

## On the Calculation of Complexity Constants by the Method of Least Squares \*

JAN RYDBERG

*University of Stockholm, and Research Institute of National Defense, Stockholm, Sweden*

Some requirements for the application of the method of least squares to equations for complex equilibria are discussed, and a review is given of what so far has been done. In a number of examples, the results of least squares calculations with high-speed digital computers are compared to results obtained by graphical methods. The application of simple statistical tests as the  $\chi^2$ - and t-tests are also demonstrated.

In complex chemistry — as in many other scientific fields — one is commonly faced with the problem of trying to fit theoretical equations to a number of experimental points. This fit may be done in a graphical or numerical manner. If the equations are complicated and the body of experimental data large, graphical methods seem more attractive and have been used by far more often in complex chemistry. However, with the increasing availability of high-speed digital computers, the complex chemist no longer has any reason to avoid numerical calculations.

The present paper shows how the method of least squares in combination with electronic computers can be applied to studies of complex equilibria. The results will be subjected to simple statistical tests and compared to results obtained graphically.

### TYPE OF ERRORS TO BE CONSIDERED

Since the method of least squares considers the errors in the experimental data, and permits the calculation of standard errors in the constants computed, it is necessary to say a few words about errors in general.

All experiments are subject to errors. Indeed, it may be observed, that it is impossible to establish the *true* value of any constant through experiments alone. However, through careful experiments, it is possible to approach the true value of the constants to any degree of closeness. This »degree of closeness» depends on the type of errors involved in the experiments.

\* Parts of this paper also appears under the same title and by the same author in *Advances in the Chemistry of Coordination Compounds*, Stanley Kirschner (Ed.), The Macmillan Co, New York 1961.

The most familiar type of error is *mistakes*, which are easily recognised because they always lie far off the main population of experimental data. In the following we shall assume that all mistakes have been eliminated from our data.

*Accidental errors* occur in all experiments, and their cause is in principle unknown. Their occurrence follows the laws of chance, and they can therefore be treated by the Mathematical Theory of Error. From the accidental errors in the experiments, this theory permits the calculation of precision in constants computed from the data. Various measures of precision are used; in this paper only the standard error (also called standard deviation or mean square error)  $\sigma$  shall be employed.

When the same system is investigated by various methods, and the results do not agree, this may either be due to the use of wrong theories, or to presence of unknown *systematic errors*. In principle such errors are determinate.

Before experimental data are used for computing constants, corrections for systematic errors should be applied, whether these corrections are based on empirical observations or true knowledge of the cause of the error. If such corrections are not made, *e.g.* because the presence of systematic errors is not realized by the investigator, the constants obtained will be much farther away from the true value than may be indicated by standard deviations given. The reader may view such results not only with suspicion but also regard them as being dangerous, because the standard deviations lend a false impression of exactness to the constants. However, it should be observed and strongly emphasized, that when errors are given for constants, in whatever field of science, these errors should be adequately defined. Standard errors, which are used in this paper, are exactly defined (*cf.*, *e.g.*, Refs.<sup>1,2</sup>) to be based only on the accidental errors in the experimental data. Thus the reader who suspects systematic errors in an investigations should, after considering the accidental errors causing the standard deviations in the constants, focus his attention on how the constants themselves have been obtained. The relative value of standard deviations in such a case will be discussed in connection with the examples at the end of the paper.

*In the following, it will be assumed that all mistakes and systematic errors have been eliminated from the experimental material, provided this question is not brought up in a particular discussion.*

#### TYPES OF EQUATIONS ENCOUNTERED IN EQUILIBRIUM STUDIES

The ultimate nature of the problem is to determine the stability constant  $\beta_{m,n,p}$  as defined by

$$\beta_{m,n,p} = \frac{[M_m A_n B_p]}{[M]^m [A]^n [B]^p} \quad (1)$$

for the complex  $M_m A_n B_p$ , where M is the central (metal) atom and A and B are two different kinds of ligands. Brackets may here indicate activities or concentrations; in the latter case it is assumed that the ratio of the activity coefficients of the various species involved in the equilibrium are constant.

More complicated complexes may be defined (*cf.*, *e.g.*, Refs.<sup>3-5</sup>, but these are almost entirely of theoretical interest because of the practical difficulties in investigating such complexes.

The mathematical formulation of the equations as well as methods for calculating the stability constants for *complexes of type MA<sub>n</sub>* have been thoroughly discussed by many authors as, *e.g.*, Bjerrum<sup>6</sup>, Sullivan and Hindman<sup>7</sup>, and Irving and Rossotti<sup>8</sup>. In the general equation

$$\sum_0^N (C_A - [A] - n \cdot C_M) \beta_n [A]^n = 0 \tag{2}$$

which for reasons given later can be written

$$\sum_0^N (y - bx - a_n) \beta_n x^n = 0 \tag{3}$$

$C_A$  and  $C_M$  are the total concentrations of ligand and metal.

Depending on the method of investigation, the symbols  $y$ ,  $b$  and  $a_n$  can be given various meanings. Thus, in spectrophotometric investigations where only the metal species show measureable extinction, one may put  $y = \epsilon_M =$  molar extinction coefficient of all metal species in solution ( $\epsilon_M = E \cdot C_M^{-1} \cdot d^{-1}$ , where  $E$  is measured extinction for cell length  $d$ ),  $b = 0$  and  $a_n = \epsilon_n =$  molar extinction coefficient of  $MA_n$  (*e.g.*, Ref.<sup>9</sup>). In partition measurements, one may put  $y = Q_M =$  partition ratio of all metal species between one phase (*e.g.* resin or organic solvent) and the aqueous phase,  $b = 0$ , and  $a_n = \lambda_n =$  partition constant for  $MA_n$  (*e.g.*, Ref.<sup>11-12</sup>).

The equations for *mixed complexes of type MA<sub>n</sub>B<sub>p</sub>* can be formulated in many ways, as given, *e.g.*, by Fronæus<sup>13</sup> and Rydberg<sup>3</sup>. The primary equations all contain double sums; a representative example is given by

$$\bar{n}_A = \frac{C_A - [A]}{C_M} = \frac{\sum_0^N n \beta_{np} [A]^n [B]^p}{\sum_0^N \sum_0^P \beta_{np} [A]^n [B]^p} \tag{4}$$

where  $\bar{n}_A$  is the average number of ligands A bound per metal atom.

Equations for the formation of *polynuclear complexes* have been elaborated by many authors, as for type  $M_2A_n$ , *e.g.*, by N. Bjerrum<sup>14</sup>, Leden<sup>15</sup> and Fronæus<sup>9</sup>, for type  $M_mA_n$ , *e.g.*, by Bodländer and Storbeck<sup>16</sup>, Brosset<sup>17</sup>, Ahrlund<sup>18</sup>, Sillén<sup>19</sup> and Hedström<sup>20</sup>, for type  $M_mH_jA_n$  (complexes between metals and acid ligands), *e.g.*, by Schwarzenbach<sup>21</sup>, for type  $H_jM_mA_n$  (polyanions neutralized by  $jH^+$ ), *e.g.*, by Rossotti and Rossotti<sup>22</sup> and for the core-and-link types by Granér<sup>23</sup> and Sillén<sup>23,24</sup>. The primary equations all contain double sums, as for example in

$$\bar{n}_A = \frac{C_A - [A]}{C_M} = \frac{\sum^m \sum^n n \beta_{mn} [M]^m [A]^n}{[M] + \sum^m \sum^n m \beta_{mn} [M]^m [A]^n} \tag{5}$$

where the summations are to be taken from 1.

A test of the validity of the models specified by equations (1)—(5) is outside the scope of the present paper, and it will be assumed that the equations are basically true for the given cases.

#### METHODS USED FOR CALCULATION OF COMPLEXITY CONSTANTS

After the experimental data have been collected, a threefold problem is encountered: (i) it is necessary to get an idea about which complexes are formed; (ii) on the basis of this idea, complexity constants are calculated; (iii) it is necessary to test if the results thus obtained are consistent with the experimental data. The first problem often involves chemical judgement. The second problem is mathematical in nature, and shall be discussed in this section. The third problem may also be a problem of judgement; it can be based both on chemical knowledge and on statistical tests, as will be further discussed in connection with the examples at the end of this paper.

At present four different approaches to the problem of computing constants in complexity equations seem to be in use. The methods are (a) the method of linear plots, (b) the method of curve fitting, (c) the method of averages, and (d) the method of least squares.

The first two methods are graphical and the last two are numerical. Each of these methods will be shortly discussed below\*. It should, however, be remembered that whatever method is applied, a graph of the experimental points should first be made, partly to eliminate mistakes (and perhaps systematic errors) and partly to see if the points in any way can be represented by a smooth curve.

(a). *The method of linear plots.* In this method (*e.g.* Ref.<sup>2</sup>) which by far is the most common, the equations are transformed so that they are linear in the variables over a certain range of one of the variables. In this range, the experimental data (variables) can be plotted as straight lines.

Under limiting conditions, the intercept of this line on one of the axes, or its slope at this point, gives a value, which may be one of the constants searched for (see, *e.g.*, Leden<sup>15</sup>) or a parameter which is a simple function of this constant (see, *e.g.*, Rydberg<sup>10</sup>). It may also be an intermediate variable obtained under fixed conditions. A number of such intermediate variables may then be plotted in a similar manner, and the intercept on one of the axes of the curve through these points may yield the constant or a new intermediate variable, etc. (see, *e.g.*, Åhrland<sup>25</sup>).

Instead of extrapolations, interpolations may be made; a set of curves drawn through the set of experimental points are cut at certain fixed values of one of the variables, and the corresponding values of the other variable is further treated in a graphical or numerical manner (see, *e.g.*, Fronæus<sup>9</sup>).

From the accidental errors in the experiments, it is possible to calculate standard errors in the final constants. However, this is rarely done, because it is so much simpler to get acceptable *estimates* of these errors from the graphs. By introducing the final constants in the equation valid for the experiments, and comparing the graph of this equation with the plot of the experimental

\* For a more comprehensive discussion, the reader is referred to the references given.

data (*cf.* next paragraph), systematic deviations between theory and experiments are easily revealed.

(b). *The method of curve-fitting.* In this method<sup>26-30</sup>, the basic equations are transformed so that for each variable there is a single relationship between that particular variable and one parameter, and this relationship is the only way in which the parameter enters the equation. Curves are then calculated for given parameter values. A fitting of these curves to the experimental data, transformed into the same type of equation, will yield the parameter values. These parameters are related to the constants searched for in a simple manner.

What is said in the foregoing paragraph on accidental errors, is also valid for the curve-fitting method. This method also easily reveals systematic deviations between theory and experiments.

(c). *The method of averages.* The method of averages (*e.g.* Ref.<sup>2</sup>) is probably the shortest and easiest numerical method for finding the constants in an empirical equation, but is limited to equations linear in the unknown constants.\* The method assumes that the best representative curve is that for which the algebraic sum of the residuals (*i.e.*, the difference between the computed and measured values of the dependent variable) is zero. When there are more equations (measured points) than unknown parameters, the equations are grouped together so that only as many equations as unknowns are obtained. This system of equations is then solved for the parameters in a conventional manner.

There is no rule for grouping together the equations, and as pointed out by Scarborough<sup>2</sup>, a careless grouping may yield quite erroneous results, unless each group comprises at least 3 to 4 points.

Though this method in its clear-cut form seems to be rarely applied in complex chemistry, the principle is used for example in Bjerrum's method<sup>6</sup> (*cf.* also Bloch and McIntyre<sup>31</sup> and many others), when  $n$  combinations  $\bar{n}([A])$  are introduced in equation (2), and the system of equations thus obtained is solved by determinants.

Properly used, the method of averages does not introduce any systematic computational errors. On the other hand it does not permit any calculation of statistical errors in the constants obtained.

(d). *The method of least squares.* This method is no doubt the most commonly used in numerical calculations in which the number of measured points exceed the number of unknown constants. The principle of least squares, which is founded on the normal probability equation (the "Gaussian curve"), says that the best or most probable value of a measured quantity is that value for which the sum of the weighted squares of the residuals is a minimum. The application of this principle to curve fitting is extensively treated, *e.g.*, by Deming<sup>32</sup>.

The method is only applicable to equations which are linear or can be reduced to a linear form in the constants. In complex chemistry this is a rather serious limitation, because it is seen that none of the equations (3)–(5) fulfill this requirement. The problem of linearity will be discussed separately in next paragraph.

---

\* See also next paragraph on least squares.

The method of least squares does not introduce any systematic computational errors. It gives the best set of constants obtainable from the data according to the Theory of Error, and it also permits the calculation of standard errors of these constants, which make it possible to subject the results to statistical tests (see, *e.g.*, Fisher<sup>33</sup>), as will be demonstrated in the examples at the end of this paper. A requirement for such a test as well as the use of the method of least squares is — as pointed out earlier — that only accidental errors occur and that they have a normal statistical distribution.

#### EQUATIONS SUITABLE FOR LEAST SQUARES TREATMENT

The determination of many parameters from a large number of experimental data with the method of least squares leads to very extensive calculations. It is therefore understandable that few chemists have used this method until very recently, when high-speed digital computers have become easily available. The following summary of what has been done in this field is therefore necessarily meager, though this situation will certainly not last very long.

(a). *Equations linear in the parameters.* If in eqn. (3),  $b = 0$  and only one value of  $a_n$  is greater than zero, the equation can be rearranged to give

$$z = \sum_0^N c_n x^n \quad (6)$$

which is linear in the parameter  $c_n$ , which is simply related to the constant  $\beta_n$ . This equation represents a number of common cases of investigations on the complex  $MA_n^-$ .

The method of least squares is straight forward in this case, and it has been applied by Irving and Rossotti<sup>8</sup> to a number of systems with  $N = 2$ , and by Kivalo and Rastas<sup>34</sup> and McMaster and Schaap<sup>35</sup> to polarographic studies with  $N = 4$ . While Irving and Rossotti, and Kivalo and Rastas made their calculations manually, McMaster and Schaap, as well as the investigators in the following references, used high-speed computer calculations. None of these authors used weighted data. Both Kivalo and Rastas, and McMaster and Schaap used an approximation technique in the calculations, and the latter authors did not calculate any standard deviations in their parameters.

Rydborg, Sullivan and Miller<sup>11,36-38</sup> have applied eqn. (2) to potentiometric and eqn. (6) to solvent extraction studies, while Zielen<sup>39</sup> has applied eqn. (6) to cation exchange data. In all cases, weighted data were used, and the computers were programmed to calculate standard deviations in the parameters.

(b). *Equations non-linear in the parameter; the Gauss transformation to a linear form.* The problem of using the method of least squares for determination of parameters, which do not enter the equations in a linear way, as is the case with eqns. (3)–(5), has been discussed, *e.g.*, by Moore and Zeigler<sup>40</sup>.

In order to obtain an equation linear in the unknowns ( $\beta_n$ ) these are replaced by a guessed or estimated value ( $\hat{\beta}_n$ ) and its deviation ( $\Delta \beta_n$ ) from the "true" value  $\beta_n$  (thus  $\beta_n = \hat{\beta}_n + \Delta \beta_n$ ). The function is then expanded in a

Taylor's series, where higher terms of  $\Delta \beta_n$  are neglected. The new function, which will take the form

$$y(\beta_n) \approx y(\hat{\beta}_n) + \frac{\partial y(\hat{\beta}_n)}{\partial \hat{\beta}_1} \cdot \Delta \beta_1 + \frac{\partial y(\hat{\beta}_n)}{\partial \hat{\beta}_2} \cdot \Delta \beta_2 + \dots \quad (7)$$

is then solved for  $\Delta \beta_n$  by the method of least squares. With the  $\Delta \beta_n$  values so obtained, new and better estimates of the constants are made, and these new constants are introduced into eqn. (7), and the procedure is repeated until the  $\Delta \beta_n$ 's are negligible to the standard errors computed for the  $\beta_n$ 's; usually very few iterations are needed<sup>40</sup>. This method may be referred to as the Gauss<sup>40</sup> or Newton-Raphson<sup>2</sup> method.

The method is based on estimated values of  $\beta_n$ . It seems not to be clear how good these estimates have to be. Moore and Zeigler report success in cases where the estimates have been as far off as a factor of  $10^5$ , and failures in cases where the estimates have been only 20 % off the correct value. In most cases, adequate estimates can probably be obtained from graphical plots.

This principle has (unknowingly?) been used by van Panthaleon van Eck<sup>41</sup> on data  $n([A])$ , where all weights have been taken equal to unity. It has also been used with weighted data in kinetic studies of the reduction of Np(VI) by Zielen, Sullivan and Hindman<sup>42</sup>, and on spectrophotometric investigations of the hydrolysis of Pu(IV) by Rabideau and Kline<sup>43</sup>.

When no approximate values of  $\beta_n$  are known or can easily be determined, reasonable such values of the parameters may be guessed and tried according to the Gauss method. Another approach is to use these guessed values, and to study how the sum  $S$  of the squares of the residuals vary with the variation of one parameter at a time. When the minimum of  $S$  ( $S_{\min}$ ) is found for one parameter, this value is retained, and the next parameter is varied until  $S_{\min}$  for that parameter is found. After  $S$  has been minimized for each parameter in turn, the procedure is started over again, and continued in this manner until  $S$  is a minimum for all of the parameters. Obviously, this technique will not guarantee that one will reach a true minimum for  $S$ . Also, there may be another set of parameters, which gives an equally small value of  $S$ <sup>40</sup>. However, if the original estimate of the parameters is not too bad, it seems likely that this method will yield correct answers.

Another more systematic method for polynuclear complexes has recently been suggested by Dyrssen, Ingri and Sillén<sup>44</sup>, who call the method "pit-mapping". In this case estimates of the parameters are obtained through graphical methods\*.

\* When mixed or polynuclear complexes are studied, the calculations will be much simplified if the experiments are carried out so that the system is studied as a function of one of the independent variables, while the other is kept constant. This situation is wellknown to all investigators in this particular field.

It was earlier indicated that eqn. (3) could be used for spectrophotometric and partition studies in which case it can be written ( $b = 0$ )

$$\sum_0^N (\epsilon_M - \epsilon_n) \beta [A]^n = 0 \quad (8a)$$

$$\sum_0^N (Q_M - \lambda_n) \beta_n [A]^n = 0 \quad (8b)$$

In these equations  $\epsilon_M$ ,  $Q_M$  and  $[A]$  are measured, while  $\epsilon_n$ ,  $\lambda_n$  and  $\beta_n$  are the unknown parameters, though usually one value of  $\epsilon_n$  and  $\lambda_n$  is known: *i.e.*  $\epsilon_0$  for  $M^{p+}$  or  $\epsilon_N$  for  $MA_N$  in spectrophotometric investigations,  $\lambda_0$  for  $M^{p+}$  in cation exchange studies, or  $\lambda_1$  for the uncharged complex  $MA$ , in solvent extraction work. By guessing reasonable values for  $\epsilon_n$  and  $\lambda_n$  the method of least squares can be used for calculation of those  $\beta_n$ 's which gives a minimum for  $S$ . This method, which has been used in one of the examples below, is obviously a variation of the principle described above.

#### SOME RESULTS OF LEAST SQUARES CALCULATIONS ON COMPLEXITY SYSTEMS

In this part, the application of the method of least squares to a number of complexity systems will be demonstrated and the discrepancies between the results and those obtained with graphical methods will be discussed. Since the method of least squares provide standard errors to the computed constants, some statistical tests have been applied and their merits analysed. It is believed that the viewpoints which are brought up in this connection are rather general in nature and shall indicate to the chemists not familiar to this field both what can be gained and what the limitations of the methods are.

(a). *The Ni(II)-NH<sub>3</sub> system; a potentiometric investigation.* The experimental data of Bjerrum<sup>6</sup> were directly introduced into eqn. (2) which was then solved for  $\beta_n$  by the method of least squares<sup>37</sup> using an IBM 704 computer. With the program used<sup>45a</sup>, one run on the computer takes about one minute. In practice, it may take a much larger time to find the  $\beta_n$ , because if proper data are not used, it may be necessary to make several runs on the computer. The reason for this is that all errors are assumed to belong to the experimental

Table 1. The Ni(II)-NH<sub>3</sub> system, 2 M NH<sub>4</sub>NO<sub>3</sub> at 30°C.

Gross constant *	Bjerrum Least squares		Step-wise constant *	Bjerrum Least squares	
	$\beta_1 \times 10^{-2}$	6.29		6.094 ± 0.048	log $k_1$
$\beta_2 \times 10^{-5}$	1.084	1.111 ± 0.020	log $k_2$	2.24	2.262 ± 0.0085
$\beta_3 \times 10^{-6}$	5.82	5.377 ± 0.173	log $k_3$	1.73	1.686 ± 0.016
$\beta_4 \times 10^{-7}$	9.00	9.254 ± 0.414	log $k_4$	1.19	1.249 ± 0.023
$\beta_5 \times 10^{-8}$	5.06	5.202 ± 0.262	log $k_5$	0.75	0.737 ± 0.029
$\beta_6 \times 10^{-8}$	5.43	5.514 ± 0.396	log $k_6$	0.03	0.026 ± 0.038

\*  $\beta_n = IIk_n$



determination of  $[A]$ , and the present program cannot handle cases where the difference  $C_A - [A] - nC_M$  is about the same or smaller than the error in  $[A]$ . If one wants to save computer time, such combinations have to be eliminated before running the program. In the Ni-NH<sub>3</sub> case, 3 points of 14 had to be excluded. Table 1 gives the results then obtained, and also the results of the semi-graphical ligand number method<sup>6</sup>. It is seen that the agreement is excellent.

The  $\chi^2$ -test<sup>33</sup> may be applied to the data. Bjerrum estimates "the limit of experimental error" in the e.m.f. measurements to 0.2–0.5 mV (p. 124)<sup>6</sup>, which here is interpreted to a standard error of  $\pm (0.1-0.25)$  mV. A standard error of  $\pm 0.1$  mV corresponds to a 0.54 % error in  $[A]$ , *i.e.*  $\sigma_{[A]} = \pm 0.0054[A]$ . Using this error of internal consistency<sup>32</sup>, a  $\chi^2$ -value is calculated (this calculation is incorporated in the computer program) to 4.25. Since the number of degrees of freedom here is 5 (11 points minus 6 parameters), the  $\chi^2$ -test shows a high degree of consistency between the experimental points and the curve calculated with the obtained constants. This consistency, found even for the smallest possible error assumed by Bjerrum, is a strong support for the basic hypothesis, *i.e.* that all stepwise complexes from Ni<sup>2+</sup> to Ni(NH<sub>3</sub>)<sub>6</sub><sup>2+</sup> exist. For larger errors in the experimental data, the  $\chi^2$  value decreases and the probability for a good fit between the calculated function and the experimental points further increases.

The same point may be illuminated by quoting Deming<sup>32</sup> "The unbiased estimate of  $\chi^2$  made by external consistency is found by calculating what value of  $\sigma$  forces  $\chi^2$  to take its mean value  $G$ ", where  $G$  is the number of degrees of freedom, which in this case, is the number of experimental points minus the number of parameters minus one. Using this criterium, one arrives at an error of 0.498 % in  $[A]$ , which obviously very closely corresponds to an error of  $\pm 0.1$  mV in the e.m.f. measurements.

These results indicate that the experiments perhaps have been even more accurate than Bjerrum assumed. Even though the result of the  $\chi^2$ -test here is striking, it must be realized that it probably is a rather favorable case. Some caution is also necessary, considering the small number of degrees of freedom in the system<sup>32</sup>.

(b). *The Th(IV), U(IV), Pu(IV) acetylacetonone system; a solvent extraction investigation.* This example may well be called "The case of the missing constants" or — as it has turned out — "The case of constant misunderstandings", as shall be explained below.

The least squares calculations of stability constants for Th(IV), U(IV) and Pu(IV) acetylacetonone complexes using eqn. (6) and the IBM 704 computer has earlier been published<sup>38</sup> and the results are collected in Table 2. In this case a program<sup>45b</sup>, slightly different from the previous one, was used. All errors are assumed to belong to the measured distribution ratio  $Q$  of the metal between the organic solvent and the aqueous phase. The speed of the computations are about the same as for the previous program. No points have to be excluded.

In the U(IV) case the agreement between the constants obtained by the two-parameter method and the least squares method is quite good. The number of experimental points in this case is 46 and the residuals show a normal distri-

Table 2. The Th(IV), U(IV) and Pu(IV) acetylacetonone systems; 25°C.

Case	System	log $k_1$	log $k_2$	log $k_3$	log $k_4$	log $\lambda_4$
1	U(IV), par. <sup>a)</sup> Least squares	9.05 ± 0.16	7.97 ± 0.17	6.90 ± 0.18	5.84 ± 0.11	3.62 ± 0.05
		9.02 ± 0.29	8.26 ± 0.15	6.52 ± 0.09	5.98 ± 0.09	3.64 ± 0.025
2	Pu(IV), Lig. <sup>b)</sup> Least squares	10.0 ± 1.	9.1 ± 0.2	8.5 ± 0.3	5.9 ± 0.3	2.50 ± 0.2
		—	17.77 ± 0.12	—	5.91 ± 0.09	2.54 ± 0.07
3	Th(IV), Lim. <sup>c)</sup> Least squares	(7.82 ± 0.16)	(7.73 ± 0.09)	6.27 ± 0.05	5.00 ± 0.04	2.52 ± 0.04
		15.57 ± 0.08	—	6.15 ± 0.04	5.14 ± 0.07	2.48 ± 0.06
4	Th(IV) Lim. <sup>d)</sup> Least squares	8.00 ± 0.17	7.48 ± 0.14	6.00 ± 0.11	5.30 ± 0.05	2.55 ± 0.03
		—	7.43 ± 0.21	5.83 ± 0.39	5.35 ± 0.42	2.53 ± 0.20

$$k_1 = \frac{a_1}{a_0} = \frac{[MA]}{[M][A]}; k_2 = \frac{a_2}{a_1} = \frac{[MA_2]}{[MA][A]}; k_3 = \frac{a_3}{a_2} = \frac{[MA_3]}{[MA_2][A]}; k_4 = \frac{a_4}{a_3} = \frac{[MA_4]}{[MA_3][A]};$$

$$\lambda_4 = \frac{1}{a_4} = \frac{[MA_4]_{org}}{[MA_4]_{aq}}; k_1 k_2 = \frac{a_2}{a_0} = \frac{[MA_2]}{[M][A]^2}; k_2 k_3 = \frac{a_3}{a_1} = \frac{[MA_3]}{[MA][A]^2}$$

a) Calculated by 2-parameter method: 0.1 M NaClO<sub>4</sub>

b) » » ligand number method: 0.1 M NaClO<sub>4</sub>

c) » » limiting value method: 0.01 M NaClO<sub>4</sub> · Benzene solvent.

d) » » » » » 0.1 M NaClO<sub>4</sub> · Chloroform solvent.

bution around the curve. Thus it may be concluded, that the two-parameter method and the least squares method of calculation applied to good solvent extraction data yields the same results, provided the basic assumptions for the two-parameter method are valid <sup>26</sup>.

Not so good agreement is found in the Pu(IV) case, which comprises 24 points. From a consideration of replicate experiments, a standard error of 20 % was assigned to the measured distribution ratio. The least squares calculations yielded the results in Table 3. The negative values of the parameters  $a_0$  and  $a_2^*$  indicate either a systematic deviation from the assumed equation for the system or that negative concentrations of the complexes are present. Since the standard errors are very large and negative concentrations of complexes are assumed not to exist, it seems reasonable to eliminate these parameters from the calculations.

This procedure may cause some concern among chemists, who may conclude that the elimination of  $a_0$  and  $a_2$  means that the corresponding chemical species (Pu<sup>4+</sup> for  $a_0$ , and PuA<sub>2</sub><sup>2+</sup> for  $a_2$ ; see Table 2) do not exist. The author feels however, that constants should not be given for equilibria involving species which one has not been able to identify \*\*. It might well be possible to identify the missing species and to determine equilibrium constants for their formation

\* The parameters  $a_n$  are defined in Table 2.

\*\* In this case, the only method of "identification" is through a fit between the experimental data and the equations involving the particular species.

Table 3. Primary parameters obtained for the Pu(IV) case.

$a_0 \times 10^{27}$	$a_1 \times 10^{26}$	$a_2 \times 10^{17}$	$a_3 \times 10^9$	$a_4 \times 10^8$	$\chi^2$	$G$
$-17 \pm 10$	$1.9 \pm 0.8$	$-0.046 \pm 0.53$	$3.3 \pm 0.6$	$2.9 \pm 0.4$	17.3	19
—	$0.61 \pm 0.15$	—	$3.6 \pm 0.5$	$2.9 \pm 0.5$	22.4	21

with a better set of experimental data. In the present experimental material,  $\text{Pu}^{4+}$  and  $\text{PuA}_2^{2+}$  cannot be identified ( $[\text{Pu}^{4+}]$  and  $[\text{PuA}_2^{2+}]$  cannot be determined) with any reasonable degree of accuracy, though the existence of  $\text{PuA}_2^{2+}$  seems highly probable from a chemical point of view, and  $\text{Pu}^{4+}$  is known to exist. The  $\chi^2$ -values indicate about equally good fits between the experimental data and the calculated curve, either  $\text{Pu}^{4+}$  and  $\text{PuA}_2^{2+}$  are assumed to be present or not. Thus the indeterminate influence of these species on the system is obvious.

A similar state of affairs exists for the Th(IV) systems in Table 2. In these systems, one expects to find the same species, even though the systems have been investigated by slightly different methods. With the graphical methods, formation constants were obtained for all stepwise complexes. Nevertheless the least squares calculations on case 3 came out with  $a_1 = (-1.3 \pm 1.3) 10^{-22}$  and on case 4 with  $a_0 = (-17 \pm 20) 10^{-30}$ . The parameters  $a_1$  and  $a_0$  are here eliminated simply because they cannot statistically be proven to differ from zero; this operation caused practically no change in the  $\chi^2$ -values. The elimination of  $a_1$  in case 3 and  $a_0$  in case 4 does not mean that  $\text{ThA}_3^+$  and  $\text{Th}^{4+}$  respectively, do not exist. In fact, the species missing in one of the systems is "found" in the other of the systems.

Tests were also carried out assuming that complexes higher than  $\text{MA}_4$  were formed, *i.e.*  $\text{MA}_5$  and  $\text{MA}_6$ , but the standard deviations of  $\beta_5$  and  $\beta_6$  far exceeded the constants. In fact, in no case where the least squares method has been tried by the author, has it been possible to identify more complexes than assumed by the original investigators.

It may thus be concluded that the method of least squares does not give any better results than can be justified by the experimental data. An advantage of this fact is that it indicates in which experimental region the data are too uncertain (too scattered, too few, *etc.*), whatever consequence this may have on the investigator.

(c). *The  $\text{VO}^{2+}-\text{F}^-$  system; a potentiometric investigation.* To this system <sup>46</sup> eqn. (2) is applicable in the same manner as for the Ni— $\text{NH}_3$  system described earlier. In Fig. 1, the original authors have plotted  $\bar{n}([\text{A}])$ . Since the points are obtained from 5 different sets of experiments, and the curve obtained with the four constants  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$  seems to fit all the 5 sets of points, a statistical *t*-test <sup>33</sup> seems applicable.

A general look at Table 4 reveals 3 things: (i) of the 5 sets of experiments some rather high values of  $\chi^2/G$  are obtained; (ii) the  $\beta_n$ -values computed by the method of least squares for the various sets of experiments do not all agree

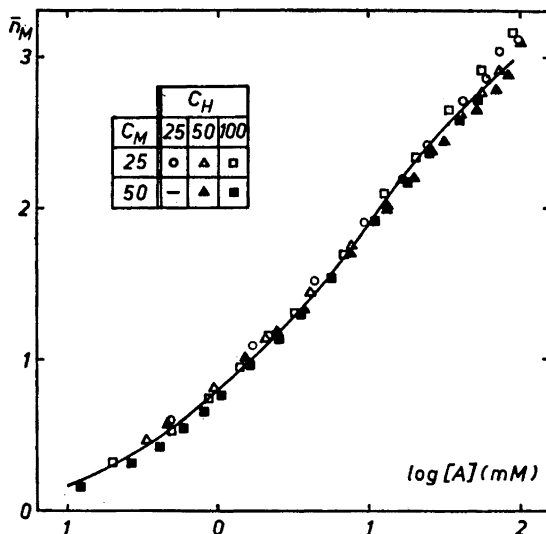


Fig. 1. The complex formation function of the vanadyl fluoride system. The points refer to titrations with different values of  $C_M$  and  $C_H$ , as indicated by the scheme given in the figure. The curve is calculated from the constants  $\beta_n$  finally found.

within their errors; (iii) Ahrlund and Noréns  $\beta_n$ -values agree approximately well with the mean values calculated by the method of least squares. These points shall be discussed in turn.

If the error of external consistency is found by calculating what value forces  $\chi^2/G = 1.0$ , one finds that the error should be around 2% in  $[A]$ , with exception for set 4, where the error must be assumed to be about 5%. This agrees rather well with the  $\pm 0.2$  mV error estimated for the e.m.f. measurements.

Table 4. The  $VO^{2+}$ -F-system. The errors given are mean square errors at the 95% confidence limit (2 times standard deviation).

Set	$[H^+] \times 10^2$	$C_M \times 10^3$	$\beta_1 \times 10^{-3}$	$\beta_2 \times 10^{-5}$	$\beta_3 \times 10^{-7}$	$\beta_4 \times 10^{-7}$	$\chi^2$	$G$
1	25	25	$2.29 \pm 0.03$	$4.84 \pm 0.07$	$1.75 \pm 0.06$	$8.97 \pm 0.64$	0.64	4
2	50	25	$2.16 \pm 0.06$	$4.14 \pm 0.16$	$1.31 \pm 0.14$	$6.14 \pm 1.31$	5.30	5
3	50	50	$2.13 \pm 0.10$	$3.67 \pm 0.28$	$1.15 \pm 0.20$	$3.12 \pm 1.73$	41.4	6
4	100	25	$1.97 \pm 0.20$	$3.49 \pm 0.60$	$1.53 \pm 0.59$	$9.65 \pm 9.23$	198	7
5	100	50	$1.53 \pm 0.03$	$2.57 \pm 0.11$	$0.93 \pm 0.07$	$3.32 \pm 0.90$	20.6	10
Mean values of least squares calculations			$2.02 \pm 0.08$	$3.74 \pm 0.24$	$1.34 \pm 0.21$	$6.2 \pm 2.8$		
Values by Ahrlund and Norén			$2.0 \pm 0.1$	$2.9 \pm 0.3$	$1.4 \pm 0.2$	3.0		

A t-test was carried out on the  $\beta_r$ -values ( $r = 1, 2, 3, 4$ ) to see if there were any significant deviations between them. The test was based on the assumption (1) that in each experimental set, the points have a normal statistical distribution, and (2) that all data were of the same relative weights. A constant percentage error  $P$  in  $[A]$  was assumed; thus  $1/W_i = (0.01 P[A]_i)^2$ .

If these assumptions are not too incorrect, the t-test indicates significant differences between the  $\beta_r$ -values. Thus for  $10^{-5}\beta_2 = 4.84 \pm 0.07$  (set 1) and  $2.57 \pm 0.11$  (set 5), the test shows that the probability is  $<1\%$  that the two  $\beta_2$ -values belong to the same normal population. As a comparison, it is found that the two values  $10^{-5}\beta_2 = 4.14 \pm 0.16$  (set 2) and  $3.67 \pm 0.28$  (set 3) may belong to the same population. After analyzing all  $\beta_r$ -values it must be concluded from a statistical point of view that the differences between these values must be caused by systematic differences among the five sets of experiments.

If all experimental points are taken together (*i.e.* are assumed to belong to the same normal population), the agreement between the mean values of the least squares calculations and the values obtained by Ahrlund and Norén is quite reasonable (see Table 4). Fig. 1 also shows how well the computed curve fits the experimental points. It is the author's opinion that graphical methods do not permit more information to be extracted from these data than has been obtained by the original authors.

(*d*). *The Ce(III)-SO<sub>4</sub><sup>2-</sup> system; a cation exchange investigation.* This is the only system<sup>47</sup> known to the author where the parameters are not linear in the equation and the trial-and-error least squares approach (see p. 1729) has been applied. In this particular case, equation 8b is applicable, and the maximum number of parameters, which have to be considered, are  $\lambda_0, \lambda_1, \beta_1, \beta_2$  and  $\beta_3$ . Of these constants  $\lambda_0$  is known directly from the experiments (see below). In the following treatment we shall assume that we have no knowledge of the results of Fronæus's calculations.

On the basis of the experiments, it can be assumed that all error can be attributed to the measured partition ratios  $Q$ , and are of the size  $\sigma_Q = \pm 0.03 Q$ . The experimental data  $Q([A])$  comprise only 8 points. The curve through these points is anchored at  $Q_0([A] = 0) = 0.752$ , which is our  $\lambda_0$ -value.

We now make a number of runs on the computer with equations based on the following assumptions:

Run:	Cerium species in solution	Cerium species on the resin
1	Ce <sup>3+</sup> , CeA <sup>+</sup> , CeA <sub>1</sub> <sup>-</sup>	Ce <sup>3+</sup>
2-5	»	Ce <sup>3+</sup> , CeA <sup>+</sup>
6	Ce <sup>3+</sup> , CeA <sup>+</sup> , CeA <sub>1</sub> <sup>-</sup> , CeA <sub>3</sub> <sup>3-</sup>	Ce <sup>3+</sup>
7-10	»	Ce <sup>3+</sup> , CeA <sup>+</sup>

Here A<sup>2-</sup> stands for SO<sub>4</sub><sup>2-</sup>. Looking at runs 1-5 in Table 5, it follows from the  $\chi^2/G$ -values that the most probable  $\lambda_1$  value is obtained for run 1 ( $\chi^2/G$  value closest to 1.0). However, even though the  $\chi^2/G$ -value for run 1 is not improbable for a fit between the assumed equation and the experimental data, one of the

Table 5. The Ce(III)–SO<sub>4</sub><sup>2-</sup> system.  $G = 5$  for runs 1–5, and 4 for runs 6–10;  $\lambda_0 = 0.752$ .

Run	$\lambda_1$	$10 \beta_1$	$100 \beta_2$	$1\ 000 \beta_3$	$Q_1$	$Q_3$	$Q_7$	$\chi^2/G$
1	0.00	$3.45 \pm 0.14$	$1.65 \pm 0.16$	0	0.424	0.201	0.0504 *	1.51
2	0.04	$3.45 \pm 0.20$	$2.90 \pm 0.20$	0	0.428	0.199	0.0514 *	2.67
3	0.0787	$3.46 \pm 0.26$	$4.10 \pm 0.23$	0	0.430 *	0.198	0.0521 *	3.90
4	0.10	$3.46 \pm 0.29$	$4.76 \pm 0.26$	0	0.432 *	0.197	0.0523 *	4.58
5	0.30	$3.48 \pm 0.60$	$11.08 \pm 1.09$	0	0.044 *	0.191 *	0.0532 *	9.94
6	0.00	$3.47 \pm 0.19$	$0.72 \pm 0.51$	$0.48 \pm 0.26$	0.418	0.205	0.0479	0.994
7	0.04	$4.01 \pm 0.21$	$1.25 \pm 0.55$	$0.95 \pm 0.32$	0.418	0.205	0.0481	0.946
8	0.0787	$4.29 \pm 0.24$	$1.86 \pm 0.58$	$1.46 \pm 0.40$	0.418	0.206	0.0482	0.904
9	0.10	$4.47 \pm 0.26$	$2.25 \pm 0.59$	$1.77 \pm 0.44$	0.417	0.206	0.0482	0.884
10	0.30	$7.05 \pm 0.60$	$8.38 \pm 0.66$	$6.37 \pm 1.21$	0.417	0.206	0.0485	0.752
Fronæus	0.0787	$4.3 \pm 0.3$	$2.2 \pm 0.5$	$1.2 \pm 0.3$	0.414 $\pm 0.012$	0.202 $\pm 0.006$	0.0483 $\pm 0.014$	—

\* These values are outside the 3 % standard error in the measured  $Q$ -values.

computed points,  $Q_7 = 0.0504$  seems somewhat too far off from the measured value 0.0483. We assume this is a significant difference, and therefore make a new set of runs, 6–10, where also  $\text{CeA}_3^{\pm}$  is considered in the equation. Since  $\chi^2/G$  has the same normal distribution as the sum  $S$  of the squares of the weighted residuals<sup>30</sup>,  $\chi^2/G$  shall show a minimum for the most probable value of  $\lambda_1$ . No such minimum is achieved in runs 6–10, and no higher  $\lambda_1$  value than 0.30 was tested, because higher values seemed very improbable from a chemical point of view. As compared to runs 1–5, the  $\chi^2/G$  values in runs 6–10 indicate that the functional form in the latter case is more probable. From this result it may be concluded that the experimental data can be best explained by considering the species  $\text{Ce}^{3+}$ ,  $\text{CeA}^+$ ,  $\text{CeA}_2^-$ , and  $\text{CeA}_3^{\pm}$  in the aqueous phase, and the species  $\text{Ce}^{3+}$  and  $\text{CeA}^+$  in the resin phase. This is the same conclusion as reached by Fronæus.

Unfortunately, as the least squares calculations were made on the present data, the  $\chi^2$ -test gives us very little help in finding the best  $\lambda_1$  value. We are therefore forced to make a subjective judgement. Since the  $\lambda_1$  value is of greater importance for the higher complexes (*i.e.* higher  $\beta_n$ ; see left part of Table 5), which dominate at the points of highest  $[A]$  value (the  $[A]$  value increases from point  $Q_0$  to  $Q_7$ ), we shall say that the best  $\lambda_1$  is that, which causes the smallest deviation between the computed and measured values of point  $Q_7$ ; this value turns out to be 0.0482 (runs 8 and 9). Since we here have two such values, we chose the one which causes the smallest deviation in the  $Q_3$ -value, which turns out to be for run 8, if the trend in  $Q_3$  is considered. This corresponds to the same  $\lambda_1$ -value as Fronæus has arrived at from somewhat different considerations. It should be observed, that the approach used here can be stated in a more strict mathematical form, because what we have done is only to increase the

weights of the points for the highest [A] values. It may be added that the electronic computer can easily be ordered to try the various  $\lambda_n$  values automatically until the  $\lambda_n$  value, which gives the smallest  $S$  value, is obtained.

If we compare the constants obtained by Fronæus and by the method of least squares for  $\lambda_1 = 0.0787$ , we find an excellent agreement. This shows that complicated graphical treatments, when carried out on good data yield results in almost complete agreement with the results of our machine calculations.

#### CONCLUSION

It is always necessary to check the computed results to see if they are reasonable from a chemical point of view. Of all tests, which can be applied to the data, this is the most crucial. When a sufficient number of experimental points are at hand, it should also be checked (by hand or using the computer) that the points have a normal distribution around the computed curve. Under these circumstances, the method of least squares in combination with high speed digital computers is both a powerful and time-saving tool for helping complex chemists in analyzing their experimental data.

#### REFERENCES

1. Cramér, H. *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946; *Sannolikhetskalkylen och några av dess användningar*, Almqvist och Wiksell, Uppsala 1961.
2. Scarborough, J. B. *Numerical Mathematical Analysis*, 4th Ed. The John Hopkins Press, Baltimore 1958.
3. Rydberg, J. *Arkiv Kemi* **8** (1955) 101.
4. Irving, H., Rossotti, F. J. C. and Williams, R. J. P. *J. Chem. Soc.* **1955** 1906.
5. Sillén, L. G. Supplement to *La Ricerca Scientifica* **28** (1958).
6. Bjerrum, J. *Metal Amine Formation in Aqueous Solution*, P. Haase and Son, Copenhagen 1941.
7. Sullivan, J. C. and Hindman, J. C. *J. Am. Chem. Soc.* **74** (1952) 6091.
8. Irving, H. and Rossotti, H. S. *J. Chem. Soc.* **1953** 3397.
9. Fronæus, S. *Komplexsystem hos koppar*. Thesis, Lund 1948.
10. Rydberg, J. *Acta Chem. Scand.* **4** (1950) 1503.
11. Rydberg, J. and Sullivan, J. C. *Acta Chem. Scand.* **13** (1959) 1057.
12. Fronæus, S. *Acta Chem. Scand.* **5** (1951) 859.
13. Fronæus, S. *Acta Chem. Scand.* **4** (1950) 72.
14. Bjerrum, N. Thesis, Copenhagen 1908.
15. Leden, I. *Potentiometrisk undersökningar av några kadmiumsalters komplexitet*. Thesis, Lund 1943.
16. Bodländer, G. and Storbeck, O. *Z. anorg. Chem.* **31** (1902) 1, 458.
17. Brosset, C. *Elektrokemisk och röntgenkristallografisk undersökning av komplexa aluminiumfluorider*. Thesis, Stockholm 1942.
18. Ahrland, S. *Acta Chem. Scand.* **3** (1949) 374.
19. Sillén, L. G. *Acta Chem. Scand.* **8** (1954) 299.
20. Hedström, B. *Acta Chem. Scand.* **9** (1955) 613.
21. Schwarzenbach, G. *Helv. Chim. Acta* **33** (1950) 947.
22. Rossotti, F. J. C. and Rossotti, H. *Acta Chem. Scand.* **10** (1956) 957.
23. Granér, F. and Sillén, L. G. *Acta Chem. Scand.* **1** (1947) 631.
24. Sillén, L. G. *Acta Chem. Scand.* **8** (1954) 318.
25. Ahrland, S. *Acta Chem. Scand.* **3** (1949) 783.
26. Dyrssen, D. and Sillén, L. G. *Acta Chem. Scand.* **7** (1953) 663.

27. Rossotti, F. J. C. and Rossotti, H. *Acta Chem. Scand.* **9** (1955) 1166.
28. Sillén, L. G. *Acta Chem. Scand.* **10** (1956) 186.
29. Rossotti, F. J. C., Rossotti, H. and Sillén, L. G. *Acta Chem. Scand.* **10** (1956) 203.
30. Sillén, L. G. *Acta Chem. Scand.* **10** (1956) 803.
31. Block, B. P. and McIntyre, Jr., G. H. *J. Am. Chem. Soc.* **75** (1953) 5667.
32. Deming, W. E. *Statistical Adjustment of Data*, J. Wiley and Son, New York 1948.
33. Fisher, R. A. *Statistical Methods for Research Workers*, 13th Ed., Hafner Publ. Co., New York 1958.
34. Kivalo, P. and Rastas, J. *Suomen Kemistilehti* **30** (1957) 128.
35. McMasters, D. L. and Schaap, W. B. *Proc. Indiana Acad. Sci.* **67** (1958) 111.
36. Rydberg, J. and Sullivan, J. C. *Acta Chem. Scand.* **13** (1959) 186.
37. Sullivan, J. C., Rydberg, J. and Miller, W. F. *Acta Chem. Scand.* **13** (1959) 2023.
38. Rydberg, J. *Acta Chem. Scand.* **14** (1960) 157.
39. Zielen, A. *J. Am. Chem. Soc.* **81** (1959) 5022.
40. Moore, R. H. and Zeigler, R. K. *The solution of the general least squares problem with special references to high-speed computers*, LA-2367, Los Alamos Scientific Laboratory, New Mexico, 1960.
41. Van Panthaleon van Eck, C. L. Thesis, Amsterdam 1958.
42. Zielen, A. J., Sullivan, J. C. and Hindman, J. C. *Private communication*.
43. Rabideau, S. W. and Kline, R. J. *J. Phys. Chem.* **64** (1960) 680.
44. Dyrssen, D., Ingri, N. and Sillén, L. G. *Acta Chem. Scand.* **15** (1961) 694.
45. a) Program 626/CHM 112. b) Program 495/CHM 104. Applied Mathematics Division, Argonne National Laboratory, Argonne, Illinois.
46. Ahrland, S. and Norén, B. *Acta Chem. Scand.* **12** (1958) 1595.
47. Fronæus, S. *Svensk Kem. Tidskr.* **64** (1952) 317.

Received May 3, 1961.