345. *The Application of Weighted Least-squares Methods to the Computation of Stability Constants*

By R. C. Lansbury, V. E. Price, and A. G. Smeeth

Two methods are described for calculating, with the aid of a digital computer, the successive formation constants of a complex system, using ligand-number data. Weighted least-squares criteria were used in both methods for determining the constants.

A large number of graphical methods have been described for calculating successive formation constants defining complex equilibria.[1] These methods tend to be subjective in nature and thus do not give objective estimates of the precision of the calculated constants; moreover, they weight the data in an arbitrary fashion, and are also often tedious to apply. More recently, several reports [2] describe the use of digital computers for computing the constants, where by objective methods tests can be applied, but few (cf. ref. 3) give detailed account of the methods for specific problems. The present Paper describes two methods which can be generally applied to one of the most frequently used experimental techniques for determining stability constants, *viz.*, the measurement of the pH of solutions of the ligands with and without the addition of the central co-ordinating ion. The methods of calculation apply specifically to mononuclear systems with a maximum co-ordination number of 2, but they can be easily extended to systems with a maximum of 3 or 4, and possibly up to 6, although a different numerical technique in which orthogonal polynomials are used would be more suitable for the problems of high order.[4]

Experimental

The present work arose from a study of complex-formation between silver ions and a number of aliphatic amino-compounds. Full experimental details and complete results will be published later, but, briefly, the experimental technique is a refined form of the pH-titration method used by Bjerrum [5] to give free-ligand concentrations [A] and the corresponding ligand number $\bar{n}$, at constant ionic strength of 0·5M. A newly developed cell with free diffusion liquid junction,[6] in conjunction with a stable electrometer (E.I.L. Vibron), were used, so that e.m.f. measurements were reproducible to $\pm 0{\cdot}1$ mv.

Weighting Procedure

The system studied involved the following equilibria:

$$Ag^+ + A \rightleftharpoons AgA^+ \quad (1)$$

and

$$AgA^+ + A \rightleftharpoons AgA_2^+ \quad (2)$$

where A is an aliphatic amino-compound, it being understood that all species are solvated. Conventionally, two equilibrium constants are used to describe this system, *viz.*, β_1 and β_2, where β_1 refers to the first equilibrium, and $\beta_2 = \beta_1 k_2$, where k_2 refers to the second equilibrium. The formation function for such a system is given by

$$\bar{n} + (\bar{n} - 1)[A]\beta_1 + (\bar{n} - 2)[A]^2\beta_2 = 0 \quad (3)$$

There will, therefore, exist a series of equations of the form

$$\bar{n}_i + (\bar{n}_i - 1)[A]_i\beta_1 + (\bar{n}_i - 2)[A]_i^2\beta_2 = R_i \quad (4)$$

where $\bar{n}_i$ and $[A]_i$ are members of a set of experimentally determined values, and R_i, which should theoretically be zero, are residuals due to experimental errors. The problem was to

[1] Rossotti and Rossotti, " The Determination of Stability Constants," McGraw-Hill, New York, 1961.
[2] Deelstra, Vanderleen, and Verbeek, *Bull. Soc. chim. belges*, 1963, **72**, 632, and references therein.
[3] Inman, Regan, and Girling, *J.*, 1964, 348.
[4] Forsythe, *J. Soc. Ind. Appl. Math.*, 1957, **5**, No. 2, 74.
[5] Bjerrum, " Metal Ammine Formation in Aqueous Solution," Haase, Copenhagen, 1941.
[6] Greczek, Ph.D. Thesis, London, 1964.

choose the constants β_1 and β_2 according to a weighted least-squares criterion. Owing to the large variations in the size of the quantities $[A]_i$, the effects of the experimental errors on the residuals will vary considerably throughout the experiment. In the normal weighted least-squares technique each value of R^2 is divided by its variance. This may be obtained from a knowledge of the variance of the measured quantities and the relations by which R is determined from them. Thus, the weighted least-squares adjustment requires that β_1 and β_2 be selected such that

$$S = \sum_{i=1}^{I} w_i R_i^2 \tag{5}$$

shall be a minimum, where I is the number of data points, and w_i is defined by

$$w_i = 1/\text{var } R_i, \tag{6}$$

var R_i being the variance of R_i. To determine the variance of R_i it is necessary to consider the possible sources of experimental errors. In the pH-titration method we have considered there to be two main sources of error. First, an error associated with the measurement of e.m.f. values, and secondly an error associated with the titre values. Some other errors, *e.g.*, those caused by weighing and the calomel electrode are considered to be negligible. It is, however, possible that liquid-junction errors are significant, but it is not possible to allow for these theoretically. Hence, each residual R is a function of the corresponding e.m.f. reading (E) and titre reading (C_a). If δE is the error associated with the measurement of E, and δC_a the error associated with the measurement of C_a, then the error δR in R will be given according to the propagation of errors, by

$$\delta R = (\partial R/\partial E)\delta E + (\partial R/\partial C_a)\delta_a \tag{7}$$

so that the variance of R will then be

$$\text{var } R = (\partial R/\partial E)^2 \text{ var } E + (\partial R/\partial C_a)^2 \text{ var } C_a \tag{8}$$

assuming no correlation exists between the errors in E and C_a.

The free-ligand concentrations [A] were determined from the measured e.m.f.s and total titrant added in the usual way,[5] so that

$$[A] = C_a \, e^{2\cdot303E/k} \tag{9}$$

where k = slope of the glass electrode, C_a = total un-ionised ligand concentration available for complex-formation, and E = difference in cell e.m.f. in the presence and absence of silver ions ($E_1 - E_2$).

The ligand number $\bar{n}$ is defined by the equation

$$\bar{n} = (C_a - [A])/C_m \tag{10}$$

where C_m = total silver concentration. It is now possible to determine $\partial R/\partial E$ and $\partial R/\partial C_a$. Differentiation of equation (4) with respect to $\bar{n}$ and [A] gives

$$\delta R = (1 + [A]\beta_1 + [A]^2\beta_2)\delta\bar{n} + \{(\bar{n} - 1)\beta_1 + 2(\bar{n} - 2)[A]\beta_2\}\delta[A] \tag{11}$$

Differentiation of equation (10) with respect to [A] and C_a gives

$$\delta\bar{n} = (1/C_m)(\delta C_a - \delta[A]) \tag{12}$$

Taking the logarithm of both sides of (9), and differentiating, gives

$$\delta[A]/[A] = \delta C_a/C_a + 2\cdot303\delta E/k \tag{13}$$

and eliminating $\delta\bar{n}$ and $\delta[A]$ from equation (11), using equations (12) and (13), we obtain

$$\begin{aligned}\delta R = \{(1/C_m)(1 + [A]\beta_1 + [A]^2\beta_2) + ([A]/C_a)(\bar{n} - 1)\beta_1 + (2/C_a)(\bar{n} - 2)[A]^2\beta_2 - \\ ([A]/C_aC_m)(1 + [A]\beta_1 + [A]^2\beta_2)\}\delta C_a + (2\cdot303[A]/k)\{(\bar{n} - 1)\beta_1 + \\ 2(\bar{n} - 2)\beta_2[A] - (1/C_m)(1 + [A]\beta_1 + [A]^2\beta_2)\}\delta E\end{aligned} \tag{14}$$

which may be written conveniently as

$$\delta R = P\delta E + Q\delta C_a \tag{15}$$

The variance of R may then be obtained from (15):

$$\text{var } R = P^2 \text{ var } E + Q^2 \text{ var } C_a \tag{16}$$

Since E represents a difference in cell e.m.f.s ($E_1 - E_2$), then

$$\text{var } E = \text{var } E_1 + \text{var } E_2 \tag{17}$$

and since the precision of measurement is the same for E_1 and E_2, *i.e.*, var E_1 = var E_2, then var E = 2 var E_1. Equation (17) now becomes

$$\text{var } R = 2P^2 \text{ var } E_1 + Q^2 \text{ var } C_a \tag{18}$$

The titre error of the micro-burette was taken as 0·01 ml. δC_a is expressed in moles/l., and since the titrant used was 0·500M-sodium hydroxide the range of the error δC_a is therefore $\pm 0{\cdot}000005$ mole/l. The range of the error δE_a is $\pm 0{\cdot}1$ mv. The variances of E_1 and C_a are proportional to the squares of the ranges of the errors associated with them.

β_1 and β_2 can now be calculated from a series of weighted equations of the type

$$\{(\bar{n}_i - 1)[\text{A}]_i\beta_i + (\bar{n}_i - 2)[\text{A}]_i^2\beta_2\}w_i = -w_i\bar{n}_i \tag{19}$$

One of the simplest methods of fitting β_1 and β_2 to these equations, using a least-squares criterion, is by matrix algebra methods. Equation (19) may be written in the matrix form

$$\mathbf{WZx} = \mathbf{WY} \tag{20}$$

where $\mathbf{x}$ is the column vector $\begin{vmatrix}\beta_1\\ \beta_2\end{vmatrix}$, $\mathbf{Y}$ is the column vector whose elements are $-n_i$, $\mathbf{Z}$ is an (I × 2) matrix whose first column consists of the elements $(\bar{n}_i - 1)[\text{A}]_i$ and whose second column consists of the elements $(\bar{n}_i - 2)[\text{A}]_i^2$, and $\mathbf{W}$ is a diagonal matrix of elements w_i. The corresponding weighted normal equations are

$$\mathbf{Z}^{\text{T}}\mathbf{WZx} = \mathbf{Z}^{\text{T}}\mathbf{WY} \tag{21}$$

where $\mathbf{Z}^{\text{T}}$ is the transpose of $\mathbf{Z}$. The solution of these equations is given formally by inverting $(\mathbf{Z}^{\text{T}}\mathbf{WZ})$, and multiplying into the column vector $(\mathbf{Z}^{\text{T}}\mathbf{WY})$, thus

$$\mathbf{x} = (\mathbf{Z}^{\text{T}}\mathbf{WZ})^{-1}(\mathbf{Z}^{\text{T}}\mathbf{WY}) \tag{22}$$

It can be seen from equations (14) and (15) that the values of P and Q depend on the magnitude of β_1 and β_2. However, the weights are not very sensitive to variations in these constants, so that approximate estimates may be used, *e.g.*, $\beta_1 = 1/[\text{A}]$ at $\bar{n} = 0{\cdot}5$, and $k_2 = 1/[\text{A}]$ at $\bar{n} = 1{\cdot}5$. Successive approximations may then be used until constant values of β_1 and k_2 are obtained, but it is rarely necessary to apply more than one iteration.

The standard deviations of the constants, β_j, are given by

$$\sigma_{\beta j}^2 = (\mathbf{Z}^{\text{T}}\mathbf{WZ})_{jj}^{-1}S_{\text{min.}}/(I - 2) \tag{23}$$

where $(\mathbf{Z}^{\text{T}}\mathbf{WZ})_{jj}^{-1}$ is the jth diagonal element of the inverse matrix $(\mathbf{Z}^{\text{T}}\mathbf{WZ})^{-1}$, $S_{\text{min.}}$ is the value of S calculated by substituting the values of β_1 and β_2 obtained from equation (22) into equation (5), and $(I - 2)$ corresponds to the number of degrees of freedom of the system.[7]

The second method for calculating the constants makes use of the symmetrical nature of the formation curve. We can fix the value of β_2 so that the theoretical and experimental values are the same at the central point where $\bar{n} = 1$, and then rotate the formation curve about it by varying β_1 until the " best " fit to the experimental data is obtained.

Rearranging equation (3) gives

$$\bar{n} = (\beta_1[\text{A}] + 2\beta_2[\text{A}]^2)/(1 + \beta_1[\text{A}] + \beta_1[\text{A}]^2) \tag{24}$$

A calculated $\bar{n}$ value may be obtained from equation (24) using the experimental [A] value and given values of β_1 and β_2. If this calculated $\bar{n}$ value be given by $\bar{n}_c$, then there will exist a residual R_i given by

$$R_i = (\bar{n} - \bar{n}_c) \tag{25}$$

[7] Dumond and Cohen, *Rev. Mod. Phys.*, 1953, **25**, 691.

where $\bar{n}$ is the experimental $\bar{n}$ value given by equation (10). Hence,

$$R_i = \left[\frac{C_a - [A]}{C_m} - \frac{\beta_1[A] + 2\beta_1[A]^2}{1 + \beta_1[A] + \beta_1[A]^2}\right]$$

If β_2 is fixed, then we require a value of β_1 such that $\sum_{i=1}^{I} w_i R_i^2$ shall be a minimum. In order to determine the weighting factors w_i, it is necessary to determine the distribution of the residuals R_i from a consideration of the function defining R_i and the errors associated with the evaluation of R_i, *viz.*, errors in E and C_a. In a similar manner as previously, it can be shown that

$$\delta R = \left[-\frac{1}{C_m} - \frac{\beta_1[A] + \beta_1\beta_2[A]^3 + 4\beta_2[A]^2}{(1 + \beta_1[A] + \beta_2[A]^2)^2}\right]\left[\frac{2{\cdot}303[A]}{k}\right]\delta E + \left[\frac{1}{C_m} + \left\{-\frac{1}{C_m} - \frac{\beta_1[A] + \beta_1\beta_2[A]^3 + 4\beta_2[A]^2}{(1 + \beta_1[A] + \beta_2[A]^2)^2}\right\}\frac{[A]}{C_a}\right]\delta C_a$$

from which var R can be determined and hence the weighting factors. The computer was programmed to evaluate R_i and w_i, and hence $\sum_{i=1}^{I} w_i R_i^2$ for a particular value of β_1, and to perform this repeatedly for different values of β_1, keeping β_2 constant. This was continued until a minimum was found for the weighted sum of the squares of the residuals. The value of β_1 which gave this minimum was then noted.

Results

Titrations of 50 ml. of 0·500M-4-aminobutan-1-ol nitrate with 0·500M-sodium hydroxide in 0·500M-potassium nitrate, in the presence of (a) 15 ml. of 0·500M-KNO_3 or (b) 15 ml. of 0·050M-$AgNO_3$ in 0·450M-KNO_3 at 20°

NaOH (ml.)	ΔpH	pA	$\bar{n}$	log (w_1)	log (w_2)	NaOH (ml.)	ΔpH	pA	$\bar{n}$	log (w_1)	log (w_2)
0·10	1·541	4·656	0·065	6·731	4·227	1·50	1·718	3·666	0·980	4·201	4·410
0·20	1·615	4·429	0·130	6·181	4·498	1·60	1·707	3·628	1·045	4·123	4·428
0·30	1·654	4·293	0·195	5·810	4·840	1·70	1·695	3·590	1·110	4·047	4·455
0·40	1·683	4·198	0·261	5·547	5·294	1·80	1·678	3·549	1·174	3·960	4·491
0·50	1·702	4·121	0·327	5·331	5·121	1·90	1·668	3·516	1·238	3·912	4·538
0·60	1·718	4·058	0·392	5·160	4·896	2·00	1·651	3·477	1·303	3·846	4·598
0·70	1·726	4·000	0·458	5·003	4·705	2·20	1·602	3·389	1·429	3·688	4·666
0·80	1·733	3·950	0·523	4·870	4·570	2·40	1·542	3·292	1·553	3·527	4·753
0·90	1·740	3·906	0·589	4·762	4·475	2·60	1·456	3·173	1·671	3·306	4·865
1·00	1·740	3·861	0·654	4·647	4·444	2·80	1·338	3·023	1·779	2·981	4·997
1·10	1·738	3·818	0·719	4·543	4·424	3·00	1·167	2·824	1·862	2·361	5·172
1·20	1·737	3·780	0·785	4·453	4·407	3·60	0·729	2·310	1·950	−0·041	5·403
1·30	1·731	3·740	0·850	4·362	4·400	4·00	0·581	2·119	1·965	−1·109	5·746
1·40	1·723	3·700	0·915	4·269	4·401						

Recorded in the Table are a representative set of $\bar{n}$ and pA values (where pA = − log [A]) obtained for a silver complex system. Also recorded is the volume of sodium hydroxide added and the ΔpH values, where ΔpH = $(E_1 - E_2)/k$, and the weighting factors for both methods recorded as their logarithmic values (w_1 = weights for first method, and w_2 for the second method). The computed values for log β_1 and log β_2 are 3·41 ± 0·05 and 7·30 ± 0·02, respectively, the ranges being given by the 95% significance level, *i.e.*, 1·96 $\sigma_{\log \beta j}$. The value for log β_1 from the second method, using log β_2 = 7·30, was 3·41.

Discussion

In the first method of calculation the values of w_i are small for those measurements towards the end of the titration. This does not indicate, however, that these experimental results are less reliable than those obtained earlier. It does indicate that the latter residuals, as calculated from equation (4), are more sensitive than the earlier ones to the effect of the experimental errors, which we have assumed lie within ranges of constant size throughout the complete experiment. One could quite easily rearrange equation (3) into

some other form with different residuals such that the distribution of weights would be quite different.

In general it is unnecessary to plot the results graphically, since any large experimental errors are exposed by a large standard deviation in the results. However, if large standard deviations are observed it is a simple matter to plot $\bar{n}$ against pA values since these have been calculated by the computer, and so determine the exact nature of these errors. Often however a print-out of the residuals will indicate this equally well.

In the second method the value of β_2 can be obtained graphically or by use of an interpolation technique. Because of the insensitivity of the value to experimental errors compared with the sensitivity of k_1 and k_2 values, this will give a reliable value.

The agreement between the β_1 values obtained using either method is good, supporting the validity of the techniques.

We should like to thank the Chemical Society for a grant towards the cost of the chemicals and glass cells.

NORTHAMPTON COLLEGE, ST. JOHN STREET, LONDON E.C.1. [*Received, July* 31*st*, 1964.]