Association Constants of 1 : 1 and 2 : 1 Molecular Complexes from Spectrophotometric Data; Experimental Design and Reliability of the Parameters

Gino Carta* and Guido Crisponi
Istituto di Chimica Generale, Inorganica e Analitica, Università di Cagliari, via Ospedale 72, 09100 Cagliari, Italy

Principles for the design of an experiment for the determination of association constants (K_1,K_2) and molar extinction coefficients (ε_1,ε_2) for molecular complexes of 2 : 1 stoicheiometry are discussed. The method employed is based on the maximization of the determinant of the information matrix |X′X| and is of general validity. It is shown that the reliability of a parameter depends on two independent factors: the available experimental apparatus and the experimental design (*i.e.* how the initial concentrations of the reagents are chosen). The correlations between parameters and to some extent the convergence of the routine of minimization are also dependent on the experimental design.

During the last twenty years much attention has been devoted to the reliable determination of association constants for molecular complexes of 1 : 1 and 2 : 1 stoicheiometry. Complexes of the 1 : 1 type in particular have been studied by many authors[1–10] and our knowledge about them is well defined, though not all the problems have been completely solved. It has not been clearly stated, for example, how to achieve optimal experimental design when experimental errors are statistically constant in absolute but not in relative value.[7,8] The situation for 2 : 1-type complexes is somewhat worse: it is not known how to achieve optimal design even if relative experimental errors are kept constant, however, over the last fifteen years there have been helpful contributions from various authors.[11–17]

Lenkinski *et al.*,[14] in particular, have proposed criteria for the best determination of each unknown parameter, and have suggested a sensible experimental approach. The difference curves of Maier and Drago[17] are of particular interest from a practical point of view; these workers, as well as Lingane and Hugus,[12] paid particular attention to partial correlation coefficients between parameters, statistical indicators of prime importance in this kind of problem.

In the present paper we describe a method, peculiar to the field of regression design,[18] which can greatly improve parameter reliability in the determination of association constants of 2 : 1 molecular complexes from spectrophotometric data. For this purpose we assume the following conditions to be valid.

(i) The interaction between the two reagents gives only molecular complexes of the 1 : 1 and 2 : 1 types [equations (i) and (ii)], the absorption is only due to the species AB and A_2B,

$$A + B \overset{K_1}{\rightleftharpoons} AB \quad \text{(i)} \qquad\qquad AB + A \overset{K_2}{\rightleftharpoons} A_2B \quad \text{(ii)}$$

(respective extinction coefficients ε_1 and ε_2), and Beer's law is valid;

(ii) The solution acts ideally (*i.e.* each species has an an activity coefficient equal to unity, or stoicheiometric stability constants are considered); furthermore autoassociation of reagents and interaction between reagents and solvent are excluded.

Though all these conditions are not always simultaneously and rigorously valid in real systems, they are perhaps not too restrictive, yet the conclusions and information drawn can be helpful in many situations, particularly where the determination of the parameters is difficult.

Calculation and Reliability of Parameters.—Some of the data used in this work have been tested with a Fortran program based on the Gauss–Newton method as modified by Marquardt.[19] This version, in fact, works better than the orginal,[20] especially with data from a poorly designed system (see later). The original Gauss–Newton method will now be outlined briefly, thus making it easier to understand the subsequent treatment.

The values of the four parameters (K_1, K_2, ε_1, ε_2 or interchangeably P_1, P_2, P_3, P_4) that best fit the assumed model can be obtained by minimizing the expression (1), where a_i and b_i

$$\text{CHISQ} = \sum_{i=1}^{N} [y_i - y(a_i,b_i,P_1,\ldots,P_m)]^2 w_i \qquad (1)$$

are the initial concentrations of the two reagents A and B, y_i and $y(a_i,b_i,P_1,\ldots,P_m)$ are respectively the experimental and calculated absorbances for unit path length, w_i is the weight of the ith point, and m and N are the numbers of unknown parameters and experimental points, respectively.

The calculated absorbance $y(a_i,b_i,P_1,\ldots,P_m)$ is given by the relation (2), where u_i and v_i are the concentrations of the species

$$y(a_i,b_i,P_1,\ldots,P_m) = u_i\varepsilon_1 + v_i\varepsilon_2 \qquad (2)$$

AB and A_2B and are defined by the equations (3) and (4).

$$K_1 = u/[(a - u - 2v)(b - u - v)] \qquad (3)$$

$$K_2 = v/[(a - u - 2v)u] \qquad (4)$$

The elimination of v between equations (3) and (4) gives equation (5), which can be solved analytically (or numerically

$$(1 - 4K_2/K_1)K_2u^3 + (1 - 2bK_2 - 4K_2/K_1)u^2 + \\ -[a + b + 1/K_1 + K_2a(a - 2b)]u + ab = 0 \qquad (5)$$

by the Raphson-Newton method) and the value of u can be substituted in equation (4) to obtain v.

The substitution of equation (2) in equation (1) would give an equation insoluble by the ordinary least-squares method, because although it is linear in ε_1 and ε_2 it is not in K_1 and K_2. However, in equation (1) by replacing $y(a_i,b_i,P_1,\ldots,P_m)$ with the Taylor series truncated to the first term, we obtain equation (6), where y_i^0 is a shortened notation that stands for

$$\text{CHISQ} = \sum_{i=1}^{N} [y_i - y_i^0 - \sum_{j=1}^{m} (\delta y_i^0/\delta P_j \Delta P_j)]^2 w_i \qquad (6)$$

$y(a_i,b_i,P_1{}^0,\ldots,P_m{}^\circ)$, superscript zero indicates initial guessed values of the parameters, and $\Delta P_j = P_j - P_j{}^0$ is the correction term to be applied to $P_j{}^0$.

Equation (6) is linear in ΔP_j and therefore if it is minimized with respect to each of these parameter increments, we finally obtain a set of four simultaneous equations (7) (Normal Equations).

$$t_k = \sum_{j=1}^{m}(c_{jk}\Delta P_j) \qquad k = 1,m \tag{7}$$

where

$$t_k = \sum_{i=1}^{N}[(y_i - y_i{}^0)\delta y_i{}^0/\delta P_k]w_i$$

and

$$c_{jk} = \sum_{i=1}^{N}(\delta y_i{}^0/\delta P_j\delta y_i{}^0/\delta P_k)w_i$$

Equation (7) can be solved easily by standard methods and gives the increments to be applied to each parameter in order to obtain a better estimate. Generally, it is necessary to iterate the process until no change is observed in CHISQ or in the parameters, the reason being that the problem can be considered approximately linear only near the minimum.

Let us define a design matrix* X of order $N \times m$ [equation (8)]. Then equation (7) can be written in more compact form as

$$X = [x_{ij}] = [\delta y_i{}^0/\delta P_j w_i{}^{\frac{1}{2}}] \tag{8}$$

(9), where the apex denotes the transpose of a matrix, and $\bar{\Delta}P$

$$X'\bar{\Delta}y = [X'X]\bar{\Delta}P \tag{9}$$

and $\bar{\Delta}y$ are column vectors of the parameter increments $P_j - P_j{}^0$ (of order $m \times 1$) and of the differences $(y_i - y_i{}^0)w_i{}^{\frac{1}{2}}$ (of order $N \times 1$), respectively.

From equations (7) and (9) we see that $[X'X] = C = [c_{jk}]$, which is known as the information matrix, while the inverse $[X'X]^{-1} = C^{-1} = [c_{jk}{}^{-1}]$ is known as the dispersion matrix.

The variance $S_j{}^2$ of the *j*th parameter, and hence its reliability, is given by equation (10), where $S^2 = \text{CHIMIN}/(N - m)$,

$$S_j{}^2 = S^2 c_{jj}{}^{-1} \tag{10}$$

$N - m$ is the number of degrees of freedom of the estimate S^2, and CHIMIN is the minimum value of CHISQ. Generally, in non-linear cases, S^2 is not an unbiased estimate of the population variance σ^2 and should be obtained in an independent way, for example from repeated observations.[21a] However, it is possible to have a measure of the linearity of the system studied and therefore of the reliability [21b] of S^2.

As can be seen in equation (10), $S_j{}^2$ is a function of two completely independent factors: S^2 and $c_{jj}{}^{-1}$. The former (S^2) is a measure of the experimental error of the data and depends on the available experimental apparatus; the latter ($c_{jj}{}^{-1}$) instead depends on which initial concentrations of reagents A and B are chosen, and therefore on the experimental design. It is obvious that both S^2 and $c_{jj}{}^{-1}$ should be kept at their minimum values within experimental constraints.

Another very important quantity for assessing the quality of the data is the multiple correlation coefficient R_j (or to a lesser degree r_{jk}, the partial correlation coefficient); $R_j = [1 - 1/(c_{jj}c_{jj}{}^{-1})]^{\frac{1}{2}}$ is a measure of how an error in P_j can be counterbalanced by a variation in the other parameters, while $r_{jk} = c_{jk}{}^{-1}/(c_{jj}{}^{-1}c_{kk}{}^{-1})^{\frac{1}{2}}$ is a measure of how an error in P_j can be compensated by a variation in P_k, the effect of the other parameters being eliminated. Both $R_j{}^2$ and $r_{jk}{}^2$ can vary between 0 and 1, and $r_{jk}{}^2 \leqslant R_j{}^2$. If $R_j{}^2 = 1$, then the same set of N values of $y(a_i,b_i,P_1,\cdots,P_m)$ could be obtained for any P_j, compensating the variation in P_j with suitable values of the other parameters; in such a situation, an infinite number of P_j values would give the same CHIMIN and clearly P_j would be completely indeterminate. If $R_j{}^2$ is only a little smaller than 1, then in the same way as before, large ranges of P_j would give about the same CHIMIN and therefore P_j would be poorly determined. On the other hand if $R_j{}^2 = 0$, then even a small change in P_j would give a well defined variation in CHIMIN, not open to compensation by the other parameters, and consequently P_j would be precisely determined. It is thus evident that a good design should keep values of $R_j{}^2$ or $r_{jk}{}^2$ as low as possible.

Experimental Design.—In general an *m*-point design is sufficient for the determination of *m* unknown parameters in a functional relationship. Clearly in such a case S^2 should be estimated in an independent way. If $w_i = 1$ then $[x_{ij}] = [\delta y_i/\delta P_j]$, as can be seen from relation (8). This means that in the *m*-dimensional space the *j*th co-ordinate ($j = 1,m$) of the *i*th point will be $x_{ij} = \delta y_i{}^0/\delta P_j$, since $\delta y/\delta P_j$ is the independent variable associated with the *j*th parameter.

If each of the *m* independent variables could vary in absolute value between 0 and 1, an excellent *m*-point design (orthogonal design) would be as shown in (11). This would amount to

$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \tag{11}$$

picking out the first point (1 0 0 0), such that all its derivatives would be zero except the one with respect to K_1. All the derivatives of the second point (0 1 0 0) should be equal to zero except the one with respect to K_2, and so on for the third and fourth points. In such a situation each point would contribute to the determination of only one parameter: the first for K_1, the second for K_2, and so on.

Both the information and the dispersion matrices in this case would be equal to the design matrix (11) and both R_j and r_{jk} values would be zero, thus showing that the parameters would be determined each independently of one another.

In practice, experiment (11) is unattainable because the independent variables are strictly correlated and cannot vary independently of one another. It is necessary, therefore, to look for some other approach. Probably the most used is based on a criterion [18,22] which maximizes the determinant of the information matrix $|X'X|$. In fact, designs which are optimal with respect to this criterion are always good from various points of view: they have in particular low variances for the parameters and low correlation coefficients.

In general for non-linear problems it is not possible to have optimally designed experiments *a priori*, because the entries in the design matrix $[x_{ij}] = [\delta y_i{}^0/\delta P_j]$ depend on the parameters of the model, which are now known in advance. It is therefore necessary to follow a sequential approach, *i.e.* an iterative procedure where the information available at a particular stage (approximate estimates of parameters) is used to design the experiment and thus to obtain better estimates which will, in turn, allow a more adequate design.†

---

* The derivatives have been calculated numerically using a computer with double precision arithmetic. The formula used was

$$\delta y_i{}^0/\delta P_j \simeq [y(a_i,b_i,P_1{}^0,\ldots,P_j{}^0+0.001P_j{}^0,\ldots,P_m{}^0) - y(a_i,b_i,P_1{}^0,\ldots,P_j{}^0-0.001P_j{}^0,\ldots,P_m{}^0)]/(2\;0.001P_j{}^0).$$

† Once the estimates of the parameters are unchanging the points can be measured and added in small groups though each point must be found singly.

Table 1. The results of some simulated experiments, for the two types of designs considered

K_1	K_2	c_{11}^{-1}/K_1^2	c_{22}^{-1}/K_2^2	$c_{33}^{-1}/\varepsilon_1^2$	$c_{44}^{-1}/\varepsilon_2^2$	R_{K1}^2	R_{K2}^2	$R_{\varepsilon1}^2$	$R_{\varepsilon2}^2$	Design
1.0	1.0	3.80	78.0	3.41	19.4	0.985	0.912	0.985	0.907	α
		1.69	25.6	1.02	3.99	0.936	0.829	0.937	0.831	β
1.0	10.0	3.88	8.39	3.43	0.886	0.984	0.863	0.980	0.691	α
		1.62	3.88	1.02	0.328	0.940	0.775	0.916	0.513	β
1.0	0.1	3.80	7 020.0	3.42	2 150.0	0.985	0.969	0.986	0.967	α
		1.68	920.0	1.03	240.0	0.936	0.902	0.953	0.931	β
1.0	0.2501	3.80	1 088.0	3.42	319.0	0.985	0.958	0.986	0.957	α
		1.67	174.4	1.03	41.2	0.935	0.870	0.949	0.904	β

Table 2. The least-squares ouput for each appropriate stage of the worked example (*i.e.* each time the design is updated)

n	log a	log b	Absorbance	K_1/l mol^{-1}	K_2/l mol^{-1}	ε_1/l mol^{-1} cm^{-1}	ε_2/l mol^{-1} cm^{-1}	c_{11}^{-1}/K_1^2	c_{22}^{-1}/K_2^2	$c_{33}^{-1}/\varepsilon_1^2$	$c_{44}^{-1}/\varepsilon_2^2$	Det.	CHIMIN × 10^6
1	−4.00	0.0	0.257										
2	−4.00	−0.60	0.225										
3	−4.00	−1.00	0.180										
4	−4.00	−1.40	0.120										
5	−1.40	−4.00	0.136										
6	−1.00	−4.00	0.223										
7	−0.60	−4.00	0.308										
8	0.0	−4.00	0.393	20.13	2.96	2 696	4 413	170	910	13	29	0.43×10^{-24}	0.12
9	−2.20	−1.80	3.880	20.02	2.98	2 699	4 409	18	700	10	25	0.32×10^{-21}	0.31
10	−2.80	−0.20	3.964	20.00	2.98	2 700	4 409	0.57	670	0.08	25	0.41×10^{-19}	0.34
11	0.0	−3.00	3.925	20.00	2.99	2 700	4 401	0.55	530	0.08	6.3	0.39×10^{-17}	0.53
12	−1.0	−2.80	3.518	20.00	3.00	2 700	4 399	0.31	7.1	0.08	0.23	0.37×10^{-15}	0.55
13	−2.20	−1.80	3.880										
14	−2.80	−0.20	3.964										
15	−1.0	−2.80	3.518										
16	0.0	−3.00	3.925	20.00	3.00	2 700	4 399	0.16	3.6	0.04	0.11	0.59×10^{-14}	0.55

A sensible approach might consist of the following steps.

(1) Select a starting design by collecting some points from Lenkinski's[14] titration curves 1 and 2 (titration curve 1 keeps the initial concentration a constant while varying b, and titration curve 2 keeps b constant while varying a).

(2) Use the available experimental points and a non-linear method of least squares (here the Gauss–Newton) to estimate the parameters and their variances. If the variances are sufficiently low stop the experiment, otherwise go to step (3).

(3) Construct the design matrix $[x_{ij}] = [\delta y_i^0/\delta P_j]$ of order $n \times m$, making use of the n points already collected and of the m parameter estimates of the last least-squares run. Find the point of the region experimentally attainable that, added to $[x_{ij}]_{n \times m}$, gives the matrix $[x_{ij}]_{(n+1) \times m}$, for which the determinant $|X'X|$ is a maximum.

(4) Measure the absorbance corresponding to the new point and go to step (2).

In Table 1 the results of some simulated experiments for two types of design and different sets of parameters are presented. The absorbances have a statistically constant relative error* (*i.e.* $w_i = 1/y_i^2$) such that the weight of an observation does not depend too much on its particular value. The experimental points (22 for each experiment) have been chosen from a 7 × 16 grid ($-3 \leqslant \log a, \log b \leqslant 0$; $\Delta\log a = 0.2$, $\Delta\log b = 0.5$) of the region supposed experimentally attainable; $\varepsilon_1 = \varepsilon_2 = 10\,000$ l mol^{-1} cm^{-1}. The type α design follows the conventional Lenkinski titration curves 1 ($\log a = -3$, $\Delta\log b = 0.5$) and 2 ($\log b = -3$, $\Delta\log a = 0.2$); type β chooses the starting design (7 points) in the same way ($\log a = -3$, $\Delta\log b = 1$ and $\log b = -3$, $\Delta\log a = 1$) and then follows the procedure of maximizing the determinant of the actual information matrix, as previously described. In all cases the values of the relative variances (apart from S^2) c_{jj}^{-1}/P_j^2 and square multiple correlation coefficients R_j^2 are better for type β than for type α designs and the difference between them increases as the indeterminability of the system increases.

The points for both α and β type designs, and $K_1 = K_2 = 1$ l mol^{-1}, are shown in the Figure.

Worked Example.—In order to illustrate better the proposed scheme an example from the literature[13] will be re-worked. The species A and B are hexamethylbenzene and fluoranil respectively, and the solvent is carbon tetrachloride; the values of K_1 and K_2 are 20.00 and 3.00 l mol^{-1}; those of ε_1 and ε_2 are 2 700 and 4 400 cm^{-1} l mol^{-1}, respectively. In the ($\log a$, $\log b$) plane a 21 × 21 grid ($-4.0 \leqslant \log a, \log b \leqslant 0.0$; $\Delta\log a = \Delta\log b = 0.2$) is made. The points in the grid that have an absorbance between 0.1 and 4.0 units (cell of unit path length) are experimentally attainable and are here indicated as candidates.

The first 8 simulated experimental points (starting design) are collected according to the titration curves 1 and 2 of Lenkinski, keeping the initial concentration of the species B constant and varying those of the species A, and then *vice versa*. In the next step the parameters and all the other quantities listed in Tables 2 and 3 are calculated, by the Fortran program GAUSS, based on the Gauss–Newton method as modified by Marquardt.[19] An

---

* Absorbances could rigorously have a constant relative error if they were all recorded at the same value (one for simplicity) by means of cells with variable path length L_i; if $y(a_i,b_i,P_1,\ldots,P_m)$ is the true value of absorbance for unit path length then $L_i = 1/y(a_i,b_i,P_1,\ldots,P_m) \simeq 1/y_i = w^{\frac{1}{2}}$, and therefore $w_i = 1/y_i^2$. In such a case $[x_{ij}] = [\delta y_i^0/\delta P_j(1/y_i)]$ and should also be $(\sum_{j=1}^{m} x_{ij}^2)^{\frac{1}{2}} \simeq 1$ for all points (*i.e.* in the m-dimensional space of the measurements, all points should lie in a hypersphere of units radius), as in problems linear in the parameters, because each point should count equally. This condition, however, is not complied with.

Table 3. The partial correlation coefficients $r(J,K)$ and the square multiple correlation coefficients R_j^2, for each appropriate stage of the worked example

n	$r(K_1,K_2)$	$r(K_1,\varepsilon_1)$	$r(K_1,\varepsilon_2)$	$r(K_2,\varepsilon_1)$	$r(K_2,\varepsilon_2)$	$r(\varepsilon_1,\varepsilon_2)$	$R_{K_1}^2$	$R_{K_2}^2$	$R_{\varepsilon_1}^2$	$R_{\varepsilon_2}^2$
8	−0.42	−0.73	0.37	0.07	−0.85	−0.18	0.70	0.80	0.65	0.76
9	−0.09	−0.99	0.09	−0.20	−0.83	−0.008	0.99	0.85	0.99	0.73
10	−0.67	−0.64	0.50	0.13	−0.84	−0.11	0.77	0.84	0.59	0.73
11	−0.66	−0.64	0.67	0.12	−0.99	−0.14	0.76	0.99	0.59	0.99
12	−0.16	−0.74	0.18	−0.16	−0.78	−0.015	0.64	0.69	0.65	0.63
16	−0.16	−0.74	0.18	−0.16	−0.78	−0.014	0.64	0.69	0.65	0.63

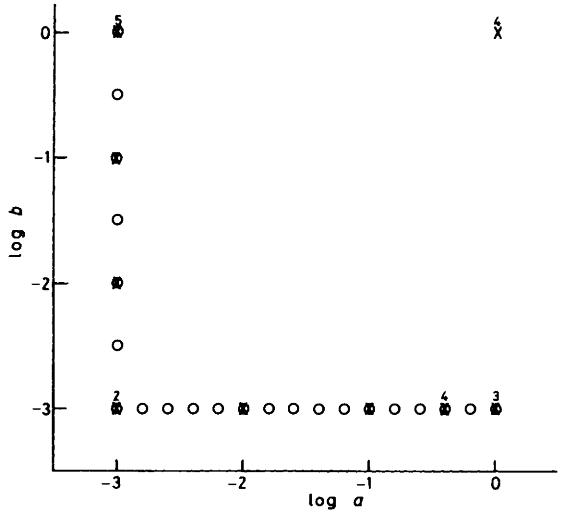

Figure. The points used in the case $K_1 = K_2 = 1$ (l mol^{-1}) for both type α (○) and type β design (×) in the plane (log *a*, log *b*). The numerals adjacent to some points indicate how many times the point is used in design β

estimate of the parameters being available, it is now possible to choose the point that, added to the experimental points already collected, gives a maximum value to the determinant $|X'X|$.

This can be done in two different but equivalent ways. The first method adds each candidate, one at a time, to the current design matrix $[x_{ij}]_{n\times m}$, and then calculates the determinant of the resulting information matrix $|X'X|$, $[x_{ij}]$ being now of order $(n + 1) \times m$. In this way as many determinants as there are candidates must be calculated. The point giving the maximum value of $|X'X|$ is the one to be experimentally determined and so actually added to the design.

The second method calculates the dispersion (or variance–convariance) matrix $[X'X]^{-1}$ of the current design $[x_{ij}]_{n\times m}$, and with this matrix * it estimates the variance of each candidate; the point having the maximum variance will be chosen and therefore experimentally measured. We have opted for this second method, making use of a Fortran program named CVAR, which, of course, must have as input the parameter estimates of the last least-squares run from which to calculate the derivatives of $[x_{ij}] = [\delta y_i^0/\delta P_j]$.

The least-squares run performed on the first 8 experimental points (*i.e.* the starting design) gives the output listed in Tables 2 and 3 corresponding to the rows having $n = 8$. The addition of the four subsequent points (by means of the program CVAR) causes each time a dramatic change in at least one value of c_{jj}^{-1}/P_j^2, as can be seen in Table 2. When n is 12 the parameters are already absolutely constant and the relative variances (apart from S^2) have decreased by a factor greater than 250. The last four points which are added simultaneously (though chosen singly; this is possible because the parameters are already stable) cause only a moderate improvement in the relative variances; thus the experiment is terminated.

In Table 3 the values of the square multiple correlation coefficients R_j^2 and of the partial correlation coefficients $r(J, K)$ for each least-squares calculation are presented. Both $r(J, K)$ and R_j^2 values at first vary erratically, but once the first fou points are added, they generally show a decrease; they remain unchanged in two cases, and $r(K_2, \varepsilon_1)$ increases though its value is still very low. The replication of the last four points has no effect on the correlation coefficients.

Conclusions.—The approach outlined, if carefully used, will extract from the system all the available information, within the experimental constraints. It is of general validity, so that it can be applied to complexes of whatever stoicheiometry; in particular it can be used for the 1 : 1 case and for n.m.r. data. Its utility is more evident in cases where some parameters are highly correlated; in fact the maximum determinant method will automatically choose those point, in the region of operability, which contribute most to a precise determination of these parameters.

In order to take full advantage of the potential of the method it is helpful to make a close-mesh grid of the whole region of operability in the (log *a*, log *b*) plane; in difficult cases, valuable information is found as to where to locate and possibly enlarge (for example by resorting to cells of different lengths) the area which is most important for parameter determination.

A square partial correlation coefficient $r_{jk}^2 \simeq 1$ implies that in the (P_j,P_k) plane (or in the P_j axis for $R_j^2 \simeq 1$) curves can be drawn of about the same value of CHIMIN, because the parameters can compensate one another to a great extent; it is evident that in such a situation, especially with data having a large experimental error, even the best computer program is likely to fail to converge. Thus it is manifest that program convergence, parameter reliability, and parameter correlation are all different aspects of the same problem and in some way dependent on the experimental design, a change in which can therefore markedly improve them.

References

1 N. J. Rose and R. S. Drago, *J. Am. Chem. Soc.*, 1959, **81**, 6138.
2 P. R. Hammond, *J. Chem. Soc.*, 1964, 479.
3 W. B. Person, *J. Am. Chem. Soc.*, 1965, **87**, 167.
4 R. A. La Budde and M. Tamres, *J. Phys. Chem.*, 1970, **74**, 4009.

---

* The point in the region of operability which maximizes $|X'X|$ is the one which has the maximum variance.[22] A point i of co-ordinates $x_{ij} = \delta y_i^0/\delta P_j w_i^{\frac{1}{2}}$ with $j = 1,m$ (*i.e.* the column vector $\bar{x}_i$) has variance $\mathrm{var}[y(a_i,b_i,P_1,\ldots,P_m)] = \bar{x}_i'\,[X'X]^{-1}\bar{x}_i$, which is easily computed. In order to find this new point to add to the design it is then sufficient to calculate the variance of a list of possible points representative of the whole region experimentally attainable.

5 S. D. Christian, E. H. Lane, and F. Gasland, *J. Phys. Chem.*, 1974, **78**, 557.

6 K. Conrow, G. D. Johnson, and R. E. Bowen, *J. Am. Chem. Soc.*, 1964, **86**, 1025.

7 D. A. Derenleau, *J. Am. Chem. Soc.*, 1969, **91**, 4044.

8 G. Carta, G. Crisponi, and V. Nurchi, *Tetrahedron*, 1981, **37**, 2115.

9 G. Carta and G. Crisponi, *J. Chem. Soc., Perkin Trans. 2*, 1982, 53.

10 G. Carta, G. Crisponi, and A. Lai, *J. Magn. Reson.*, 1982, **48**, 341.

11 D. A. Derenleau, *J. Am. Chem. Soc.*, 1969, **91**, 4050.

12 P. J. Lingane and Z. Z. Hugus, *Inorg. Chem.*, 1970, **9**, 757.

13 B. Dodson, R. Foster, and A. A. S. Bright, *J. Chem. Soc. B*, 1971, 1283.

14 R. E. Lenkinski, G. A. Elgavish, and J. Reuben, *J. Magn. Reson.*, 1978, **32**, 367.

15 G. D. Johnson and R. E. Bowen, *J. Am. Chem. Soc.*, 1965, **87**, 1655.

16 B. L. Shapiro and M. D. Johnston, *J. Am. Chem. Soc.*, 1972, **94**, 8185.

17 T. O. Maier and R. S. Drago, *Inorg. Chem.*, 1972, **11**, 1861.

18 G. E. P. Box and H. L. Lucas, *Biometrika*, 1959, **46**, 77.

19 F. R. Bevington, 'Data Reduction and Error Analysis for the Physical Sciences,' McGraw-Hill, New York, 1969, pp. 232—242.

20 F. R. Hartley, C. Burgess, and R. Alcock, 'Solution Equilibria,' Ellis Horwood, Chichester, 1980, pp. 93—95.

21 N. Draper and H. Smith, 'Applied Regression Analysis,' Wiley, New York, 1966; (*a*) pp. 26—31; (*b*) pp. 273—274, 282—284.

22 O. Dykstra, *Technometrics*, 1971, **13**, 682.

Received 28*th January* 1985; *Paper* 5/148