High-Speed Calculation of Least-Squares Best Consecutive Formation Constants of Metal-Ion Complexes Using an Electronic Computer
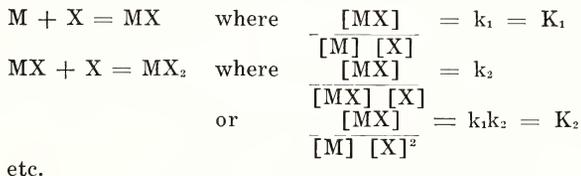
DONALD L. MCMASTERS and WARD B. SCHAAP, Indiana University[1]

Today, high-speed electronic computers are used in calculations that would be almost impossible and certainly impractical to do in any other way. In the field of chemistry, this applies mostly to the calculations involved in X-ray diffraction, electron diffraction, quantum mechanics, and so forth. Some of these calculations actually require many hours of electronic computer time.

There has been relatively little work done using computers for calculations which require much shorter periods of time, i.e., for problems requiring several tedious hours on a desk calculator. Electronic computers not only allow one to perform errorless calculations in the minimum amount of time, but also the use of electronic computers should encourage one to reconsider the basic equations used in his research to see whether or not he could use the more exact, but usually more mathematically complicated, fundamental equations to increase the accuracy of his results. This sometimes involves merely including more terms in the calculations or using more experimental data, or both. Also, certain statistical routines can usually be included to obtain the statistically best results.

As an example, in our research a least-square routine has proven to be a powerful tool in the calculation of the consecutive formation constants of inorganic complexes. This technique can be made applicable to other equilibrium processes in which it is desired to calculate all the equilibrium constants.

Let us consider an equilibrium process involving the consecutive addition of a group X to a central metal ion M. This can be represented in the following way:

$$M + X = MX \quad \text{where} \quad \frac{[MX]}{[M]\,[X]} = k_1 = K_1$$

$$MX + X = MX_2 \quad \text{where} \quad \frac{[MX]}{[MX]\,[X]} = k_2$$

$$\text{or} \quad \frac{[MX]}{[M]\,[X]^2} = k_1 k_2 = K_2$$

etc.

For example, the metal ion may be Cu^{++}, Cd^{++}, Ni^{++}, etc., and the ligand, X, may be monodentate Cl^-, NH_3, pyridine, etc., or a polydentate ligand, X, may be monodentate Ce^-, NH_3, pyridine, etc., or a polydentate number of four of these ligands can be coordinated to the metal ion.

The method to be discussed for the calculation of these formation constants involves the use of a power equation in which the coefficients of the variable are the overall formation constants of the complexes.

$$F_o(X) = k_o + K_1[X] + K_2[X]^2 + K_3[X]^3 + K_4[X]^4 \qquad (1)$$

This $F_o(X)$ function is calculated from polarographic data and includes the half-wave potential of the simple metal ion, the half-wave potential of the complexed metal ion and the ratio of the diffusion currents.

$$F_o(X) = \text{antilog} \left\{ \frac{n}{0.0591} \left[(E_{1/2})_s - (E_{1/2})_c \right] + \log \frac{(I_d)_s}{(I_d)_c} \right\} \qquad (2)$$

This relationship involving polarographic data was first derived by DeFord and Hume in 1951 [1] by combining the Nernst equation with the Ilkovic equation and making simplifying assumptions concerning activities and concentrations of the various species in the region of the mercury drop.

The formation constants themselves are usually determined by a graphical technique. The graph of $F_o(X)$ versus the concentration of the complexing agent has an extrapolated zero intercept which is k_o, the formation constant of the zeroth complex. Of course, k_o has a theoretical value of 1.0 and this then serves as a check on the data. This value of the intercept is subtracted out and the result is then divided through by the concentration of the complexing agent to get a next lower order equation. This $F_1(X)$ function is then plotted versus the concentration of the complexing agent and the extrapolated zero intercept is the formation constant K_1 of the first complex, MX.

$$\frac{F_o(X) - k_o}{[X]} = K_1 + K_2[X] + K_3[X]^2 + K_4[X]^3 = F_1(X) \qquad (3)$$

This process is repeated until all the constants have been determined.

$$\frac{F_1(X) - K_1}{[X]} = K_2 + K_3[X] + K_4[X]^2 = F_2(X) \qquad (4)$$

$$\frac{F_2(X) - K_2}{[X]} = K_3 + K_4[X] = F_3(X) \qquad (5)$$

$$\frac{F_3(X) - K_3}{[X]} = K_4 = F_4(X) \qquad (6)$$

It is important to realize that there is an easily recognizable relation between the order of the $F(X)$ equations (or the shapes of their plots) and the maximum number of complexes which can form. The last equation for the highest complex does not contain the concentration of the complexing agent and thus $F_4(X)$ should be a constant and independent of concentration. It can be considered a straight line horizontal to the concentration axis. The next to the last curve is a straight line of positive slope. The next lower curve represents a quadratic equation, the next a cubic and the last curve a quartic equation. (We are still assuming the presence of four complexes in the above discussion.)

Another point which should be mentioned is that the uncertainty in the values for the various constants increases as one goes from the first formation constant to the last. This is partly due to the uncertainty in the values of the extrapolated intercepts, since they are used in the later calculations, and partly due to the fact that the concentration of the complexing agent is reintroduced in each step of the calculations.

The obvious point to begin a least-squares calculation is with the $F_o(X)$ equation involving the formation of the zeroth complex, since the data involved in this calculation are the most accurate. Also, if the earlier constants can be calculated more accurately, since they are themselves used in the calculation of the rest of the constants, this will tend to increase the accuracy of the remaining constants.

Notice (Equation 1.) that the $F_o(X)$ equation contains all the constants. Theoretically, therefore, by solving for the coefficients of just this one quartic equation from the data by the method of least-squares, it should be possible to obtain all the formation constants directly. It will be explained later why this does not always work out in practice.

The least-squares routine (IBM 6.0.006) used in this work was originally programmed for the IBM 650 by a research group at General Electric and is made available to users of IBM 650's throughout the country by a central library maintained at the headquarters of the IBM Data Processing Division.

This routine requires about two minutes to calculate all the least-squares best coefficients for the linear, quadratic, cubic and quartic equations for any number of data points up to 100. That is, the program is set up to try automatically all four orders of equations with each set of supplied data. It also calculates y values for each given x value using the least-squares best coefficients in each order equation. These last data are useful in estimating the standard deviations of the results. The experimental data to be considered as an example in this discussion are from an actual metal-ion complex system and the authors are indebted to Mr. Frank A. Guthrie for his permission to use his data in this paper.

The procedure is straightforward. The values of $F_o(X)$ and $F_1(X)$ at different concentrations of the complexing agent are calculated according to equations (2) and (3), respectively. These functions are then first plotted graphically versus the concentration of the complexing

TABLE I

Calculated Least-Squares Coefficients for F (X) *vs.* [X] Data
Assuming Linear, Quadratic, Cubic and Quartic Equations.
Equation   Coefficients (Formation Constants):

Data	Type	k_o	K_1	K_2	K_3	K_4
$F_o(X)$	linear	0.3573	46.21			
	quadratic	0.9706	21.56	157.5		
	cubic	0.9860	20.52	173.8	−68.44	
	quartic	0.9157	27.82	−29.85	1937	−6358
$F_1(X)$	linear		16.58	215.7		
	quadratic		20.47	144.8	171.8	
	cubic		21.19	115.9	357.9	−299.0
$F_2(X)$	linear			145.4	178.6	
	quadratic			121.6	249.0	−36.31
$F_2(X)$	linear (constant)				174.4	3.723

agent to obtain an approximate value of the extrapolated zero intercept for k_0 and K_1, respectively. (k_0 should theoretically be 1.0 and in our example K_1 is approximately 20.) The $F_0(X)$ data are then used in the IBM 650 least-squares routine. The results of the least-squares calculation of the $F_0(X)$ data are then checked to see how the k_0 and K_1 constants agree with the graphically extrapolated values of 1.0 and 20. The value of K_2 should also be noted and kept for future comparisons.

Assuming that four complexes are present, k_0 and K_1 should fit the least-squares coefficients of the quartic equation best. The least-squares results for the quartic equation indicate that k_0 is 0.9157 and K_1 is 27.82. (See Table I.) However, if we consider the results of all the least-squares calculated coefficients, we see that the values for k_0 and K_1 of the cubic equation approach the expected values more closely. (The value of K_2 is also more reasonable.) Therefore, we may begin to suspect that only three complexes are formed.

This, then, is an important step of the procedure: If the graphically extrapolated values of k_0 and K_1 fit the least-squares coefficients of the quartic equation best, then there will be four and only four complexes. If the values fit the cubic coefficients best, then only three complexes are formed; if the quadratic coefficients fit best, then only two complexes are indicated; and if the linear gives the best fit, then only one complex is indicated.

In our example, k_0 and K_1 fit the cubic equation best and therefore only three complexes are indicated. Furthermore, the results of the next calculation can also be used to support this. When the $F_1(X)$ data are used in the least-squares calculation, the results indicate that these data best fit a quadratic equation with a value of 20.47 for K_1. (See Table I.) This then confirms the fact that only three complexes can be present since a next lower order equation is theoretically expected and the values of the constants check with the graphical results. The quadratic coefficients for the $F_1(X)$ data give 144.8 for K_2 and 171.8 for K_3.

The calculations can then be continued with a definite knowledge of the shapes of the remaining curves and also the approximate values of the remaining constants. In our example, there are only two more lines. One of these is linear and the other is constant (i.e., horizontal). The intercept coefficient, K_2, obtained from the linear equation coefficients calculated from the $F_2(X)$ data by the least-squares method is 145.4, which agrees well with the value of 144.8 obtained from the K_2 coefficient of the quadratic equation for the $F_1(X)$ data. The K_3 coefficient of the quadratic $F_1(X)$ equation is 171.8; the K_3 coefficient from the linear $F_2(X)$ equation is 178.6. Both of these agree well with the intercept coefficient, $K_3 = 174.4$, for the approximately horizontal $F_3(X)$ line. However, in this case it is undoubtedly better to use an average of the constant values of $F_3(X)$ rather than the intercept value, for there is no point in magnifying any false trend in the supposedly constant values of $F_3(X)$ by extrapolating to zero concentration. The mean value for $F_3(X)$ is 177.3, and we shall take this as the statistically best value for K_3. Ideally, since K_3 is the formation constant of the highest complex, the value of K_4 should be zero. Actually, the values of K_3 from the $F_3(X)$

calculations do exhibit a slight slope and the least-squares best value of K_4 is 3.7, which is certainly close enough to zero to be within the limits of experimental error.

Estimation of the standard deviations of the various coefficients of the polynomial equations is a fairly complex statistical problem and will not be discussed here. Since values of the lower coefficients are used in calculations of the higher coefficients, the errors will be cascaded.

It was mentioned previously that it should be possible in theory to obtain all the values of K_n, the formation constants, from the least-squares best coefficients of the $F_o(X)$ data, but that this did not work out in actual practice. The reason for this is that certain 'assumed constants' such as activity coefficients and liquid junction potentials are not actually constant, but gradually change when the concentration of the complexing agent changes. These changes in 'assumed constants' cause the half-wave potential to shift. But since the total observed half-wave potential shifts are used in calculating the formation constants, the apparent values of these formation constants will be different when calculated from data in widely different concentration regions. Self-consistent and concordant results are obtained only when data in a more limited concentration range are used in the calculations This then serves as an explanation for the disagreement in the values of the higher least-squares coefficients in the $F_o(X)$ calculation and the final calculated results.

We have chosen to include data only from concentration ranges where a particular complex is present in significant amounts in the calculation of the formation constant of that complex. Thus, only the data from the ten lowest concentrations were used in the final calculations of k_o. Data from the lowest fourteen concentrations were used in the final calculation of K_1, while data from the eleven most concentrated solutions were used in the final calculation of both K_2 and K_3.

Another point also becomes obvious when one applies a statistical treatment to data from which a number of constants are to be evaluated. This is that a larger amount of raw data (in this study, half-wave potentials at a given concentration of complexing agent) should be used than are usually used in making such calculations. For example, half-wave potentials at six different concentrations would provide only two degrees of freedom if four independent constants are to be evaluated, so the results would show relatively large standard deviations. In most published applications of this method, a total of only about eight to ten points are used, which certainly would be the minimum acceptable, even if all points were used in the calculation of all the formation constants, which is normally not the case. In the results discussed here 21 concentrations are used and at each concentration the half-wave potential is the average of at least four and sometimes five or six half-wave potentials. We feel, then, that our results are relatively statistically reliable.

In summary, the reliability of calculated consecutive formation constants of complexes can be greatly increased by applying statistical techniques. The use of electronic computors allows the necessary calculations to be made quickly and easily and thus encourages applica-

tion of mathematical relationships that would not ordinarily be used because of the great amount of time that would be involved in doing the same calculations with a desk calculator.

Literature Cited

1. DeFord, D. and D. Hume. 1951. J. Am. Chem. Soc. **73** : 5321.