

$$\Delta(\sin \theta) = 0.35\lambda/4R = 0.14\lambda/L_a. \quad (18)$$

This is close to the same peak shift obtained previously by Warren (1941).

The results obtained here are rigorous within the conditions assumed. We have postulated disk shaped graphite layers all of one size, with no allowance for a size distribution. The results are applicable only for graphite layers which are large enough to give well developed hk reflections for which it is possible to make a breadth measurement which is not influenced by neighboring peaks. It is also postulated that the layers have random orientation about the layer normal, with no modulations in the peak due to neighboring layers assuming the graphite orientation.

For samples in which the layers are very small, it is necessary to develop the diffraction pattern from the general Debye scattering equation. The problem has been treated in this way by Diamond (1957) and Warren & Bodenstein (1965). Both of these treatments follow Diamond's convention, which in terms of areas of disks, amounts to defining an $L_a = 1.9R$. This differs from the average dimension $L_a = (\pi/2)R$ used in this paper. The difference is about 17%, and it must be

kept in mind in comparing results from the two treatments.

This work was done in part at the Computation Center at Massachusetts Institute of Technology, Cambridge. We are indebted to Mr F. A. Heckman of Cabot Corporation for discussions which suggested the need for a more rigorous treatment of the shape of the hk reflections. We are also indebted to Professor F. B. Hildebrand of the M.I.T. Mathematics Department, who pointed out the use of the Struve function in this treatment. The contribution by one of us (P.B.) has been sponsored by the South African Atomic Energy Board.

References

- CLARKE, J. S. (1964). Private communication.
 DIAMOND, R. (1957). *Acta Cryst.* **10**, 359.
 WARREN, B. E. (1941). *Phys. Rev.* **59**, 693.
 WARREN, B. E. (1959). *Prog. Metal Phys.* Vol. VIII, page 199.
 WARREN, B. E. & BODENSTEIN, P. (1965). *Acta Cryst.* **18**, 282.
 WATSON, G. N. (1952). *Theory of Bessel Functions*. Cambridge Univ. Press.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 245.

Acta Cryst. (1966). **20**, 605

Correlation of Observations in the Least-Squares Calculation for the Analysis of Gas Electron Diffraction Data

BY YOSHITADA MURATA AND YONEZO MORINO

Department of Chemistry, Faculty of Science, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

(Received 17 October 1964 and in revised form 29 July 1965)

In an ordinary least-squares analysis of gas electron diffraction, the standard deviation for the most probable value of a parameter for a molecule depends on the interval of measurements Δs on a microphotometer recording, that is, on the number of observed points on an intensity curve.

It is shown that a reasonable estimate of the standard deviation of a parameter can be obtained by taking into account the effect of 'correlation' among the points of observations. A general method and its simplified form for dealing with the correlation are developed by introducing off-diagonal elements into the weight matrix used in the least-squares fit of the observed values. The simplified method is applied, as an example, to the electron-diffraction data of silicon tetrachloride, and it is shown that the standard deviation of the most probable value estimated by using an infinite number of observations does not approach zero. At the same time, it is shown that when a diagonal weight matrix is used for simplicity there is an optimum interval for measurements in order to get a correct standard deviation. The optimum interval in the example given was about $\Delta s = \pi/10$.

Introduction

In a usual method of least squares, the standard deviation for the most probable value is approximately proportional to the inverse of the square root of the number of observed points used in the calculation. Since an arbitrary number of points can be chosen on a continuous curve, such as the microphotometer trace

obtained by experiments of gas electron diffraction, the standard deviation for a parameter estimated by the least-squares analysis can be made unlimitedly small if the number of observations is infinitely increased. Such an argument, however, is based on a wrong assumption that all observed points remain mutually independent even when the interval of measurements becomes small. It would thus be desirable to have a

formulation taking account of the correlation among the observations. The importance of this kind of consideration has often been overlooked, except in a discussion by Bastiansen, Hedberg & Hedberg (1957).

There are two ways through which correlation is introduced into the observations. It may come in directly from the process of the measurement; it may also be introduced through a process of transformation of the observations, such as Fourier transformation.

It was pointed out in an earlier paper (Morino, Kuchitsu & Murata, 1965) that the correlations which are brought into indirect observations through a transformation of the measured quantities (direct observations) can be treated by introducing off-diagonal terms into the weight matrix and by minimizing the weighted squared sum of the residuals for the indirect observations. On the other hand, the points which are closely located on an intensity curve of gas electron diffraction are not mutually independent in any experiment. For example, the movement of a pen in a recorder has a finite time constant, and in fact it has been found in the least-squares analysis that many of the residuals of the observed values have the same sign as those of their neighboring points. In the present paper, the correlation among the direct observations will be discussed by introducing off-diagonal terms into the weight matrix for the least-squares calculation. The primary object is to estimate a reasonable interval of measurements when a diagonal weight matrix can be used in place of the non-diagonal weight matrix.

General theory

Error function

As was described in the earlier paper (Morino, Kuchitsu & Murata, 1965), if the correlations come in only through a transformation from direct observations, the most probable values can be obtained by applying the usual condition of least squares to the indirect observations:

$$\mathbf{V}^*\mathbf{P}\mathbf{V} \rightarrow \text{minimum} \quad (1)$$

where \mathbf{V} denotes the residuals referred to the resulting indirect observations. \mathbf{P} is a non-diagonal weight matrix for the indirect observations and is given by

$$\mathbf{P} = (\mathbf{G}\mathbf{P}_0^{-1}\mathbf{G}^*)^{-1} \quad (2)$$

where \mathbf{G} is a matrix whose elements are the coefficients of the linear transformation from direct observations to indirect ones, and \mathbf{P}_0 is a weight matrix for the original direct observations. If a point of the direct observation contributes to two or more points in the indirect observations, the matrix \mathbf{P} has off-diagonal elements even if the matrix \mathbf{P}_0 is diagonal.

The consideration of the above result seems to suggest a method for treating correlations among the direct observations. One may assume that a condition similar to equation (1) can be used for obtaining the most probable value even in the case where the direct

observations have correlations due to some unavoidable sources in the process of measurements. As a natural extension of the ordinary Gaussian distribution of errors, the error function for this case may be defined by a multivariate normal distribution, as follows,

$$F(V_1 V_2 \dots V_n) = \frac{1}{(2\pi\sigma^2)^{-n/2} |\mathbf{P}|^{1/2}} \exp(-\mathbf{V}^*\mathbf{P}\mathbf{V}/2\sigma^2), \quad (3)$$

where \mathbf{P} is a positive definite matrix and n is the number of observations.

Weight matrix

The weight matrix \mathbf{P} can be obtained as follows. The expectation of the product of residuals is defined as

$$\langle V_i V_j \rangle = \int_{-\infty}^{\infty} V_i V_j F(V_1 V_2 \dots V_n) dV_1 dV_2 \dots dV_n, \quad (4)$$

where $F(V_1 V_2 \dots V_n)$ is the error function given by equation (3). The integration can be performed by the diagonalization of the weight matrix \mathbf{P} . Since \mathbf{P} is supposed to be symmetric, the matrix \mathbf{P} can be diagonalized by a unitary matrix \mathbf{U} , that is

$$\mathbf{U}^*\mathbf{P}\mathbf{U} = \mathbf{Q} \quad (5)$$

where \mathbf{Q} is a diagonal matrix. Putting

$$\mathbf{W} = \mathbf{U}^*\mathbf{V} \quad (6)$$

it is shown that

$$\mathbf{V}^*\mathbf{P}\mathbf{V} = \sum_i Q_i W_i^2 \quad (7)$$

where Q_i is not zero, since \mathbf{P} is defined to be non-singular. Thus equation (4) may be simplified as

$$\begin{aligned} \langle V_i V_j \rangle &= \int_{-\infty}^{\infty} \sum_k \sum_l U_{ik} U_{jl} W_k W_l (2\pi\sigma^2)^{-n/2} |\mathbf{Q}|^{1/2} \\ &\quad \exp\left(-\sum_i Q_i W_i^2 / 2\sigma^2\right) dW_1 dW_2 \dots dW_n \\ &= \sigma^2 \sum_k (1/Q_k) U_{ik} U_{jk} \\ &= \sigma^2 (P^{-1})_{ij}. \end{aligned} \quad (8)$$

If the matrix $\langle V_i V_j \rangle$ is denoted as \mathbf{D} , it is proportional to the inverse matrix of \mathbf{P} by equation (8), and, therefore, it follows that

$$\mathbf{P} = \sigma^2 \mathbf{D}^{-1}. \quad (9)$$

The matrix \mathbf{D} can be obtained in an empirical way by using many observations at each point, and the weight matrix \mathbf{P} can be calculated by equation (9), where the coefficient σ^2 may be taken arbitrarily because σ^2 is cancelled in the normal equation and in the standard deviation for the parameter. Since the residual V_i actually depends on the weight function, \mathbf{P} may be made self-consistent by iteration, if necessary.

Least-squares analysis

The condition $\frac{\partial}{\partial \mathbf{X}} (\mathbf{V}^* \mathbf{P} \mathbf{V}) = 0$ (10)

gives the normal equation

$\mathbf{A}^* \mathbf{P} \mathbf{A} \mathbf{X} = \mathbf{A}^* \mathbf{P} \mathbf{M}$, (11)

where \mathbf{A} is the matrix whose element is given by the first derivative of an observed quantity with respect to a parameter, and \mathbf{X} denotes the difference between the most probable value and the estimated value for parameters, and \mathbf{M} the difference between the observed and the calculated values. In the process of the iteration of the least-squares calculation the \mathbf{M}_2 in the second step are equal to the residuals \mathbf{V}_1 in the first step. On the other hand, the relation of $\mathbf{V}^* \mathbf{P} \mathbf{V} = \mathbf{M}^* \mathbf{P} \mathbf{M} - \mathbf{X}^* \mathbf{A}^* \mathbf{P} \mathbf{A} \mathbf{X}$ is obtained from equation (11) by $\mathbf{M} = \mathbf{A} \mathbf{X} + \mathbf{V}$. Thus, the weighted squared-sum of the residuals in the second step, $\mathbf{V}_2^* \mathbf{P} \mathbf{V}_2$, reduces to $\mathbf{V}_1^* \mathbf{P} \mathbf{V}_1 - \mathbf{X}_2^* \mathbf{A}_2^* \mathbf{P} \mathbf{A}_2 \mathbf{X}_2$. Since $\mathbf{X}^* \mathbf{A}^* \mathbf{P} \mathbf{A} \mathbf{X}$ is positive, $\mathbf{V}^* \mathbf{P} \mathbf{V}$ decreases after applying one cycle. Therefore, equation (11) is equivalent to equation (1).

The variance or the covariance for the estimate is given by the same equation as in the case of the diagonal weight matrix,

$\sigma_{ij}^2 = (\mathbf{B}^{-1})_{ij} \sigma^2$ (12)

where $\mathbf{B} = \mathbf{A}^* \mathbf{P} \mathbf{A}$, $\sigma^2 = \mathbf{V}^* \mathbf{P} \mathbf{V} / (n - m)$, and m is the number of parameters.

A simplified non-diagonal weight matrix

The weight matrix can be obtained empirically by equation (9) when the correlation among observations is taken into account. The process to be followed for this purpose, however, involves considerable practical inconvenience, since the determination of the \mathbf{D} matrix calls for a large number of measurements on many photographic plates and a tedious processing of the data by the use of an electronic computer. Thus it would be more convenient if the correlation could be dealt with by using a simplified weight matrix.

Generally speaking, all off-diagonal terms do not vanish. However, if the intervals of measurements are large, the correlations among the observations may become small. Therefore, it is likely that the off-diagonal elements decrease gradually as they depart from the diagonal. Then, it seems to be reasonable to assume the weight matrix of the following form,

$$\mathbf{P} = \begin{bmatrix} h & -p & 0 & 0 & \dots & \dots & \dots \\ -p & h & -p & 0 & \dots & \dots & \dots \\ 0 & -p & h & -p & \dots & \dots & \dots \\ 0 & 0 & -p & h & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 & -p & h & -p \\ \dots & \dots & \dots & \dots & 0 & 0 & -p & h \end{bmatrix}$$
 (13)

The diagonal elements are taken to be all equal to h and the next off-diagonal elements to $-p$, whereas other off-diagonal elements are all equal to zero.

The expectation value of the product of residuals, $D_{i, i+k}$, is approximately proportional to $(p/h)^k$, and hence, the correlations gradually decrease as k increases*. This means that the matrix (13), which has non-vanishing elements only among the nearest neighbors, is a useful form to deal with the correlations which may practically occur.

The matrix is easily diagonalized in an explicit form, and the eigenvalues and eigenvectors are given by

$Q_j = h - 2p \cos \frac{\pi}{n+1} j$, $j = 1, 2, \dots, n$, (14)

$U_{jk} = \left(\frac{2}{n+1}\right)^{\frac{1}{2}} \sin \frac{\pi}{n+1} jk$, $k = 1, 2, \dots, n$. (15)

By equations (8), (14) and (15), we have

$D_{ij} = \sigma^2 \sum_k \frac{2}{n+1} \sin \frac{\pi}{n+1} ik \sin \frac{\pi}{n+1} jk$
 $\left(h - 2p \cos \frac{\pi}{n+1} k\right)^{-1}$. † (16)

Putting

$\alpha = \frac{1}{n} \sum_{i=1}^n D_{ii} = \bar{D}_{ii}$, (17)

and

$\beta = \frac{1}{n-2} \sum_{i=2}^{n-1} \frac{1}{2} (D_{i, i+1} + D_{i, i-1}) = \bar{D}_{i, i\pm 1}$, (18)

the trace of the matrix \mathbf{D} is then given by

$n\alpha = \sigma^2 \sum_{k=1}^n \left(h - 2p \cos \frac{\pi}{n+1} k\right)^{-1}$
 $\simeq \sigma^2 \left[\frac{n+1}{(h^2 - 4p^2)^{\frac{1}{2}}} - \frac{h}{h^2 - 4p^2} \right]$. ($h > |2p|$, $n \gg 1$) (19)

Therefore, the averaged diagonal element is approximated by

$\alpha = \sigma^2 (h^2 - 4p^2)^{-\frac{1}{2}}$. (20)

On the other hand, the average of the off-diagonal elements on the first nearest neighbors of the diagonal is given by

$\beta \simeq (h\alpha - \sigma^2) / 2p$. (21)

The ratio of the off-diagonal term to the diagonal term of the weight matrix is given by

$p/h = \alpha\beta / (\alpha^2 + \beta^2)$. (22)

* $D_{i, i+k}$ is proportional to $\{(h - \sqrt{h^2 - 4p^2}) / 2p\}^k$ and for $|p/h| < 0.4$, it can be approximated as $(p/h)^k$.

† If the integration is performed at this step, equation (16) is given by

$D_{i, i+k} \simeq \frac{\sigma^2}{\sqrt{h^2 - 4p^2}} A^k (1 - A^{2i})$
 $D_{i, i-k} \simeq \frac{\sigma^2}{\sqrt{h^2 - 4p^2}} A^k (1 - A^{2(i-k)})$,

where $k \geq 0$ and $A = (h - \sqrt{h^2 - 4p^2}) / 2p$. Then α and β are easily calculated from these equations and are given in the same form as in equations (20) and (21).

Equation (14) leads to the condition

$$|p/h| < \frac{1}{2}, \quad (23)$$

because all eigenvalues of a positive definite matrix are positive. Thus the weight matrix obtained by applying equation (22) is positive definite.

The weight matrix of the simplified form can easily be obtained by the following steps. (a) The most probable values of parameters are estimated from a molecular intensity curve by use of the usual method of least squares. (b) The residual of the observation for each q -value, V_i , is obtained by taking the difference between the observed and calculated values of $qM(q)$. (c) The averaged diagonal elements and the averaged off-diagonal elements of the \mathbf{D} matrix, α and β respectively, are calculated by using equations (17) and (18) respectively. This procedure, which is based on the assumption that all points on $qM(q)$ have the same reliability, seems to be superior to the laborious direct determination of each element of \mathbf{D} , because α and β can easily be obtained from one set of $qM(q)$. (d) The simplified weight matrix is given by equation (22) in terms of α and β obtained above; $-p/h$ may be taken equal to the off-diagonal element of the weight matrix (13) whose diagonal elements are taken as unity. As described in the preceding section, the diagonal element of the weight matrix can be taken arbitrarily.

In order to visualize the practical meaning of the correlations in the analysis of gas electron diffraction data, the process described above is applied to the analysis of the molecular intensity curve for silicon tetrachloride. The observed values were obtained by the method described in our previous paper (Morino & Murata, 1965). The points of observations were taken at an interval $\Delta q = 0.2$ over the range from $q = 17.4$ to $q = 96.6$. Six parameters, the atomic distances, the mean amplitudes and the indices of resolution for two kinds of atom pairs, Si-Cl and Cl-Cl, were taken in the least-squares analysis. Many of the residuals obtained in the least-squares fit of the molecular intensity have the same sign as those of their close neighbors; the situation is illustrated in Fig. 1 for $\Delta q = 1$ ($q = 10s/\pi$).

The ratio p/h was obtained by the use of equation (22) from the residuals for the molecular intensity curve; one set corresponds to $\Delta q = 0.2$, two sets to $\Delta q = 0.4$ (one set measured at $q = 17.4, 17.8, 18.2, \dots$ and the other at the intermediate points, $q = 17.6, 18.0, \dots$), three sets to $\Delta q = 0.6$, and so on. A mean value is marked by a dot for each interval, and a smooth curve was drawn through the dots, as shown in Fig. 2. It is reasonable that p/h increases to 0.5 when the interval Δq decreases to zero, as is expected from equations (22) and (23).

The standard deviations calculated from the fit of the molecular intensity curve by the application of the off-diagonal terms thus obtained are shown in Fig. 3 for the atomic distances of the Si-Cl and Cl-Cl pairs. The mean amplitudes, $l(\text{Si-Cl})$ and $l(\text{Cl-Cl})$, and the indices of resolution, $k(\text{Si-Cl})$ and $k(\text{Cl-Cl})$, were

shown to have nearly the same trend. It should be emphasized that, when the interval decreases to zero, the standard deviation for the most probable value does not approach to zero but remains finite when the correlation is properly taken into consideration.

Optimum interval of measurements

When the observations have correlations, a non-diagonal weight matrix should be used in a least-squares fit. In that case, the standard deviation decreases with a decrease in the interval of measurements and converges to a certain minimum value, as shown in Fig. 3. Since the interval of measurements makes no problem if a proper non-diagonal matrix is used, the best estimate of the standard deviation can be obtained as the small-interval limit, σ_{\min} , in Fig. 3. However, this procedure requires laborious calculations even when the above mentioned 'simplified matrix' is assumed. It would therefore be practically advantageous if a diagonal weight matrix could be used in the least-squares process as an effective substitute for the non-diagonal matrix. Evidently, the most probable values of the parameters are correctly estimated by this diagonal matrix. In order to derive a proper standard deviation, σ_{\min} , the interval of measurements may be chosen as Δq_0 in Fig. 3. It is also evident that the optimum interval thus obtained depends on the experimental conditions such as the camera distance; the magnifying ratio in the microphotometer tracing and the feature of the intensity curve.

Let us again consider the case of silicon tetrachloride for a practical example. The standard deviations obtained by the application of the diagonal weight matrix

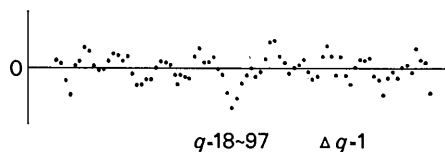


Fig. 1. The residuals of the least-squares fit (obs. - calc.), on an arbitrary scale, for the molecular intensity curve of silicon tetrachloride against q .

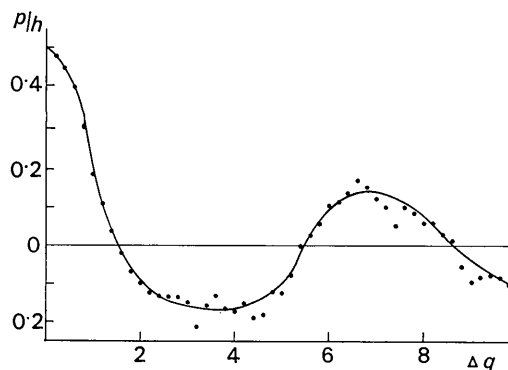


Fig. 2. The ratios of the off-diagonal term p to the diagonal term h for various intervals. The dots show empirical values obtained from the residuals.

are illustrated in Fig. 3. Here a unit matrix is used for the weight matrix, because equation (13) will simply reduce to a unit matrix if the off-diagonal terms are put equal to zero. It is easily seen that the optimum intervals, Δq_0 , are about 1.0 and 0.8 in the q scale for the Si-Cl and Cl-Cl atom pairs, respectively. This indicates that the interval $\Delta q_0 = 1$ which has often been used in the analysis of the molecular intensity curves of SiCl_4 and also of other molecules, has been an appropriate choice.

Discussion

In the previous report on the analysis of silicon tetrachloride, it was concluded, by comparing two kinds of standard deviations, σ_1 and σ_2 , that the interval of $\Delta q_0 = 1$ was suitable for the analysis (Morino & Murata, 1965). The development described above now provides a more direct support for this interval. Even though the present treatment is a crude approximation, it is sufficient for the purpose of demonstrating the effect of correlation, because the standard deviation is not sen-

sitive to the extent of the correlation, as shown in Fig. 3. The standard deviation for the most probable value estimated by using the rigorous weight matrix should be invariant under the change in the interval of measurements. In the present case, however, as can be seen from Fig. 3, the standard deviations for $r(\text{Si-Cl})$ and $r(\text{Cl-Cl})$ depend slightly on Δq for smaller Δq . The same trend is also observed for other parameters. Hence, it would be better to introduce the second nearest neighbors in the off-diagonal weight matrix. In such a case the optimum interval might be given from the plateau of the curve instead of the minimum value, although the difference between obtained values would be very small.

The interval, Δq_0 , for Si-Cl is found to be slightly larger than that for Cl-Cl. This may correspond to the fact that the period of the molecular intensity for Si-Cl is larger than that for Cl-Cl because of the longer Cl-Cl distance than that of Si-Cl.

The above discussion concerns the analysis of only one photographic plate. If the analysis is made for n plates, the random error becomes $n^{-\frac{1}{2}}$ times that derived from one curve. Therefore, as the number of curves analyzed increases, the random error decreases unlimitedly. It does not mean that the accuracy of the final result can be increased indefinitely: it is limited by the presence of systematic errors which have finite values depending on the experimental conditions.

It is to be noted that a strong correlation may enter into the analysis if irregular readings of the intensity curve are smeared out. In order to show this effect, an observed molecular intensity curve was Fourier-transformed into a radial distribution curve and transformed back into a molecular intensity curve. As a result of such a double conversion, most of the irregular features in the original experimental readings disappeared, and an apparently smooth curve was obtained. A strong correlation, however, was found in the smoothed curve since the p/h value became 0.435, in contrast to 0.140 of the original curve. This result gives a warning against an unduly small estimate of the standard deviation by smoothing out the observations in the course of the analysis, and by ignoring the effect of correlation among the observations.

The authors are indebted to Professor J. Kakinoki and Drs T. Ino and K. Katada of Osaka City University for their valuable criticism and advice.

References

- BASTIANSEN, O., HEDBERG, L. & HEDBERG, K. (1957). *J. Chem. Phys.* **27**, 1311.
 MORINO, Y., KUCHITSU, K. & MURATA, Y. (1965). *Acta Cryst.* **18**, 549.
 MORINO, Y. & MURATA, Y. (1965). *Bull. Chem. Soc. Japan*, **38**, 104.

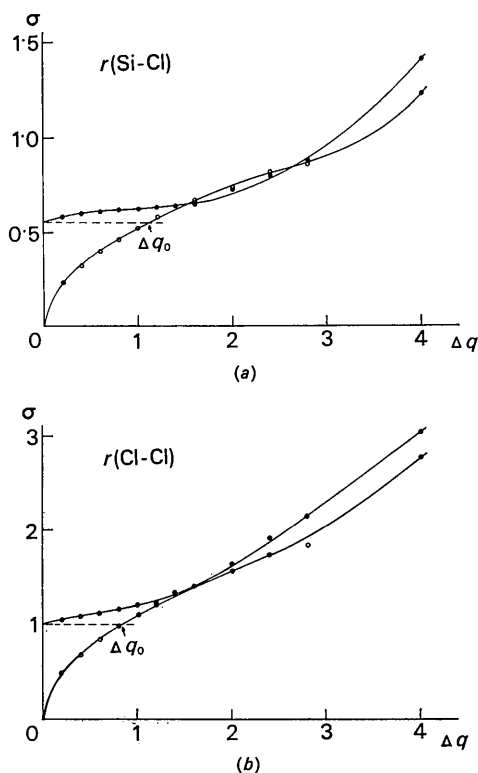


Fig. 3. Standard deviations of the atomic distances of SiCl_4 obtained from the least-squares fit of a molecular intensity curve. Filled circles represent the outputs of the analysis in which the effect of correlation is taken into account. Open circles represent those for which the effect of correlation is ignored. Δq_0 denotes the optimum interval when the diagonal weight matrix is assumed.