Statistical Methods for the Computation of Stability Constants. Part I. Straight-line Fitting of Points with Correlated Errors

By **G. L. Cumming,** Department of Geology and Mineralogy, University of Oxford, Parks Road, Oxford OX1 3PR
J. S. Rollett, Computing Laboratory, University of Oxford, Parks Road, Oxford OX1 3PL
F. J. C. Rossotti * and **R. J. Whewell,** Inorganic Chemistry Laboratory, South Parks Road, Oxford OX1 3QR

A rigorous method is described for fitting a straight line to a set of independent points subject to correlated errors. The algorithm is briefly stated, and includes expressions for the estimated standard deviations of the resultant parameters. One application of this general method is detailed: the computation of stability constants from precise potentiometric data for systems of dibasic acids HA, H_2A, and mononuclear metal-ion complexes BA, BA_2. Some other applications are indicated.

A VARIETY of methods, both graphical [1,2] and computerised,[3] have been described for obtaining values of stability constants from the experimentally determined formation curve.[1] Rigorous statistics have, however, seldom been used to obtain the ' best ' values of the constants and to assess realistic limits of error. The procedure described here, for systems comprising two complexes, is rigorous and yet simple to use.

For a system containing only the species B, A, BA, and BA_2 the average value of n is given by equation (1)

$$\bar{n} = \frac{[BA] + 2[BA_2]}{b + [BA] + [BA_2]} = \frac{A - a}{B} \qquad (1)$$

where the symbolism follows that in ref. 1, and charges are omitted. The stepwise constants [equations (2)] may

---

[1] F. J. C. Rossotti and H. S. Rossotti, ' The Determination of Stability Constants,' McGraw-Hill, New York, 1961; Mir, Moscow, 1965.

[2] F. J. C. Rossotti and H. S. Rossotti, *J. Phys. Chem.*, 1959, **63**, 1041.

[3] F. J. C. Rossotti, H. S. Rossotti, and R. J. Whewell, *J. Inorg. Nuclear Chem.*, 1971, **33**, 2051.

refer either to the successive additions of, *e.g.*, ligands A to a metal ion B, or to stepwise addition of protons, A, to a base, B. Equations (1) and (2) can be combined

$$K_1 = [BA]b^{-1}a^{-1} \qquad (2a)$$

$$K_2 = [BA_2][BA]^{-1}a^{-1} \qquad (2b)$$

to give a number of different (but dependent) linear transforms, *e.g.* (3)—(5). An unweighted least-squares

$$\frac{(\bar{n}-1)a}{\bar{n}} = \frac{K_2(2-\bar{n})a^2}{\bar{n}} - \frac{1}{K_1} \qquad (3)$$

$$\frac{(\bar{n}-1)}{(2-\bar{n})a} = K_2 - \frac{\bar{n}}{(2-\bar{n})a^2}\cdot\frac{1}{K_1} \qquad (4)$$

$$\frac{\bar{n}}{(1-\bar{n})a} = K_1\,K_2\,\frac{(2-\bar{n})a}{(\bar{n}-1)} - K_1 \qquad (5)$$

treatment of equation (5) has been used[4] to determine values of K_1 and K_2. However, treatments based upon equations (3) and (4) are preferable, as the infinity at $\bar{n} = 1$ in equation (5) precludes the use of the data in the range $0{\cdot}95 < \bar{n} < 1{\cdot}05$. Equations (3) and (4) have been treated by an unweighted least-squares procedure[5] and equivalent general linear least-squares methods have been used.[3] This type of treatment is only appropriate when one co-ordinate of each point is exact, whereas there is a constant standard deviation for the other co-ordinate. In practice, the linear plots involve sets of points for which both co-ordinates have standard deviations, which vary from point to point by several orders of magnitude, so that an unweighted least-squares treatment will give incorrect results. In order to deal with variations of standard deviation, weighting factors[6-8] are introduced, *i.e.*, equation (6)

$$W_{X_i} = 1/\sigma^2(X_i) \qquad (6)$$

where W_{X_i} is the weighting factor appropriate to the co-ordinate X_i and $\sigma^2(X_i)$ is the estimated variance of X_i. This results in a distribution of unit standard deviation, which is necessary for the application of the chi-squared test.

Regression of y upon x can thus be made appropriate to data where $\sigma(y)$ is variable and x is exact, but the method is invalid when both co-ordinates are likely to be in error. Only a change in the function to be minimised will overcome this difficulty.

A rigorous treatment must also take account of the correlation of errors in the co-ordinates, which are both functions of errors in $\bar{n}$ and a, *cf.* equations (3) and (4).

Input for the programme FAJAF45 for computing stability constants from potentiometric data are volumes and concentrations of reagents, and cell potentials; the formation function $\bar{n}(a)$ is calculated by use of equation (1) and hence co-ordinates X_i and Y_i for the linear plots based on equations (3) and (4). Random errors are estimated as standard deviations for each measurement made, and these errors combined to give estimates of the variances $\sigma^2(\bar{n})$, $\sigma^2(a)$, and their co-variance $\mathrm{cov}(\bar{n}, a)$. (Covariance is the expected value of the product of the deviations of two variates from their means.) Further combination of these estimates gives $\sigma^2(X_i)$,$\sigma^2(Y_i)$, and $\mathrm{cov}(X_i,Y_i)$ for each linear plot. This lengthy procedure will not be detailed, but an illustrative example is given in Appendix 1.

Systematic errors, *e.g.*, in analyses of stock solution, affect the accuracy, but not the precision, of the calculated stability constants. Their effects can be assessed separately,[1] but cannot be incorporated into the calculation of weighting factors.

The Line-fitting Procedure.—Introduction. The following symbols are used: p, the number of experimental points; X_i,Y_i, the co-ordinates of the ith experimental point; x_i,y_i, the co-ordinates of a point on the line fitted, closest to the ith experimental point in the sense of yielding the minimum value for the ith term on the right-hand side of equation (12) below; and r, s, t, parameters of the line $rx + sy + t = 0$ fitted to

$$|V_i| = \sigma^2(X_i)\sigma^2(Y_i) - \mathrm{cov}(X_i,Y_i)^2 \qquad (7)$$

$$W_{X_i} = \sigma^2(Y_i)|V_i|^{-1} \qquad (8)$$

$$W_{Y_i} = \sigma^2(X_i)|V_i|^{-1} \qquad (9)$$

$$W_{XY_i} = -\mathrm{cov}(X_i,Y_i)|V_i|^{-1} \qquad (10)$$

the points. From the variances input we can calculate equations (7)—(10). The straight line (11) is obtained

$$rx + sy + t = 0 \qquad (11)$$

so that equation (12) is minimised subject to the following subsidiary conditions (13) and (14).

$$S = \sum_{i=1}^{p}\{W_{X_i}(x_i - X_i)^2 + W_{Y_i}(y_i - Y_i)^2 + 2W_{XY_i}(x_i - X_i)(y_i - Y_i)\} \qquad (12)$$

$$r^2 + s^2 = 1 \qquad (13)$$

$$rx_i + sy_i + t = 0 \text{ for all } i \qquad (14)$$

Equation (14), the redundant form for equation (11) is used so that neither r nor s becomes infinite for any slope. The function S is proportional to the logarithm of the product of the probability densities at (x_i,y_i), given bivariate normal distributions of error centred on (X_i,Y_i) with variances defined by $\sigma^2(X_i)$, $\sigma^2(Y_i)$, and $\mathrm{cov}(X_i,Y_i)$. The line obtained is therefore that which maximises the likelihood that the given points (X_i,Y_i) deviate from the collinear points (x_i,y_i) by experimental errors alone, provided that these errors are correctly estimated by $\sigma^2(X_i)$, $\sigma^2(Y_i)$, and $\mathrm{cov}(X_i,Y_i)$, and their distributions are normal.

Algorithm. We start with un-normalised values of r and s, $r = s = 1$ (*i.e.*, $r^2 + s^2 = 1$ is not satisfied; this ensures that the test for convergence will fail on the first iteration) and compute a weight (15) for each

$$w_i = 1/[r^2\sigma^2(X_i) + 2rs\,.\,\mathrm{cov}(X_iY_i) + s^2\sigma^2(Y_i)] \qquad (15)$$

---

[4] H. Irving and H. S. Rossotti, *J. Chem. Soc.*, 1953, 3397.

[5] L. D. Pettit, A. Royston, C. Sherrington, and R. J. Whewell, *J. Chem. Soc.* (*B*), 1968, 588.

[6] E. Whittaker and G. Robinson, 'The Calculus of Observations,' London, Blackie, 1924.

[7] F. S. Acton, 'Analysis of Straight-line Data,' Wiley, New York, 1959.

[8] W. C. Hamilton, 'Statistics in Physical Science,' Ronald, New York, 1964.

point, the centroid (16) of the points, and co-ordinates

$$\bar{X} = \sum_i (w_i X_i) / \sum_i (w_i) \qquad (16a)$$

$$\bar{Y} = \sum_i (w_i Y_i) / \sum_i (w_i) \qquad (16b)$$

referred to the centroid, given by equations (17). Except for the first iteration (where $\Delta_i = 0$), we define (18).

$$X_i' = X_i - \bar{X} \qquad (17a)$$

$$Y_i' = Y_i - \bar{Y} \qquad (17b)$$

$$\Delta_i = rX_i' + sY_i' \qquad (18)$$

Equations (19) and (20) (derived in Appendix 2) are

$$\sum_{i=1}^{p} [w_i X_i'^2 - \sigma^2(X_i) w_i^2 \Delta_i^2] r + \sum_{i=1}^{p} [w_i X_i' Y_i' - \mathrm{cov}(X_i, Y_i) w_i^2 \Delta_i^2] s = 0 \qquad (19)$$

$$\sum_{i=1}^{p} [w_i X_i' Y_i' - \mathrm{cov}(X_i, Y_i) w_i^2 \Delta_i^2] r + \sum_{i=1}^{p} [w_i Y_i'^2 - \sigma^2(Y_i) w_i^2 \Delta_i^2] s = 0 \qquad (20)$$

obtained for new values of r and s. There is no exact solution of these equations except $r = s = 0$ unless the determinant of their matrix is zero. We therefore take r and s to be the elements of the latent vector corresponding to the smaller latent root of the matrix, normalised so that $r^2 + s^2 = 1$. If the sums of squares of differences between the old and new values of r and s is less than 10^{-12}, calculation ceases (this cannot happen on the first iteration). Otherwise the new values of r and s are substituted for the old and the calculation repeated from equation (15). In all experimental tests so far four iterations have been sufficient to give convergence, and the smaller latent root of the matrix of equations (19) and (20) has been $<10^{-6}$.

These operations give the position of the centroid $(\bar{X},\bar{Y})$ through which the line passes, and its slope $-r/s$. If the line is required in the form of equation (11), $t = -(r\bar{X} + S\bar{Y})$ can be computed.

The co-ordinates of the collinear points (x_i,y_i) are obtained from equations (21) and (22).

$$x_i = X_i - w_i \Delta_i [r\sigma^2(X_i) + s \,.\, \mathrm{cov}(X_i, Y_i)] \qquad (21)$$

$$y_i = Y_i - w_i \Delta_i [r \,.\, \mathrm{cov}(X_i, Y_i) + s\sigma^2(Y_i)] \qquad (22)$$

The algorithm has been extended to provide estimates of error for various quantities, but since we have been dealing with sloping lines this part of the work has not been made valid for vertical and horizontal cases. A test has been included to by-pass the error estimate calculation for exactly vertical and horizontal lines, so avoiding division by zero.

A difficulty is that the expressions for most quantities include the variances $\sigma^2(w_i)$ of the weights of the points, and these depend in turn on a number of the other variances. To avoid the complications which would result from an attempt to derive all the error estimates directly, a rapidly convergent iteration procedure has again been adopted. The effect of the $\sigma^2(w_i)$ has been small in our applications and in many cases it might perhaps be ignored.

We set, initially, the condition (23) and compute

$$\sigma^2(w_i) = 0 \quad i = 1, \ldots, p \qquad (23)$$

equations (24)—(32) with $\sigma^2(w_i)$ from (23) or later from equation (48).

$$\sigma^2(\bar{X}) = \sum_{i=1}^{p} [w_i^2 \sigma^2(X_i) + X_i'^2 \sigma^2(w_i)] / (\sum_{i=1}^{p} w_i)^2 \qquad (24)$$

$$\sigma^2(\bar{Y}) = \sum_{i=1}^{p} [w_i^2 \sigma^2(Y_i) + Y_i'^2 \sigma^2(w_i)] / (\sum_{i=1}^{p} w_i)^2 \qquad (25)$$

$$\mathrm{cov}(\bar{X},\bar{Y}) = \sum_{i=1}^{p} [w_i^2 \mathrm{cov}(X_i, Y_i) + X_i' Y_i' \sigma^2(w_i)] / (\sum_{i=1}^{p} w_i)^2 \qquad (26)$$

$$\mathrm{cov}(X_i,\bar{X}) = w_i \sigma^2(X_i) / \sum_{i=1}^{p} w_i \qquad (27)$$

$$\mathrm{cov}(Y_i,\bar{Y}) = w_i \sigma^2(Y_i) / \sum_{i=1}^{p} w_i \qquad (28)$$

$$\mathrm{cov}(X_i,\bar{Y}) = \mathrm{cov}(Y_i,\bar{X}) = w_i \mathrm{cov}(X_i, Y_i) / \sum_{i=1}^{p} w_i \qquad (29)$$

$$\sigma^2(X_i') = \sigma^2(X_i) + \sigma^2(\bar{X}) - 2\mathrm{cov}(X_i,\bar{X}) \qquad (30)$$

$$\sigma^2(Y_i') = \sigma^2(Y_i) + \sigma^2(\bar{Y}) - 2\mathrm{cov}(Y_i,\bar{Y}) \qquad (31)$$

$$(\mathrm{cov} X_i', Y_i') = \mathrm{cov}(X_i, Y_i) + \mathrm{cov}(\bar{X},\bar{Y}) - \mathrm{cov}(X_i,\bar{Y}) - \mathrm{cov}(Y_i,\bar{X}) \qquad (32)$$

Since r and s satisfy equations (19) and (20), we obtain equation (33) and hence equations (34)—(42).

$$\left(\frac{r}{s}\right)^2 = \frac{\sum_{i=1}^{p} [w_i Y_i'^2 - w_i^2 \sigma^2(Y_i) \Delta_i^2]}{\sum_{i=1}^{p} [w_i X_i'^2 - w_i^2 \sigma^2(X_i) \Delta_i^2]} \equiv \frac{U}{L} \qquad (33)$$

$$(\partial U / \partial Y_i') = 2w_i Y_i' \qquad (34)$$

$$(\partial U / \partial \Delta_i) = -2w_i^2 \sigma^2(Y_i) \Delta_i \qquad (35)$$

$$(\partial U / \partial w_i) = Y_i'^2 - 2w_i \sigma^2(Y_i) \Delta_i^2 \qquad (36)$$

$$(\partial L / \partial X_i') = 2w_i X_i' \qquad (37)$$

$$(\partial L / \partial \Delta_i) = -2w_i^2 \sigma^2(X_i) \Delta_i \qquad (38)$$

$$(\partial L / \partial w_i) = X_i'^2 - 2w_i \sigma^2(X_i) \Delta_i^2 \qquad (39)$$

$$\sigma^2(\Delta_i) = r^2 \sigma^2(X_i') + 2rs \,.\, \mathrm{cov}(X_i', Y_i') + s^2 \sigma^2(Y_i') \qquad (40)$$

$$\mathrm{cov}(X_i', \Delta_i) = r\sigma^2(X_i') + s \,.\, \mathrm{cov}(X_i', Y_i') \qquad (41)$$

$$\mathrm{cov}(Y_i', \Delta_i) = r \,.\, \mathrm{cov}(X_i', Y_i') + s\sigma^2(Y_i') \qquad (42)$$

We have ignored $\mathrm{cov}(X_i', w_i)$, $\mathrm{cov}(Y_i', w_i)$, and $\mathrm{cov}(\Delta_i, w_i)$ and computed equations (43)—(47). It can be shown

$$\sigma^2(U) = \sum_{i=1}^{p} \{(\partial U / \partial Y_i')^2 \sigma^2(Y_i') + (\partial U / \partial \Delta_i)^2 \sigma^2(\Delta_i) + (\partial U / \partial w_i)^2 \sigma^2(w_i) + 2(\partial U / \partial Y_i')(\partial U / \partial \Delta_i) \mathrm{cov}(Y_i', \Delta_i)\} \qquad (43)$$

$$\sigma^2(L) = \sum_{i=1}^{p} \{(\partial L / \partial X_i')^2 \sigma^2(X_i') + (\partial L / \partial \Delta_i)^2 \sigma^2(\Delta_i) + (\partial L / \partial w_i)^2 \sigma^2(w_i) + 2(\partial L / \partial X_i')(\partial L / \partial \Delta_i) \mathrm{cov}(X_i', \Delta_i)\} \qquad (44)$$

$$\mathrm{cov}(U,L) = \sum_{i=1}^{p}\{(\partial U/\partial Y_i')(\partial L/\partial X_i')\mathrm{cov}(X_i',Y_i') + (\partial U/\partial\Delta_i)(\partial L/\partial\Delta_i)\sigma^2(\Delta_i) + (\partial U/\partial w_i)(\partial L/\partial w_i)\sigma^2(w_i) + (\partial U/\partial Y_i')(\partial L/\partial\Delta_i)\mathrm{cov}(Y_i',\Delta_i) + (\partial U/\partial\Delta_i)(\partial L/\partial X_i')\mathrm{cov}(X_i',\Delta_i)\} \quad (45)$$

$$\sigma^2[(r/s)^2] = \sigma^2(U)/L^2 + U^2\sigma^2(L)/L^4 - 2U\,.\,\mathrm{cov}(U,L)/L^3 \quad (46)$$

$$\sigma^2(r^2) = s^8\,.\,\sigma^2[(r/s)^2] \quad (47)$$

that $\sigma^2(w_i)$ is given by equation (48). The calculation

$$\sigma^2(w_i) = w_i^4\sigma^2(r^2)\{[\sigma^2(X_i) - \sigma^2(Y_i)] + [(s/r) - (r/s)]\mathrm{cov}(X_i,Y_i)\}^2 \quad (48)$$

from equation (24) to equation (48) is repeated until two successive values of $\sigma^2[(r/s)^2]$ agree to one part in 10^4. For this iteration again no more than four cycles have proved necessary, so that the convergence of the algorithm has been in no doubt. Having obtained convergence, we get equation (49) where $\sigma^2(r/s)$ is the

$$\sigma^2(r/s) = 0{\cdot}25(s/r)^2\sigma^2[(r/s)^2] \quad (49)$$

variance of the slope of the line.

When $\mathrm{cov}(\bar{X},r/s)$ and $\mathrm{cov}(\bar{Y},r/s)$ are neglected, equation (50) can be calculated, to obtain the variance

$$\sigma^2(y) = \sigma^2(\bar{Y}) + (r/s)^2\sigma^2(\bar{X}) + (x - \bar{X})^2\sigma^2(r/s) + 2(r/s)\mathrm{cov}(\bar{X},\bar{Y}) \quad (50)$$

of the intercept on a vertical line at a given value of x. We have also computed S, of equation (12), for comparison with $(p - 2)$, the number of degrees of freedom, in a χ^2 test of the hypothesis that the (X_i,Y_i) do not deviate significantly from the fitted line. The validity of this test depends on the correctness of the variances $\sigma^2(X_i)$, $\sigma^2(Y_i)$, and $\mathrm{cov}(X_i,Y_i)$ assigned to the (X_i,Y_i) at the outset, since the weights assigned to the points are not altered in the light of their actual deviations from the line.

Since these expressions involve w_i^4, it is unwise to use data with variances differing from unity by many orders of magnitude. This may lead to floating-point overflow unless the computer used has capacity for an exceptionally large range of floating point numbers. The co-ordinates are thus scaled suitably before input to the line-fitting procedure, and the variances and covariances scaled accordingly.

Conclusion.—The straight-line-fitting procedure described above is an improvement on the least-squares cubic method suggested by York[9] for two reasons: (*a*) correlation between the co-ordinate errors is allowed for, and (*b*) vertical and horizontal lines will not cause programme failure. The algorithm has been tested by comparison with a programme which uses the York algorithm, which ignores correlation between the errors of X_i and Y_i. The resulting parameters were identical, but there were cases in which the York programme failed because it attempted to compute three real roots for the least-squares cubic equation, when two of them were complex. This is a criticism of a deficiency in the programme rather than of the York algorithm. However, the method given in this paper is not subject to such a difficulty.

York[10] subsequently published a method during the course of the present work in which allowance was made for correlation. Our own method is equivalent, but more easily extended to situations with more than two parameters (see Part II).

When our method was used on data with correlation between the errors in X_i and Y_i, it produced larger values of χ^2 when the points were more likely to be in error parallel to the line than at right angles to it than it did for calculations in which correlation was neglected. For such data our method gave a more critical test of the agreement between the experimental points and the calculated line than did York's method.[9]

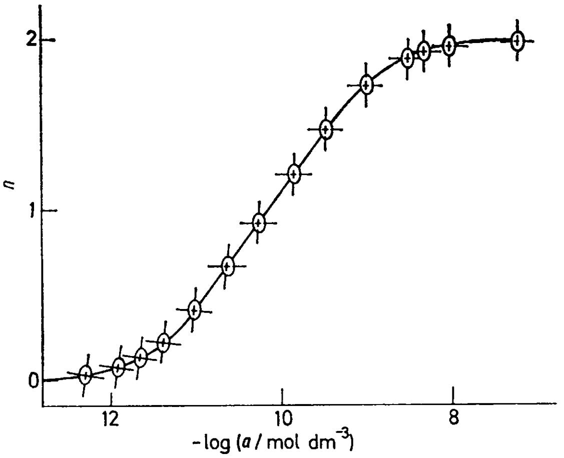

FIGURE 1 The formation curve for the copper(II)–ethylenediamine system in aqueous methanol (ref. 11) (experiment 1 of Table 1). The calculated error ellipses are enlarged to 10σ for clarity

When line-fitting techniques are applied to the computation of two stability constants, whether for proton complexes or metal-ion complexes, it is instructive to compare the results obtained by using both equations (3) and (4). As an example we recompute some published data[11] to obtain the two protonation constants of ethylenediamine and the stability constants of the two mononuclear ethylenediaminecopper(II) complexes in mixed solvents. A formation curve is shown in Figure 1 and the linear plots based upon equations (3) and (4) are shown in Figures 2 and 3; see also Tables 1 and 2. Although the intercept is extremely small in each example by comparison with the values of Y_i, our method gives values of the constants and their standard deviations which agree to about 0·2%. Where the intercept is larger, the constants agree to the four significant figures printed. The constant determined from identical data (where $0{\cdot}05 \leqslant \bar{n} \leqslant 1{\cdot}95$) by an unweighted least-squares adjustment may vary by as much as a power of ten

---

[9] D. York, *Canad. J. Phys.*, 1966, **44**, 1079.

[10] D. York, *Earth and Planetary Sci. Letters*, 1969, **5**, 320.

[11] G. Faraglia, F. J. C. Rossotti, and H. S. Rossotti, *Inorg. Chim. Acta*, 1970, **4**, 488.

between the two calculations, although relatively reliable values of the constants are obtained from the gradients of the graphs (*cf.* Figures **2** and **3**). Inclusion of data restricted to $0{\cdot}2 \leqslant \bar{n} \leqslant 1{\cdot}8$ improves the results of the unweighted squares treatment somewhat (see Table 1). The estimates of errors in the constants calculated by standard unweighted least-squares formulae are evidently misleading, since they are uniformly below 0·01 log unit and frequently as low as 0·003 log unit.

TABLE 1

Examples of calculated stability constants with standard deviations ($\times 10^3$) in parentheses. Potentiometric titration data at 25·00 ± 0·05 °C in 0·5M-$LiClO_4$ from ref. 11

	log K_1				log K_2			
	Experiment 1		Experiment 2		Experiment 1		Experiment 2	
	Intercept [c]	Gradient [d]	Intercept [c]	Gradient [d]	Gradient [c]	Intercept [d]	Gradient [c]	Intercept [d]
Ethylenediamine in 48·1% (w/w) aqueous dioxan								
Projection strip (ref. 2)	9·805(10)		9·805(10)		6·995(10)		6·995(10)	
FAJAF45 [a,e]	9·811(6)	9·810(6)	9·820(6)	9·819(6)	6·994(6)	6·994(6)	6·995(6)	6·995(6)
Unweighted least sq.[a]	9·260(5)	9·744(6)	9·441(20)	9·733(8)	6·998(1)	8·301(8)	6·995(1)	8·439(2)
Unweighted least sq.[b]	9·924(4)	9·803(2)	10·114(43)	9·810(2)	6·991(1)	7·306(9)	6·991(1)	7·386(4)
Copper(II)–ethylenediamine in 54·3% (w/w) aqueous methanol								
Projection strip (ref. 2)	10·835(10)		10·825(10)		9·465(10)		9·455(10)	
FAJAF45 [a,e]	10·833(4)	10·834(4)	10·828(6)	10·828(6)	9·464(4)	9·460(4)	9·457(6)	9·454(6)
Unweighted least sq.[a]	11·152(40)	10·841(1)	(K –ve)	10·821(2)	9·420(3)	9·405(3)	9·348(9)	9·470(4)
Unweighted least sq.[b]	10·860(6)	10·831(2)	10·866(3)	10·832(2)	9·455(2)	9·469(2)	9·447(3)	9·447(7)

[a] For $0{\cdot}05 \leqslant \bar{n} \leqslant 1{\cdot}95$. [b] For $0{\cdot}2 \leqslant \bar{n} \leqslant 1{\cdot}8$. [c] From equation (3); gradient = K_2, intercept = $-1/K_1$. [d] From equation (4); gradient = $-1/K_1$, intercept = K_2. [e] FAJAF45 sets $\delta(E)$ = 0·12 mV, $\sigma(E_0)$ = 0·10 mV, σ(concentration) = 0·15%, σ(volume) = 0·005 ml for a grade A 10 ml burette, 0·02 ml for a 25 ml burette.

These data have also been treated by graphical methods. Although the orders of magnitude of the co-ordinates of the linear plots makes these difficult to treat graphically, no such restriction applies to the projection strip method.[1,2] This well-proven technique, if used properly, yields results of high precision [11] (although limits of error are inevitably subjective), and the agreement between the constants obtained by using the projection strip, from our programme FAJAF45 and indeed from LETAGROP VRID [12] is most satisfactory (see Table 2). The equality in size of the error ellipses (enlarged to 10σ in Figure 1 for clarity) for the formation curve of the copper(II)–ethylenediamine system accounts for the success of the projection strip method. However, $\sigma(\bar{n})$ is more nearly proportional to $\bar{n}$ for systems of weaker complexes.[13] Moreover, the effects of the errors are by no means uniform in linear plots derived from the formation curve. Experimental points at high values of $\bar{n}$ in equation (3) (see Figure 2) and at low values of $\bar{n}$ in equation (4) (see Figure 3) have large associated uncertainties. Although such points are commonly arbitrarily omitted (*cf.* ref. 4), an unweighted regression is capable of producing a ' best ' straight line considerably different from that corresponding to our definition (see Figures 2 and 3).

TABLE 2

Further examples of calculated stability constants with standard deviations ($\times 10^3$) in parentheses. Protonation constants of ethylenediamine in 48·1% (w/w) aqueous dioxan. Combined data for experiments 1 and 2 of Table 1

	log K_1		log K_2	
	Intercept	Gradient	Intercept	Gradient
FAJAF45	9·815(4)	9·814(4)	6·994(4)	6·994(4)
Projection strip (ref. 2)	9·805(10)		6·995(10)	
LETAGROP (ΔH) [a] VRID	9·815(3)		6·996(7)	
LETAGROP (ΔE) [b] VRID	9·806(5)		6·998(11)	

[a] Minimising $\Sigma(H_{\text{calc}} - H_{\text{obs}})^2$, where H is the total analytical hydrogen-ion concentration, *cf.* equation (1). [b] Minimising $\Sigma(E_{\text{calc}} - E_{\text{obs}})^2$; E is proportional to log h.

Other major programmes for stability-constant calculations, such as LETAGROP VRID [12] and SCOGS,[14] are more general than our FAJAF45, but the simplicity of the last permits greater rigour. Nevertheless, it must be stressed that no statistical treatment, however sophisticated, can improve poor experimental data.

The present method has been applied to problems other than those of computing formation constants for dibasic acids HA, H_2A, and mononuclear metal-ion complexes BA and BA_2. For example, the Gran plot [1,15] used particularly in the standardisation of glass electrodes,[16,17] has been programmed so that a rigorously determined value of E_0 can be obtained in a few minutes *via* a remote link to a computer. It has also been applied in radiometric dating calculations by Cumming.[18] In fact, the line-fitting procedure was originally written for use in the latter field, where it is also often important to allow for correlated errors between the co-ordinates.

[12] L. G. Sillén and B. Warnqvist, *Arkiv Kemi*, 1969, **31**, 377.
[13] V. S. Jacewicz and F. J. C. Rossotti, to be published.
[14] I. G. Sayce, *Talanta*, 1969, **15**, 1397.
[15] F. J. C. Rossotti and H. S. Rossotti, *J. Chem. Educ.*, 1965, **42**, 375.
[16] R. P. Henry, J. E. Prue, F. J. C. Rossotti, and R. J. Whewell, *Chem. Comm.*, 1971, 868.
[17] F. J. C. Rossotti and R. J. Whewell, to be published.
[18] G. L. Cumming, *Canad. J. Earth Sci.*, 1969, **6**, 719.

Our original programme FAJAF45 was written in 1968 for the English Electric–Leo Marconi (International Computers Ltd) KDF9 computer. It is now also available as part of programme MERCURY written for the newer 1906A computer. An extension of the method is in progress, whereby a hyperplane is fitted to a set of co-ordinates in several dimensions. This has enabled a rigorous method to be applied to systems of complexes BA_n, H_jA (n, $j > 2$), or BA_n + BHA (see Part II).

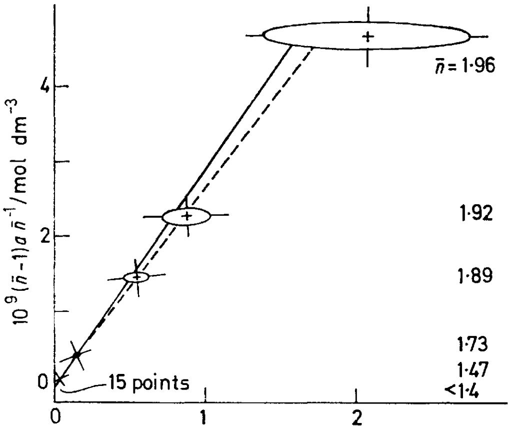

FIGURE 2 The linear plot of equation (3) derived from experimental points for the copper(II)–ethylene diamine system shown in Figure 1. Note that 15 of the 20 experimental points bunch together on the scale used. Error ellipses correspond to 3σ. The broken line is the unweighted least-squares 'best' fit. The full line is obtained by using the present treatment

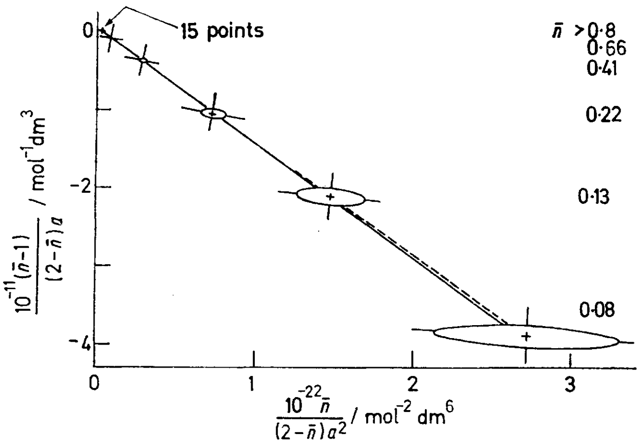

FIGURE 3 The linear plot of equation (4) derived from the experimental points for the copper(II)–ethylenediamine system shown in Figure 1. Note that 15 of the 20 experimental points bunch together on the scale used. Error ellipses correspond to 3σ. The broken line is the unweighted least squares 'best' fit. The full line is obtained by using the present treatment

APPENDIX 1

The variances $\sigma^2(\bar n)$, $\sigma^2(a)$ of $\bar n$ and a respectively, together with their covariance $\mathrm{cov}(\bar n,a)$, are calculated for each experimental point from the variances estimated for the measurements. A similar calculation then gives $\sigma^2(X)$, $\sigma^2(Y)$, and $\mathrm{cov}(X,Y)$ for the co-ordinates. From equation (3) we obtain (51) and thus (52) and

$$Y = (\bar n - 1)a/\bar n \tag{51}$$

$$\mathrm{d}Y = \frac{(\bar n - 1)}{\bar n}\mathrm{d}a + \frac{a}{\bar n^2}\,\mathrm{d}\bar n \equiv f_1\mathrm{d}a + f_2\mathrm{d}\bar n \tag{52}$$

$$X = (2 - \bar n)a^2/\bar n \tag{53}$$

$$\mathrm{d}X = \frac{(2 - \bar n)}{\bar n}2a\mathrm{d}a - \frac{2a^2}{\bar n^2}\,\mathrm{d}\bar n \equiv f_3\mathrm{d}a + f_4\mathrm{d}\bar n \tag{54}$$

similarly (53) and (54), whence equations (55)—(57) follow. The co-ordinates corresponding to equation (4)

$$\sigma^2(Y) = f_1^2\sigma^2(a) + f_2^2\sigma^2(\bar n) + 2f_1f_2\mathrm{cov}(\bar n,a) \tag{55}$$

$$\sigma^2(X) = f_3^2\sigma^2(a) + f_4^2\sigma^2(\bar n) + 2f_3f_4\mathrm{cov}(\bar n,a) \tag{56}$$

$$\mathrm{cov}(X,Y) = f_1f_3\sigma^2(a) + f_2f_4\sigma^2(\bar n) + (f_1f_4 + f_2f_3)\mathrm{cov}(\bar n,a) \tag{57}$$

are treated in the same way; were the errors uncorrelated, $\mathrm{cov}(X,Y)$ would be zero.

APPENDIX 2

In deriving equations (19) and (20) matrix notation is used for brevity. We set $\mathbf{r} = (r,s)$, $\mathbf{x}_i = (x_i,y_i)$, $\mathbf{x}_i' = \mathbf{x}_i - \bar X$, $\mathbf{X}_i = (X_i,Y_i)$, $W_i = \{W_{jki}\}$ and so on, and use the equation $\mathbf{r}^T.\,\mathbf{x}_i' + t' = 0$ for the line. Hence we obtain equation (58). At the minimum of S,

$$S = \sum_{i=1}^{p}[(\mathbf{x}_i' - \mathbf{X}_i')^T W_i(\mathbf{x}_i' - \mathbf{X}_i')]^2 \tag{58}$$

we have equations (59) and (60) for all i. We multiply

$$\mathrm{d}S = 2\sum_{i=1}^{p}[(\mathbf{x}_i' - \mathbf{X}_i')W_i\mathrm{d}\mathbf{x}_i'] = 0 \tag{59}$$

$$\mathbf{r}^T.\,\mathrm{d}\mathbf{x}_i' + \mathbf{x}_i'^T.\,\mathrm{d}\mathbf{r} + \mathrm{d}t' = 0 \tag{60}$$

each equation (60) by a Lagrangian multiplier λ_i, add to equation (59), and then set equal to zero the coefficients of $\mathrm{d}t'$ and of the elements of $\mathrm{d}\mathbf{a}$ and $\mathrm{d}\mathbf{x}_i'$. This yields equations (61)—(63) for all i, and hence (64)

$$\sum_{i=1}^{p}\lambda_i = 0 \tag{61}$$

$$\sum_{i=1}^{p}\lambda_i\mathbf{x}_i' = 0 \tag{62}$$

$$W_i(\mathbf{x}_i - \mathbf{X}_i') + \lambda_i\mathbf{r} = 0 \tag{63}$$

$$V_iW_i(\mathbf{x}_i' - \mathbf{X}_i') = -\lambda_iV_i\mathbf{r} \tag{64}$$

where $V_i = \{V_{jki}\}$ from the matrix of variances. But $V_iW_i = I$, the identity matrix, and hence we obtain equation (65). Since $\mathbf{x}_i'$ lies on the fitted line, equations

$$\mathbf{r}^T.\,\mathbf{x}_i' - \mathbf{r}^T.\,\mathbf{X}_i' = -\lambda_i\mathbf{r}^TV_i\mathbf{r} \tag{65}$$

(66) and (67) hold. If we set (68) we obtain (69),

$$\mathbf{r}^T\cdot\mathbf{x}_i' + t' = 0 \tag{66}$$

$$\mathbf{r}^T\cdot\mathbf{X}_i' + t' = \lambda_i\mathbf{r}^TV_i\mathbf{r} \tag{67}$$

$$w_i = (\mathbf{r}^TV_i\mathbf{r})^{-1} \tag{68}$$

$$\lambda_i = w_i(\mathbf{r}^T.\,\mathbf{X}_i' + t') \tag{69}$$

and we now eliminate λ_i and $\mathbf{x}_i'$. From equation (64) we obtain (70) and (71). Hence, from (62) and (61)

$$\mathbf{x}_i' = \mathbf{X}_i' - \lambda_i V_i \mathbf{r} \tag{70}$$

$$\lambda_i \mathbf{x}_i' = \lambda_i \mathbf{X}_i' - \lambda_i^2 V_i \mathbf{r} \tag{71}$$

respectively we obtain equations (72) and (73), but because of equation (74) $t' = 0$ and the line passes through the centroid $\bar{X}$.

$$\sum_{i=1}^{p} [w_i(\mathbf{r}^T . \mathbf{X}_i' + t')\mathbf{X}_i' - w_i^2(\mathbf{r}^T . \mathbf{X}_i' + t')^2 V_i \mathbf{r}] = 0 \tag{72}$$

$$\sum_{i=1}^{p} w_i(\mathbf{r}^T . \mathbf{X}_i' + t') = 0 \tag{73}$$

$$\sum_{i=1}^{p} w_i X_i' = \sum_{i} w_i Y_i' = 0 \tag{74}$$

Also we can reduce equation (72) to equation (75) which is equivalent to equations (19) and (20).

$$\sum_{i=1}^{p} [w_i(\mathbf{X}_i'^T . \mathbf{r})\mathbf{X}_i' - w_i^2(\mathbf{X}_i'^T . \mathbf{r})^2 V_i \mathbf{r}] = 0 \tag{75}$$

We thank the S.R.C. for a research studentship (to R. J. W.), Miss E. A. Miller for the preparation of the Algol and Fortran library subroutines, and Dr. B. Warnqvist of the Royal Institute of Technology, Stockholm, for performing the check calculation using LETAGROP VRID included in Table 2. Copies of these subroutines or of our programmes FAJAF45 or MERCURY are available on request (from F. J. C. R.).

[1/1606 *Received, 3rd September,* 1971]