, 270, 1/3). Hence, in accordance with the rules of approximate calculations, the number of valuable digits in the results cannot exceed the same value 1–2 that means the low precision of results. Usually any additive schemes are mostly convenient for evaluation of constants varying almost linearly within series (enthalpies and entropies of formation or phase transitions [4], partition coefficients [22], etc.), or those presented in specially linearized scales (e.g., retention indices in GC or high performance liquid chromatography [23, 24]).

The comparison of all methods indicates both advantages and disadvantages of them, namely restricted set of variables (i), empirical choice of equations (ii, iv), low precision (v), or non-linearity (iii). Two typical non-linear dependencies $A(n_C)$ are presented in graphical form on Figs. 1 and 2, namely:

- boiling points (T_b , °C) of perfluoro- n -alkanes versus n_C (Fig. 1);
- refractive indices (n_D^{20} , non-dimensional constant) of 1-alkanols versus n_C (Fig. 2);

Similarly, these non-linear dependencies for other properties can be ascending (prominent) or descending (concave). In the first case they can tend to the asymptotes (Fig. 2) or not (Fig. 1).

It is interesting to note that within methods of all mentioned groups (i–v) no attempts to evaluate any constants of homologues using the values of the same properties for previous members of the same series are known up to present. The algorithms of calculations of any constants with the use of their values for simpler homologues should be equivalent to the application of so-called recurrent relationships [or, by other words, equations with shifted argument(s)], namely in the following simplest

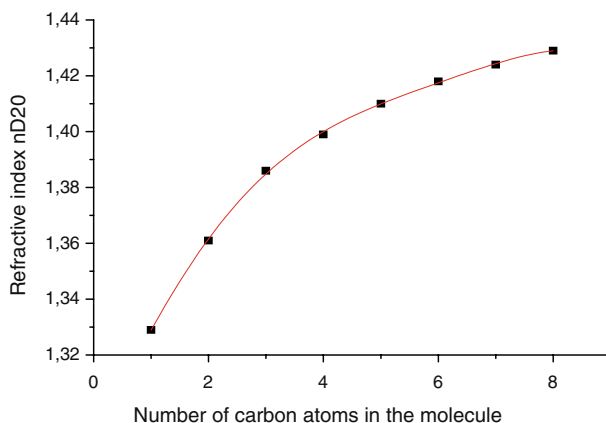


Fig. 2 Typical non-linear dependence of refractive indices (n_D^{20} , non-dimensional constants) of $n - 1$ -alkanol versus number of carbon atoms in the molecule (tends to a limit approx. 1.457 ± 0.003 at $n_C \rightarrow \infty$)

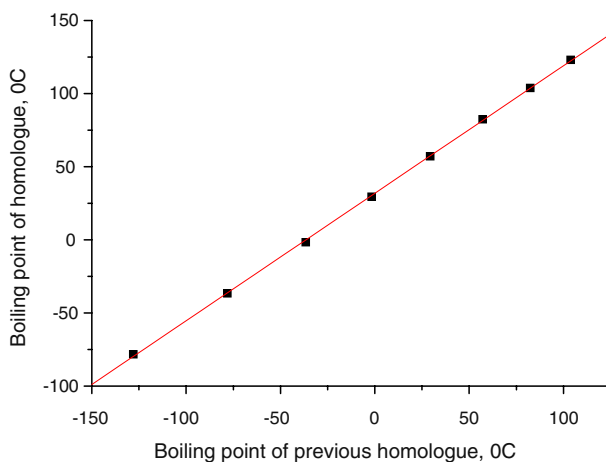


Fig. 3 Linear dependence $T_b(n+1) = a T_b(n) + b$ of boiling points ($^{\circ}\text{C}$) of perfluoro- n -alkanes $C_n F_{2n+2}$: $a = 0.872 \pm 0.005$; $b = 31.8 \pm 0.4$; $\rho = 0.9999$; $S_0 = 1.1$

form of recurrent equation of the first order:

$$A(n+1) = a A(n) + b \quad (5)$$

where the coefficients a and b are calculated by LSM.

Surprisingly, this relationship provides an excellent linear approximation practically for all known physicochemical constants of homologues within various series, but never been proposed or used in organic or physical chemistry up to now. It can be illustrated by examples of boiling points of perfluoro- n -alkanes (Fig. 3, compare with non-linear dependence on Fig. 1) and refractive indices (n_D^{20}) of n -alkanes (Fig. 4, compare with Fig. 2). In both cases correlation coefficients (ρ) exceed 0.999.

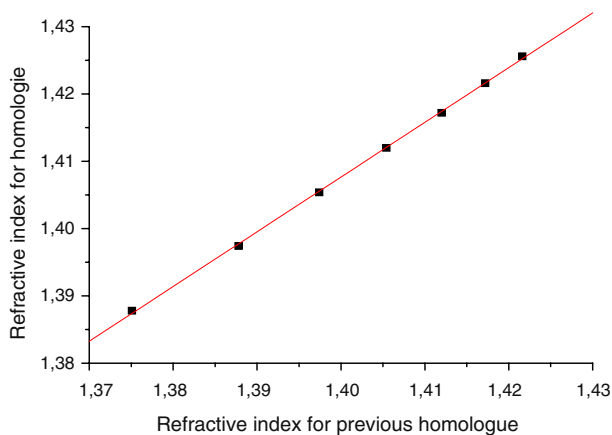


Fig. 4 Recurrent linear dependence $n_D^{20}(n+1) = a n_D^{20}(n) + b$ of refractive indices (non-dimensional constants) of n -alkanes $C_n H_{2n+2}$: $a = 0.804 \pm 0.008$; $b = 0.154 \pm 0.006$; $\rho = 0.9997$; $S_0 = 0.0007$

This first time revealed fact means that two problems can be solved very simply using this approach. One of them is prediction of A -values for previously non-characterized members of series. Another one is the possibility of verification of known data by their comparison with data for other (neighbor) homologues.

This paper is devoted to the discussion of principal mathematical properties of recurrent equations and application of approach based on these relationships for approximation the various constants of organic compounds.

2 Processing and presentation of physicochemical data (experimental)

The values of physicochemical constants of organic compounds were taken preferably from Beilstein reference edition [2], as well as from all others available reference books [25–28]. Usually few alternative values of any constant of the same compound are not equal one each others and can be presented with unequal precision. The unit of the last valuable digits was considered as the measure of data precision (if it is not indicated specially). For instance, boiling point of 1-fluorohexane 93.15°C corresponds to the precision $93.15 \pm 0.01^\circ\text{C}$, whilst boiling point of n -decyl benzene (298°C)—only $298 \pm 1^\circ\text{C}$. If any constants are presented as intervals, the average values within them were chosen, namely like in the case of methyl undecanoate boiling point (247 – 249°C): accepted value is 248°C . However, if these intervals were too broad, additional corrections with the use of the method under discussion were made. For example, it was actual for butyl carbonochloridate (138 – 145°C), or octanal (163 – 174°C). The values 130 and 171°C correspondingly were accepted for these compounds, because they are in the better coincidence with T_b values for neighbor homologues.

Hence, when few data for the same constant were available, most frequent values or those with maximal number of valuable digits have been chosen preferably.

The calculation of the parameters of linear regression was performed by the plotting software ORIGIN 4.1. Solution of recurrent equation (5) was achieved with the use of MAPLE software (version 7).

Retention times of *n*-alkanes C₅–C₁₃ were experimentally measured using Biochrom-1 gas chromatograph (Moscow, Russia) with flame ionization detector and quartz WCOT column (51 m × 0.23 mm) with OV-101 at isothermal conditions (110, 120, and 130 °C).

3 Some mathematical properties of recurrent relationship

$$A(n + 1) = a A(n) + b$$

For chemical purposes (prediction or verification of physicochemical constants) it seems simpler and more convenient to use Eq. 5 in its native form. Nevertheless, for better understanding its mathematical sense various forms of this equation should be characterized.

The Eq. 5 has the following algebraic solution, which can be easily found using MAPLE software:

$$A(n) = k a^n + b(a^n - 1)/(a - 1), \quad (6)$$

where *n* the number of carbon atoms in the molecule of homologue, *a*, *b*, *k* constants; $(a^n - 1)/(a - 1) = a^{n-1} + a^{n-2} + \dots + 1$.

Few noticeable consequences on the application of approach under consideration can be derived from this solution. At first, if $a \equiv 1$, $b \neq 0$, Eq. 6 transforms into $A(n) = k + bn$, that is simple arithmetical progression. At the same time, at $0 < a \neq 1$ and $b \equiv 0$ this equation converts into expression for geometric progression, $A(n) = ka^n$. Hence, initial recurrent function (5) generalizes the properties of both kinds of progressions, that is one of the explanation of its high approximating power.

Secondly, if the Eq. 6 can be presented as a row $A(n) = ka^n + b[a^{n-1} + a^{n-2} + \dots + 1]$, the degree of this polynomial (*n*) is different for every member of series $A(1), \dots, A(n)$. It is the principal difference of data approximation by recurrences comparing with well known approximation by polynomials of constant degrees $A(n) = 1 + a_1n + \dots + a_m n^m$ ($m = \text{const}$).

Thirdly, two limiting values of any properties *A* seem important in chemistry, namely those at the hypothetical zero number of carbon atoms in the molecule (low limit) and at the hypothetically unrestricted increasing of number of carbon atoms (high limit). For the hypothetical “zero” members of any homologous series ($n = 0$) $A(0) = k$.

The presentation of formula (6) in the form of the row explains us the existence of limits of $A(n)$ values at $n \rightarrow \infty$ only for $a < 1$. Hence, depending on the values of coefficients *a*, at the hypothetical increasing the number of carbon atoms in the molecules of homologues ($n_C \rightarrow \infty$), the values of function $A(n_C)$ can tend to infinity (at $a > 1$), or to a limit (only at $a < 1$):

$$\lim A(n_C \rightarrow \infty) = b/(1 - a) \quad (0 < a < 1) \quad (7)$$

This mathematical property of function (2) seems very important, because it allows us to subdivide all physicochemical constants of organic compounds onto two types: some of them theoretically tend to infinity at $n_C \rightarrow \infty$, but another part—to non-infinite limits, namely:

Having no limits at $n_C \rightarrow \infty (a > 1)$	Having limits at $n_C \rightarrow \infty (a < 1)$
Boiling point [11, 15]*	Refractive index [11, 15]
Critical temperature	Density [11, 15]
Viscosity	Surface tension
Partition coefficient	Critical pressure
Chromatographic retention index	Ionization potential
	Dielectric constant
	Dipole moment

Boiling point (*) is the single exception from this general rule. In accordance with physicochemical sense, these constant has no limits at $n_C \rightarrow \infty$ (practically T_b values are restricted by thermal decomposition of compounds), while for most of series of normal linear homologues the values of coefficients a are slightly less than 1, that is equivalent to the existence the “false” limits of boiling points around 550 °C.

As it was mentioned in Introduction, non-linear dependencies $A(n_C)$ for various physicochemical properties of organic compounds can be ascending or descending. The ascending dependencies $A(n)$ are characterized by inequality $A(n+1) - A(n) > 0$. The subtraction relation (6) for $A(n)$ from those for $A(n+1)$ gives the inequality $ka^{n+1} + ba^n > ka^n$. Dividing both parts of this inequality on $a^n > 0$ we receives $ka + b > k$, or $k(a - 1) + b > 0$. Keeping in mind that $k = [A(1) - b]/a$ the last expression can be transformed into the following condition of ascending:

$$A(1)(1 - 1/a) + b > 0 \quad (8)$$

Besides of the application of general relationship (5), another recurrent procedure can be proposed for the calculation of constants of higher homologues using the data of previous members of series. Equation 5 can be rewritten for any two consecutive pairs of homologues, as:

$$A(n) = a A(n - 1) + b; \quad A(n - 1) = a A(n - 2) + b,$$

It gives the following expressions for coefficients a , and b :

$$a = [A(n) - A(n - 1)]/[A(n - 1) - A(n - 2)];$$

$$b = [A(n - 1)^2 - A(n) \times A(n - 2)]/[A(n - 1) - A(n - 2)]$$

Finally we receive another recurrent formula for calculation of any properties of homologues with the use of the data for three previous members of series:

$$A(n + 1) = \{A(n - 1)^2 + A(n) \times [A(n) - A(n - 1) - A(n - 2)]\}/[A(n - 1) - A(n - 2)] \quad (9)$$

However, the relationship based on the data only for three homologues provides less precision of extrapolation comparing with general equation (5). Nevertheless, the expression for coefficient a (the ratio) allows for an important conclusion for practical application of approximations based on the recurrent relationship (5): all data under processing can be expressed in any units and it is not necessary to convert them into standard scales. For example, boiling temperatures can be presented not only in Celsius ($^{\circ}\text{C}$), but in Kelvin (K), as well as Fahrenheit or Reaumur degrees.

Another noticeable feature of recurrences should be pointed out. If recurrent relationship of the first order $A(n+1) = a A(n) + b$ is correct for any set of data, all recurrences of higher orders $A(n+k) = a A(n) + b(k > 1)$ should be correct, as well. If the set of data can be approximated by second order recurrences (see Sect. 5), the relationships of only even orders are correct in these cases.

4 Mathematical applications

Besides simplest arithmetical and geometric progressions, recurrent function (5) is applicable for other dependencies of equidistant values of arguments. Equation 6 at $k = 0$, $b = 0$ and $a = e = 2.71828\dots$ can be used in processing of exponential data, $y = e^x$ or $y = a^x$ in general case ($a \neq e$). Thus, the same recurrent relationship (5) is fulfilled simultaneously both for dependencies $y = x$ and for $y = a^x$. This fact leads to very interesting conclusion: the same equation should be correct for function $y = \log(x)$, which is illustrated by following examples.

It easy to verify that recurrent approximation is applicable for factorials $y = n!$. For more uniform distribution of points, this function is better to consider in the form $y = \log(n!)$, that is recurrent function by definition, $\log(n+1)! = \log(n)! + \log(n+1)$. Few dozens points at the plot on Fig. 5 illustrate this linear dependence ($\rho = 1$). This recurrent mode of factorial's evaluation at high n sometimes seems to be more

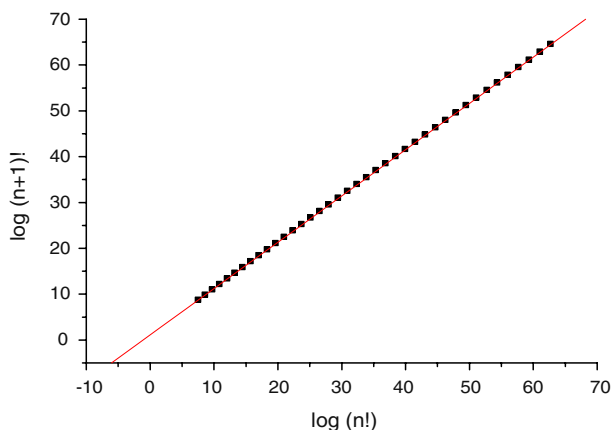


Fig. 5 Recurrent linear dependence $\log(n+1)! = a \log(n)! + b$: $11 \leq n \leq 49$; $\rho = 1.0$; $S_0 = 0.042$; $a = 1.010 \pm 0.000$; $b = 1.113 \pm 0.015$

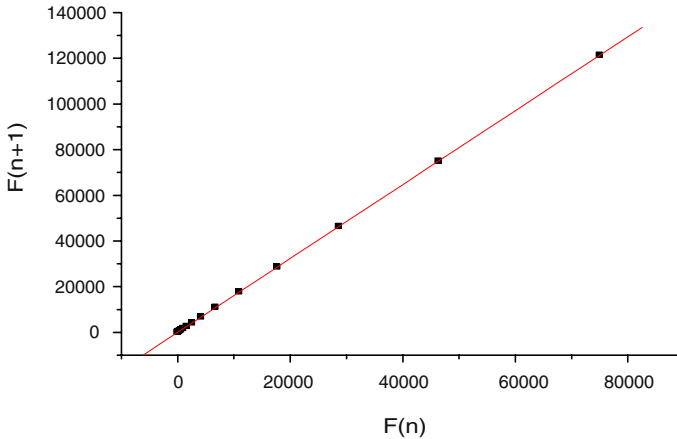


Fig. 6 Recurrent linear dependence of Fibonacci numbers $F(n+1) = aF(n) + b$: 25 points; $\rho = 1.0$; $S_0 = 0.16$; $a = 1.618\dots$; $b = 0.018 \pm 0.036 \rightarrow 0$

convenient then well known Stirling's approximation (see, e.g., <http://mathworld.wolfram.com/stirlingApproximation.html>).

Additionally, recurrent equation (5) is suitable for the linear approximation of Fibonacci numbers (F) [29] belonging to the row 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...; [$F(i+2) = F(i+1) + F(i)$ at $F(0) = 0$, $F(1) = 1$, see, e.g., http://en.wikipedia.org/wiki/Fibonacci_number]. The plot illustrating linear dependence $F(n+1) = aF(n) + b$ is presented on Fig. 6. At the increase of the number of points taken into consideration the coefficient a tends to the value of so-called "golden cut" (1.61803...), and coefficient b tends to zero, $b \rightarrow 0$. Similarly to factorials, the same type of linear recurrent dependence observed not only for Fibonacci numbers themselves, but for their logarithms (excluding first zero point), that is illustrated by plot on Fig. 7

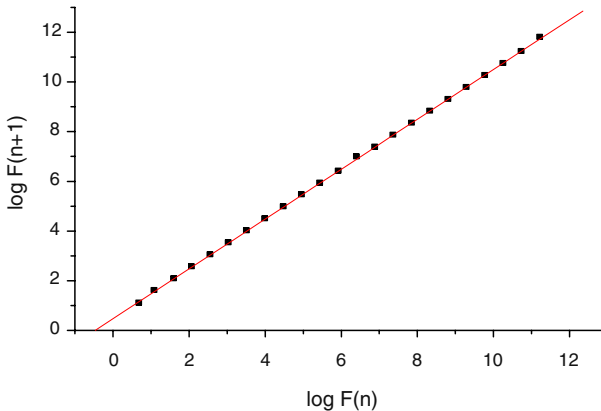


Fig. 7 Recurrent linear dependence $\log F(n+1) = a \log F(n) + b$: $3 \leq n \leq 25$; $\rho = 0.99996$; $S_0 = 0.03$; $a = 1.003 \pm 0.002$; $b = 0.466 \pm 0.013$

($\rho = 0.99996$). Various analogues of Fibonacci numbers (Tribonacci, Tetranacci, etc.) can be approximated by the same recurrences. For example, parameters of Eq. 5 for the series of so-called Lucas numbers [the row 2, 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, 199, ...; $L(i + 2) = L(i + 1) + L(i)$ at $L(0) = 2, L(1) = 1$, see, e.g., http://en.wikipedia.org/wiki/Lucas_number] are the same as those for Fibonacci numbers: $a \rightarrow 1.61803\dots$ (golden cut), $b \rightarrow 0$. The logarithms of Lucas numbers can be approximated by the same way.

The possibility of approximation of Fibonacci numbers it's not only of the theoretical significance. An important kind of molecular topological parameters, namely Hosoya's indices (Z) [18, 19] possess by the properties of Fibonacci numbers. For the organic compounds of simplest molecular topology— n -alkanes— Z -values are equal to $F(n)$ -values. Therefore, discussed approach based on recurrent relationships can be also applied in estimation of topological molecular parameters. Recurrent relationship (5) provides an appropriate linear approximation of Wiener indices [16, 17]. Namely, for n -alkanes C_7 – C_{15} their values are 56(7), 84(8), 120(9), 165(10), 220(11), 286(12), 364(13), 455(14), 560(15), ..., that corresponds to the following parameters of recurrent equation: $\rho = 0.99992$; $S_0 = 2.1$; $a = 1.185 \pm 0.007$; $b = 23.3 \pm 1.8$.

An example of more chemical nature is the number of structural isomers of homologues, which can be approximated by linear recurrent regression of general type (5) as well. For instance, in the alkane series the quantity of isomers depending on the number of carbon atoms in the molecule $N(n_C)$ is 1(1), 1(2), 1(3), 2(4), 3(5), 5(6), 9(7), 18(8), 35(9), 75(10), 159(11), 355(12), 802(13), 1858(14), 4347(15), ... [3]. To further advocate our approach we note the work by Henze and Blair [30] devoted to the elaboration of algorithm to predict the number of isomers $N(n_C)$ of alkane series. The authors have proposed the complex calculation procedure and pointed out that it includes the elements of recurrent calculations. Hence, it is the ground for general recurrent relationship (5) to be correct in this case. The plot illustrating linear dependence $N(n + 1) = a N(n) + b$ is presented on Fig. 8, that can be considered as a reason

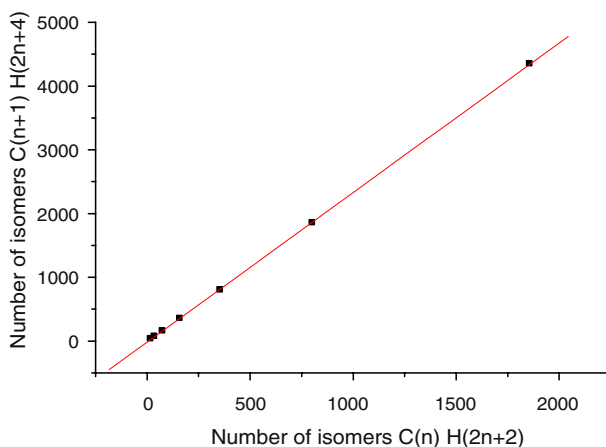


Fig. 8 Recurrent linear dependence of number of isomers of alkanes $N(n + 1) = a N(n) + b$: $8 \pm n \pm 14$; $\rho = 0.99998$; $S_0 = 9.5$; $a = 2.345 \pm 0.006$; $b = -16 \pm 4$

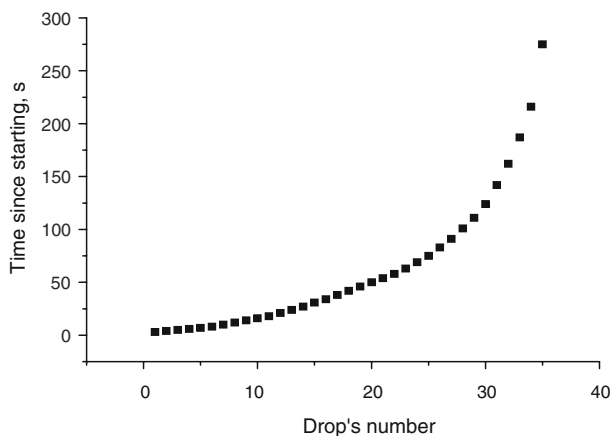


Fig. 9 Typical non-linear dependence of wine dropping time from empty bottle versus number of drop (the sample of red Crimean dessert wine “Massandra”, *experimental data*)

for new way of evaluating of the number of isomers for high members of series. If the linear recurrent dependence (5) is correct for $N(n_C)$ values, it should be correct for $\log N(n_C)$ values.

It's easy to demonstrate that if the set of data can be approximated by polynomials of degrees $m \geq 2$: $A(n) = 1 + a_1n + \dots + a_m n^m$, recurrent relationships like (5) are incorrect in these cases. Hence, the application of recurrent approximation seems to be most important for data sets which cannot be effectively approximated by polynomials, like it is for most of physicochemical constants of organic compounds. In general, there are no contradictions between mentioned approaches and they supply each other.

The predestination of the last illustration (not mathematical) is to demonstrate the possibility of extrapolation of real complex functional dependence, when we know nothing about it. One of the most complex examples is the linearization of non-linear dependence of the time of wine dropping from empty bottles. The plot of this dependence for red Crimean dessert wine “Massandra” (35 drops were observed during the time of about five minutes after the bottle became empty) is presented on Fig. 9. Indeed, it is impossible to propose *ab initio* any reasonable approximation function for this process. However, it is not necessary to choose any special function for these data, because the equation $t_{\text{drop}}(n + 1) = at_{\text{drop}}(n) + b$ seems to be correct even in this case. The linear plot presented on Fig. 10 confirm this conclusion unambiguously ($\rho = 0.9995$). It means that we can predict the time of next drop falling for any liquids with high reliability with the use of the information about previous dropping times.

5 Chemical applications

The success of the application of recurrent equations in the estimation of physicochemical properties is due to the fact that the latter are depended on the number of carbon atoms in the molecule that is the integer argument by definition.

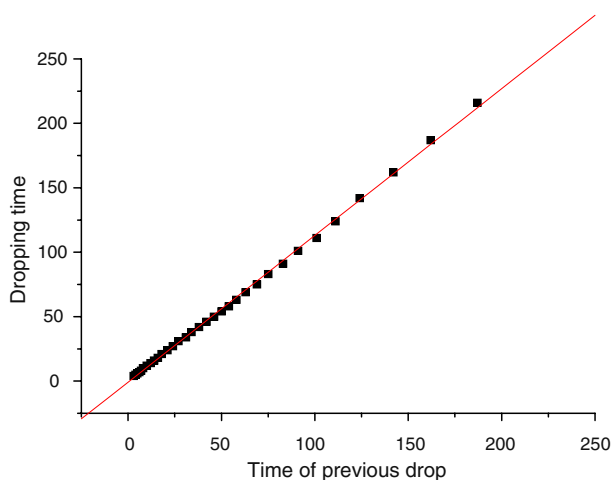


Fig. 10 Recurrent linear dependence $t_{\text{drop}}(n+1) = at_{\text{drop}}(n) + b$ of wine dropping time from empty bottle (compare with Fig. 9): $a = 1.139 \pm 0.007$; $b = -0.8 \pm 0.5$; $\rho = 0.9995$; $S_0 = 1.8$

So far as the recurrent function $A(n+1) = aA(n) + b$ possesses the high “approximation power”, it is not surprising that it provides the linearization of the dependencies of any physicochemical constants versus number of carbon atoms in the molecules of organic compounds within homologous series. Two examples (boiling points of perfluoro- n -alkanes and refractive indices of n -alkanes) have been depicted in Figs. 3 and 4. The similar practically “ideal” linear dependencies are observed for other properties of organic compounds (critical temperatures, critical pressures, densities, viscosities, surface tensions, vapor pressures, dielectric constants, ionization potentials, partition coefficients in heterophaseous systems, solubilities, chromatographic retention indices, etc.) [31]. In accordance with the above mentioned features, recurrent relationship (5) should be correct not only for dependencies of the type of $A(n_C)$, but also for $\log[A(n_C)]$, and can be used in this form for all properties of organic compounds, if necessary.

Some examples of this approach have been collected in Table 1. A large amount of the properties is represented by the data of n -alkanes because this series is characterized by the most reliable constants. More than hundred homologous series have been characterized by recurrent dependencies of boiling points, and no exceptions were revealed. In all cases correlation coefficients exceed 0.999. Moreover, the approach under discussion has been already used for correction of erroneous or misprinted values of constants [31, 32], meaning that it provides an effective way for the calculation of any constants for higher homologues using available data for previous members of series *without special selection of approximation functions*.

An illustrative example of the proposed approach will be given by estimation the boiling points of some perfluorocarboxylic acids $C_nF_{2n+1}CO_2H$. For the first eight members of series they are [33]: 72.4 °C(CF_3CO_2H), 96.4 (C_2F_5), 120–122 (C_3F_7), 144.6 (C_4F_9), 162.7(C_5F_{11}), 178.4 (C_6F_{13}), 190 (C_7F_{15}), and 202.4 (C_8F_{17}). No data for perfluorodecanoic acid $C_9F_{19}CO_2H$ is available at present. Surprisingly, the

Table 1 Results of approximation of various physicochemical constants of organic compounds by recurrent equation $A(n+1) = aA(n) + b$

Property	Homologous series	Range n_C	ρ	S_0	a	b	Existence of limit (Eq. 7)	Increasing (+) or decreasing (-) (Eq. 8)
Boiling points (T_b)	<i>n</i> -Alkyl benzenes	6–17	0.9999	1.0	0.939 ± 0.005	33.4 ± 1.0	+ ^a	+
	Alkanenitriles	2–10	0.9998	0.9	0.950 ± 0.007	29.5 ± 1.4	+	+
	Perfluoro- <i>n</i> -alkanes	1–9	0.9999	1.1	0.872 ± 0.005	31.8 ± 0.4	+	+
Critical temperatures (T_{crit})	<i>n</i> -Alkanes	2–19	0.9997	2.9	0.885 ± 0.005	64.1 ± 1.9	+	+
	<i>n</i> -Alkanes	2–12	0.9998	0.19	0.84 ± 0.06	1.9 ± 0.2	+	-
Critical pressures (P_{crit})	<i>n</i> -Alkanes	6–13	0.9998	0.0003	0.813 ± 0.008	0.270 ± 0.011	+	+
Refractive indices (n_D^{20})	1-Iodoalkanes	2–10	0.9989	0.0003	0.772 ± 0.016	0.338 ± 0.024	+	-
	<i>n</i> -Alkanes	6–13	0.9997	0.0007	0.804 ± 0.008	0.154 ± 0.006	+	+
Relative densities (d_4^{20})	1-Bromoalkanes	2–16	0.9992	0.004	0.811 ± 0.010	0.19 ± 0.01	+	-
	<i>n</i> -Alkanes	5–17	0.9999	0.02	1.204 ± 0.006	0.05 ± 0.01	-	+
Dynamics viscosities (η)	<i>n</i> -Alkanes	6–18	0.9996	0.08	0.848 ± 0.008	4.6 ± 0.2	+	+
Surface tensions (σ)	<i>n</i> -Alkyl benzenes	7–14	0.9998	0.05	0.322 ± 0.003	0.02 ± 0.02	+	-
Vapor pressures (P, 20 °C)	1-Alkanols	1–7	0.995	0.6	0.73 ± 0.04	2.1 ± 0.8	+	-
Dielectric constants (ϵ)	<i>n</i> -Alkanes	1–7	0.996	0.06	0.66 ± 0.03	3.3 ± 0.3	+	-
Ionization potentials (I)	<i>n</i> -Alkyl benzenes	1–7	0.9991	0.05	1.06 ± 0.02	0.34 ± 0.08	-	+
Partition coefficients (log P)								

^a “False” limit for boiling points (see comments in the text)

reference T_b value for perfluoroundecanoic acid $C_{10}F_{21}CO_2H$ seems too high (245 °C) [33] comparing to data of previous members of series. The application of recurrent relationship (5) confirms the existence of linear dependence $T_b(n+1) = aT_b(n) + b$ for C_2 – C_9 homologues ($a = 0.845 \pm 0.011$; $b = 40.8 \pm 1.8$; $\rho = 0.9996$, $S_0 = 0.9$). After that we can precalculate unknown T_b value for $C_9F_{19}CO_2H$ by simple way using reference T_b value for previous homologue

$$T_b(C_9F_{19}CO_2H) \approx 0.845 \times 202.4 + 40.8 \approx 212^\circ C$$

and, moreover, to correct suspicious boiling point for $C_{10}F_{21}CO_2H$ by applying the same recurrent calculations twice:

$$T_b(C_{10}F_{21}CO_2H) \approx 0.845 \times 212 + 40.8 \approx 220^\circ C$$

Hence, the reference boiling point of this compound (245 °C) [33] seems erroneous and should be reconsidered.

Besides that, within T_b interval for perfluorobutanoic acid ($C_3F_7CO_2H$, 120–122 °C) the point providing the best agreement with data for other homologues can be defined more precisely, that is 122 °C.

Only one kind of the important physicochemical constants of organic compounds is not presented in Table 1. It is melting point (T_m) demonstrating strong altering effects for homologues with even and odd carbon number in the molecules. No attempts to approximate these dependencies by polynomials $A = f(n_C)$ are known up to present owing to their objective complexity. Moreover, the homologous variations of these constants cannot be approximated by recurrent equation (5) of the first order.

The plot of initial non-linear dependence $T_m(n_C)$ on the example of melting points of strait chain carboxylic acids $C_{n-1}H_{2n-1}CO_2H$ is presented on Fig. 11. It is easy to observe two “sub-families” of points belonging to two sub-groups of homologues. The upper set of points (highest melting points) belongs to compounds with even number

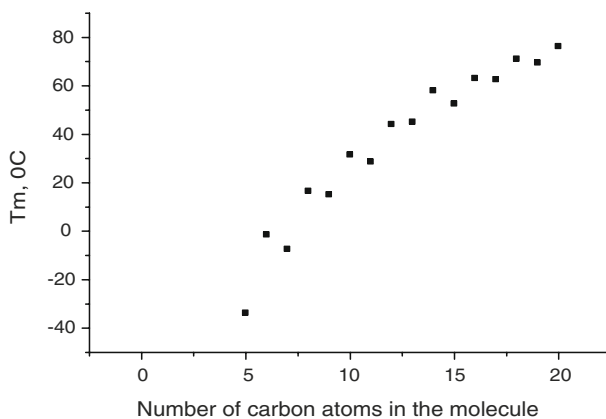


Fig. 11 Typical dependence of melting points of carboxylic acids (T_m , °C) versus number of carbon atoms in the molecule. No linearity and strong altering effects are observed

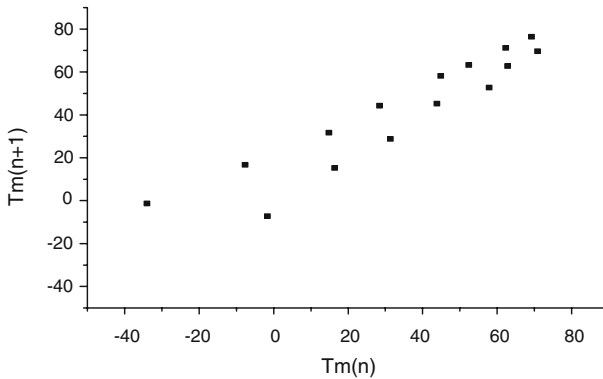


Fig. 12 Graphical illustration the absence of linear dependence of melting points of carboxylic acids (T_m , °C) versus melting points of previous homologues. Two linear sub-dependences for compounds with even and odd carbon atoms in the molecules are observed

of carbon atoms in the molecule, whilst the low set (smallest melting points)—to the compounds with odd number of carbon atoms. The approximation of all these data by recurrent function of the first order (5) gives two unequal sub-dependencies, which can be presented (if necessary) by two separate linear regressions with different coefficients a and b (Fig. 12). The example becomes more interesting when members of the same homologous series exhibit so different properties. However, even this complex dependence can be approximated by single recurrent equation not the first, but the second order:

$$A(n+2) = a A(n) + b \quad (10)$$

The linear plot of this recurrent dependence ($\rho = 0.998$) is presented on Fig. 13. In this form both “sub-families” of data are characterized by the same regression coeffi-

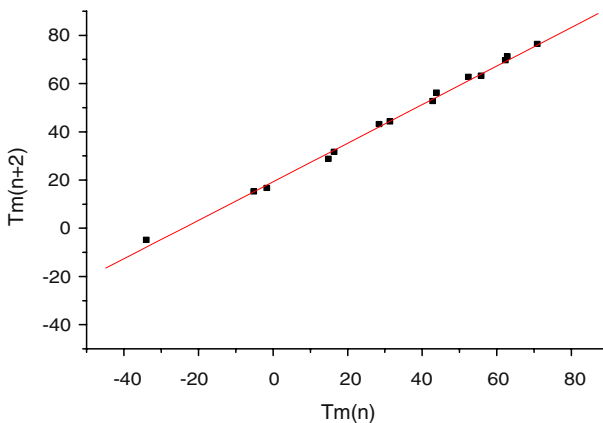


Fig. 13 Single linear recurrent dependence $T_m(n+2) = a T_m(n) + b$ of melting points of carboxylic acid: $a = 0.799 \pm 0.014$; $b = 19.3 \pm 0.6$; $\rho = 0.998$; $S_0 = 1.5$

icients, that permits us to precalculate melting points for higher member of this series and correct previously known values with errors not more than $\pm 1\text{--}2^\circ\text{C}$ (the value of dispersion $S_0 = 1.5$ can be used as the average approximate measure of estimated precision of interpolation or extrapolation).

The algebraic solution of equation (10), which can be easily obtained using MAPLE software, is much more complex than that for standard recurrent equation (5) and it is not discussed here. However, its mathematical analysis gives the same (as that for Eq. 7) limiting values at $n_C \rightarrow \infty$, namely $\lim A(n_C \rightarrow \infty) = b/(1 - a)$.

The existence of general recurrent function (5) explains us a lot of new possibilities of approximation of physicochemical constants of organic compounds. It can be used not only within the series of so-called normal linear homologues (any types of compounds R-X with variable n -alkyl fragments in the molecules), but in the cases of multi-row homology (any types of compounds of general formulas R_2Y , R_3Z , etc., when two, three, or more alkyl chains are varied simultaneously). Moreover, the same approach can be generalized onto homologues with branched carbon skeletons and on the so-called “cyclic” homology (when the variable structural parameter is the fragment of carbo- or heteroatomic rings in the molecules). It remains to be correct not only for compounds with homologous difference CH_2 , but for compounds with other variable structural fragments, namely CF_2 , CCl_2 , $\text{CH}_2\text{CH}_2\text{O}$, SiF_2 , $\text{Si}(\text{CH}_3)_2\text{O}$, etc. However, all these possibilities (important for organic chemistry) will be the subject of future studies.

Numerical chemical applications of recurrent relationships require the explanation of general problem: what is the physicochemical sense of this approach? At this moment there are no theoretical grounds to explain so universal character of recurrences in relation to approximation of various constants of homologues. Of course, the nature of various physicochemical properties is different. However, the regularities of their variations within homologous series seem as much as identical, that they can be approximated by single simplest recurrent equation.

6 Chromatographic applications

One of the clearest examples on the advantages of recurrent algorithms is provided by the typical problem of the evaluation of the data obtained from gas chromatography. The determination of retention indices requires reference compounds (n -alkanes) to be added into analyzed samples. Sometimes (especially for complex mixtures separated using capillary columns) it is difficult to recognize chromatographic signals of some of these reference compounds among peaks of other components. The method of evaluation of their retention times should be used.

Theoretically this method is well known [34]. There is general dependence between retention times of homologues and number of carbon atoms in the molecules:

$$\log(t_R - t_0) = an_C + b \quad (11)$$

where t_0 is the hold-up time of chromatographic system.

Table 2 Results of the calculation of retention times (min) of *n*-alkanes C₁₁–C₁₃ using the data for *n*-alkanes C₅–C₁₀ in isothermal conditions of gas chromatographic analysis by different methods

Compound	t_R (Isotherm 110 °C)	t_R (Isotherm 120 °C)	t_R (Isotherm 130 °C)
Experimental data for <i>n</i> -alkanes C ₅ –C ₁₀			
C ₅	6.33	6.14	6.56
C ₆	6.50	6.26	6.66
C ₇	6.80	6.49	6.85
C ₈	7.37	6.88	7.19
C ₉	8.41	7.60	7.77
C ₁₀	10.31	8.87	8.77
Parameters of linear regressions and evaluation of hold-up times from experimental data			
t_0	6.11 *	6.01 *	6.45 *
ρ	0.9999	0.9997	0.9997
S_0	0.008	0.013	0.016
a	0.257 ± 0.002	1.848 ± 0.005	0.265 ± 0.004
b	-1.950 ± 0.013	-5.20 ± 0.03	-2.27 ± 0.03
	I	II	I
	Exp.	II	Exp.
	I	II	II
Predicted (I, II) and experimental (Exp.) retention times of <i>n</i> -alkanes C ₁₁ –C ₁₃			
C ₁₁	13.64	13.86	11.11
	**	11.48	10.87
C ₁₂	19.72	20.31	15.18
	**	16.15	14.57
C ₁₃	**	32.59	22.05
	**	22.45	**
		18.74	18.52

I, using Eqs. 11 and 12; and II, based on recurrent relationship (5)

* Evaluation of t_0 is not required

** Unacceptable values

Using the data set for known (reliably identified) reference components, coefficients a and b can be calculated by LSM and used for evaluation of retention times of other members of the same series. However, this method requires preliminary evaluation parameter t_0 using the data for three consecutively eluted homologues (t_{R1} , t_{R2} , and t_{R3}):

$$t_0 \approx (t_{R2}^2 - t_{R1} \times t_{R3}) / (t_{R1} + t_{R3} - 2t_{R2}) \quad (12)$$

It is highly surprising that these two theoretically correct equations lead to receiving the results deviated by inevitable experimental errors, that can be illustrated by data presented in Table 2 (method I).

An alternative recurrent procedure of calculations (II) looks much simpler. It implies direct processing of initial t_R data set using relationship (5), i.e., $t_R(n+1) = at_R(n) + b$. No calculations of hold-up time, conversion of raw t_R -values into net retention times ($t_R - t_0$) and calculation of their logarithms are required. As a consequence of this simplification of the calculation procedure the results become much more precise, that can be illustrated by comparison with experimental data (method II, Table 2). Absolute errors of the prediction of retention times of n -undecane by the method (I) are -0.22 , $+0.37$, and $+0.37$ min, and by the method (II)— $+0.05$, $+0.03$, and 0 min, respectively. The application of both methods for next homologue (n -dodecane) gives more dramatic picture: -0.59 , -1.03 , and $+1.06$ min (method I), that is absolutely unacceptable for capillary columns, against only $+0.19$, $+0.02$, and $+0.02$ min in method (II). It is undoubtedly, that only recurrent extrapolation should be recommended for evaluation of retention times of next homologue (n -tridecane, the difference in the number of carbon atoms from last recognized compounds— n -decane—is three), when the errors are $+0.10$, $+0.40$, and $+0.22$ min.

It is important to note, that only recurrent procedure permits us to solve the same problem in temperature programming regimes, when Eqs. 11 and 12 have no sense.

The new modification of the known method of chromatographic quantitation—that of double internal standard—based on the recurrent interpolation of the properties of homologues has been proposed recently [35]. It permits us to neglect significant losses both analytes and standard at all procedures of sample preparation and provide the high precision of results in spite of these losses.

The content of this paper has been presented in the lecture on ICCMSE-2005 [36].

7 Conclusion

The number of applications of recurrent relationships in chemistry seems to be highly large. This approach can be used not only for approximation of physicochemical constants of organic compounds, but for processing other data without information about the origin of functional dependencies. The single demand for its application is equidistant sets of argument values.

The applicability of recurrences for various properties of organic compounds can be interpreted by different manners. At first, this kind of dependencies can be considered as universal mathematical regularity in chemistry. Secondly, it is equivalent

to the existence of new “hyper-property” of organic compounds. Finally, summarizing all aspects of this approach, the following general chemical regularity should be postulated, for example in the following form: *the values of most properties of organic compounds indicate the linear dependences upon the values of the same properties of previous members of homologous series.*

Acknowledgments Author acknowledges Ph.D. Anatoly N. Marinichev (Chemical Research Institute of St. Petersburg State University) who pointed out the existence of algebraic solution of recurrent equation.

References

1. Organikum, *Organisch-Chemisches Grundpraktikum* (VEB Geutscher Verlag, Berlin, 1976)
2. *Beilstein Handbuch der Organischen Chemie*, (4-te Auflage, Berlin, 1918)—present time, Bd. I-XXXI
3. Z. Slanina, *Teoretické Aspekty Fenomeny Chemické Isomerie* (Akademia, Praha, 1981)
4. R.C. Reid, T.K. Sherwood, *The Properties of Gases and Liquids*, 2nd edn. (McGraw-Hill, New York, 1966)
5. B.V. Ioffe, *Refractometric Methods of Chemistry (in Rus.)*, 2nd edn. (Khimia Publ. House, Leningrad, 1974)
6. M.Kh. Karapet'yants, *Methods of Comparative Calculation of Physico-Chemical Properties (in Rus.)* (Nauka Publ. House, Moscow, 1965)
7. I.G. Zenkevich, *Fresenius' J. Anal. Chem.* **365**, 305 (1999)
8. I.G. Zenkevich, *J. Phys. Chem. (Rus.)*. **70**, 33 (1996)
9. A.G. Morachevsky, I.B. Sladkov, *Physico-Chemical Properties of Molecular Inorganic Compounds. Experimental Data and Methods of Calculation (in Rus.)*, 2nd edn. (Khimia Publ. House, St. Petersburg, 1996)
10. A. Kreglewsky, B.J. Zwolinsky, *J. Phys. Chem.* **65**, 1050 (1961)
11. B.V. Ioffe, I.G. Zenkevich, *J. Phys. Chem. (Rus.)*. **74**, 2101 (2000)
12. K. Heberger, *Anal. Chim. Acta.* **223**, 161 (1989)
13. K. Heberger, *Chromatographia.* **29**, 375 (1990)
14. I.G. Zenkevich, L.M. Kuznetsova, *Collect. Czech. Chem. Commun.* **56**, 2042 (1991)
15. I.G. Zenkevich, B. Kranicz, *Chemometr. & Intel. Lab. Systems.* **67**, 51 (2003)
16. H. Wiener, *J. Am. Chem. Soc.* **69**(17), 2636 (1947)
17. H. Wiener, *J. Phys. Chem.* **52**(425), 1082 (1948)
18. H. Hosoya, *Bull. Chem. Soc. Jap.* **44**, 2332 (1971)
19. H. Hosoya, K. Kawasaki, K. Mizutami, *Bull. Chem. Soc. Jap.* **45**, 3415 (1972)
20. C.R. Kinney, *J. Am. Chem. Soc.* **60**, 3032 (1938)
21. I.B. Sladkov, *J. Appl. Chem. (Rus.)*. **72**, 1266 (1999)
22. A. Leo, C. Hansch, D. Eilins, *Chem. Rev.* **71**, 525 (1971)
23. R. Kaliszán, *Quantitative Structure—Chromatographic Retention Relationships* (Wiley, New York, 1987)
24. R. Kaliszán, *Structure and Retention in Chromatography. A Chemometric Approach* (Harwood Acad. Publ, Amsterdam, 1997)
25. B.P. Nikolsky (ed.), *Chemistry Reference Book (in Rus.)*, vol. 1 (Khimia Publ. House, Leningrad, 1971)
26. R. C. Weast (ed.), *Handbook of Chemistry & Physics*, 63rd edn. (CRC Press, Boca Raton, 1982–1983)
27. A. A. Potekhin (ed.), *Properties of Organic Compounds. Reference Book (in Rus.)*, (Khimia Publ. House, Leningrad, 1984)
28. F.I. Vilesov, First adiabatic ionization potentials and innermolecular induction interactions. In *Advances in Photonics (in Rus.)*, vol. 1, (Publ. Leningrad State Univ., 1969)
29. M. Gardner, *Mathematical Games. From Scientific American. (in Russian transl.)* (Mir Publ. House, Moscow, 1974)
30. H.R. Henze, C.M. Blair, *J. Am. Chem. Soc.* **53**, 3042 (1931)
31. I.G. Zenkevich, *Rus. J. Org. Chem.* **42**, 9 (2006)
32. I.G. Zenkevich, *Rus. J. Phys. Chem.* **80**, 1636 (2006)

33. G. Siegemund, A. Feiring, R. Smart, F. Behr, B. McKusick, in *Fluorine compounds. Organic. Ullmann's Encyclopedia of Industrial Chemistry*, 5th edn. (V. A11. Verlagsgesellschaft mbH. Weinheim 1988)
34. E. Leibnitz, H.G. Struppe, *Handbuch der Chromatographie* (Academische Verlag, Leipzig, 1984)
35. I.G. Zenkevich, E.D. Makarov, *J. Chromatogr. A.* **1150**, 117 (2007)
36. I.G. Zenkevich, *Lect. Ser. Comput. Comput. Sci.* **4**, 1574 (2005)