Derivation of Thermodynamic Functions of Ionisation from Acidic Dissociation Constants

The Sigma Plot Method *vis-à-vis* Polynomial Regression Procedures

By D. J. G. Ives*† and P. G. N. Moseley‡

For $\Delta G°(T)$ data at equally spaced T, classical difference tables are invaluable for assessing the quality of the measurements and in setting limits to their exploitation by currently used statistical methods. A return to the principles of the Everett and Wynne-Jones [3] calculation is advocated; first differences are used to proceed directly to a best estimate of an average ΔC_P°. This allows the original data to be treated by the Σ-plot method, based on a chosen definition of ideal acidic dissociation behaviour, which gives idealised values of the required thermodynamic functions. If experimental accuracy permits, systematically distributed residuals from the Σ-plot will be available to study deviations from the defined ideal behaviour and to secure closer approximations to the thermodynamic functions, including $d\Delta C_P^\circ/dT$.

In so far as the interpolations used are rectilinear, the method minimises distortions due to the "smoothing-in" of incoherent errors, and is preferred to methods based on initial curvilinear data-fitting, the hazards of which are demonstrated in detail. Comparisons with, and correction of, some previously published data require the use throughout of cal units (1 cal = 4.184 J).

Derivation of standard thermodynamic functions of ionisation from acidity constants at several temperatures can be conducted by either of the van't Hoff or Gibbs–Helmholtz routes, which lead to identical results from a hypothetical set of error-free data.

The situation is complicated for experimental data because these are limited in number and are subject, at best, to a chance distribution of errors. Both routes then have the same disability, which is the great amplification of error associated with each differentiation step. Unless the primary measurements are of adequate accuracy, all significance may be lost *en route*, and this may escape detection by statistical tests. Such a mishap can lead to an equation for a derived function that is dominated by an adventitious distribution of errors and not at all by physical facts. This applies to the finding by Feates and Ives [1] that $\Delta C_P^\circ(T)$ for the ionisation of cyanoacetic acid in aqueous solution has the form of a distorted parabola. Clarke and Glew,[2] rightly critical of the polynomial data-fitting used in this work, advocated their own Taylor's series expansion method (taking the van't Hoff route), but failed to dismiss the erroneous result.

All "initial curve-fitting by means of a series" methods are hazardous if multiple differentiation is to follow interpolation. Termination of the series assumes such critical importance, and must often be made so early in the series, that the methods become pointless.

We therefore urge a return to the principles of the Everett and Wynne–Jones treatment [3] of acidic dissociation. The principles concerned are not tied to these

---

† Emeritus Professor of Electrochemistry, Birkbeck College, University of London. Present address: 1 Taswell Close, Dover, Kent.

‡ Department of Chemistry, Faculty of Science, University of Khartoum, Sudan.

authors' own equation; on the contrary, they are the basis of a general method of deriving thermodynamic functions from equilibrium constants, or of dealing with any $y(x)$, $y'(x)$, $y''(x)$ ... problem arising from a set of $y(x)$ data of suitable experimental design. This justifies *ab initio* development of the commended method in the following section. The virtues of a data-treatment scheme some three centuries old are emphasised; it is used in conjunction with a procedure commonly used in another field of applied thermodynamics, the " sigma plot method ", which is suitable for adoption because it implies the use of none but rectilinear interpolations.

THE SIGMA PLOT METHOD

(i) DIFFERENCE TABLES [4]

Values of a dependent variable, y, determined at equally spaced values of an independent variable x constitute the first column of a difference table. In the second column are entered " first differences ", namely, those between consecutive values of y. The third column contains differences between consecutive first differences, and so on; each column consists of differences between the entries in the column immediately preceding it. Given a sufficient number of original entries, and provided that y is a continuous function of x, the table ends with a column in which the entries are all the same. If a given column is expressed by a polynomial in x of degree r, the succeeding column conforms to a polynomial of degree $r-1$.

If, on the other hand, the original values of y are subject to random incidence of errors, and are therefore no longer a strictly continuous function of x, compilation of the table proceeds differently. Sooner or later a column is reached in which the entries consist of random numbers, and all significance has been lost. This little used procedure offers a means of scrutinising any experimental (y, x) data for their true worth. It is almost identical with successive differentiations, and shows, without concealment, the sensitivity to errors which must be faced. The debated cyanoacetic acid data [1] are subjected to this test in table 1.

TABLE 1.—DIFFERENCE TABLE FOR CYANOACETIC ACID ΔG_{obs}°; 5 TO 45°C

ΔG_{obs}°/cal	Δ_1	Δ_2	Δ_3	Δ_4
3 111.382				
	58.677			
3 170.059		4.589		
	63.266		−1.779	
3 233.325		2.810		+3.305
	66.076		+1.526	
3 299.401		4.336		−3.578
	70.412		−2.052	
3 369.813		2.284		+3.206
	72.696		+1.154	
3 442.509		3.438		−1.813
	76.134		−0.659	
3 518.643		2.779		+1.726
	78.913		+1.067	
3 597.556		3.846		
	82.759			
3 680.315				
mean	—	3.440	−0.124	+0.569
		±0.700	±1.373	±2.612

Somewhere along the sequence of difference columns, $\Delta_1, \Delta_2 \ldots$, significance is lost, leaving only random numbers generated by the incoherent errors. This is to be looked for in the vicinity of a column containing entries of alternating sign, followed immediately by another column breaking the systematic sequence of falling mean differences. Columns Δ_3 and Δ_4 in table 1 exemplify these effects, and it therefore seems likely that, were it not for the errors, the third differences, Δ_3, would be zero. This being so, Δ_2 would be constant, Δ_1 linear with T, and $\Delta G°$ would be a quadratic function of T. The implication is that the data conform to the Harned and Robinson equation,[5] and so they do, to a good approximation. The difference table alone has excluded the possibility of any strong dependence of $\Delta C_P°$ on temperature, such as inferred by Feates and Ives [1] on the basis of faulty statistics.

The difference table offers options for the initial interpolation necessary to secure statistical advantage. Everett and Wynne–Jones [3] chose first differences, Δ_1. To illustrate the benefit of this choice, use has been made of the fundamental equation

$$\Delta G° = \Delta H_0° + \int \Delta C_P° \, dT - T \int \frac{\Delta C_P°}{T} \, dT - T \Delta S_0° \tag{1}$$

with adoption of realistic values for $\Delta H_0°$, $\Delta S_0°$ and $\Delta C_P°(T)$, to generate a hypothetical,
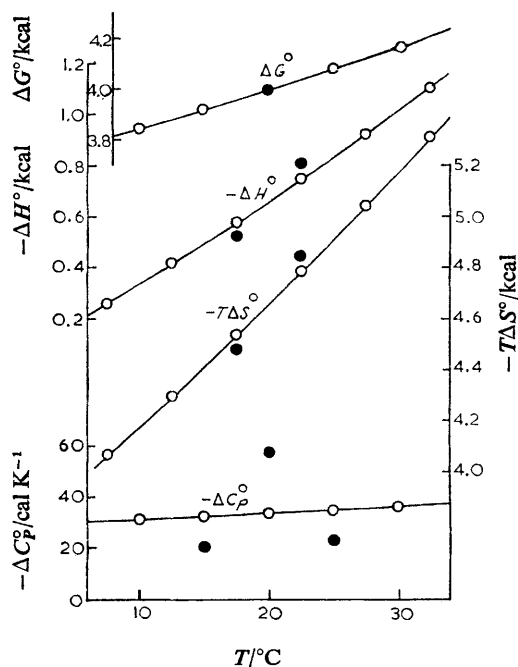

FIG. 1.—Formal representation of the effects of one error in $\Delta G°$ on derived thermodynamic functions of ionisation.

error-free set of thermodynamic functions of ionisation. These are represented by the smooth lines in fig. 1. Introduction of an error of 1 cal (~ 0.025 %; barely perceptible on the scale of plotting) in one $\Delta G°$ has the indicated consequences for the other functions. The characteristic " up and down " double errors in the first derivatives from $\Delta G°$ are noteworthy for their magnitude and diagnostic value.

(ii) EXPLOITATION OF FIRST DIFFERENCES

Note that sporadic errors in y remain incoherent in the first differences, Δy. Equation fitting to Δy, rather than to y (in effect, one step along the differentiation chain), is therefore more effective in minimising distortion due to chance errors of experiment. It weights the systematic part of $y(x)$ heavily in relation to the unsystematic part, and favours the extraction of the best approximation to the true continuous function. Fig. 1 shows the advantage to legitimate smoothing of the fact that one error in y gives rise to two larger errors of opposite sign to each other in Δy. The Everett and Wynne–Jones procedure [3] and the modification of it proposed below exploit this advantage.

All our measurements have been made at precisely equal intervals of temperature. Each experimental $\ln K$ is converted to ΔG° by the identity $\Delta G^\circ = -RT \ln K$. First differences, $\Delta(\Delta G^\circ)$, are tabulated and inspected, and $-\Delta(\Delta G^\circ)/\Delta T$, where $\Delta T = 5$, is equated to ΔS°. This is tantamount to identifying the slope of a chord to the $\Delta G^\circ(T)$ curve between abscissae T and $T+5$ with the slope of the tangent at $T+2.5$, and is an approximation. It involves an error, shown by model calculations, that is confined to the sixth significant figure of ΔS° (i.e., <0.004 %) and varies vary little with temperature, with a consequently imperceptible effect on $d(\Delta S^\circ)/dT$ (~ 0.001 %).

If, alternatively, $\ln K$ were to be retained as dependent variable, the corresponding approximation involving $\Delta(\ln K)/\Delta T$ is less satisfactory because $\ln K(T)$ can be of comparatively high curvature, especially when the temperature of maximum dissociation ($\Delta H^\circ = 0$) occurs within the experimental range. In contrast, the temperature of minimum ΔG° of dissociation must lie some 70 K lower; it is only exceptionally accessible, so that an extremum in $\Delta G^\circ(T)$ is rare and the curvature is, in any case, slight. Of the alternative routes of calculation, the Gibbs–Helmholtz is accordingly preferred. Nevertheless, if there is any point in taking the van't Hoff path, it may be followed, with obvious adaptations, without appreciable penalty.

Values of ΔS° of ionisation obtained by the " first differences method " from the original cyanoacetic ΔG° data [1] are shown as a function of $\ln T$ in fig. 2. The independent variable is chosen to comply with the assumption that $d\Delta C_P^\circ/dT = 0$. If this is untrue in physical fact, and the data are capable of showing that $\Delta C_P^\circ/dT \neq 0$, this will appear later.

The signs of incoherent errors in ΔG° are visible in fig. 2, but their magnitude and distribution suggest that the slope of a least squares straight line through these eight unsmoothed points will give the best available assessment of ΔC°_P; its constant value, or its average value, as the case may be.* The result in this case is $\Delta S^\circ = \Delta S_0^\circ + a \ln T = 211.529\,5 - 39.627\,6 \ln T$ cal K^{-1}, from which the plotted ΔS° values show a mean deviation of ± 0.075 cal K^{-1}, with 5 changes of sign.

(iii) THE SIGMA PLOT

If $\Delta C_P^\circ = a$, eqn (1) becomes

$$\Delta G^\circ = \Delta H_0^\circ - aT \ln T - (\Delta S_0^\circ - a)T. \tag{2}$$

Of the three parameters in this equation, two have been determined and one of them, a, to best advantage. The Σ-plot provides the others. In this adaptation, the device

---

* Everett and Wynne-Jones [3] on the same basis chose $\Delta(T \ln T)/\Delta T$ as independent variable. This eliminates our chord/tangent slope approximation but enhances rounding errors to an extent unacceptable with the facilities at our disposal. In the limit, $d/dT(T \ln T) = \ln T + 1$, with $\Delta T = 5$ K, the differences between $\ln(T+2.5)$ and $(T+5) \ln (T+5) - T \ln T$ is less than 0.000 5 %, so that none but insignificant approximations are involved in choosing the simpler independent variable.

consists of a rearrangement of eqn (2);

$$\Sigma = \Delta G^\circ + aT \ln T = \Delta H_0^\circ - (\Delta S_0^\circ - a)T \tag{3}$$

which should give a good straight line. That it does so is best illustrated by the fact that the first differences, $\Delta\Sigma$ (in which divergences are magnified), are constant to $\pm 0.025 \%$.
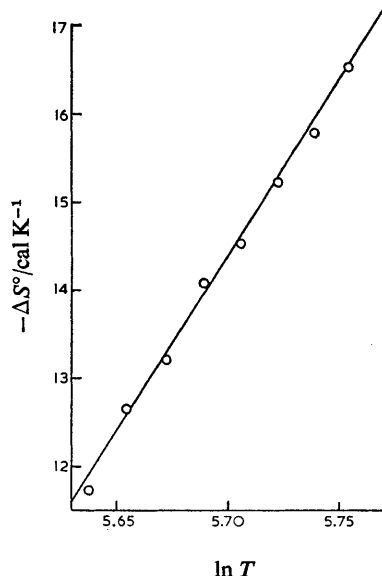

FIG. 2.—Entropies of ionisation of cyanoacetic acid from the "first differences method" as functions of $\ln T$

This gives for cyanoacetic acid an equation for $\Delta G^\circ(T)$, valid for 5 to 45°C;

$$\Delta G^\circ = 10\,932.039\,7 + 39.627\,6\,T \ln T - 251.147\,8\,T \text{ cal} \tag{4}$$

and, for other functions,

$$\Delta H^\circ = 10\,932.039\,7 - 39.627\,6\,T \text{ cal} \tag{5}$$

$$\Delta S^\circ = 211.520\,2 - 39.627\,6 \ln T \text{ cal K}^{-1}. \tag{6}*$$

Results of the completed calculations are shown in table 2.

The square of the correlation coefficient for $\Delta G^\circ(T)$ is $r^2 = 0.999\,999\,975$; it is reasonable to conclude that this derivation method provides a close first approximation. The values of ΔS_0° from the Σ-plot calculation [eqn (6); 211.520 2] and from the "first differences plot" (211.529 5) are so close that the difference between them might be considered negligible, and therefore, the Σ-plot superfluous. In this case, the parameters a and ΔS_0° already to hand without the Σ-plot could be used in conjunction with ΔG_{obs}° in a rearranged eqn (2) to get a set of values for ΔH_0° which could then be averaged. This is not a satisfactory procedure.

* Readers may be critical of the large number of significant figures that are being used. They are necessary to preserve the interrelation of the parameters in such equations, especially when these involve large terms alternating in sign. This mathematical requirement has nothing to do with physical significance. In relation to measured quantities, the policy is to retain digits extending into total insignificance in order to guard every trace of significance in the measurements. Realistic rounding comes at the end.

The Σ-plot is important both in principle and practice because:

(i) it is desirable to call upon the first differences plot only for a slope, and not for an intercept as well;

(ii) the Σ-plot returns to the original data and incorporates them in a rectilinear interpolation that incoherent errors are unable to distort;

(iii) use of the Σ-plot is obligatory if eqn (2) is to be furnished with numerical parameters enabling it to represent the experimental data with the maximum likelihood associated with the least squares principle. This is desired in order to bring the method in line with the well established policy of describing observed behaviour in terms of a dominant ideal " law " on which subordinate deviations are superimposed. The Σ-plot regression line based on $d\Delta C_P^\circ/dT = 0$ is used to define ideal acidic dissociation, and the distribution of the residuals offers a means to study deviations from this ideal. Since the residuals incorporate the incoherent errors, delusions about the severity of experimental limitations to such a study are unlikely to occur.

TABLE 2.—PROVISIONALLY REVISED IONISATION FUNCTIONS OF CYANOACETIC ACID

T/K	ΔG_{obs}°/cal	ΔG_{calc}°/cal	δ/cal	ΔH°/cal	ΔS°/cal K^{-1}
278.15	3 111.382	3 111.212	+0.170	−90.4	−11.51
283.15	3 170.059	3 170.534	−0.475	−288.5	−12.22
288.15	3 233.325	3 233.353	−0.028	−486.7	−12.91
293.15	3 299.401	3 299.612	−0.211	−684.8	−13.59
298.15	3 369.813	3 369.251	+0.562	−882.9	−14.26
303.15	3 442.509	3 442.214	+0.295	−1 081.1	−14.92
308.15	3 518.643	3 518.444	+0.199	−1 279.2	−15.57
313.15	3 597.556	3 597.887	−0.331	−1 477.3	−16.21
318.15	3 680.315	3 680.495	−0.180	−1 675.5	−16.84

$$\pm 0.272$$

$$\Delta C_P^\circ = -39.6 \text{ cal K}^{-1}$$

(iv) STUDY OF DEVIATIONS

A set of residuals heavily loaded with sporadic errors might be written off as unprofitable for study if it were not for the fact that only one kind of distribution is to be looked for as of possible significance. This has been realised as a result of the following model calculations.

Two sets of error-free $\Delta G^\circ(T)$ data, covering an extended range of temperatures symmetrical about 298.15 K, have been generated by adoption of the following parameters;

(A) $d\Delta C_P^\circ/dT = 0$, $\Delta C_P^\circ = -40$ cal K^{-1}, $\Delta H_0^\circ = 11\,000$ cal and $\Delta S_0^\circ = 210$ cal K^{-1}.

(B) $d\Delta C_P^\circ/dT = -0.1$ cal K^{-2}, $\Delta C_P^\circ = -10.185 - 0.1\,T$ cal K^{-1} (= -40 cal K^{-1} at 298.15 K), $\Delta H_0^\circ = 6\,575.051\,6$ cal, and $\Delta S_0^\circ = 70.007\,926$ cal K^{-1}.

The dissociation of hypothetical acid A is ideal (tacitly $d\Delta C_P^\circ/dT = 0$), that of B is not, but their computed $\Delta G^\circ(T)$ values differ from each other only by an average ± 0.6 cal over a temperature range of 80 K. The parameters for B had been chosen to achieve this maximum similarity with the initial purpose of finding out, in principle, how well, or ill, the Σ-plot method could be expected to discriminate between ideal and non-ideal dissociations.

There was no difficulty in recovering the original parameters from which the A set of data points had been generated. The B data, now best regarded as simulating a set of experimental ΔG_{obs}° (T) points, were also subjected to the whole of the Σ-plot

calculational routine, starting with the "first differences method", and continuing with rectilinear least squares regressions of $\Delta(\Delta G^\circ_{obs})/\Delta T$ against $\ln T$ and of $\Delta G^\circ_{obs} + aT \ln T$ against T. The parameters that emerged from this "ideal dissociation treatment" (again tacitly $d\Delta C^\circ_P/dT = 0$) of non-ideal dissociation data naturally "aped" the A parameters; they were $\Delta C^\circ_P = -39.858\,042$ cal K^{-1}, $\Delta H^\circ_0 = 10\,959.570\,6$ cal and $\Delta S^\circ_0 = 209.197\,05$ cal K^{-1}.

These were taken to account for the ideal part of the overall dissociation of B (ignoring the non-ideal part), and were used to derive a set of ΔG°_{calc} (T) figures. The residuals, $\Delta G^\circ_{obs} - \Delta G^\circ_{calc}$, are shown graphically in fig. 3, and are of cubic distribution. The corresponding ΔS° residuals are also shown; consistent with $d(\delta \Delta G^\circ)/dT = -\delta \Delta S^\circ$, they are quadratic in distribution.
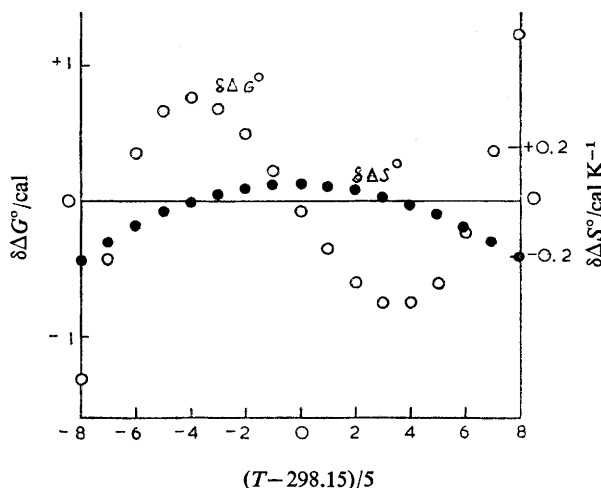

FIG. 3.—Residuals (obs-calc) of non-ideal ΔG° and ΔS° data subjected to ideal interpolation treatment.

This situation arises because the rectilinear regression lines we have used, expressed generally as $y = mx + c$, pass through $\bar{x}$, $\bar{y}$. If the $y(x)$ points being fitted in fact lie on an uninflected shallow curve, this curve will be cut twice by the straight regression line. The residuals, as functions of $(x - \bar{x})$ will then be of predominantly quadratic distribution. This applies to our "first differences method" of assessing a constant or mean ΔC°_P; if the measurements are significantly indicative of temperature dependence of ΔC°_P, this will lead at the end of the Σ-plot calculation to a cubic, or near-cubic distribution of the ΔG° residuals. It is this that is to be looked for.

The question of what further action to take when such a distribution is found has been explored by further use of the hypothetical acid B data in the following way.

Suppose that the Σ-plot treatment has given the ideal components, ΔX°_{ideal}, of the required thermodynamic functions—or the parameters of the analogues of eqn (4)-(6) by which they may be calculated. The ΔG° residuals should provide the means of deriving "thermodynamic excess functions", ΔX^E, for satisfactory completion of the determinations in terms of the summation $\Delta X^\circ_{over-all} = \Delta X^\circ_{ideal} + \Delta X^E$. Viewed alternatively, it should be possible to arrive at the necessary changes to the equations defining each $\Delta X^\circ(T)$. A successful method must, in the imaginative exercise contemplated, recover the synthetic "ΔG°_{obs}" data accurately, and of course, the other functions precisely related to them.

In the absence not only of theoretical guidance, but also of sporadic experimental errors, it is considered justifiable to use empirical orthogonal polynomials (EOP) to express ΔG^E (*i.e.*, the residuals of fig. 3) as a function of T. A 3rd degree EOP accounted for ΔG^E to within ± 0.02 cal ($\sim \pm 0.000\,5\,\%$ of ΔG°_{obs}) over the whole temperature range, but the very small remainders were of unmistakably quartic distribution. These were reduced to less than ± 0.001 cal by a 4th degree EOP fitting, which was adopted, although with some compunction. This was justified by recovery of $d\Delta C^\circ_P/dT = -0.100$ cal K^{-2} at the median temperature of 298.15 K; elsewhere recovery was less exact, with errors rising to 0.005 cal K^{-2} at the extreme temperatures. This not unexpected defect is a limitation of the empirical fitting, which leaves second order residuals of systematic but non-significant distribution, to be accommodated by any reasonable assessment of errors. The result is therefore regarded as satisfactory. Whether or not such a calculational experiment can have any practical application can now be discussed.

(v) ATTEMPTED EXTENSION OF THE Σ-PLOT METHOD

Inspection of the column of residuals in table 1 shows them to have a roughly cubic distribution similar to that of fig. 3, but inverted, suggesting that $d\Delta C^\circ_P/dT$ is positive. To explore this, a 3rd degree EOP has been fitted to the residuals, as before. This requires $\Delta C^\circ_P(T)$ to take the form

$$\Delta C^\circ_P = a + bT + cT^2. \tag{7}$$

It is best to build this information into an extended Σ-plot, with

$$\Sigma = \Delta G^\circ + aT \ln T + b/2T^2 + c/6T^3 \tag{8}$$

to revise the parameters ΔH°_0 and ΔS°_0 of the analogue of eqn (2), namely,

$$\Delta G^\circ = \Delta H^\circ_0 - aT \ln T - (\Delta S^\circ_0 - a)T - b/2T^2 - c/6T^3. \tag{9}$$

This, with $\Delta H^\circ_0 = 13\,241.292$, $a = -39.627\,6$, $\Delta S^\circ_0 = 235.065\,5$, $b = -0.159\,85$ and $c = 0.000\,541\,94$ in appropriate units, gives the results in table 3, to be compared with those in table 2.

TABLE 3.—IONISATION FUNCTIONS OF CYANOACETIC ACID FROM THE EXTENDED Σ PLOT

T/K	ΔG°_{recalc}/cal	δ/cal	ΔH°/cal	ΔS°/cal K^{-1}	ΔC°_P/cal K^{-1}
278.15	3 111.201	+0.181	−77.2	−11.46	−42.2
283.15	3 170.389	−0.330	−286.3	−12.21	−41.4
288.15	3 233.235	+0.090	−491.6	−12.93	−40.7
293.15	3 299.613	−0.212	−693.1	−13.62	−39.9
298.15	3 369.395	+0.418	−890.7	−14.29	−39.1
303.15	3 442.458	+0.051	−1 084.2	−14.93	−38.3
308.15	3 518.677	−0.034	−1 273.5	−15.55	−37.4
313.15	3 597.931	−0.375	−1 458.4	−16.15	−36.5
318.15	3 680.103	+0.212	−1 638.8	−16.72	−35.6

± 0.211

$d\Delta C^\circ_P/dT \simeq +0.16$ cal K^{-2}.

Reduction of the mean deviation is not impressive, but the δ column, with doubled number of sign changes, looks more like a list of irreducible incoherent errors. Significant changes in ΔH° and ΔC°_P suggest that the ideal " constant ΔC°_P model " is not as close an approximation as might have been supposed. It would be unrealistic to quote more than a mean value of $d\Delta C^\circ_P/dT$. With improved accuracy of

measurement, the difference table would give better evidence; as it stands, least squares treatment of second differences gives $d\Delta C_P^{\circ}/dT \sim +0.12$ cal K^{-2}, in fair agreement.

Of two other comparable studies, one (di-isopropylcyanoacetic acid [4]) responded similarly to this treatment, indicating $d\Delta C_P^{\circ}/dT \sim +0.17$ cal K^{-2}. The other (dimethylcyanoacetic acid [7]) gave no identifiably systematic distribution of residuals.

It is evident that the study of residuals does not, as yet, give definitive results, and will not do so until more accurate, numerous and, especially, replicated data become available. We believe that our approach offers the best prospect for progress, owing to the benefits of the first differences plot, and postponement of interpolation of the original data until it can be conducted on the basis of a rectilinear (unextended) Σ-plot, thus isolating an ideal contribution to $\Delta G^{\circ}(T)$ and leaving residuals for the study of deviations, disciplined by the retention of incoherent errors.

If the study of residuals is thought supererogatory, it should be emphasised that such study is, in principle, complementary to the freedom of choosing an arbitrary definition of the ideal contribution. If a definition based on $d/dT(\Delta C_P^{\circ}/T) = 0$, in line with the Harned and Robinson equation [5] is chosen to define ideal behaviour, the appropriate Σ-plot treatment gives, for example, an appreciably different table 2 for the cyanoacetic acid data: on the other hand, the extended Σ-plot gives an inappreciably different table 3. The choice of equation to define ideal behaviour is immaterial in relation to the results of the completed calculations, and the limitations of existing literature data preclude any general preference based on superior conformity with observations.

These studies underline the significance of the distribution of errors. Albeit random in incidence, the errors associated with a small set of unreplicated data-points cannot be described as random in distribution. It is an essential contribution of this paper to show the crucial influence of this fact on the initial curve-fitting procedures that have led and still lead to false conclusions.

INTERPOLATION ERRORS

For empirical interpolation of a set of $y(x)$ data, it does not matter whether or not the polynomial used is orthogonal, how the independent variable may be coded, or where its arbitrarily chosen zero may be placed. The results are the same, and identical with those obtained by alternative procedures (e.g., 3rd to 5th order determinantal, and 2nd to 4th degree EOP solutions coincide). The critical operation is the termination of the series used.

The situation here has been explored by model calculations based on idealised cyanoacetic acid data ($\Delta G_{\text{calc}}^{\circ}$ in table 1), in accord with the Everett and Wynne-Jones equation.[3] These "EWJ data" have been subjected to 2nd to 5th degree EOP fitting, with results shown in the left-hand side of table 4.

TABLE 4.—MEAN DEVIATIONS ($\pm$cal OR $\pm$cal K^{-1}) FROM EOP INTERPOLATIONS OF ERROR-FREE DATA AND EWJ DATA WITH ONE INJECTED ERROR

degree of EOP	error-free EWJ data			EWJ data with one injected error			
	ΔG° id-calc	ΔH° id-calc	ΔC_P° id-calc	ΔG° obs-calc	ΔG° id-calc	ΔH° id-calc	ΔC_P° id-calc
2	0.107 4	8.21	1.48	0.135 4	0.061 1	1.57	0.09
3	0.001 7	0.21	0.06	0.097 8	0.079 7	2.83	0.48
4	0.000 5	0.04	0.01	0.087 5	0.092 7	8.02	2.27
5	0.000 1	0.08	0.03	0.061 6	0.095 7	15.03	5.57

The fitting of the error-free ΔG_{id}° values improves with increasing degree of the EOP; convergence is satisfactory. ΔH° and ΔC_{P}° derived from the interpolation equations also converge, except for the last entries, but these are very sensitive to rounding errors.

Next, one error has been injected: 0.5 cal added to the second data-point of the nine, increasing ΔG° by 0.016 %. The impaired data have been treated as " ΔG_{obs}° " and subjected to the same EOP operations.

The first result is shown in the 5th column of table 4 in terms of the mean $\Delta G_{obs}^{\circ} - \Delta G_{calc}^{\circ}$. The fitting again improves with increasing degree, and if not for the comparison provided by the second column, this might be taken to represent satisfactory convergence. The F-test should dispose of this. The last three columns of the table contain mean deviations relative to the true, unimpaired ΔX_{id}° data. The full extent of the damage is not apparent from the mean deviations but is better shown by the $\Delta C_{P}^{\circ}(T)$ plots in fig. 4(a). For comparison, figs. 4(b), (c) and (d) show
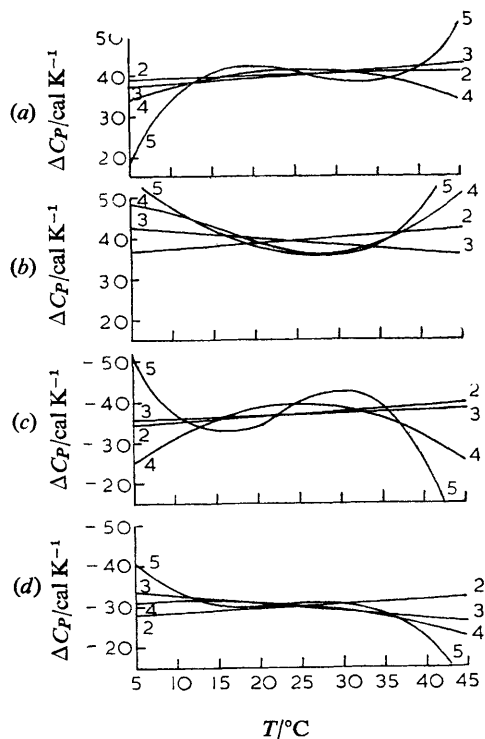

FIG. 4.—$\Delta C_{P}^{\circ}(T)$ derived from 2nd to 5th order EOP fitting of (a) idealised EWJ cyanoacetic acid ΔG° data impaired by one " experimental " error, (b) experimental cyanoacetic acid data,[1] (c) experimental dimethylcyanoacetic acid data,[7] and (d) experimental di-isopropylcyanoacetic acid data.[6]

the results of renewed EOP calculations on the experimental data for cyanoacetic,[1] dimethyl-[7] and di-isopropylcyanoacetic[6] acids respectively. The original arithmetic of Feates and Ives[1] is confirmed, and each example has its own eccentricities arising from 4th and 5th degree fittings.

The matter is settled by putting some figures to the model calculation. The 5th degree orthogonal polynomial can be written

$$\Delta G^{\circ} = \Delta G^{\circ}\xi_0 + b_1\xi_1 + b_2\xi_2 + b_3\xi_3 + b_4\xi_4 + b_5\xi_5. \qquad (10)$$

The coefficients of the ξ_0 to ξ_5 terms for the error-free EWJ data are, in sequence 3 380.333 6, 71.199 6, 1.662 67, -9.352×10^{-3}, 7.05×10^{-5}, 9.94×10^{-6}. As a result of introducing the one error of $+0.5$ cal, they become 3 380.389 1, 71.174 6, 1.666 46, -6.406×10^{-3}, -298.9×10^{-5}, $1\ 772.8 \times 10^{-6}$, from which it is evident that 4th and 5th degree terms have suffered complete subversion, but remain innocuous in relation to the interpolation of $\Delta G°$.

To obtain the derivatives of $\Delta G°$ with respect to temperature, eqn (10) is transformed to

$$\Delta G° = \Delta G_0° + b_1'\theta + b_2'\theta^2 + b_3'\theta^3 + b_4'\theta^4 + b_5'\theta^5 \tag{11}$$

where $\theta = (T - 298.15)/5$, $\Delta G_0°$ is the value of $\Delta G°$ at $\theta = 0$, and the main contributions to the coefficients b_1' to b_5' come from their counterparts b_1 to b_5 in eqn (10). The successive derivatives of $\Delta G°$ with respect to θ contain the terms $5\,b_5'\,\theta^4$, $20\,b_5'\,\theta^3$, and so on. These will be of increasing significance in the $\Delta S°$ and $\Delta C_p°$ values that emerge. In the event that b_5' or b_4' are wrong by several orders of magnitude and even in sign, the effects may be disastrous.

The problem of the distribution of errors remains and is best approached by the ξ_i functions in eqn (10). Except that $\xi_0 = 1$, each is a function of θ: $\xi_1 = \theta$, $\xi_2 = \theta^2 - (n^2 - 1)/12$, where n is the number of data-points, and the others are given by the recursion formula [8]

$$\xi_{r+1} = \xi_r \xi_1 - \frac{r^2(n^2 - r^2)}{4(4r^2 - 1)} \xi_{r-1}.$$

The values of ξ for $n = 9$ are shown in table 5; even and odd members are symmetrical and antisymmetrical respectively about $\theta = 0$.

TABLE 5.—ORTHOGONAL POLYNOMIALS FOR $n = 9$

θ	ξ_1	$3\xi_2$	$5\xi_3$	$7\xi_4$	$3\xi_5$
-4	-4	28	-84	168	-80
-3	-3	7	42	-252	220
-2	-2	-8	78	-132	-80
-1	-1	-17	54	108	-180
0	0	-20	0	216	0

Inspection of this table suggests that where a given error may occur along the temperature scale is not insignificant, although it makes no difference to a computed " standard error ".

This has been studied, using the same EWJ data, by injecting a single $\Delta G°$ error of $+0.2$ cal successively at $\theta = -4$ to 0, and following the least squares procedure to extract the new values of b_1 to b_5 necessary for best fitting.

The effects of the error on b_1 and b_2 are too small for its placing to matter much; the $\pm$percentage changes are 0.01 and 0.2 respectively. For b_3 on the other hand, the percentage changes resulting from insertion of the error at $\theta = -4$ to 0 are, in sequence, -25.2, $+12.6$, $+23.4$, $+16.2$ and zero. For b_4 the mean percent change over all θ exceeded $\pm 1\ 000$ %, so the study has not been pursued further.

Therefore the distribution of errors is of importance, and requires closer examination than that afforded by the statistical formalism of the F-test. One unfortunately placed, rather larger than normal, error might well lead to a deceptively favourable variance ratio.

These difficulties must always be associated with a restricted number of data-points. Table 5 shows that the polynomial term of rth degree changes sign r times over the range of θ, and that there is an unacceptable hazard to significance (or of

gross distortion of derived functions) unless $n \gg r_{max}$. The implication is that in the context of the present field of $\Delta G°(T)$ or $\ln K(T)$, termination of the polynomial fitting must be made at no higher than 3rd degree.

It is of interest that a 3rd degree EOP fits the ideal eqn (2) with considerable accuracy. If a polynomial of the same degree also fits non-spurious residuals, it should also fit the original data and lead to acceptable final results. This is true for the small number of examples studied, but such a routine is not recommended because it avoids the critical assessment of data, and lacks the safety factors and versatility of the Σ-plot approach.

Although this discussion mainly concerns the EOP procedure, it is believed to apply, *mutatis mutandis*, to the fitting of any equation by whatever method to a small number of data-points subject to incoherent errors. This includes the method proposed by Clarke and Glew,[2] which could have been developed along the Gibbs–Helmholtz, rather than the van't Hoff route. Had this been done, the primary equation would have been a Maclaurin series, namely,

$$\Delta G_\theta° = \Delta G_0° + \theta \frac{d\Delta G_0°}{d\theta} + \frac{\theta^2}{2!} \frac{d^2 \Delta G_0°}{d\theta^2} + \cdots \tag{12}$$

which is identical with eqn (11).

As few as nine unreplicated data-points cannot contain information beyond, if as far as, $d\Delta C_p°/dT$. It is regretted that there has been failure in the past to heed the warning that " dependence on statistical methods . . . where the necessary conditions for the application of the mathematical theory of probability are not present, may not only make the results obtained useless but completely misleading." [9] Unfortunately, this was almost certainly the case for previously reported [10] high dependences of $\Delta C_p°$ on temperature.

We suggest that in this field, and perhaps others, initial curve fitting could well be replaced by a simple operation in conformity with the precept " Differentiate then smooth ; never smooth then differentiate."

[1] F. S. Feates and D. J. G. Ives, *J. Chem. Soc.*, 1956, 2798.
[2] E. C. W. Clarke and D. N. Glew, *Trans. Faraday Soc.*, 1966, **62**, 539.
[3] D. H. Everett and W. F. K. Wynne-Jones, *Trans. Faraday Soc.*, 1939, **35**, 1380.
[4] E. Whittaker and G. Robinson, *The Calculus of Observations*, (Blackie, London, 1944), 4th edn.
[5] H. S. Harned and R. A. Robinson, *Trans. Faraday Soc.*, 1940, **36**, 973.
[6] D. J. G. Ives and P. D. Marsden, *J. Chem. Soc.*, 1965, 649.
[7] D. J. G. Ives and P. G. N. Moseley, *J. Chem. Soc.*, 1970, 1655.
[8] C. A. Bennett and N. L. Franklin, *Statistical Analysis in Chemistry and Chemical Industry* (Wiley, London, 1954), p. 257.
[9] ref. 8, p. 5.
[10] D. J. G. Ives and D. Prasad, *J. Chem. Soc.*, 1970, 1652.