ACTA CHEMICA SCANDINAVICA 13 (1959) 186—202

Short Communications

Least Square Method for Computor Calculation of Stability Constants *

JAN RYDBERG** and JAMES C. SULLIVAN

Chemistry Division, Argonne National Laboratory Lemont, Illinois, USA

Although a number of investigators have been concerned with the problem of calculation of stability constants, the methods presented have in general been approximation techniques (*cf. e. g.* Ref.[1]). With the growing use of high-speed computors, it becomes feasible to re-examine some of the data in the literature using a least square technique to calculate such stability constants. In addition, the use of such a computor enables the experimentalist to assign standard deviations to constants so calculated. This feature of computor calculation is a valuable contribution, since it provides an objective measure of the precision associated with such constants.

The necessity for such an approach is rather pointedly illustrated when one considers the compilation prepared by Bjerrum, Schwarzenbach and Sillén [2], where, for example, in the system Fe(III)-SCN^- there are 35 different investigations tabularized. In the absence of any measure of precision (not considering systematic errors) of the calculated stability constants, the user of this compilation is forced to *ad hoc* considerations.

The authors have therefore started a series of studies in order to reveal the advantage and precision of using least square methods in combination with high-speed digital computors for the calculation of stability constants. The results of our studies will be given in a series of papers dealing with the application of the method to systems containing simple complexes (one central atom and one type of ligand, *e. g.* MA_n), composite complexes (one central atom and several different types of ligands, *e. g.* MA_nB_p) and polynuclear complexes (several central atoms of same kind for each kind of ligand, *e. g.* M_mA_n) as studied by means of solvent extraction, ion exchange, spectrophotometric and e.m.f. methods.

In the present paper, as an example, the treatment of a simple system will be briefly described.

The U(IV)-acetylacetone system

This system, which has been studied by means of solvent extraction [3], was chosen partly because of its familiarity to one of the authors, and partly because of its regular behavior in the present treatment.

Outline of the mathematical treatment. U(IV), here abbreviated to M^{+4}, is assumed to form a series of complexes, MA_n^{4-n}, with acetylacetone, HA, in aqueous solutions. The uncharged complex MA_4 can be extracted into inert organic solvents like C_6H_6, and it is assumed that this complex is the only extractable one of M. The distribution ratio, q, of all species of M between the organic and aqueous solvents will then be

$$q = \frac{[M]_{org}}{[M]_{aq}} = \frac{\lambda_N \cdot \varkappa_N [A^-]^N}{\sum_0^n \varkappa_n [A^-]^n} \qquad (1)$$

where λ_N is the distribution constant of the uncharged complex MA_N between the organic and aqueous phases, and $\varkappa_n$ is the over-all stability constant for the n:th complex:

* Based on work performed under the auspices of the U.S. Atomic Energy Commission.

** On leave of absence from the University of Stockholm, and Research Institute of National Defence, Stockholm, Sweden.

$$\varkappa_n = \frac{[MA_n^{N-n}]}{[M^N]\,[A^-]^n} \qquad (2)$$

If we introduce the following notations

$$y = \frac{[M]_{aq}}{[M]_{org}}\ ,\ x = [A^-],\ a_n = \frac{\varkappa_n}{\lambda_N \varkappa_N}\ ,\ \text{and}\ z = \frac{[M]_{aq}}{[M]_{org}}\,[A]^N = y\,x^N$$

into eqn. (1), it will yield

$$z = \sum_0^n a_n x^n \qquad (3)$$

This can be solved for a_n, provided n combinations of $z(x)$ are available. Usually the number L of experimental points is $\gg n$, and the L equations of type 3 can therefore be solved with a least square technique. However, x and z are experimentally determined quantities and therefore have associated with them errors, which makes it necessary to properly weight the equations. The details of the procedures will be described in a later paper.

Computer operations. In the U(IV) case it was assumed that the errors in $[A^-]$ could be neglected compared to the errors in $[M]_{org}$ and $[M]_{aq}$, which will be called σ_{org} and σ_{aq}. The computer was therefore programmed for these five kinds of data: $[A^-]$, $[M]_{org} \pm \sigma_{org}$ and $[M]_{aq} \pm \sigma_{aq}$, which were punched on cards together with information of N, L and maximum number of n. In the programming, the computor was ordered, 1) to print the best combination of positive a_n for a minimum number of n, 2) to print the standard deviations of a_n, 3) to print the z_i values, calculated according to eqn. (1) with the a_n given by the computor, for the various input x_i values, 4) to print the standard deviations of these z_i, 5) to print the over-all accuracy of the computation. This latter information is given in the form of S_{min}/k*, which should have a value between 1.5 and 0.5 for reliable data.

For $n = N = 4$, and $L = 35$, the time of the whole operation (*i. e.* putting in program and data cards, running the computor and printing out) is about 5 min for the computer used (an I.B.M. 704).

Results of the computations. In the U(IV)-acetylacetone system [3], it was felt that there was no obvious way to unambigously establish the analytical errors in the determinations of $[M]_{org}$ and $[M]_{aq}$. Therefore, the error in log q was estimated from the spread of the individual points; see Fig. 1. For this purpose, points with very close lying log $[A^-]$ values were chosen. From 6 such groups, comprising altogether 14 points, at various log $[A^-]$, an average error of ± 0.04 in log q was estimated, which corresponds to a ± 10 % error in q on a linear scale. For the sake of simplicity, and also in order to make the data more suitable to the input form of the computor, the whole error was attributed to $[M]_{aq}$. Thus the data put into the machine were $[M]_{org} = q \pm 0.00$, $[M]_{aq} = 1.00 \pm 0.10$, and $[A^-]$.

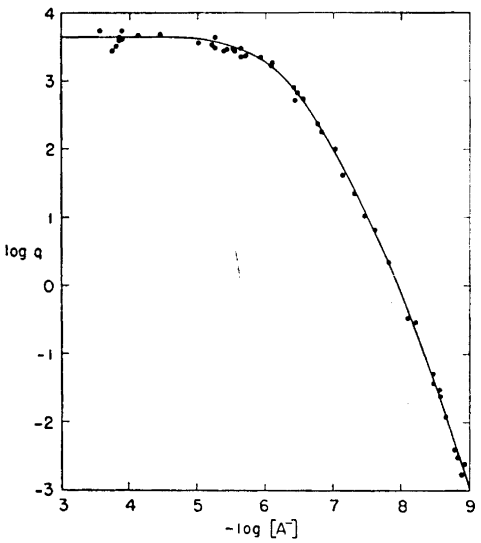

Fig. 1.

* The significance of the function

$$S_{min}/k = \frac{\sum \omega_i \sum (a_n x_i^n - z_i)^2}{L - N - 2},$$

where ω_i is the weight of the point i, will be fully discussed in a subsequent paper.

Table 1. Equilibrium constants calculated by the least square method.

Reaction	Constants
$U^{4+} + A^- \rightleftharpoons UA^{3+}$	$(1.04 \pm 0.69)\ 10^9$
$UA^{3+} + A^- \rightleftharpoons UA_2^{2+}$	$(1.80 \pm 0.64)\ 10^8$
$UA_2^{2+} + A^- \rightleftharpoons UA_3^{+}$	$(3.31 \pm 0.72)\ 10^6$
$UA_3^{+} + A^- \rightleftharpoons UA_4$	$(0.95 \pm 0.19)\ 10^6$
UA_4 (aq.) $\rightleftharpoons UA_4$ (C_6H_6)	$(4.38 \pm 0.25)\ 10^3$

In addition to the 10 % error in $[M]_{aq}$, a number of other errors were also tested. The S_{min}/k values of these runs were 15.5 (for a 5 % error), 3.9 (10 % error), 1.0 (20 % error) and 0.4 (30 % error), indicating that the 10 % error first assigned to the experimental points was too small.

In this particular system, it turned out that the best fit of the experimental data to the curve given by eqn. (1) is obtained when all constants $\varkappa_1$,.....$\varkappa_4$ and λ_4 are present. The stepwise formation constants k_n as well as the distribution constant λ_4 are given in Table 1. It is seen that the standard deviation is very large for k_1 (about 70 %), but diminishes successively up to k_4 (about 20 %). The precision in λ_4 is unusually good. With these values (or rather $\varkappa_n = \prod_1^n k_n$), the curve in Fig. 1 is calculated.

In Table 2, the log $\varkappa_n$ and log λ_4 values are given in order to compare them with the constants manually calculated in the original paper [3]. There, two different sets of constants were given for the two-parameter method; here, the average values are given, and the errors are the spread between these two values. When comparing the constants, the agreement between the constants calculated with the least square method and those of the two-parameter method is rather good, while it is less good for the ligand number method. Since both these latter methods are graphical in nature, the better agreement of the two-parameter method probably stems from the fact that the experimental data in that method are used directly for the estimation of the stability constants, and that the sequence-values of the stability constants are of the suitable two-parameter type. In the ligand number method, the graphical derivation of the smoothed curve through the experimental points makes possible the introduction of large errors.

From Table 2 it is further seen that the precision assigned to the constants obtained by the two graphical methods is far too small. While all errors in the primary data are considered in the least square computations, only errors in a very narrow range of data can be regarded for estimating the errors of the constants obtained by the graphical methods. This illuminates one of the statements made in the introduction, namely the inherent difficulty of the approximation methods in making valid estimates of the precision of the stability constants.

Table 2. Comparison of equilibrium constants (log $\varkappa_n$) obtained with three different calculation procedures.

Reaction	Least square method	Ligand number method	Two-parameter method
$U^{4+} + A^- \rightleftharpoons UA^{3+}$	9.02 ± 0.29	8.6 ± 0.2	9.05 ± 0.1
$U^{4+} + 2A^- \rightleftharpoons UA_2^{2+}$	17.27 ± 0.26	17.0 ± 0.2	17.02 ± 0.15
$U^{4+} + 3A^- \rightleftharpoons UA_3^{+}$	23.79 ± 0.26	23.4 ± 0.1	23.92 ± 0.1
$U^{4+} + 4A^- \rightleftharpoons UA_4$	29.77 ± 0.29	29.5 ± 0.1	29.76 ± 0.1
UA_4 (aq.) $\rightleftharpoons UA_4$ (C_6H_6)	3.64 ± 0.025	3.62 ± 0.05	3.62 ± 0.05

1. Sullivan, J. C. and Hindman, J. C. *J. Am. Chem. Soc.* **74** (1952) 6091.
2. Bjerrum, J., Schwarzenbach, G. and Sillén, L. G. *Stability Constants. Part I: Organic Ligands. Part II: Inorganic Ligands.* The Chemical Society, Burlington House, London W. 1, 1957 and 1958.
3. Rydberg, J. and Rydberg, B. *Arkiv Kemi* **9** (1956) 81.

Received November 12, 1958.