THE ANALYST

A General Graphical Method for Evaluating Experimental Results that Should Fit a Linear Equation

By H. M. N. H. IRVING

(*Department of Inorganic and Structural Chemistry, The University, Leeds* 2)

When the results of an experiment should be representable by a linear equation of the general form $y = mx + c$ the errors inevitably associated with the values of x_j and y_j ($j = 1$ to $n > 2$) give rise to a set of n simultaneous equations that will be inconsistent. As an alternative to a least-squares computation, a graphical procedure is described for finding the best values of the constants m and c. The method is quite general and it is illustrated by worked examples of problems encountered in analytical chemistry, *e.g.*, the potentiometric titration of a dibasic acid, the spectrophotometric determination of the stability of a weak complex, the evaluation of redox potentials and the liquid - liquid extraction of a weak monobasic acid.

WITH the increasing availability and use of computers, graphical methods of handling data are tending to drop out of favour. There are experiments, however, in which graphical procedures have advantages, even if only as a preliminary to more rigorous computational methods; sometimes the precision of the experimental data is such that conclusions from graphical methods are adequate.

Many chemical problems yield experimental data in terms of sets of two variable quantities x_j and y_j that can be related by a linear equation

$$y_j = mx_j + c \quad .. \qquad .. \qquad .. \qquad .. \qquad .. \qquad (1),$$

which can be presented graphically as a straight line passing through the points (x_j, y_j). The problem is to determine the values of the constants m and c. Provided only two sets of observations have been made ($j = 0$ and 1) there will be a unique solution to the resulting pair of simultaneous equations, irrespective of the experimental errors made in the measurements. When, as in practice, there may be more than two sets of observations, each of which will be subject to experimental error, the corresponding linear equations will probably not be consistent, and the problem is now to find the "best" values for the constants m and c. This can be done graphically by drawing the "best straight line" through the sets of points (x_j, y_j) or by computing the slope and intercept of this line by the "least squares method." It should be pointed out that in this procedure it is common to assume that all the random errors are associated with one of the observations (*e.g.*, with measurements of x_j), while measurements of the other variable, y_j, are free from error.

The object of the present paper is to present an alternative graphical procedure for the solution of sets of (possibly inconsistent) linear equations. No novelty is claimed for this procedure, which appears to have been suggested by Professor Finsler and has been used extensively by Schwarzenbach, Willi and Bach.[1] However, the method is not widely known and experience has shown that many research workers have not found the description in the German text easy to follow. I have therefore taken the opportunity of generalising the procedure and showing, by worked examples, how it can be applied in several different situations.

BASIC PRINCIPLES

Instead of plotting the data (x_j, y_j) as a series of points through which the "best straight line" is to be drawn, the basic principle is to use a co-ordinate system such that each associated pair of observations (x_j and y_j) is used to define a straight line; the various straight lines for all the observations are then to intersect in a common point whose co-ordinates are a measure of the desired constants m and c.
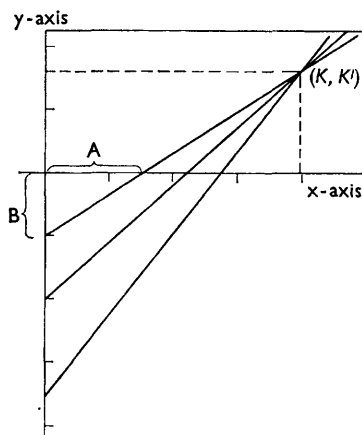
Fig. 1.   Plots of lines conforming to the general equation

$$\left(\frac{1}{A}\right)K + \left(\frac{1}{B}\right)K^1 = 1$$

Suppose a straight line is drawn to pass through the points (A,O) and (O,B) situated along the x-axis and y-axis, respectively, of a rectilinear co-ordinate system (Fig. 1).   If this line also passes through the point (K,K') it is easy to show that its equation must be

$$\left(\frac{1}{A}\right)K + \left(\frac{1}{B}\right)K' = 1. \quad .. \qquad .. \qquad .. \qquad .. \quad (2).$$

The graphical procedure now involves several stages—

(*i*) The exact linear equation representing the process under investigation is written down in its appropriate symbols.

(*ii*) The form of this equation is adjusted to conform exactly with equation (2), *i.e.*, with the desired constants on the left-hand side, and only unity on the right.

(*iii*) By direct comparison with equation (2), A is expressed in terms of the experimentally determinable quantities and the same procedure is carried out for B. The relationship of K and K' to the two desired constants is also written down.

(*iv*) The points (A,O) and (O,B) are plotted with suitable scales and joined to give a straight line.

(*v*) Similar straight lines are drawn for values of A and B calculated from all the other experimental results.

(*vi*) The values of K and K' are determined from the point of intersection and those of the desired constants are deduced from them.

A few examples will make the procedure clear.

EXAMPLE 1—

A dibasic acid, H_2A, has been titrated potentiometrically with alkali, with the object of determining the values of the two acid-dissociation constants

$$K_1 = \frac{[H^+][HA^-]}{[H_2A]}$$

and

$$K_2 = \frac{[H^+][A^{2-}]}{[HA^-]}.$$

If the total amount of acid is known,

$$C_A = [H_2A] + [HA^-] + [A^{2-}]$$

and, if observations are made of the pH after adding various amounts of alkali, we have, from considerations of electro-neutrality,

$$[OH^-] + [HA^-] + 2[A^{2-}] = [H^+] + aC_A$$

where a is the degree of neutralisation. The last term on the right-hand side, therefore, refers to the concentration of univalent cation (*e.g.*, K^+) introduced during the titration. By elimination we arrive at the general equation

$$\left[\frac{\{(2-a)C_A - [H^+] + [OH^-]\}}{[H^+]\{(a-1)C_A + [H^+] - [OH^-]\}} \right] K_2 - \left[\frac{[H^+]\{aC_A + [H^+] - [OH^-]\}}{\{(a-1)C_A + [H^+] - [OH^-]\}} \right] \frac{1}{K_1} = 1 \qquad (3).$$

Comparison with the standard equation (2) shows that

$$K_2 = K, \; K_1 = \frac{1}{K'} \;,$$

$$A = \frac{[H^+]\{(a-1)C_A + [H^+] - [OH^-]\}}{\{(2-a)C_A - [H^+] + [OH^-]\}}$$

and that

$$B = -\frac{\{(a-1)C_A + [H^+] - [OH^-]\}}{[H^+]\{aC_A + [H^+] - [OH^-]\}} \cdot$$

Table I (rows 1 to 3) gives experimental data for the potentiometric titration of 100 ml of about 10^{-3} M anthranilic acid-NN'-diacetic acid, o-HOOC.C$_6$H$_4$.N(CH$_2$COOH)$_2$, with carbonate-free 0·1 M potassium hydroxide.[1]

TABLE I
THE POTENTIOMETRIC TITRATION OF ANTHRANILIC ACID-NN'-DIACETIC ACID

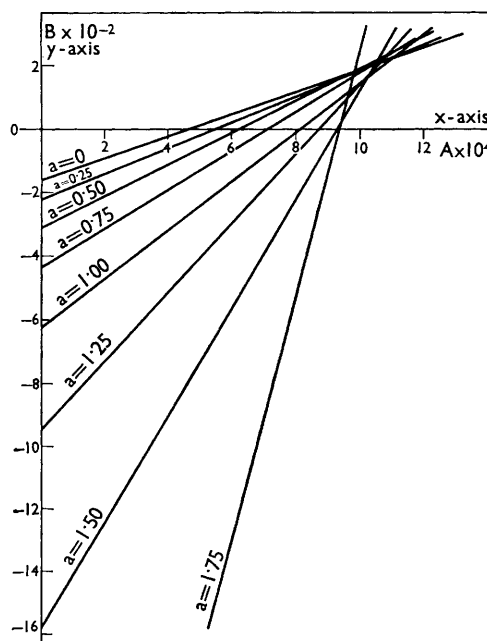| 10^3C_A			1·031	1·029	1·026	1·023	1·021	1·018	1·016	1·013
a			0	0·25	0·50	0·75	1·00	1·25	1·50	1·75
pH			2·886	2·955	3·032	3·123	3·228	3·369	3·561	3·886
$10^3[H^+]$			1·301	1·110	0·929	0·753	0·592	0·428	0·275	0·130
10^4A (calculated)			4·616	5·436	6·207	7·122	8·169	8·707	9·241	9·385
$10^{-2}B$ (calculated)			−1·595	−2·229	−3·097	−4·344	−6·200	−9·379	−15·83	−35·97

Fig. 2.   Data from Table I relating to the titration of anthranilic acid-NN'-diacetic acid. Calculation of K_1 and K_2

The various straight lines are plotted in Fig. 2. Apart from that for $a = 1.75$, the convergence to a single point is quite satisfactory and leads to the values $K = 10.4 \times 10^{-4}$ and $K' = 2.0 \times 10^2$, from which we deduce $K_1 = 5.00 \times 10^{-3}$ and $K_2 = 1.04 \times 10^{-3}$. The more tedious treatment of these experimental data by least-squares computation leads to

$$K_1 = 5.06 \ (\pm 0.05) \times 10^{-3} \text{ and } K_2 = 1.04 \ (\pm 0.01) \times 10^{-3}.$$

However, as Schwarzenbach, Willi and Bach point out,[1] values of these constants obtained from a completely independent titration may show an absolute error, which exceeds that calculated from the same titration curve. For example, in a second titration of anthranilic acid-NN'-diacetic acid the values $K_1 = 4.80 \times 10^{-3}$ and $K_2 = 1.01 \times 10^{-3}$ were obtained graphically and the values $K_1 = 4.25 \ (\pm 0.25) \times 10^{-3}$ and $K_2 = 1.05 \ (\pm 0.02) \times 10^{-3}$ by computation. Greater accuracy in the determination of K_1 and K_2 can only be achieved by paying much greater attention to points of experimental technique (*e.g.*, the more exact measurement of the concentration of the titrant and the constancy of the potential of the reference electrode).

In the present instance the graphical method suggests visually that the results for $a = 1.75$ are inconsistent with the rest, and this reading is automatically omitted when locating the point (K, K'). Calculation shows that a pH reading of 3.876 (*i.e.*, only one hundredth of a unit higher) would have given the "correct" result. Of course, in the least-squares calculation the less reliable data for $a = 1.75$ will have been included with the same weighting as all the others.

EXAMPLE 2—

The stability constant, K_{ML}, governing the formation of a weak complex ML according to the equation

$$M + L \rightleftharpoons ML \quad .. \qquad .. \qquad .. \qquad .. \qquad .. \qquad (4),$$

can often be obtained from spectrophotometric measurements of the optical density, E, of mixtures containing different initial concentrations, C_M and C_L, of the components, provided the molecular extinction coefficients, ϵ_M, ϵ_L and ϵ_{ML}, of the three species are known. Clearly, for a 1-cm cell length

$$E = [M]\epsilon_M + [L]\epsilon_L + [ML]\epsilon_{ML} = (C_M - [ML])\epsilon_M + (C_L - [ML])\epsilon_L + [ML]\epsilon_{ML}.. \quad (5)$$

and

$$(C_M - [ML]) (C - [ML]) = K_{ML}[ML] \quad .. \qquad .. \qquad .. \qquad (6).$$

When K_{ML} is small it is often necessary to add a considerable excess of one component (*e.g.*, L) to give appreciable amounts of ML, and with $C_L \gg [ML]$ we find, on eliminating terms in [ML], that

$$K_{ML} = \frac{C_L \ (C_M\epsilon_{ML} + C_L\epsilon_L - E)}{(E - C_M\epsilon_M - C_L\epsilon_L)} \quad .. \qquad .. \qquad .. \qquad .. \qquad (7).$$

This readily leads to values of K_{ML} (which should be identical within the limits of experimental error) on inserting values of E appropriate to different initial concentrations C_M and C_L. When using 1-cm cells we can replace $C_M\epsilon_M$ by E_M°, the optical density when $C_L = 0$, and $C_L\epsilon_L$ by E_L°, the optical density when $C_M = 0$. We can also write $E_\infty = C_M\epsilon_{ML}$ for the limiting case, in which the whole of the component M has been transformed into ML by a large excess of L. In practice, the actual value of E_∞ may not be determinable by

TABLE II

SPECTROPHOTOMETRIC STUDIES OF ADDUCT FORMATION

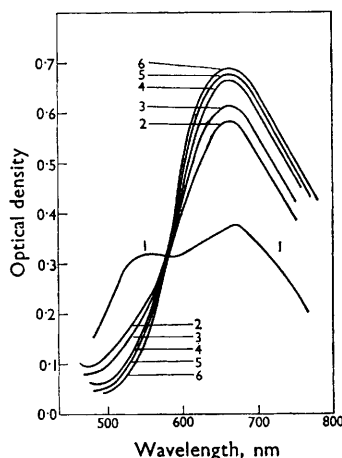| $10^3 C_L$	E	A (calculated)	B (calculated)
0	0.388	—	0.388
39.70	0.431	−0.397	0.431
79.41	0.460	−0.507	0.460
119.2	0.482	−0.611	0.482
159.1	0.497	−0.726	0.497
199.0	0.513	−0.756	0.513
278.7	0.544	−0.972	0.544
398.5	0.568	−1.258	0.568
598.3	0.600	−1.693	0.600
798.1	0.618	−2.145	0.618

Fig. 3.   Absorption spectra (1-cm cells) for mixtures of 2·5 × 10⁻³M copper bisacetylacetonate with varying amounts of isoquinoline in benzene.   Curves 1 to 6 correspond to 0·0, 0·12, 0·20, 0·40, 0·60 and 0·80 M isoquinoline, respectively
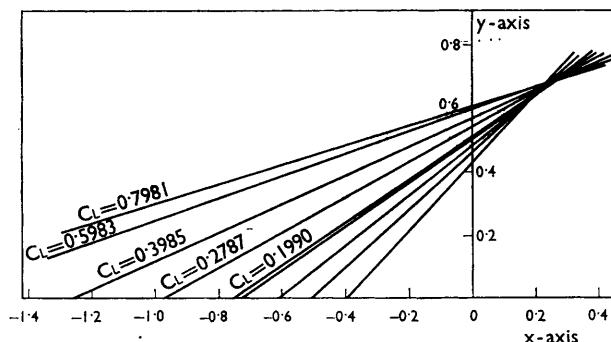

Fig. 4.   Data from Table II relating to adduct formation between copper bisacetylacetonate and isoquinoline in benzene.   Calculation of K_{ML} and ϵ_{ML}

experiment, for this limiting value is only slowly approached asymptotically if the complex is weak.   Indeed, the necessary large excess of the reagent L may so modify the nature of the solvent (specifically its dielectric constant and solvating power) as to yield spurious results. It now becomes necessary to treat both K_{ML} and ϵ_{ML} as unknowns.

Table II summarises data for mixtures of copper bisacetylacetonate, Cu(acac)₂, (=M) and quinoline (=L) in benzene when a weak 1 + 1 adduct, Cu(acac)₂ . (quinoline), is formed.[2] The absorption spectra of all the mixtures show an isosbestic point at 580 nm (Fig. 3), but the absorption caused by the 1 + 1 adduct and unchanged copper chelate has still not reached a constant value, even when the ratio of C_L to C_M is as high as 320 : 1.

Optical density measurements (1-cm cells) were made with a Unicam SP500 spectrophotometer at 670 mμ, at which wavelength $\epsilon_L = 0$, $E_M^0 = 0.388$, C_M was always 2·5 × 10⁻³ M and C_L ranged from 0 to 0·8 M.   By writing $\epsilon_L = 0$ in equation (7) and re-arranging we obtain

$$\frac{(E_M^0 - E)}{E C_L} K_{ML} + \left(\frac{1}{E}\right) E_\infty = 1 \qquad .. \qquad .. \qquad .. \qquad (8),$$

and when this is compared with the standard form of the equation [equation (2)] we realise that

$$K_{ML} = K, \quad E_\infty = K',$$

and that the new variables A and B are to be calculated from

$$A = \frac{E C_L}{(E_M^0 - E)} \text{ and } B = E \quad .. \qquad .. \qquad .. \qquad .. \qquad (9).$$

The calculated values of A and B are given in Table II and the corresponding lines are shown plotted in Fig. 4.   The intersection leads to the values $K = 0.27 \pm 0.01$ and $K' + 0.69$, whence $K_{ML} = 3.70 \pm 0.15$ and $E_\infty = 0.690$.   The far more laborious method for handling such data is that suggested by Graddon and Watton.[3]   Here a trial value for E_∞ is assumed and values for K_{ML} calculated for every experimental value of E.   The average value of K_{ML} and its standard deviation is then calculated.   The procedure is repeated for other assumed values of E_∞, and the approved value is taken as that for which the spread of values of K is least.   The principle is illustrated by the calculations shown in Table III.

TABLE III

CALCULATIONS OF THE STABILITY CONSTANT OF THE WEAK ADDUCT $Cu(ACAC)_2.(QUINOLINE)$

E_∞	0·678	0·688	0·691	0·693	0·694	0·695	0·698	0·708	0·738
K_{ML}	4·200	3·870	3·810	3·760	3·714	3·700	3·630	3·400	3·860
Standard deviation	0·290	0·210	0·200	0·190	0·177	0·190	0·200	0·230	0·380

The close agreement between the graphical and the computed value $K_{ML} = 3·71 \pm 0·18$ is fortuitous in this instance. However, the graphical method quickly indicates a reasonable approximate value to use for E_∞.

EXAMPLE 3—

When a monobasic acid, HL, with acid-dissociation constant, K_a, partitions between an organic phase and water

$$p = \frac{[HL]_{organic}}{[HL] + [L^-]}$$

$$= \frac{p^0}{\left(1 + \dfrac{K_a}{[H^+]}\right)} \quad .. \quad .. \quad .. \quad .. \quad (10),$$

where $p^0 = \dfrac{[HL]_{organic}}{[HL]}$, the partition coefficient for the undissociated species.

On re-arrangement, equation (10) gives

$$\left(\frac{1}{p}\right) p^0 - \left(\frac{1}{[H^+]}\right) K_a = 1 \quad .. \quad .. \quad .. \quad .. \quad (11),$$

and comparison with the standard equation (2) shows that $p^0 = K$, $K_a = K'$ and that we must set $A = p$ and $B = [H^+]$ as shown in Table IV, which summarises unpublished results for the partition of a weak acid between cyclohexane and buffer solutions adjusted to constant ionic strength ($\mu = 0·1$) with potassium chloride.

TABLE IV

PARTITION OF A WEAK BASE BETWEEN CYCLOHEXANE AND BUFFER ($\mu = 0·1$)

pH	5·71	5·23	5·10	5·05	4·99	4·70	4·52	4·30	4·02
p (≡ A)	0·350	0·745	0·880	0·940	1·01	1·32	1·48	1·60	1·80
$10^5 [H^+]$ (≡ B) ..	0·195	0·589	0·794	0·891	1·023	1·995	3·020	5·012	9·550

Fig. 5 shows the graphical solution, which leads to $p^0 = 1·93 \pm 0·03$ and $pK_a = 5·02 \pm 0·02$. Two points are especially noteworthy in these equations. Although the range of pH values covers only 1·7 units, $[H^+]$ varies 50-fold, and this makes it impossible to accommodate all the B-values along the ordinate axis without intolerably "crowding" the values for $0·2 < 10^5[H^+] < 2·0$. Nevertheless, the results for values of B greater than 2×10^{-5} can readily be incorporated by writing the equation to the line through the points (A,O) and (O,B), viz., $y = \left(\dfrac{B}{A}\right) x + B$, and calculating the value of x corresponding to the smallest accessible ordinate (in this example $y = -2 \times 10^{-5}$), plotting this point and joining up to the point (A,O), as shown in Fig. 5 for $10^5[H^+] = 3·020$, 5·012 and 9·550. The same device has been used without explanation in Figs. 2 and 4. The second, and rather obvious, point is that this graphical procedure will break down if the partition coefficient p^0 is more than an order of magnitude greater than the measured values of p. This imposes too lengthy an extrapolation to locate the point (K,K') with any accuracy. Indeed, it is just this condition which limits the value of the graphical procedure in the general case. The size of the area which, in practice, generally replaces the single unique point of intersection of all the lines through (A,O) and (O,B) gives an immediate indication of the quality of the experimental data and often shows whether the value of one or other of the constants K and K' is particularly influenced by experimental error.
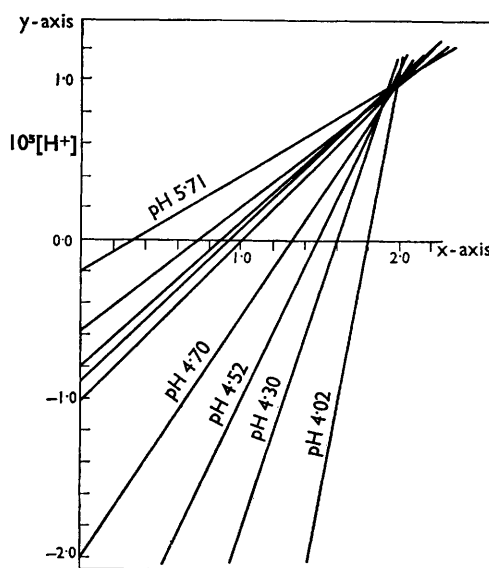
Fig. 5.   Data from Table IV relating to the partition of a monobasic acid between cyclohexane and buffer solutions. Calculation of $\rho°$ and K_a

EXAMPLE 4—

The stability of complexes of EDTA and other complexones with iron(II) and with iron(III) ions can be determined from measurements of redox potentials.[4] To take a simple example, the redox potential, E_M, of the couple Fe(III) - Fe(II), in the presence or absence of a complexone as desired, can be obtained from measurements of the potential E (relative to a standard hydrogen electrode) of a gold electrode dipping into V ml of a solution of an iron(II) salt to which varying volumes, v, of a standardised oxidant such as iodine have been added.   If v_e is the volume needed to oxidise the Fe^{2+} completely we have

$$E = E_M + \left(\frac{RT}{F}\right) \ln \frac{v}{(v_e - v)} \quad .. \quad .. \quad .. \quad .. \quad (12)$$

where the unknowns to be determined are the end-point volume, v_e, and the mid-point potential, E_M.   Writing $\left(\frac{RT}{F \log e}\right) = s = 0.059517$ at 25° C, and re-arranging we obtain

$$\left(\frac{1}{v}\right) v_e - \left(\frac{1}{10^{E/s}}\right)10^{E_M/s} = 1 \quad .. \quad .. \quad .. \quad .. \quad (13).$$

Comparison with the standard equation (2) shows that
$$v_e = K \quad E_M = s.\log K',$$
and that the variables A and B are to be calculated from A $= v$, and B $= -10^{E/s}$.

Table V shows typical experimental data for the titration of 100 ml of about 8×10^{-4} M iron(II) sulphate containing 2×10^{-3} M EDTA at a pH of 4.625 with a constant salt background, $\mu = 0.1$, potassium chloride.[5]   The titrant was about 2.2×10^{-2} M iodine.

TABLE V

TITRATION OF IRON(II) SULPHATE WITH IODINE IN THE PRESENCE OF EDTA

v, ml	..	..	1.40	1.70	2.60	3.20	3.52	3.60
E, mV	..	..	39.6	46.5	71.6	92.2	115.2	124.8
$10^{E/s}$	..	..	4.671	6.109	15.02	36.19	88.59	128.7

v, ml	..	..	3.66	3.68	3.70	3.72	3.74
E, mV	..	..	135.3	139.6	144.5	152.8	159.9
$10^{E/s}$	..	..	193.6	229.0	277.1	382.8	504.7

The results are shown graphically in Fig. 6, which only displays the top right-hand portion of the whole large-scale graph.   By inspection, $v_e = 3.8$ and $10^{(E_M/s)} = 7.5$ whence $E_M = 0.0521$.   In this graph the precise location of the ordinate K' is subject to considerable error and the uncertainty is about ± 0.002 volt, which is scarcely acceptable.   More confidence could be attached to the result of plotting v^{-1} against $10^{-E/s}$ for, as shown by equation (13), this gives a straight line of intercept v_e^{-1} and slope $-v_e^{-1}10^{(E_M/s)}$.   Alternatively, a precise value of v_e can be obtained by Gunnar Gran's method[6]; this involves plotting $V.10^{17(k-E)}$ against V (k is any arbitrary number such that $10^{17(k-E)}$ forms a convenient range of positive numbers) and noting the values of v at which the resulting straight line cuts the x-axis. This value of v_e can then be used in equation (12) to calculate the value of E_M for all relevant values of E and v.
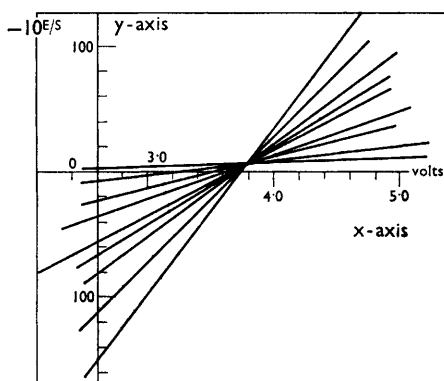

Fig. 6.   Data from Table V relating to the redox potentials of Fe(II) - Fe(III) in a buffer solution containing EDTA. Calculation of V_e and E_M

In conclusion, I would again point out that the graphical procedures described above have their proper function.   In some graphs difficulties of interpolation will not give results of acceptable precision, even when a large-scale graph is used; here more refined computational methods are essential.   In others, the procedure will lead to numerical results entirely adequate for the precision of the calculations in train.   There is little justification for using an elaborate programme and expensive computer time if the primary experimental results are insufficiently well defined, and the graphical procedure outlined above often provides a simple means of applying this criterion.

REFERENCES

1. Schwarzenbach, G., Willi, A., and Bach, R. O., *Helv. Chim. Acta*, 1947, **30**, 1303.
2. Al-Niaimi, N., D.Phil. Thesis, Oxford, 1964.
3. Graddon, E. P., and Watton, E. C., *J. Inorg. Nucl. Chem.*, 1961, **21**, 4.
4. Schwarzenbach, G., and Heller, J., *Helv. Chim. Acta*, 1951, **34**, 576.
5. Sharpe, K., Ph.D. Thesis, Leeds, 1968.
6. Gran, G., *Analyst*, 1952, **77**, 661.