Some Graphical Methods for Determining Equilibrium Constants. IV. On Methods for Three-Variable Data $w(x, y)$

LARS GUNNAR SILLÉN

Department of Inorganic Chemistry, Royal Institute of Technology, Stockholm 70, Sweden

The treatment in part II [1] is extended to indicate graphical methods of determining a number of unknown constants (parameters) p_1, p_2 ... from a set of three-variable data (w, x, y) when the function $w(x, y, p_1, p_2 ...)$ is known.

Sometimes, by transformation of variables a true reduction to a two-variable problem is possible. In other cases the function can be transformed to $w(x, f_1(y), f_2(y) ...)$. By considering sections at constant values of y, the parameters may then be found by two successive two-variable treatments. The method is illustrated by application to extraction equilibria of complexes of the general formula $B(HA)_p(H)_{-n}$.

When a reduction to two variables is not possible, methods of "surface-fitting" may be applied, if the parameters can be eliminated by normalizing the variables. For instance, with two variables and two parameters one may fit a calculated projection map $Y(X)_w$ to an experimental projection map $y(x)_w$. Convenient normalized variables are given for a system with species B, AB, and A_PB_Q, as an example.

Applications will be given in following papers on hydrolysis equilibria.

In part II of this series [1] various ways were discussed of treating two-variable experimental data, *i. e.* data that can be represented by a planar diagram $y(x)$, where y and x are functions that can be obtained directly from the experimental data.

More complicated cases arise if the experimental points are sets of three variables. For example, in the study of polynuclear complexes formed between two reagents A and B, the average number of A bound per B, Z, is often measured together with a, the concentration of free A, and B, the total concentration of B, so that each experimental point is a set (Z, a, B). The data can thus be represented by a curved surface in three-dimensional space, $Z(a, B)$.

We shall consider a general case where the variables determined experimentally are w, x, and y (we shall not use the symbol z to avoid confusion with Z), and the constants to be determined (parameters) are denoted by p_1, p_2 ... . We know in advance the function

$$w = w(x, y, p_1, p_2 ...) \qquad (1)$$

relating the experimental variables with the parameters. As in part II, w, x, and y need not be the quantities directly determined, although it should be possible to calculate them from the data; nor need the p_i be the constants ultimately required. The most suitable functions w, x, y and parameters p_i for each case may not be immediately obvious.

SYMBOLS

a, b	concentrations of free A and B
B	total concentration of B
c	concentration of free HA (5)
f_i, f_n	functions (2), (8)
h	concentration of H^+ (5)
j	constant (16)
K_{pn}, K_{11}, K_{PQ}	equilibrium constants (3), (12)
L_{pN}	distribution constants (4)
N	charge of metal ion B^{N+}
n, p, P, Q	integer numbers (3), (12)
p_1, p_2, p_2	unknown constants (parameters) to be determined
q	over-all distribution ratio of B between organic and aqueous phase
u, v	auxiliary functions (15), (16)
w, x, y	variables determined experimentally
W, X, Y	normalized values for w, x, y
x_0, y_0	values for x and y coinciding with $X = 0$, $Y = 0$ at best fit (18, 19)
Z	average number of A bound per B
subscript)$_x$	indicates that x is constant

REDUCTION TO A TWO-VARIABLE PROBLEM

True reduction

In some cases an analysis of the data shows that one may definitely reduce the problem to one with only two variables. Three simple examples will be given.

In the formation of mononuclear complexes A_nB by a central group B and a ligand A, the experimental variables may be, for instance, (Z, a, B) or (a, b, B). Now, if only mononuclear complexes are formed, it is easily shown (N. Bjerrum [2]) that both Z and bB^{-1} are functions of a only. (Also for "homonuclear" complexes [3] Z (but not bB^{-1}) is a function of a only.) So the total concentration B is eliminated, and we are left with the two-variable problem of finding the equilibrium constants from the functions $Z(a)$ or $bB^{-1}(a)$. For refined mathematical treatment, further transformations of the variables are usually necessary [1,4].

If, again, all complexes formed are polynuclear and can be written in the "core + links" form, $B(A_tB)_n$, then Z and bB^{-1} can be written [3,5] as functions of the single variable $x = t \log a + \log B$.

Another example is the distribution of a metal ion B, between an aqueous solution and an organic solvent, in the presence of an acid HA, forming complexes like A_nB. We shall denote by q the ratio between the total concentrations of B in the organic and aqueous phases. One might first consider this as a problem of four variables, say q (the distribution ratio), B (the total concentration of B in the aqueous phase), and the concentrations of free H^+ and HA in the aqueous phase. If only mononuclear complexes are formed, the total concentration B should, however, have no influence on $q([H^+], [HA])$. If, moreover, all complexes in the aqueous phase are of the form A_nB and only the uncharged A_NB is extracted, then q should be a function of the concentration of free A only, $a = [A] = [HA][H^+]^{-1}K_a$, so that we really have the two-variable problem [6,7] of studying $q(a)$. If, on the other hand, not A_NB but a single complex $A_NB(HA)_{p-N}$ is extracted, then $q[HA]^{N-p}$ is a function of a only.

For each of these cases, one may apply the methods for two-variable data given in part II [1].

Reduction to two-variable sections

Sometimes it proves impossible by any transformation, truly to reduce three-variable data (w, x, y) to two variables. However one may cut sections at constant values of y (say) and treat each of these as a two-variable problem. Suppose that we can transform equation (1) to the form

$$w = w(x, f_1(y), f_2(y), \ldots) \tag{2}$$

where the functions $f_i(y)$ contain the variable y, and usually a number of unknown constants. If a section $w(x)_y$ at a constant value of y is considered, the functions f_i are constants, which can be found by some method for two-variable problems given in part II [1]. Thus one may determine the values of f_1, f_2 etc. for a number of values of y. Finally, the functions $f_i(y)$ are treated separately; each of these functions contains one or more parameters, which can now be determined by some two-variable method.

As an example, we shall consider a case previously treated somewhat differently by Rydberg [8]. A metal ion B is distributed between an aqueous solution and an organic solvent, in the presence of an acid HA, which may form complexes with B. All those complexes present may be written in the form $B(HA)_p(H)_{-n}$, since their content of organic solvent, water or ions from the ionic medium cannot be determined by equilibrium measurements. (We replace Rydberg's symbols x and y by p and n to avoid confusion in the present paper.)

If the charge of the ion B is N^+, it is reasonable to assume that only uncharged complexes, with $n = N$, are extracted by the organic solvent. We may then write, assuming only mononuclear complexes:

$$[B(HA)_p(H)_{-n}]_{aq} = [B][HA]^p[H]^{-n}K_{pn} = bc^ph^{-n}K_{pn} \tag{3}$$

$$[B(HA)_p(H)_{-N}]_{org} = [B][HA]^p[H]^{-N}K_{pN}L_{pN} = bc^ph^{-N}K_{pN}L_{pN} \tag{4}$$

where K_{pn} are the formation constants, and L_{pN} the distribution constants of the complexes, and

$$c = [HA] \text{ and } h = [H^+] \tag{5}$$

The over-all distribution ratio q is given by

$$q = \frac{\Sigma[B(HA)_p(H)_{-N}]_{org}}{\Sigma\Sigma[B(HA)_p(H)_{-n}]_{aq}} = \frac{\Sigma c^p h^{-N} K_{pN} L_{pN}}{\Sigma\Sigma c^p h^{-n} K_{pn}} \tag{6}$$

Thus $q(c, h)$ is a function of c and h. We may, however, obtain two-variable sections by studying instead the variable

$$q^{-1}h^{-N} = \frac{\Sigma\Sigma h^{-n}c^p K_{pn}}{\Sigma c^p K_{pN} L_{pN}} = \Sigma h^{-n} f_n(c) \tag{7}$$

$$f_n(c) = \frac{\Sigma c^p K_{pn}}{\Sigma c^p K_{pN} L_{pN}} \; ; \; f_0(c) = \frac{\Sigma c^p K_{p0}}{\Sigma c^p K_{pN} L_{pN}} \tag{8}$$

If c is kept constant, (7) gives a polynomial in h^{-1}. Each section can thus be used to find the values for the functions $f_n(c)$ at the value of c chosen. Rydberg [8] has used a single "sweep" [1], with successive linear extrapolations. However, it may be preferable to study the function $w(x)_c$

$$w = \log(q^{-1}h^{-N}) = \log \Sigma h^{-n} f_n; \; x = \log h \tag{9}$$

and find the best values of $f_n(c)$ by use of the curve-fitting methods for polynomials, described in part II [1].

The treatment of the data $f_n(c)$ so obtained, will depend on the accuracy of the data and the range of concentrations studied.

It seems a reasonable assumption [8] that complexes $B(HA)_p^{N+}$ can be neglected, so that $K_{p0} = 0$ for $p \neq 0$. By definition, $K_{00} = 1$. Then, as indicated by Rydberg (whose symbols are slightly different), it is profitable to study the functions

$$f_0^{-1}(c) = \Sigma K_{pN} L_{pN} c^p \tag{10}$$

$$f_n(c) f_0^{-1}(c) = \Sigma K_{pn} c^p \tag{11}$$

To find the individual * constants K_{pn} and L_{pN} one may use the curve-fitting methods for polynomials, for instance preparing plots of $\log f_0^{-1}$ (or, say, $\log(c^{-N} f_0^{-1})$), and $\log f_n f_0^{-1}$ (or, say, $\log(c^{-n} f_n f_0^{-1})$), as functions of $\log c$, and comparing them with families of normalized curves for polynomials [1]. Such methods seem preferable to linear plots, for the same reasons as given in part II [1]. Hök-Bernström [9] has recently applied this method to extraction data for UO_2^{2+}.

As a rule, partition data will allow an independent determination of only a few equilibrium constants, and it may be wise to start by searching for the *smallest set* of complexes $B(HA)_p(H)_{-n}$ needed to explain the data. With the curve-fitting methods given, one may compare the same graphical representa-

---

* If $[A_pBH_{p-N}]$ is important in the organic phase but negligible in the aqueous solution, only the product $K_{pN} L_{pN}$ may be obtained.

tion of the data with curves for a number of possible combinations. One may then judge — more directly than with linear plots — which combinations can be excluded, which are just compatible with the data, and which will give the best fit.

METHODS WITHOUT REDUCTION

Elimination

If one is not successful in reducing a three-variable problem to two-variable plots, the problem becomes slightly more complicated. A survey of the possible methods is given in Table 1, which is analogous to Table 1 of part II [1].

With only one parameter, p_1, a value for this parameter is obtained by inserting the coordinates of each experimental point (w, x, y) into eqn. (1). To test for the spread and any trend in the values of p_1, one would need a three-dimensional diagram (x, y, p_1), which could be given in the form of curves $p_1(x)_y$. If p_1 is really a constant, then these curves will all coincide in a single horizontal line.

Similarly, with two or three parameters, the graphs will not be more complicated than for two-variable data with two or three parameters, provided one does not want to test for a trend. With two parameters, one may for instance use curves $p_2(p_1)_{x,y}$.

Normalized graphs

With three variables, more serious complications occur in using normalized graphs, as indicated in Table 1. For instance, with two parameters, and two normalized variables, one can no longer use a single curve $Y(X)$ to compare with the data $y(x)$, but must instead compare a curved surface in space $w(x, y)$ with another, $w(X, Y)$. In practice this is done by preparing a *projec-*

Table 1. Methods for finding unknown constants p_i *from experimental data* w(x, y).

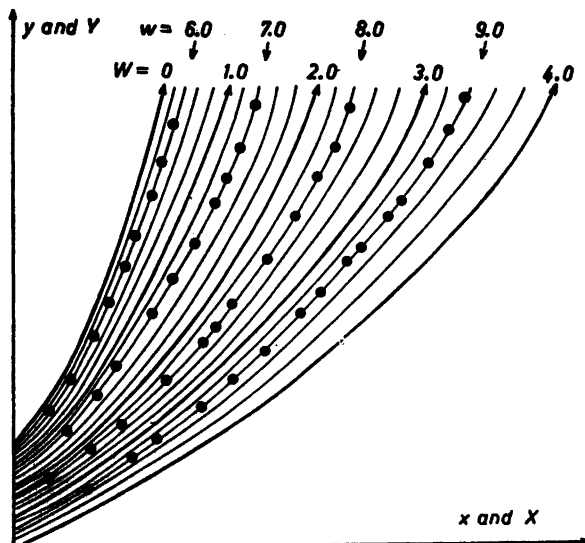| Number of		Compare (w, x, y) graph with calculated graph	Eliminate w to give graph	Two-dimensional projections, e. g.
parameters	normalized variables			
1 (p_1)	$\begin{cases}1 & (x) \\ 0 &\end{cases}$	w, X, y	x, y, p_1	$y(X)_w$ and $y(x)_w$ $p_1 \ (x)_y$
2(p_1, p_2)	$\begin{cases}2 & (x, y) \\ 1 & (x) \\ 0 &\end{cases}$	w, X, Y w, X, y, p_2	x, y, p_1, p_2	$Y(X)_w$ and $y(x)_w$ — intersection of curves $p_2(p_1)_{x,y}$
3(p_1, p_2, p_3)	$\begin{cases}3 & (w, x, y) \\ 2 & (x, y) \\ 1 & (x) \\ 0 &\end{cases}$	W, X, Y w, X, Y, p_3 w, X, y, p_2, p_3	x, y, p_1, p_2, p_3	$Y(X)_W$ and $y(x)_w$ — — —

Fig. 1. "Surface-fitting" with three parameters. Curves = calculated projection map Y(X)w for W = 0, 1, 2, 3, 4 (thick curves) and intermediate values, with a spacing of 0.2 units (thin curves). Circles: experimental points y(x)w for w = 6.0, 7.0, 8.0, and 9.0, in the position of best fit. The parameter p_3 = W — w in this case is around —5.6. — From the differences between normalized and experimental variables one could also find the two other parameters, p_1 = X — x and p_3 = Y — y. These coordinate scales have not been marked out in the figure.

tion map with $Y(X)_w$ curves for a number of values of w, and comparing it with the experimental data, given as curves $y(x)_w$ with the same values for w. The additional work involved should be a strong incentive to reduce the problem to a two-variable one. However, it will be shown in subsequent papers that if necessary such "surface-fitting" operations may be carried through successfully.

With three variables, a problem with one parameter and one normalized variable also requires the "surface-fitting" of a projection map $y(x)_w$ of the data and a normalized one $y(X)_w$ with calculated curves.

With three parameters and three normalized variables, one should fit a normalized surface $W(X, Y)$ to the plot $w(x, y)$ of the experimental data. This could possibly be carried out on three-dimensional models. When one tries to find the best fit of two-dimensional projections $Y(X)_W$ for a number of values of, W and $y(x)_w$ for a number of values of w, a difficulty arises: in the best fit, the experimental points may in general fall somewhere midway between the calculated curves. It might be practical to use a constant spacing of w in the experimental plot, and draw a set of thicker calculated curves with the same spacing in W; between these one may insert a series of thinner calcu-lated curves for intermediate values of W (Fig. 1).

The other types of systems in Table 1 lead us to consider four-dimensional or five-dimensional super-surfaces, so that successive approximations and "sweeping" must be resorted to [1].

Application to systems with two equilibrium constants

As a simple example, let us assume that the reagents A and B form the two complexes AB and $A_P B_Q$, where P and Q are known, and that we want to determine the formation constants K_{11} and K_{PQ} from experimental data $Z(a, B)$.

From the definitions of B and Z, and the law of mass action we have

$$B = [B] + [AB] + Q[A_P B_Q] = b + abK_{11} + Qa^P b^Q K_{PQ} \qquad (12)$$

$$BZ = [AB] + P[A_P B_Q] = abK_{11} + Pa^P b^Q K_{PQ} \qquad (13)$$

Now, it is convenient to use the experimental variables

$$w = Z; \quad x = \log a; \quad y = \log B \qquad (14)$$

and the normalized variables

$$X = x + p_1; \quad Y = y + p_2 \qquad (15)$$

Let us first introduce the auxiliary variables

$$u = 10^X = a \cdot 10^p; \quad v = bj \qquad (16)$$

where j is a constant. We wish to give such values to p_1, p_2 and j that all equilibrium constants are removed from (12) and (13), thus

$$Y = \log(v + uv + Qu^P v^Q) \qquad (12a)$$

$$Y + \log w = \log(uv + Pu^P v^Q) \qquad (13a)$$

Comparison with equations (12) to (16) gives us

$$p_1 = \log K_{11}; \quad p_2 = \frac{1}{Q-1} (\log K_{PQ} - P \log K_{11}) = \log j \qquad (17)$$

The procedure is then to construct a map $Y(X)_w$ for a series of values of w, using (12a), (16), and (13a), and to plot the experimental data $y(x)_w$, i. e. $\log B(\log a)_Z$, for the same values of w. The plots are moved parallel to the coordinate axes until the best fit is found. One may then read the coordinates of the point (x_0, y_0) coinciding with $(X = 0, Y = 0)$, which give us

$$x_0 = -p_1 = -\log K_{11} \qquad (18)$$

$$y_0 = -p_2 = -\frac{1}{Q-1} (\log K_{PQ} - P \log K_{11}) \qquad (19)$$

The curve $Y(X)_w$ may be calculated in two ways. One may plot curves $w(v)_u$ and $Y(v)_u$ inserting a series of values for u and v into (12a) and (13a). Then, for each value of w considered one may take sets of corresponding values $v(u)_w$, and insert them into the

Y plot which gives $Y(u)_w$ or $Y(X)_w$ since $X = \log u$ (16). One may also eliminate v from (12a) and (13a), which gives

$$(Q-1)Y = (Q-1)\log(P+u(P-Q)) + \log(w+uw-u) - Q\log(P-Qw)-P\log u \qquad (20)$$

Thus for each set of values of w and $X = \log u$ one may calculate Y directly.

If experimental data (a, b, B) are available instead, we may use the same experimental, normalized and auxiliary variables as in (14, 15, 16, 17) with the exception of w:

$$x = \log a; \quad X = x + p_1; \quad p_1 = \log K_{11} \qquad (21)$$

$$y = \log B; \quad Y = y + p_2; \quad p_2 = \frac{1}{Q-1}(\log K_{PQ} - P\log K_{11}) \qquad (22)$$

$$u = 10^X = a \cdot 10^{p_1}; \quad v = bj = b \cdot 10^{p_2} \qquad (23)$$

$$Y = \log(v + uv + Qu^Pv^Q) \qquad (12a)$$

As the third experimental variable it is convenient to use the dimensionless quantity $\log(Bb^{-1})$ denoted by η in previous papers [3,5]. From (22), (23), and (12a)

$$w = \log\frac{B}{b} = y + p_2 - \log v = Y - \log v \qquad (24)$$

$$w = \log(1 + u + Qu^Pv^{Q-1}) \qquad (25)$$

The normalized curve $Y(X)_w$ may be constructed by inserting a series of values of $X = \log u$ and w into (25); this gives v immediately, and Y is obtained from (24): $Y = w + \log v$.

The projection map $y(x)_w$ of the experimental data is then moved parallel to the axes until the best fit with the normalized projection map $Y(X)_w$ is found, which gives p_1 and p_2, and thus K_{11} and K_{PQ}.

One may proceed similarily for any set of equilibria with only two equilibrium constants, thus either free B and two complexes, or three complexes and a negligible concentration of free B. With four species containing B, thus three equilibrium constants, one may draw a series of projection maps, in each of which the ratio between two equilibrium constants (or some other suitable expression) is kept constant.

An advantage of the choice (14; 21, 22, 24) of log a and log B as planar coordinates and Z or $\log\dfrac{B}{b}$ as "height" in the experimental maps is that several alternative explanations may be compared directly on the same plot of experimental data.

Examples will be given in subsequent papers on hydrolysis equilibria.

REFERENCES

1. Sillén, L. G. *Acta Chem. Scand.* **10** (1956) 186 (part II).
2. Bjerrum, N. *Kgl. Danske Videnskab. Selskab. Skrifter, Naturvidenskab. math. Afdel.* (7) **12** (1915) 206.
3. Sillén, L. G. *Acta Chem. Scand.* **8** (1954) 299.
4. Rossotti, F. J. C. and Rossotti, H. S. *Acta Chem. Scand.* **9** (1955) 1166 (part I).
5. Sillén, L. G. *Acta Chem. Scand.* **8** (1954) 318.
6. Rydberg, J. *Acta Chem. Scand.* **4** (1950) 1503.
7. Dyrssen, D. and Sillén, L. G. *Acta Chem. Scand.* **7** (1953) 663.
8. Rydberg, J. *Arkiv Kemi* **8** (1955) 101, 113.
9. Hök-Bernström, B. *Acta Chem. Scand.* **10** (1956) 163.