Recurrent Relationships as an Important Class of Mathematical Equations in Chemistry

By Igor G. Zenkevich

St. Petersburg State University

Abstract- The unique potential and different areas of applications of recurrent (synonymous: recursive) relationships in chemistry and chromatography are considered. Recurrent relations can be used in two forms: as functions of integer arguments, $y(x + 1) = ay(x) + b$, and as functions of equidistant argument values, $A(x + \Delta x) = aA(x) + b$, $\Delta x = $ const. The first form applies to all physicochemical properties of homologs in organic chemistry, because the number of carbon (and other) atoms in a molecule can be integer only. The second one applies to chemical variables depending on temperature, pressure, concentrations, etc., when the chemists should provide equal "steps" of their variations.

Recurrent relations combine the properties of arithmetic and geometric progressions, which accounts for their unique approximation abilities. This was illustrated by approximating the number of isomers of alkanes, the boiling points of homologs (nonlinear dependencies), the melting points of homologs (alternation effects), the temperature dependence of the solubility of inorganic salts in water, and by revealing the anomalies of gas chromatographic retention indices and retention times in reversed-phase high performance liquid chromatography.

Keywords: recurrent relations, chemistry, chromatographic retention, chemical variables, number of isomers, physicochemical properties, linear approximation.

GJSFR-B Classification: LCC CODE: QD453.2

RECURRENTRELATIONSHIPSASANIMPORTANTCLASSOFMATHEMATICALEQUATIONSINCHEMISTRY

Strictly as per the compliance and regulations of:

Recurrent Relationships as an Important Class of Mathematical Equations in Chemistry

Igor G. Zenkevich

Abstract- The unique potential and different areas of applications of recurrent (synonymous: recursive) relationships in chemistry and chromatography are considered. Recurrent relations can be used in two forms: as functions of integer arguments, $y(x + 1) = ay(x) + b$, and as functions of equidistant argument values, $A(x + \Delta x) = aA(x) + b$, $\Delta x =$ const. The first form applies to all physicochemical properties of homologs in organic chemistry, because the number of carbon (and other) atoms in a molecule can be integer only. The second one applies to chemical variables depending on temperature, pressure, concentrations, etc., when the chemists should provide equal "steps" of their variations.

Recurrent relations combine the properties of arithmetic and geometric progressions, which accounts for their unique approximation abilities. This was illustrated by approximating the number of isomers of alkanes, the boiling points of homologs (nonlinear dependencies), the melting points of homologs (alternation effects), the temperature dependence of the solubility of inorganic salts in water, and by revealing the anomalies of gas chromatographic retention indices and retention times in reversed-phase high performance liquid chromatography. In the latter case, specific anomalies in the recurrent approximation of retention times allow detecting the reversible hydration of analytes.

The restricted text cannot represent all the examples of the applications of recurrent relations in chemistry and chromatography. However, even this short review allows us to conclude that the potential of recurrences should not be underestimated.

The author confirms that the objective difficulties of the perception of recurrences seem to be the main reason for their rare use in chemistry. In such a situation, any examples of demonstrating their capabilities are significant. Consequently, the author is very grateful to the Editorial Board of GJSFR for the kind offer to submit a manuscript on this topic.

Keywords: recurrent relations, chemistry, chromatographic retention, chemical variables, number of isomers, physicochemical properties, linear approximation.

I. Introduction

Mathematics is an essential tool for data processing in all natural sciences, including chemistry [1–4]. However, in all of them, there are several stereotypes of presenting different mathematical equations. Most of them are the following: in equations of the form $y = f(x, \ldots)$ the argument(s) $(x,$ …) are typically on the right side, while the function is on the left. The number of possible examples is so large that it is difficult to estimate it. For instance, the well known Antoine equation which has many chemical applications relates the absolute temperature (T, argument), and the vapor pressure of pure liquids (P, function): $\log P = a/T + b$ (the coefficients a and b should be precalculated). If necessary, the argument and function can be swapped and this relationship can be used to calculate the temperature from the pressure. Besides that numerous analogues of Antoine equation are used in different areas. For instance, the dependence of retention times *vs.* temperature of gas chromatographic column is described with Antoine-like equation $\log(t_R') = a/T + b$.

Numerous other forms of representing dependences (e.g., parametric) are known for mathematicians, but in chemistry they are perceived as rather unusual.

Recurrent equations are precisely such unusual forms of mathematical relationships. The main feature of them is the absence of arguments in their writing: recurrent equations relate the current value of a function to its previous value(s). The rare use of these relations is confirmed by the fact that many contemporary manuals have no information about them. Recurrent (recursive) progressions are mentioned in [2, P. 174] only as a particular case for comparing with other progressions (arithmetic, geometric, etc.). Such an attitude towards recurrences cannot be accepted, since the properties of recurrences make them indispensable mathematical object in chemistry. The main reason of that is, probably, just unusual mathematical form of recurrences, namely the absence of arguments.

The purpose of this article is to illustrate the application of recurrences with examples from various areas of chemistry and chromatography.

Generally, the most straightforward (first-order) linear recurrences can be represented in two ways. *The first kind of recurrence* combines the values of functions of discrete integer arguments ($n + 1$ and n) in relation (1):

$$y(n + 1) = ay(n) + b \tag{1}$$

The coefficients a and b should be precalculated by the least-squares method using data for a preselected training set.

Author: Address for correspondence: Universitetskii ave., 26, St. Petersburg 198504, Russia; St. Petersburg State University, Institute for Chemistry. e-mail: izenkevich@yandex.ru

At first glance, the scope of application of such relations with integer arguments seems to be somewhat limited. However, it is not; in organic chemistry many properties of organic compounds are considered as the functions of their positions in the corresponding homologous series, or, in other words, of the number of carbon (or other elements) atoms in the molecule. Meanwhile, *the number of atoms of any element in a molecule can take integer values only*. It is the principal reason for the applying of recurrent dependences to various properties of organic compounds. One can only be surprised that this was not been done before 2006, but it may only be explained by the unfamiliarity of these relations.

The second form of recurrent dependences applies to any functions for which it is possible to provide the equidistant argument values, namely, in the form $(x + \Delta x)$, where $\Delta x = $ const (relation 2). In other words, the experimenter must select appropriate argument values from their possible multitude:

$$y(x + \Delta x) = ay(x) + b, \quad \Delta x = \text{const} \qquad (2)$$

Such an expansion of the domain of definition allows applications of recurrences to functions of continuous variables such as temperature (T), pressure (P), concentrations (C), etc., i.e., not only to properties of homologs in organic chemistry, but also to characteristics of chemical systems in physical, analytical, inorganic chemistry, etc.

The general feature of recurrences is the absence of argument values in both parts (left and right) of equalities. Every point on the plots of recurrences is specified by two values of functions: current and previous, or current and subsequent ones. The second feature of equation (2) is the appearance of the additional (artificial) variable Δx: the shape of the graph begins to depend on this value.

As a result, both forms of recurrences are somewhat difficult to visualize, which explains in some way their rare use in scientific practice. To illustrate this, let us consider as an example the recurrent representation of any well-known function $y(x)$. Let us select 36 points within the domain of its definition with the increment $\Delta x = 10$ and build a plot corresponding to the equation $y(x + 10) = ay(x) + b$, shown in Fig. 1.
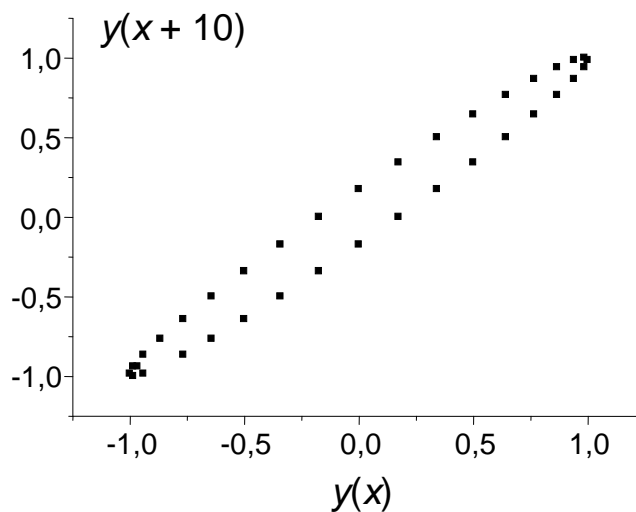

Fig. 1: Recurrent representation of a well-known mathematical function in the form $y(x + 10) = ay(x) + b$. It is not so easy to recognize it

We obtain some kind of closed ellipsoidal curve, which has no analogues among the plots of commonly used functions. If we change the value Δx and select $\Delta x = 1$ (or less) instead of $\Delta x = 10$, the ellipsoidal curve will turn into a practically straight segment located between points $(-1, -1)$ and $(+1, +1)$. So as not to continue the intrigue, let us note that this function is $y = \sin x$, but it is rather difficult to "recognize" it from the recurrent graph without special preparation.

First-order recurrent relations like (1) or (2) look like simple mathematical expressions only at a first glance. These equations have the non-recurrent algebraic solution that is an unusual type of polynomial

of variable degree, because the values of the argument are the power:

$$y(x) = y(0)a^x + b(a^x - 1) / (a - 1) \qquad (3)$$

This solution can be easily found using MAPLE software. The value of the auxiliary parameter $y(0)$ should be predetermined, for example, using the value $y(1)$, because at $x = 1$ $y(1) = y(0)a + b$, and $y(0) = [y(1) - b]/a$.

This solution is the row, because $(a^x - 1) / (a - 1) = a^{x-1} + a^{x-2} + \ldots + 1$. The variable degree of this polynomial is a crucial explanation of the usefulness of

applying recurrences to approximation of chemical variables. This fact deserves special comments.

Example 1: Let us try to approximate the nonlinear dependence of normal boiling points of 1-chloroalkanes $C_nH_{2n+1}Cl$ on the number of carbon atoms in a molecule: 46.2°C ($n = 3$), 78.4°C ($n = 4$), 107.8°C ($n = 5$), 135.6°C ($n = 6$), 160.6°C ($n = 7$), 183.8° C ($n = 8$), 205.4°C ($n = 9$), and so on. The commonly used way to do this is choosing the fixed-degree polynomial (e.g., second or third). At the same time, the polynomial of variable degree (3) will have the degree $x = 3$ for propyl chloride ($n = 3$), $x = 4$ for butyl chloride ($n = 4$), $x = 5$ for pentyl chloride ($n = 5$), etc.

In addition, a noticeable consequence of the application of this approach can be derived from this solution. If $a \equiv 1$ and $b \neq 0$, equation (3) transforms into the relation for a simple arithmetic progression, $y(x) = k + bx$. At the same time, if $0 < a \neq 1$ and $b \equiv 0$, this equation transforms into the expression for geometric progression, $y(x) = ka^x$. Hence, in the general case (at arbitrary values of coefficients a and b), the recurrent equation (1) and/or (2) combines the mathematical properties of both kinds of progressions in variable proportions. It explains us the excellent approximating "power" of recurrences for numerous chemical variables. Thus, we must conclude that recurrences are not the particular case of other progressions [2] but are their generalization.

The second noticeable consequence from equation (3) is the behavior of function $y(x)$ or $A(n_C)$ with a hypothetical increase in the number of carbon atoms in a molecule, $x \rightarrow \infty$, or $n_C \rightarrow \infty$. If the value of coefficient a obeys the inequality $a < 1$, the limit of polynomial (3) exists, and it is equal to:

$$\lim A(n_C)|_{nC \to \infty} = b / (1 - a) \qquad (4)$$

If $a > 1$, the initial numerical sequence has no limit and tends to infinity.

The author cannot cite all his papers devoted to the properties and chemical applications of recurrences due to the necessity to restrict the self-citation. Only six of them published within the period from 2009 to 2024 are mentioned as references [5–10].

To conclude the Introduction, it should be noted that, probably, the most famous and widely mentioned examples of recurrences are so-called Fibonacci numbers, F_n, known since ancient times [11–13]:

$$F_n = F_{n-2} + F_{n-1}, \qquad (F_0 = F_1 = 1) \qquad (5)$$

A special journal named *The Fibonacci Quarterly* has been published since 1963 (see ref. [11]).

II. Experimental

This semi-review article does not imply specific experimental conditions. The necessary data are available from the references cited. In general, recurrent relationships can be applied to any non-linear set of equidistant variables not only in chemistry or chromatography. For example, the temperatures of the cooling kettle, $T(°C) = f(t, min)$ are: 72(0), 62(5), 55(10), 50(15), 46(20), etc. All these data correspond to the linear recurrent dependence $T(t + 5\ min) = aT(t) + b$ with the following parameters: $a = 0.72 \pm 0.01$. $b = 10.0 \pm 0.8$, $R = 0.9996$, $S_0 = 0.2$.

Another important example is the approximation of the carbon dioxide content in the atmosphere of Earth (data from site http://climate.gov),(CO_2 concentration (ppm)) = f(year): 324(1970), 337(1980), 353(1990), 369(2000), 390(2010), and 412(2020). Parameters of the recurrence $C(year + 10) = aC(year) + b$ are: $a = 1.14 \pm 0.01$, $b = -33 \pm 8$, $R = 0.9994$, $S_0 = 1.2$. After that we can easily (because the dependence is linear) estimate the concentration of CO_2 in the atmosphere at 2030: it should be $1.14 \times 412 - 33 = 437$ ppm.

The values of different physicochemical properties of organic compounds (boiling points, melting points, etc.) were taken from all available sources of reference information, e.g., [14], as well as from Internet. The necessary stage of processing the initial information was the rejection of outliers, followed by averaging the data. It is required because recurrences are very "sensitive" even to minor errors in data sets.

The simplicity of recurrent calculations requires special comments. The use of, e.g., Origin software implies the following stages. First, the initial data on the function of integer (equation 1) or equidistant (equation 2) arguments are entered into column A[X]. The next step is copying all data from this column excluding the value from the first line into column B[Y] with shifting one line up. Just such shifting is the sense of recurrences. After that, we can operate with the obtained two-dimensional data array in the usual way, i.e., plot this dependence and calculate all the parameters of the linear regression $y(x+1) = ay(x) + b$. For example, let us make the recurrent approximation of the squares of natural numbers from 3 to 10. Hence, the first column (Origin or Excel software) contains the values 9, 16, 25, 36, 49, 64, 81, and 100 (eight values). The second column (shifted) in the same lines contains the values 16, 25, 36, 49, 64, 81, and 100. In the result we obtain seven pairs of numbers for calculating the parameters of linear regression by least squares method: $a = 1.16 \pm 0.01$, $b = 4.5 \pm 0.5$, $R = 0.9998$, $S_0 = 0.7$. The plot of this dependence looks like "ideal" straight line. Such calculations are very simple, and the application of no special functions is required.

III. RESULTS AND DISCUSSION

a) Chemical Applications

i. Number of Isomers

In this section, it seems advisable to avoid considering the simplest artificial examples and start directly with real chemical problems. The first is the dependence of the number of structural isomers (N) of homologs on the number of carbon atoms in their molecules (n_C). The number of such isomers for n-alkanes C_5–C_{18} is listed in Table 1; it is the rapidly increasing function [15]. It is noteworthy that estimates of the number of isomeric alkanes with n_C = 18 and above, obtained by different methods, appeared to be slightly different, e.g., 60524 vs. 60523 for $C_{18}H_{38}$. However, such small differences for many practical purposes are negligible.

Table 1: Number of structural isomers of alkanes C_nH_{2n+2}

Number of Carbon Atoms, N_c	Number of Structural Isomers, N	Log n
5	3	0.477
6	5	0.699
7	9	0.954
8	18	1.255
9	35	1.544
10	75	1.875
11	159	2.201
12	305	2.550
13	802	2.904
14	1858	3.269
15	4347	3.638
16	10359	4.015
17	24894	4.396
18	60524 (60523)*	4.782

The plot of this function $N = f(n_C)$ for alkanes with $5 \le n_C \le 17$ is presented in Fig. 2(a). It seems rather challenging to precalculate the number of isomeric alkanes with a higher number of carbon atoms because we do not know the type of this function (it should be evaluated or assumed preliminarily). However, if we transform this dependence to the linear form, the problem's solution enormously simplifies, because we can evaluate other values using the simplest linear extrapolation. Just recurrent approximations allow us to implement this approach.
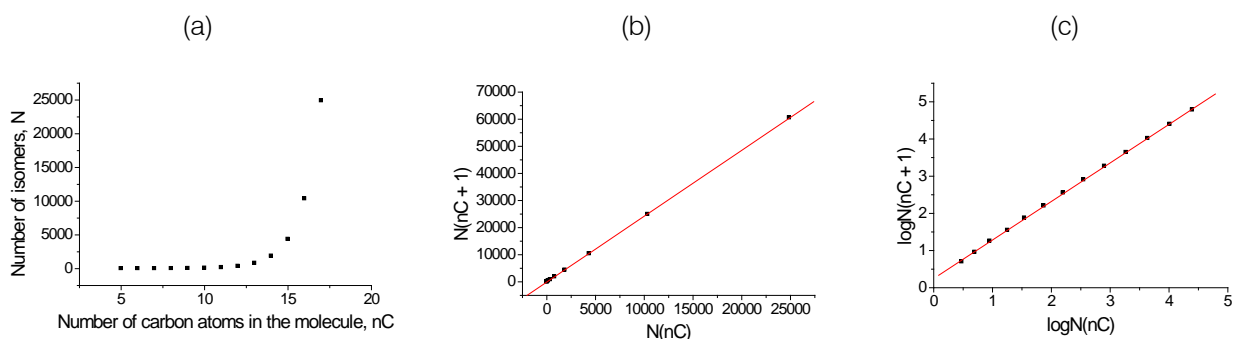
(a)   (b)   (c)

Fig. 2: (a) Plot of the number of structural isomers of alkanes C_nH_{2n+2} (N) vs. the number of carbon atoms in the molecule (n_C); (b) plot of the recurrent dependence $N(n_C + 1) = aN(n_C) + b$; parameters of the linear regression: a = 2.429 ± 0.004, b = –62 ± 30, R = 0.99999, S_0 = 98; (c) plot of the recurrent dependence of logarithms, log$N(n_C + 1)$ = alog$N(n_C)$ + b; parameters of the linear regression: a = 1.036 ± 0.005, b = 0.248 ± 0.012, R = 0.9998, S_0 = 0.02

Approximating the data set in the form $N(n_C + 1)$ vs. $N(n_C)$ gives the plot presented in Fig. 2(b). It is a linear dependence with the correlation coefficient R = 0.99999. Other parameters of the linear regression are indicated in the footnote to Fig. 2. If the correlation coefficients exceed 0.999, the plots of such dependencies are visually perceived as absolutely straight lines. It is an essential moment in the

characterization of recurrent dependencies. For numerous chemical variables, these dependencies are linear, but if not, the reasons of deviations from linearity should be investigated. If we will use this result for evaluating the number of isomers of alkane $C_{18}H_{38}$ using the number of isomers for the previous alkane $C_{17}H_{36}$, we should make the following simple calculations:

$$24894 \times 2.429 - 62 \approx 60406$$

For the example considered the evaluation of uncertainty appeared to be approximately ± 98. The correct answer is 60524 or 60523 isomers (Table 1). Hence, the relative error is approximately 0.19%. Such accuracy seems to be enough for many practical purposes.

This example illustrates another feature of recurrent approximation. If it is linear for variable x, it remains linear for any monotonic functions of this variable, e.g., x^n, $\log(x)$, $\exp(x)$, etc. Hence, we can approximate not the number of isomeric alkanes (N) but the $\log N$ values. Fig. 2(c) illustrates the linear recurrent dependence for logarithms of isomers' number.

The number of isomers of homologs of other series [16, 17] can be approximated using recurrent relations without any restrictions.

ii. *Approximation of Boiling Points of Homologs*

To illustrate the potential of recurrent relations in form (1) (functions of integer argument), let us consider the approximation of values of any physicochemical properties of homologs. The simplest example is normal (at the atmospheric pressure) boiling points (T_b) of n-alkanes C_nH_{2n+2} within the set of compounds C_2–C_{14} (if necessary, this range can be expanded); the reference data are readily available [14]. The plot in Fig. 3(a) illustrates the initial nonlinear dependence $T_b(n_C)$; the linear recurrent approximation $T_b(n_C + 1) = aT_b(n_C) + b$ is illustrated in Fig. 3(b). The parameters of the linear regression are listed in the footnote to this figure.
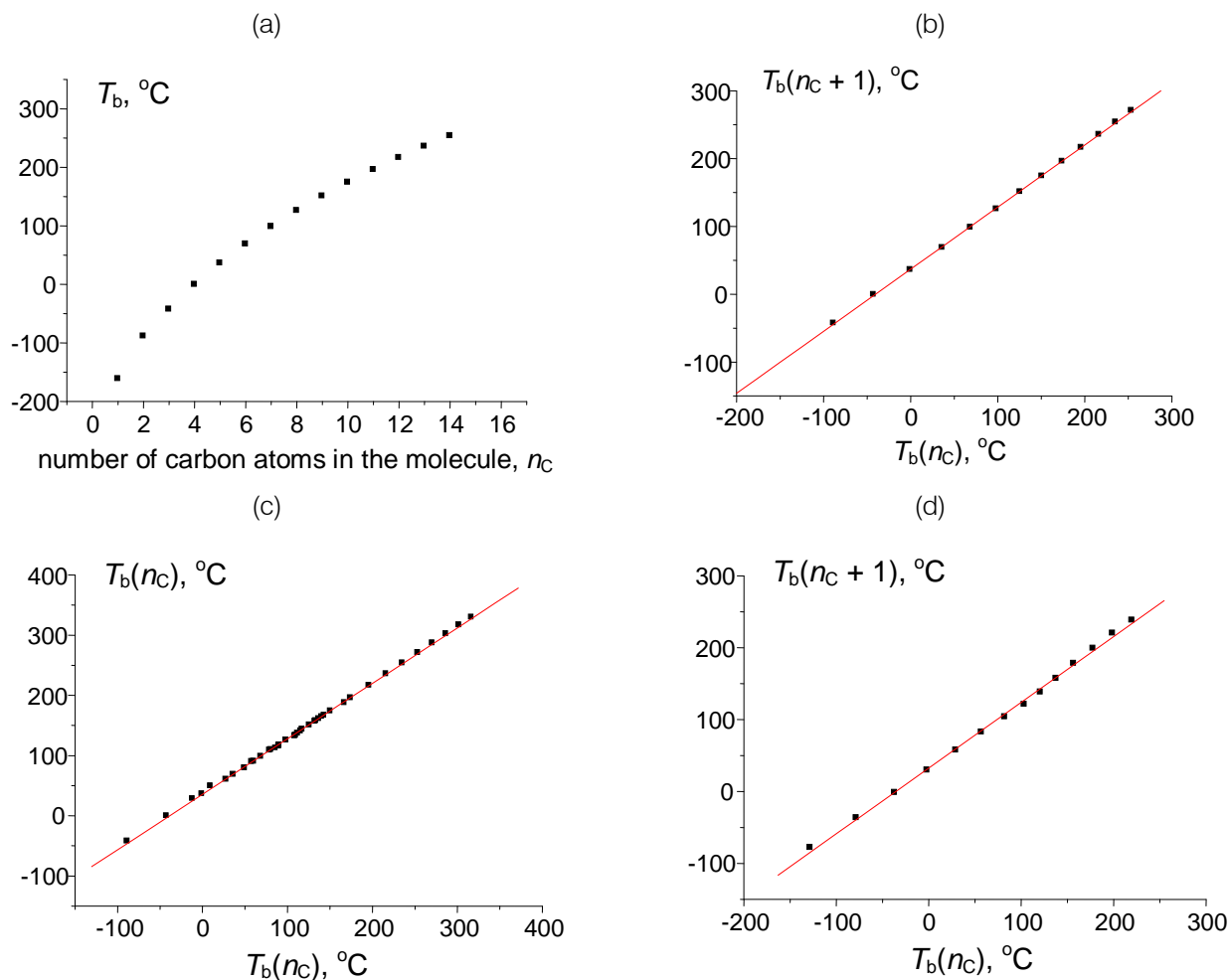

Fig. 3: (a) Dependence of the normal boiling points of n-alkanes C_1–C_{14} (°C) on the number of carbon atoms in the molecule; (b) recurrent approximation of the normal boiling points (equation 1); parameters of linear regression: a = 0.916 ± 0.003, b = 36.9 ± 0.5, R = 0.99993, S_0 = 1.2; (c) recurrent approximation of the normal boiling points for all alkanes (including branched isomers); parameters of the linear regression: a = 0.922 ± 0.003, b = 35.6 ± 0.5, R = 0.9997, S_0 = 2.0; (d) recurrent approximation of the normal boiling points for perfluoro-1-alkanes C_nF_{2n+2}; parameters of the linear regression: a = 0.91 ± 0.01, b = 32.5 ± 1.4, R = 0.9991, S_0 = 4.2

The next exciting feature of recurrences does not follow from their mathematical properties but has a critical chemical sense. Namely, if the recurrent linearization is valid for normal alkanes (with non-branched carbon skeleton), the equation with practically the same coefficients will be valid for all isomeric structures of the same class, i.e., for alkanes with branched carbon skeletons. It is illustrated by Fig. 3(c) (39 points for *iso*-alkanes were added). Comparing the coefficients of recurrences (b) and (c) confirm their identity: 0.916 and 0.922 (values of coefficient *a*), as well as 36.9 and 35.6 (values of coefficients *b*). Moreover, combining the data for 189 homologs of 10 different homologous series confirms the existence of a single linear relationship for all of them with the following parameters: $a = 0.930 \pm 0.002$, $b = 33.5 \pm 0.3$, $R = 0.9995$, $S_0 = 2.4$ [7]. It allows us to conclude that coefficients *a* and *b* or recurrent regressions are close to each other for all homologous series with the same homologous differences, at first CH_2.

The following example illustrates the applicability of recurrences not to alkanes but compounds of other chemical nature. The plot of the recurrent approximation of boiling points of perfluoro-*n*-alkanes C_1–C_{14} is presented in Fig. 3(d).. The correlation coefficient of this dependence exceeds 0.999.

The linearity of recurrent relations makes it possible to evaluate the boiling points of any next homologs using the data for previous ones.

Example 2: Let us evaluate the normal boiling point of 1-chlorodecane using T_b value for 1-chlorononane (205.4°C, see Example no. 1). The values of recurrent coefficients for the series of chloroalkanes are $a = 0.923 \pm 0.002$, $b = 35.7 \pm 0.3$, $R = 0.99999$, and $S_0 = 0.3$. The next step of calculations is simple: $205.4 \times 0.923 + 35.7 = 225.3°C$ (the reference value is 225.8°C).

Example 3: Let us evaluate the normal boiling point of *n*-butylisocyanate using T_b value for *n*-propylisocyanate (88°C). As the data for the series of alkyl isocyanates are insufficient to calculate the coefficients of equation (1), let us use the "universal" values of these coefficients for any homologous series (see above). The result is $88 \times 0.930 + 33.5 = 115.6°C$ (the reference value is 115°C).

Recurrent relations provide linear approximations of not only normal boiling points but also the values of other physicochemical properties of organic compounds within homologous series, $A(n_C)$. These include, for example, relative density (d_4^{20}), refractive index (n_D^{20}), dynamic viscosity (μ), surface tension (σ), ionization potential (I), acidity constant (pK_a), dielectric permittivity (ε), water solubility (w, pS), and many others. The fundamental requirement to the dependences $A(n_C)$ is the same: these functions should be monotonic (the signs of their first derivatives should be constant). The example of nonmonotonic dependencies is, for instance, the temperature dependence of the solubility of organic compounds in water.

Due to the problem of the monotony of changes in the values of physicochemical properties of organic compounds, it is interesting to consider another form of recurrent relations, namely, second-order recurrences:

$$y(n + 2) = ay(n) + b \qquad (5)$$

iii. Melting Points of Homologs

Similarly to the previous ones, the coefficients *a* and *b* of equation (5) should be precalculated by the least-squares method. The algebraic solution of equation (5) can be found in the same way as the solution of equation (1) using MAPLE software. Surprisingly, it turned out to be much more complex than the solution of (1):

$$y(x) = \frac{(y(0)a - y(1))\sqrt{a}(-\sqrt{a})\uparrow x - (y(1)\sqrt{a} - ay(0))(\sqrt{a})\uparrow x}{2a} - \frac{b}{a-1} - \frac{b(\sqrt{a}-a)(-\sqrt{a})\uparrow x + b(\sqrt{a}+a)(\sqrt{a})\uparrow x}{2a(a-1)}$$

(6)

Relation (6) obviously cannot be recommended for any calculations due to its complexity, but it allows us to make important conclusions. First, this solution contains negative terms to the *x* power, namely (-sqr(*a*)) ↑*x*. It means that these terms have different signs depending on whether the *x*-values are even or odd. Therefore, the presence of terms with alternating signs in the equation makes it applicable for approximation of variables with pronounced alternation effects. The most important of these variables are the melting points of homologs with even and odd numbers of carbon atoms in the molecules. The alternation of the melting point is well known for *n*-alkanes, *n*-alkenes, 1-alkanols, carboxylic acids, 1-alkylamines, numerous 1,ω-

difunctionalized alkanes, etc. Fig. 4 illustrates different kinds of the recurrent data approximation for 1-alkylamines. Plot (a) represents the set of initial data, "distorted" with alternation effects. The first-order linear recurrent approximation (equation 1) converts this plot to two nonparallel straight lines (b). Finally, the application of equation (5) to this data set gives a single straight line with the correlation coefficient $R = 0.9996$ (no less than the *R*-values in the other cases mentioned above), but with irregular arrangement of points (c). The parameters of linear regression (c) are listed in the footnote.
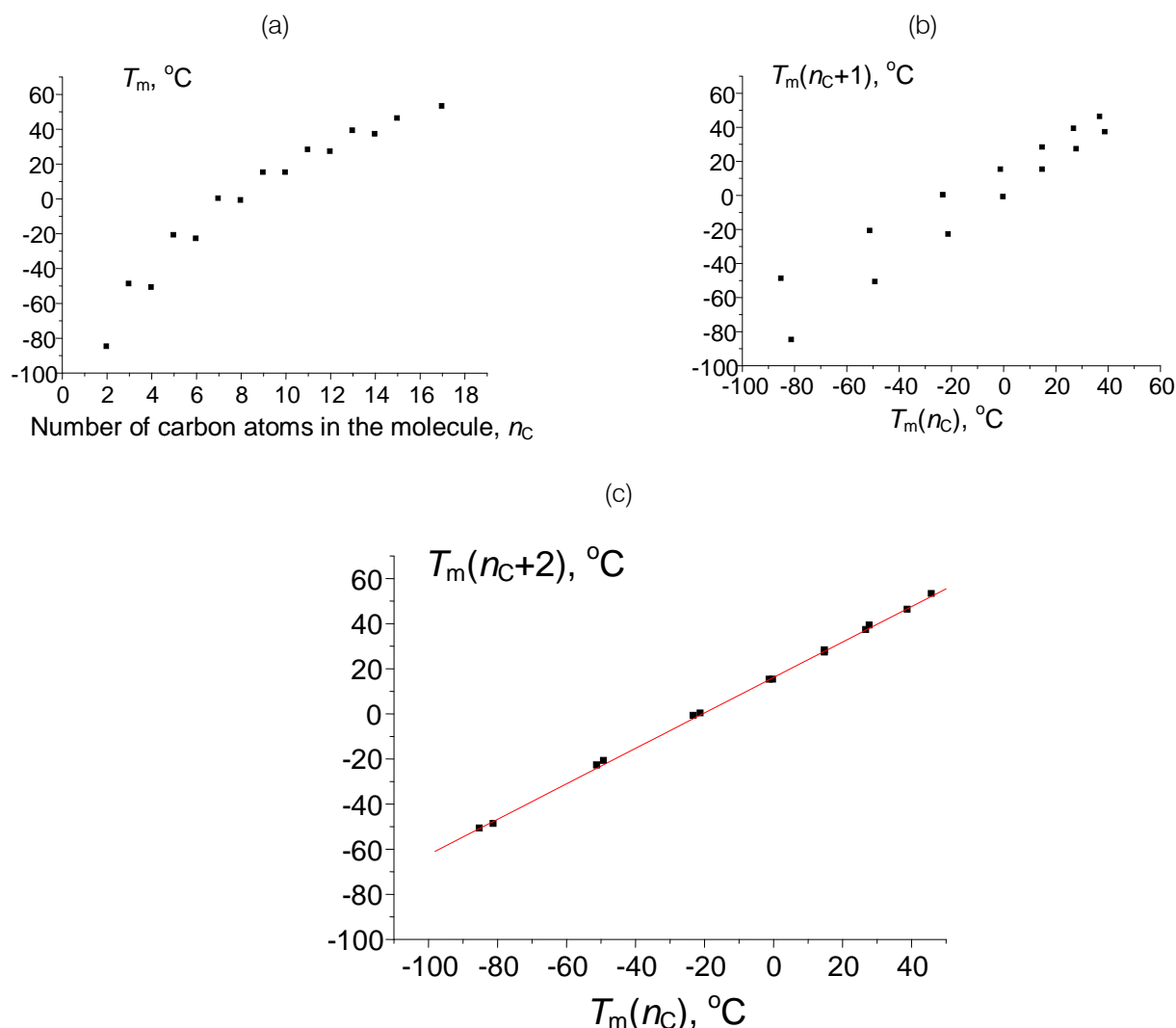
(a)

(b)

(c)

Fig. 4: (a) Dependence of the melting points of n-alkylamines C_2–C_{18} (°C) on the number of carbon atoms in the molecule; the alternation effect is easily noticeable; (b) recurrent approximation of the melting points (equation 1); all points belong to two non-parallel lines; (c) recurrent approximation of the melting points for n-alkyl amines (equation 4); parameters of the linear regression: a = 0.786 ± 0.006, b = 16.1 ± 0.2, R = 0.9996, S_0 = 0.9

Secondly, despite the complexity of equation (6), it is easy to conclude that the limiting value of $y(x)$ at $x \rightarrow \infty$ is the same as that of the polynomial (3), namely, $b/(1 - a)$. Hence, we can evaluate the melting point for hypothetical 1-alkylamine with $n_C = \infty$, i.e.,16.1×(1 − 0.786) ≈ 75°C. The highest-molecular-mass 1-alkylamine for which the melting point has been determined is 1-octadecylamine with T_m 53°C [14].

Another example with alternating melting points of homologous *n*-alkane carboxylic acids is considered in [5].

iv. *Temperature Dependence of Solubility of Inorganic Salts in Water*

The next possible application of recurrences is approximating the temperature dependence of the solubility of inorganic salts in water, $w(T)$, g/100 g of water. It is known that numerous inorganic and some organic compounds form hydrates in aqueous solutions. Using the recurrences allows us to detect important features of the data sets on solubility, namely, the formation of hydrates. If the dissolved compound exists in the single chemical form (either hydrated or non-hydrated) at different temperatures of the solution, the recurrent approximation of $w(T)$ dependence has no anomalies; i.e., it is linear. This case can be illustrated with plots for the solubility of copper sulfate within the temperature range $0 \leq w(T) \leq 100$°C presented in Fig. 5. Despite the nonlinearity of the initial dependence $w(T)$ (a), its recurrent approximation has the correlation coefficient $R = 0.9998$ (b).
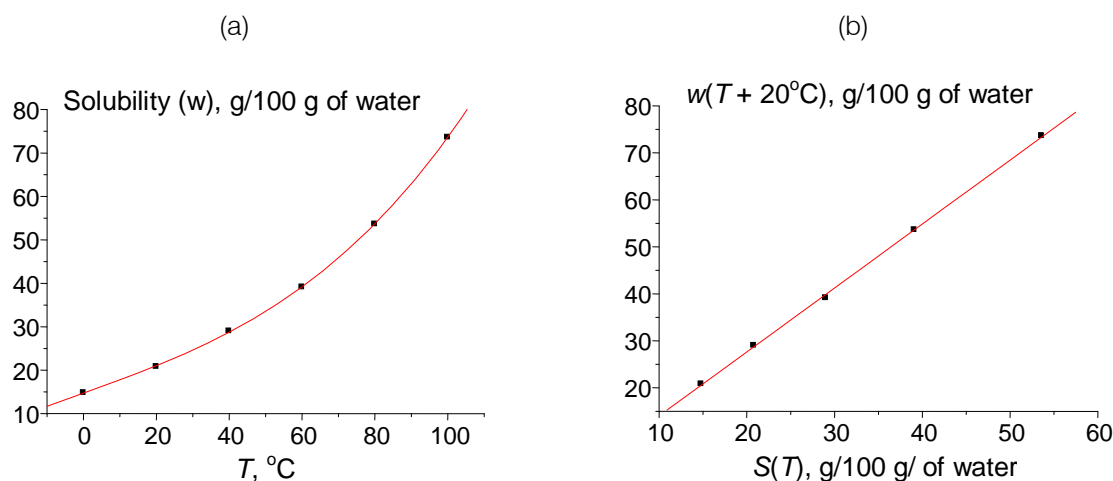
(a)   (b)

Fig. 5: (a) Nonlinear temperature dependence of the solubility of copper sulfate pentahydrate ($CuSO_4·5H_2O$); (b) linear recurrent approximation of this dependence, $w(T + 20°C) = aw(T) + b$; parameters of the linear regression: $a = 1.36 \pm 0.02$, $b = 0.4 \pm 0.6$, $R = 0.9998$, $S_0 = 0.5$
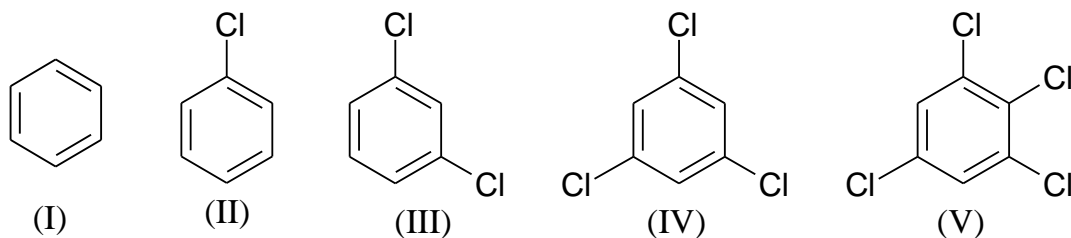
Another example is the solubility of potassium acetate (Fig. 6). Considering the initial $w(T)$ dependence (a) does not allow us to decide whether the leftmost point is weird or belongs to a general nonlinear dependence. The red line in plot 6(a) results from approximation $w(T)$ by a fourth-degree polynomial. However, if we present the same data in the recurrent form, we obtain the linear dependence for all the points ($R = 0.9998$), excluding the leftmost one, which is an obvious outlier. It follows from this fact that, at low temperatures (e.g., 0°C), this salt exists in another form, less soluble than we can expect from the dependence $w(T)$ at other temperatures.

b) Applications of Recurrences in Chromatography

  i. Revealing the anomalies of gas chromatographic retention indices of chlorinated benzenes
      Gas chromatographic retention indices (RI) are important chromatographic characteristics of analytes and are informative for their GC-MS identification. Let us consider the sequence of GC retention indices (RI) for benzene and its chlorinated derivatives with chlorine atoms in unhindered positions (the series of so-called congeners). It includes data for only four compounds, namely, unsubstituted benzene (I), chlorobenzene (II), 1,3-dichlorobenzene (III), and 1,3,5-trichlorobenzene (IV) (all chlorine atoms are in *meta*-positions relative to each other). It should be noted that all RI values were taken from an independent source of reference information, namely, the NIST RI Database [18]:

(I)   (II)   (III)   (IV)   (V)

RI values on standard non-polar polydimethylsiloxane stationary phases [18]:

654 ± 7   839 ± 7   989 ± 10   1117 ± 14   1322 ± 11

Adding the next chlorine atom (1,2,3,5-tetrachlorobenzene, V) will cause it to be in an *ortho*-position to two other chlorine atoms, which gives rise to steric tension in the molecule. However, it is difficult to identify this effect directly from the dependence RI = $f(n_{Cl})$ (Fig. 7(a)). At first glance, only the second point from the right on this plot seems to be located slightly below the regression line, which is not understandable. At the same time, plotting the recurrent dependence $RI(n_{Cl} + 1) = aRI(n_{Cl}) + b$ (Fig. 7(b)) immediately reveals the anomaly just for the rightmost point corresponding to tetrachlorobenzene (V). The steric hindrance in this tetrachlorobenzene increases its retention index above the value extrapolated from data for other (non-

hindered) congeners. Similar effects are also observed for retention indices of bromo- and methyl-substituted benzenes.
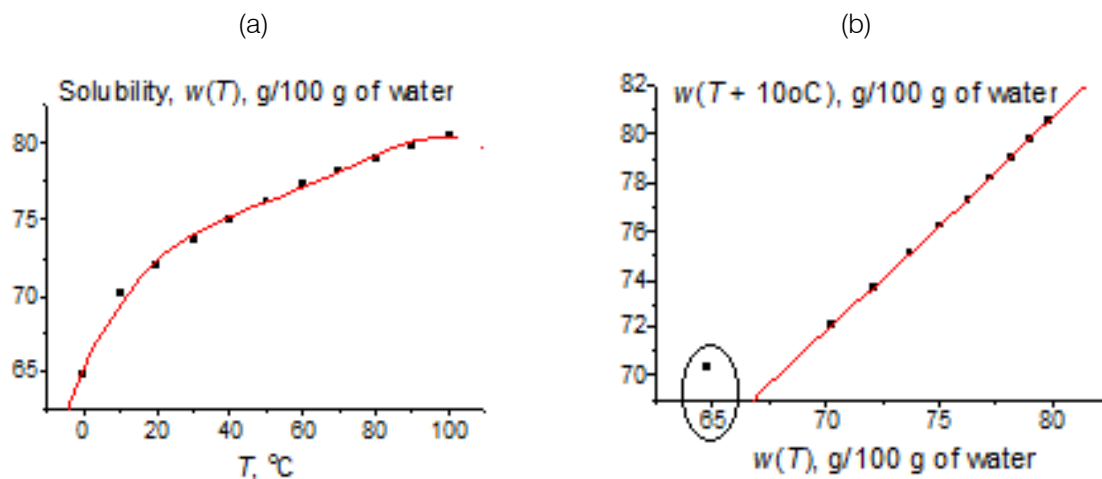
(a)   (b)

Fig. 6: (a) Nonlinear temperature dependence of the solubility of potassium acetate (CH_3CO_2K) in water; (b) recurrent approximation of this dependence, $w(T + 10°C) = aw(T) + b$; parameters of the linear regression without the leftmost point: $a = 0.882 \pm 0.005$, $b = 10.1 \pm 0.4$, $R = 0.9998$, $S_0 = 0.5$.

(a)   (b)
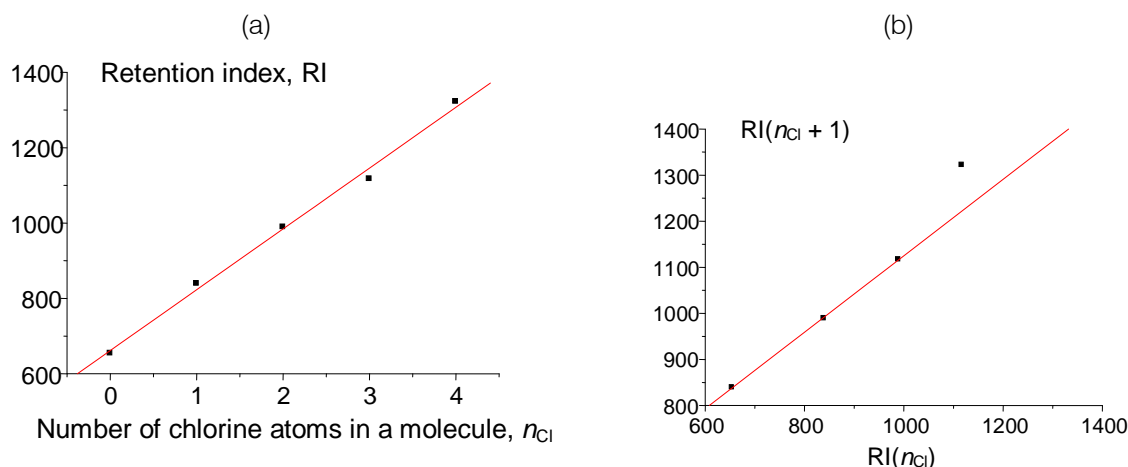

Fig. 7: (a) Dependence of the retention indices of benzene and its four chloroderivatives (see text) on the number of chlorine atoms (standard nonpolar polydimethylsiloxane stationary phase); (b) recurrent approximation of this dependence, $RI(n_{Cl} + 1) = aRI(n_{Cl}) + b$; parameters of the linear regression without the rightmost point: $a = 0.83 \pm 0.01$, $b = 296 \pm 10$, $R = 0.9998$, $S_0 = 2.9$

The last example illustrates the next important feature of recurrences. In many cases, the deviations of recurrent approximations from the linearity seem to be no less informative than the linearity of these dependencies. The explanation of this fact appears to be relatively simple: Recurrent relations (1) or (2) allow linear approximations of numerous monotonic dependencies, but preferably if they can be described with a single equation. If these dependencies are distorted by additional effects, the high "linearization ability" of recurrences appeared to be insufficient for their linearization. This feature allowed us to detect the

steric effects of the fourth chlorine atom in RI values of chlorinated benzenes.

ii. *Revealing the anomalies of retention parameters in reversed-phase high-performance liquid chromatography*

Considering the results of the last example, let us discuss the unusual applications of recurrent relations in reversed-phase high-performance liquid chromatography (RP HPLC). It is the contemporary analytical method based on the different distribution of smallest amounts of analyzed compounds (analytes) between moving liquid phase (eluent) and stationary

phase fixed in a chromatographic column (modified silica gels).

The dependences of retention times of analytes on the content of organic solvent in an eluent, $t_R(C)$, should be considered first. Establishing the regularities of these dependencies seems to be the general approach in characterization both analytes and sorbents in chromatographic columns [19-22].

First, let us illustrate the general statement mentioned in the previous subsection: If the dependence $t_R(C)$ has no anomalies, its recurrent approximation is linear with a high R-value. The absence of anomalies means the single law of variations of retention times *vs.* concentration of organic modifier in an eluent throughout the concentration range under consideration. It can be illustrated with data for acetophenone (methanol-water eluents). Chromatographic analyses were performed using a Shimadzu LC-20 Prominence liquid chromatograph equipped with a diode-array detector and Phenomenex C18 columns 250 mm long and 4.6 mm i. d. with a sorbent particle size of 5 μm. Water–methanol mobile phases were used in several isocratic modes with 5 or 10% concentration steps of methanol at an eluent flow rate of 1.0 mL min⁻¹ and column temperature of 30°C. The samples were injected using an SIL-20A/AC autosampler; the sample volume was 20 μL. To prepare eluents, we used deionized water (resistivity 18.2 MΩ cm) prepared using a Milli-Q device (Millipore, USA), and methanol (analytical grade, Kriokhrom, St. Petersburg). The pH values of the eluent were 6.2–6.3. The number of replicate injections of each sample was 2–3. The interinjection variations of the retention times in all the cases did not exceed 0.01–0.02 min.

The plot in Fig. 8(a) presents the initial nonlinear (close to exponential or hyperbolic) dependence of retention times of acetophenone on the methanol content in the eluent. The recurrent approximation of these data (Fig. 8(b)) is linear with $R = 0.99999$. Such R-value means that 99.999% of this function is a linear component, and only 0.001% is its distortion.
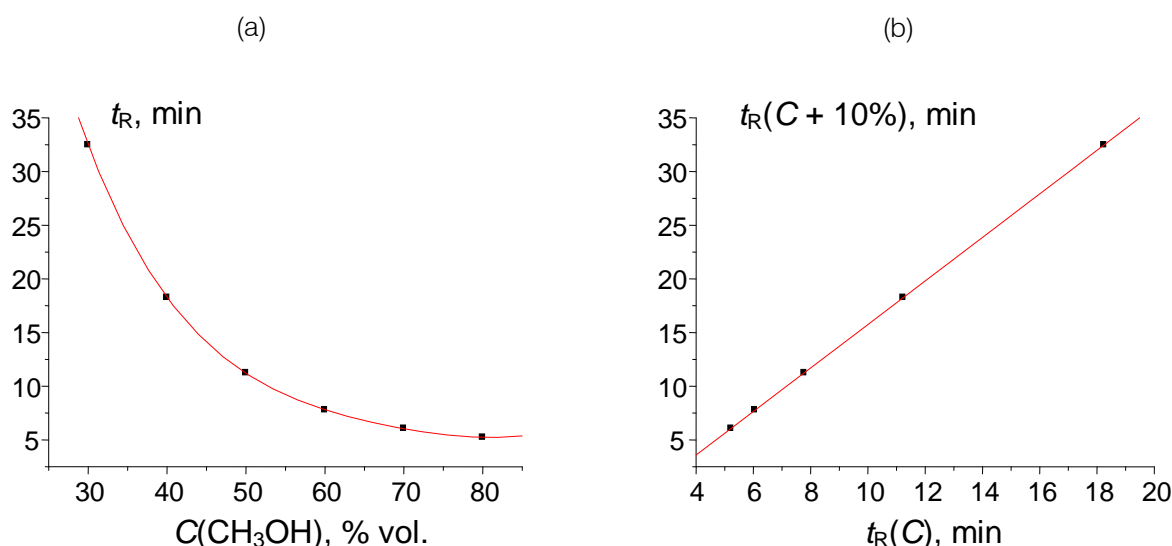
(a)

(b)

Fig. 8: (a) Plot of the acetophenone retention time (min) *vs.* methanol concentration in the eluent, $t_R(C)$: 32.464(30), 18.244(40), 11.230(50), 7.774(60), 6.052(70), and 5.226(80); (b) recurrent approximation of these data: $t_R(C + 10\%) = at_R(C) + b$. Parameters of the linear regression: $a = 0.4932 \pm 0.0005$, $b = 2.231 \pm 0.008$, $R = 0.99999$, $S_0 = 0.01$ (no outliers)

Let us combine two theses. *First*, some organic compounds form hydrates and some of these hydrates are pretty stable. Several examples of such compounds are, for example:

Compound	CAS no. of anhydrous form	CAS no. of hydrate
Glycine	56-40-6	130769-54-9
Oxalic acid	144-62-7	856335-90-5 (mono) 6153-56-6 (di)
Toluene	108-88-3	112270-21-0
Phenol	108-52-2	217182-78-0 144796-97-4
Nicotinamide	98-92-0	917925-73-6
Caffeine	58-08-2	5743-12-4
Methane (!)	74-82-8	14476-19-8

CAS numbers obviously are different both for anhydrous and hydrated forms of the compounds mentioned. Such a number for a hydrate does not confirm its stability or isolation; it is the confirmation of interest in this hydrated form. It is interesting that this list even includes the simplest hydrocarbon, methane. *Second*, the separation of analytes in RP HPLC takes place in water-containing eluents, when the probability of hydrate formation is high.

To identify the area in which anomalies should be expected, let us look again at Fig. 6. So far as the formation of hydrate of potassium acetate is most probable at low temperature of the aqueous solution, one anomalous leftmost point corresponds to the minimal values of solubility, $w(T)$. In HPLC eluents, the hydrate formation is most probable at the maximal water content of an eluent. Hence, the weird point in the $t_R(C)$ plot should be located on the right. The plot in Fig. 9 represents the recurrent approximation of the set of retention times $t_R(C)$ for 2-methoxybenzaldehyde oxime at retention times within the range of approx. 5–40 min. Four points belong to the single straight line with the correlation coefficient $R = 1.000$, but the rightmost point deviates downward from the regression line [10]. The formation of hydrates was confirmed experimentally for some oximes of aromatic carbonyl compounds [23, 24].
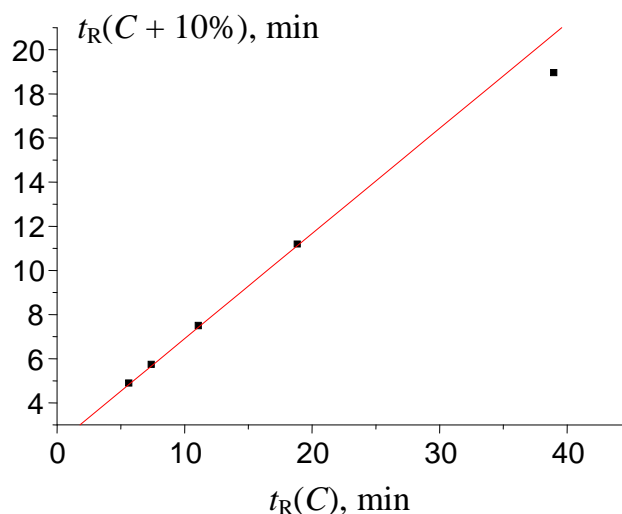

Fig. 9: The plot illustrating the anomaly (the rightmost point below the regression line) in the recurrent approximation of the retention times for 2-methoxybenzaldehyde oxime. Parameters of the linear regression (without outlier): $a = 0.4763 \pm 0.0006$, $b = 2.144 \pm 0.007$, $R = 1.000$, $S_0 = 0.006$.

It is worth noting that deviations of the rightmost points in the plots of recurrent approximations of retention times in RP HPLC were detected for the first time for complex organic compounds, namely, synthetic antitumor sulfonamide drugs containing the polar functional group $–SO_2\text{-}NRR'$ [9]. Most sulfonamides form stable hydrates, which can be isolated in a solid form [25–28].

Thus, the recurrent approximation of retention times in RP HPLC allows us to detect the reversible formation of hydrates of analytes. The possibility is all the more important considering that detecting hydrates using other methods is much more difficult.

IV. CONCLUSION

This manuscript provides a brief exploration of the applications of recurrent relations in chemistry and chromatography, acknowledging that it cannot encompass all possible examples. Nonetheless, the review highlights the significant potential of recurrent relations in simplifying calculations through the linearization of diverse chemical dependencies.

The linearization achieved by recurrent relations not only streamlines mathematical manipulations but also reveals valuable insights from deviations in chemical dependencies. For instance, in reversed-phase high performance liquid chromatography (RP HPLC), non-linear variations in retention times can indicate hydration phenomena during analyte separation.

Formally, two alternative conclusions emerge from our discussions. Firstly, a broad hypothesis posits that many monotonic chemical variables follow first-order recurrent dependencies — a potential general law in chemistry. However, a more modest yet preferable conclusion emphasizes the exceptional extrapolation capabilities of recurrent relations, which justify their widespread adoption in approximating a wide array of chemical variables.

In conclusion, to make a reader smile, one humorous application of recurrences should be

mentioned. There is a very nonlinear dependence of the time of drop falling of liquid (including wine) from an empty bottle, $t(n)$, on the serial number of the drop (n). No idea concerning the analytical form of this function can be proposed (or, at least, it is rather difficult). However, the recurrent representation of this dependence, $t(n+1) = at(n) + b$, demonstrates the excellent linearity [5].

REFERENCES RÉFÉRENCES REFERENCIAS

1. Gutman I., Polanskii O.E. *Mathematical Concepts in Organic Chemistry.* Springer-Verlag, Berlin, 1986. 212 p.
2. Cockett M., Doggett G. *Maths for Chemists.* 2nd Edn. RSC Publ, Cambrige, 2012. 417 p.
3. Cunningham A., Whelan R. *Math for Chemists.* University of Birmingham, University of Leeds, 2014. 119 p.
4. Oriakhi C.O. *Chemistry in Quantitative Language.* 2nd Edn. Oxford Univ. Press, 2021. 528 p.
5. Zenkevich I.G. Approximation of any physicochemical constants of homologues with the use of recurrent functions. *J. Math. Chem.* 2009. Vol. 46. P. 913-933.
6. Zenkevich I.G. Application of recurrent relationships in chromatography. *J. Chemometr.* 2009. V. 23. P. 179-187. doi: 10.1002/cem.1214.
7. Zenkevich I.G. Application of recurrent relations in chemistry. *J. Chemometr.* 2010. V. 24. P. 158-167. doi: 10.1002/cem.1297.
8. Zenkevich I.G. Recurrent Relationships in Separation Science. In "Chemometrics in Chromatography" Eds. L. Komsta, Y.V. Heyden, J. Sherma. New York, Taylor & Francis, 2017, Ch. 24. P. 449-468.
9. Zenkevich I.G., Derouiche A., Nikitina D.A. Detection of organic hydrates in reversed phase high performance liquid chromatography using recurrent approximation of their retention times. *J. Liquid Chromatogr. Related Technol.* 2021. V. 44. № 11-12. P. 588-598. doi: 10.1080/10826076.2021.1998905.
10. Zenkevich I.G., Derouiche A. Features and advantages of the recurrent approximation of retention times in reversed-phase high performance liquid chromatography. *Separations.* 2024. V.11. № 69. doi: 10.3390/separations11030069.
11. Basin S.L. The Fibonacci sequence as it appears in nature. *The Fibonacci Quarterly.* 1963. V. 1. № 1. P. 53-56.
12. Singh P. The so-called Fibonacci numbers in ancient and medieval India. *Historia Mathematica.* 1985. V. 12. № 3. P. 228-244. doi: 10.1016/0315-0860(85)90021-7.
13. Brasch T., Bystrom J., Lystad L.P. Optimal control and the Fibonacci sequence. *J. Optimization Theory Applns.* 2012. V. 154. № 3. P. 857-878. doi: 10.1007/s10957-012-0061.
14. Lide D.R. Ed. *Handbook of Chemistry and Physics.* 84th Edn. CRC Press, Boca Raton, FL. 2003-2004.
15. Henze H.P., Blair C.M. The number of the isomeric hydrocarbons of the methane series. *J. Am. Chem. Soc.* 1931. V. 53. P. 3077-3085.
16. Henze H.P., Blair C.M. The number of structurally isomeric alcohols of the methanol series. *J. Am. Chem. Soc.* 1931. V. 53. P. 3042-3046.
17. Perry D. The number of structural isomers of certain homologs of methane and methanol. *J. Am. Chem. Soc.* V. 54. P. 2918-2920.
18. The NIST Mass Spectral Library (NIST/EPA/NIH EI MS Library, 2017 Release). Software/Data Version; NIST Standard Reference Database, Number 69, August 2017. National Institute of Standards and Technology, Gaithersburg, MD 20899: http://web book.nist.gov (Accessed: June 2024).
19. Soczewinski E. Quantitative retention-eluent composition relationships in liquid chromatography. *J. Liquid Chromatogr.* 1980. V. 12. P. 1781-1806.
20. Snyder L.R., Kirkland J.J., Dolan J.W. Introduction to modern liquid chromatography. 3rd Edn. New York: J. Wiley & Sons, 2009.
21. Saifutdinov B.R., Davankov V.A., Il'in M.M. Thermodynamics of the sorption of 1,3,4-oxadiazole and 1,2,4,5-tetrazine derivatives from solutions on hypercrosslinked polystyrene. *J. Phys. Chem. A.* 2014. V. 88. № 3. P. 358-364. doi: 10.1134/S0036024414030224.
22. Saifutdinov B.R., Buryak A.K. The influence of water-acetonitrile solvent composition on thermodynamic characterization of adsorption of aromatic heterocycles on octadecyl-bonded silica gels. *Colloid J.* 2019. V. 87. № 6. P. 754-761. doi: 10.1134/S1061933X1 9060176.
23. Jerslev B., Larsen S. Hydrogen bonding and stereochemistry of ring-hydroxylated aromatic aldehyde oximes, Crystal structure of three 4-OH-substituted benzaldehyde oximes. *Acta Chem. Scand.* 1991. V. 45, P. 285-291.
24. Jerslev B., Larsen S. Crystallization and crystal structure at 122 K of anhydrous (E)-4-hydroxybenzaldehyde oxime. *Acta Chem. Scand.* 1992. V. 46. P. 1195-1199.
25. Guo M.-L. 4-Acetamido-N-(4-methylpiperazinyl) benzenesulfonamide monohydrate. *Crystallogr. Commun.* 2004. V. 60. № 4. P. 574-575. doi: 10.1107/S1600536804005446.
26. Suchetan S.F., Foro S., Gowda B.T., Prakash M.S. 4-Chloro-N-(3-methylbenzoyl)benzenesulfonamide monohydrate. *Acta Crystallogr. Sect. E Struct. Rep. Online.* 2012. V. E68. № 1. P. o46. doi: 10.1107/S1600536811051932.
27. Kompella A., Kasa S., Balina V.S., Kusumba S., Adibhatla B.R.K., Muddasani P.R. Large scale

process to prepare highly pure Bosentan monohydrate. *Science J. Chem.* 2014. V. 2. № 6-1. P. 9-15. doi: 10.11648/j.sjc.s.2014020601.12.

28. Jatezak M., Sidoryk K., Kossykowska M., Luniewski W., Zagrodzka J., Lipiec-Abramska E. Development and validation of a UHPLC UV method for the in-process control of bosentan monohydrate synthesis. *Chromatographia.* 2016. V. 79. P. 1131-1141. doi: 10.1007/s10337-016-3124y.