CUSTERS, J. F. H. & RATHENAU, G. W. (1941). *Physica, 's Grav.*, **8**, 759.
GRENINGER, A. B. (1935). *Z. Krystallogr.* **91**, 424.
LACOMBE, P. (1948). *Report of a Conference on the Strength of Solids.* London: Physical Society.
LACOMBE, P. & BEAUJARD, L. (1944). *C.R. Acad. Sci., Paris*, **219**, 66.
LACOMBE, P. & BEAUJARD, L. (1945). *C.R. Acad. Sci., Paris*, **221**, 414.
LACOMBE, P. & BEAUJARD, L. (1947). *J. Inst. Met.* **74**, 1.
LACOMBE, P. & BERGHEZAN, A. (1949). *C.R. Acad. Sci. Paris*, **228**, 93.
LACOMBE, P. & YANNAQUIS, N. (1948). *Rev. Métall.* **45**, 68.
MAHL, H. & STRANSKI, I. N. (1942). *Z. Phys. Chem.* (B), **52**, 257.
MAY, W., TIEDEMA, T. J. & BURGERS, W. G. (1948). *Nature, Lond.*, **162**, 740.
O'NEILL, H. (1928). *J. Iron Steel Inst.* no. 1, 689.
PETERSEN, C. (1947). *Metallforschung*, **11**, 289.
SANDEE, J. (1942). *Physica, 's Grav.*, **9**, 741.
SCHIEBOLD, E. & SACHS, G. (1926). *Z. Krystallogr.* **63**, 34.
SEUMEL, G. (1936). *Z. Krystallogr.* **93**, 249.
WALTON, C. J. (1944). *Trans. Electr. Soc.* **85**, 239.

Acta Cryst. (1949). **2**, 154

The Accuracy of Atomic Co-ordinates Derived by Least-Squares or Fourier Methods

BY D. W. J. CRUICKSHANK

St John's College, Cambridge, England

(*Received* 23 *December* 1948)

Recent work by Cruickshank on the accuracy of atomic co-ordinates obtained in Fourier syntheses is here related to that of Cochran on the connexions between the least-squares and Fourier methods. Except when the atomic peaks are not resolved, it is shown that the same co-ordinates and estimated errors are obtained by the method of least squares and by appropriately weighted Fourier series (corrected for finite summation). An expression is obtained for the relative accuracies of the weighted and unweighted Fourier methods.

1. Introduction

The purpose of this paper is to relate and extend two of the approaches which have recently been made to the problem of the determination of atomic co-ordinates and their accuracy. In one approach (Cox & Cruickshank, 1948; Cruickshank, 1949) it has been shown how to estimate the standard deviations of co-ordinates derived by the Fourier method, when corrected for finite summation. In the other approach (Cochran, 1948*a*, *b*) relationships have been obtained between the Fourier and least-squares methods.

According to the former approach the standard deviation of an atomic co-ordinate (in an orthogonal cell) derived from a Fourier series, after correction for finite summation by Booth's (1946) method, is

$$\sigma(x_{rj}) = \frac{1}{V}\frac{2\pi}{a_j}\left\{\sum_3 h_j^2 \Delta F^2\right\}^{\frac{1}{2}} \bigg/ \left|\frac{\partial^2 \rho}{\partial x_j^2}\right|, \qquad (1\cdot1)$$

where x_{rj} is the jth co-ordinate of the rth atom ($j=1$, 2, or 3), a_j the cell side, h_j the plane indices, $\Delta F = F_o - F_c$, $\partial^2\rho/\partial x_j^2$ is the curvature of the peak in the x_j direction, and $\sum_3$ denotes summation over all planes included in the series. This estimate and the finite-series correction are valid assuming

(1) the ΔF's may be treated as independent;

(2) the shapes of the observed and calculated peaks near the maxima are the same; and

(3) the peaks are resolved and the finite-series corrections small.

It has since been realized that (1·1) is incorrect for certain positions in the unit cell in systems with symmetry relations which permute the plane indices, e.g. if $F_{hkl} = F_{lhk} = F_{klh}$. The correct formula in this case is given in § 4.

2. Determination of atomic co-ordinates

Cochran's relationships between the Fourier and least-squares methods may be obtained in the following way.

Let u be the total number of independent planes observed, and let $\sum_u$ denote a summation over those u planes, a summation which must be distinguished from $\sum_3$, which includes all planes dependent on these u planes.

In Hughes's (1941) application of least squares the required atomic co-ordinates are those for which $\sum_u w(F_o - F_c)^2$ is a minimum, w being the weight given to one independent 'observation'. This expression may be rewritten $\sum_3 w_1(F_o - F_c)^2$, where, if s is the multiplicity of a plane, $w = w_1 s$. The co-ordinates x_{rj} satisfy

$$\sum_3 w_1(F_o - F_c)\frac{\partial F_c}{\partial x_{rj}} = 0, \qquad (2\cdot1)$$

$\partial F_c/\partial x_{rj}$ denoting differentiation of the calculated structure factor with respect to one of the atomic parameters, so that, if t atomic co-ordinates are determined by symmetry from this one parameter, $\partial F_c/\partial x_{rj}$

is a sum of t terms of the type (if there is a centre of symmetry) $\mp(2\pi/a_j)\,h_jf_r\sin\theta$. But, by the symmetry, all of the t sums

$$\sum_3 \mp w_1\frac{2\pi}{a_j}\,h_j f_r \sin\theta$$

are equal, and on taking the one involving x_{rj} we have

$$-\frac{2\pi}{a_j}\sum_3 w_1h_jf_r(F_o-F_c)\sin\theta_r=0, \tag{2·2}$$

where $$\theta_r=2\pi\left\{\sum_j\frac{h_j}{a_j}x_{rj}\right\}.$$

Hence

$$-\frac{1}{V}\frac{2\pi}{a_j}\sum_3 w_1h_jf_rF_o\sin\theta_r=-\frac{1}{V}\frac{2\pi}{a_j}\sum_3 w_1h_jf_rF_c\sin\theta_r, \tag{2·3}$$

so that the slopes at x_r of the 'density' functions

$$p_o=\frac{1}{V}\sum_3 w_1f_rF_o\cos\theta \tag{2·4}$$

and $$p_c=\frac{1}{V}\sum_3 w_1f_rF_c\cos\theta \tag{2·5}$$

are equal.

Thus, with the obvious generalization of assumption (2), the positions of the maxima of the rth peaks of p_o and p_c are the same; and, with assumption (3), the co-ordinates derived either by least squares or by the Fourier series (2·4), with correction for finite summation (and peak overlapping) by (2·5), are the same. If the f's of the different atoms are not proportional for varying reciprocal radii, more than one weighted Fourier synthesis will be necessary to get the best determinations of all the co-ordinates. However, there will probably be little loss of accuracy by using the same form of f curve as a weighting factor for all the peaks. When $w_1=1/f_r$ the p's become the electron densities ρ_o and ρ_c. Cochran has also pointed out that, when $w_1=1$, the p's may be regarded as the electron densities obtained when the F's are multiplied by the artificial temperature factor $\alpha=f_r$ (Brill, Grimm, Hermann & Peters, 1939).

Booth (1947), for ease of calculation by automatic machines, has suggested that atomic co-ordinates may be found by minimizing the function

$$R_2=\sum_u w''(F_o^2-F_c^2)^2.$$

Assuming the errors ΔF are small compared with the F's, when $w''=w/F_o^2$, the co-ordinates and the estimated errors obtained by minimizing R_2 or $\sum_u w(F_o-F_c)^2$ are the same.

3. Standard deviations of the co-ordinates

In the following discussion some standard results of least-squares theory are quoted; they will be found in many of the treatises on statistics (e.g. Whittaker & Robinson, 1940, Chap. 9).

To estimate the errors of the least-squares determination introduce assumption (1), that the ΔF's may be treated as independent. The arguments for this are the same as those in § 14 of Cruickshank's (1949) paper. The estimate of the standard deviation σ of an observation with unit weight is

$$\sigma^2=\sum_u w\Delta F^2/(u-v)=\sum_3 w_1\Delta F^2/(u-v), \tag{3·1}$$

where v is the number of unknown parameters and u again the number of independent F values. Denoting the unknowns by $x_1, x_2, \ldots, x_r, \ldots, x_v$, define D as the determinant

$$D=\begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1v} \\ b_{12} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ b_{1v} & \cdots & \cdots & b_{vv} \end{vmatrix}, \tag{3·2}$$

where

$$b_{ij}=\sum_u w\left(\frac{\partial F_c}{\partial x_i}\right)\left(\frac{\partial F_c}{\partial x_j}\right)=\sum_3 w_1\left(\frac{\partial F_c}{\partial x_i}\right)\left(\frac{\partial F_c}{\partial x_j}\right). \tag{3·3}$$

If B_{rr} is the rth principal minor of D, then the estimate of the standard deviation $\sigma(x_r)$ of x_r is

$$\sigma^2(x_r)=\sigma^2B_{rr}/D. \tag{3·4}$$

This is a general result. If the terms $b_{rj}(j\neq r)$ are small compared with b_{rr} (implying that the peaks are well separated in an orthogonal cell)

$$B_{rr}/D=1/b_{rr}. \tag{3·5}$$

For structures not involving symmetries of the type $F_{hkl}=F_{lhk}$, and assuming that the co-ordinates of this peak are general, we have, on taking the mean value of $\sin^2\theta=\frac{1}{2}$,

$$b_{rr}\doteqdot\frac{4\pi^2}{a_j^2}\sum_3 w_1h_j^2f_r^2\left(\frac{2t}{s}\right)^2\left(\frac{s}{2}\right)\left(\frac{1}{2}\right), \tag{3·6}$$

which is obtained by considering the number of non-identical $\sin\theta$ terms which occur in $\partial F_c/\partial x_r$ (t and s are defined close to equation (2·1)). Hence

$$\sigma^2(x_r)=\sigma^2\Big/\frac{4\pi^2}{a_j^2}\,t^2\sum_3(w_1h_j^2f_r^2/s). \tag{3·7}$$

By the definition of the weight w, the standard deviation $\sigma(\Delta F)$ of ΔF satisfies

$$\sigma^2=w\sigma^2(\Delta F)=w_1s\sigma^2(\Delta F), \tag{3·8}$$

σ being the standard deviation of an observation with unit weight. The generalization of (1·1) for the p density is

$$\sigma(x_{rj})=\frac{1}{V}\frac{2\pi}{a_j}\left\{\sum_3 w_1^2h_j^2f_r^2\Delta F^2\right\}^{\frac{1}{2}}\Big/\left|\frac{\partial^2p}{\partial x_j^2}\right|. \tag{3·9}$$

On substituting $\sigma(\Delta F)$ for ΔF from (3·8) we get

$$\sigma(x_{rj})=\frac{\sigma}{V}\frac{2\pi}{a_j}\left\{\sum_3 w_1h_j^2f_r^2/s\right\}^{\frac{1}{2}}\Big/\left|\frac{\partial^2p}{\partial x_j^2}\right|. \tag{3·10}$$

Now

$$\frac{\partial^2p}{\partial x_j^2}=-\frac{1}{V}\frac{4\pi^2}{a_j^2}\sum_3 w_1h_j^2f_r\cos\theta_{rj}\,(\Sigma f_i\cos\theta_i).$$

If we neglect cross terms in which $\theta_i \not\equiv \theta_{rj}$ as in (3·5) and proceed as in (3·6) we have approximately

$$\frac{\partial^2 p}{\partial x_j^2} \doteqdot -\frac{1}{V}\frac{4\pi^2}{a_j^2}\sum_3 \frac{t}{s} w_1 h_j^2 f_r^2. \qquad (3{\cdot}11)$$

The factor t/s, which might not have been expected, allows for those planes which are absent from the summation owing to symmetry.

Substituting (3·11) in (3·10) we get (3·7), thus checking that, when the peaks are resolved, the errors calculated either from least squares or the Fourier method are the same. Since the same co-ordinates and estimated errors are obtained by the two methods (when the peaks are resolved), the choice of method is purely one of ease in calculation.

4. Further statistical problems

We shall now discuss the errors in the slope of the electron density when the symmetry relations involve permutations of the plane indices. Suppose

$$F_{hkl} = F_{lhk} = F_{klh},$$

and that we are considering the error in the slope at (x, x, x). In equation (11·7) of Cruickshank (1949), the coefficient of ΔF_{hkl}^2 will be

$$(h+k+l)^2\,[\sin\theta_1 - \sin(-\theta_1) + \ldots]^2,$$

which has a mean value in the cell equal to $(h+k+l)^2 m^2$, where m is the number of planes related to F_{hkl} by relations involving negative indices. The error in the slope will be given by summing

$$\frac{2\pi}{aV}\Sigma(h+k+l)^2 m^2 \Delta F^2$$

over all independent ΔF's.

Two points concerning estimates of standard deviations and significance tests will now be dealt with.

Equations (1·1) and (3·9) are strictly accurate only when $\sigma(\Delta F)$ is substituted for ΔF. If the ΔF's are the final residual ΔF's obtained from the least-squares calculations, then $\sum_u w\Delta F^2$ will be an estimate, not of $u\sigma^2$ but of $(u-v)\,\sigma^2$, because the least-squares process imposes v constraints on the ΔF's. This is the reason why $(u-v)$ appears in the denominator of (3·1). Consequently estimates of $\sigma(x_{rj})$ from (3·9) using ΔF's given by the final atomic co-ordinates should be multiplied by a factor $[u/(u-v)]^{\frac{1}{2}}$. When the number of parameters is small compared with the number of observations, and when the ΔF's are estimated from F's calculated for positions before the final correction for finite series (so that they are a little larger than the final ΔF's), no great error will be made by omitting the factor $[u/(u-v)]^{\frac{1}{2}}$.

In applying significance tests to discussions of differences in bond lengths (see Cruickshank, 1949, § 3) the statistic $\delta l/\sigma(l)$ is used, where δl is the difference in the bond lengths and $\sigma(l)$ is the estimated standard deviation of δl. When $(u-v)$ is large $\sigma(l)$ may be treated as an accurate estimate of the standard deviation, and since δl is normally distributed, the significance levels of the normal law (given in the above-mentioned paper) should be used. When $(u-v)$ is small, say less than 30, $\sigma(l)$ is an uncertain estimate of the standard deviation; accordingly the values of $\delta l/\sigma(l)$ for the different levels of significance have to be increased. A discussion of this type of problem is given by Kendall (1946, §§ 21·24–21·31).

5. Standard deviations by the weighted and the unweighted Fourier methods

We shall now calculate the ratio of the errors $\sigma_L(x_{rj})$ of the weighted Fourier or least-squares methods to those $\sigma_F(x_{rj})$ of the ordinary unweighted Fourier method. Substituting $\sigma(\Delta F)$ from (3·8) for ΔF in (1·1), and using (3·11), we have for the standard deviation $\sigma_F(x_{rj})$ of the ordinary Fourier method

$$\sigma_F^2(x_{rj}) = \left[\sigma^2 \sum_3 h_j^2/(w_1 s)\right] \Bigg/ \left[\frac{4\pi^2}{a_j^2}\left(\sum_3 \frac{t}{s} h_j^2 f_r\right)^2\right]. \qquad (5{\cdot}1)$$

Hence, using (3·7), the ratio is

$$\frac{\sigma_L^2(x_{rj})}{\sigma_F^2(x_{rj})} = \frac{\left(\sum_3 h_j^2 f_r/s\right)^2}{\left(\sum_3 h_j^2/(w_1 s)\right)\left(\sum_3 w_1 h_j^2 f_r^2/s\right)}, \qquad (5{\cdot}2)$$

which is perhaps more easily appreciated if written in the form

$$\frac{\sigma_L^2(x)}{\sigma_F^2(x)} = \frac{\left(\sum_u h^2 f\right)^2}{\left(\sum_u h^2/w_1\right)\left(\sum_u w_1 h^2 f^2\right)}. \qquad (5{\cdot}3)$$

(N.B. the summations include w_1 not w.)

By the algebraic properties of the right-hand side of (5·3), except when $w_1 = 1/f_r$, $\sigma_L(x_r) < \sigma_F(x_r)$, as is intended by the least-squares method. In my earlier paper (Cruickshank, 1949, § 16) I suggested that the errors in least-squares determinations would be a little larger than those of the ordinary Fourier determinations. This is correct when applied to the r.m.s. error over the whole cell of the density or slope of the calculated structure; it is incorrect, by the above argument, when applied to the atomic co-ordinates.

6. Discussion

We can now comment on methods of obtaining accurate atomic co-ordinates. Apart from considerations of actual computing technique, there seems no reason in many cases why the ordinary process of Fourier refinement (with final correction for finite-series effects) should be replaced by weighted syntheses or least-squares methods at the beginning of a structure determination. Besides giving the atomic co-ordinates, the electron-density plots give much useful information, which is not given by weighted density plots. The standard deviations of the co-ordinates obtained at this stage can be calculated by (1·1). If the accuracy is insufficient for the required purpose, (5·3) can be evaluated to find out whether the increased accuracy of the weighted series would be large enough to justify their use.

The first step is to assign weights to the observations. Since the F_c's are directly involved in the process of determining the atomic co-ordinates, the weights must be estimated from the ΔF's ($\Delta F = F_o - F_c$). For this the ΔF's at the last stage of the ordinary Fourier refinement may be used, since the small shifts which may follow later will reduce them but little (see Cruickshank, 1949, § 14). The actual scheme of weighting will depend on the particular data, and it should be carefully checked for suitability. The frequency distribution of $\Delta F \sqrt{w}$ should be approximately normal; this may be checked by Pearson's χ^2 test of goodness of fit (see, for example, Weatherburn, 1947, § 79). If we define q by

$$\sum_u w\Delta F^2 = uq^2,$$

it is also important that

$$q^2 \sum_u h^2 \doteqdot \sum_u wh^2 \Delta F^2, \tag{6·1}$$

with similar expressions involving k and l. These help to ensure a proper variation of w with reciprocal radius. Having assigned the weights, we can calculate (5·3) to decide whether it is worth proceeding with the weighted synthesis.

Approximate calculations with the dibenzyl data (Jeffrey, 1947; Cruickshank, 1949) indicate that an improvement of the order of 25% in the standard deviations might be obtained in this case by using weighted syntheses; i.e. the standard deviation of a co-ordinate would be reduced from 0·0074 A. to about 0·0056 A. It would be very desirable to pronounce with confidence on bond-length differences of 0·010 A.; for this the co-ordinate standard deviations would have to be 0·0025 A. or less. If the dibenzyl figures are typical, despite improvements by using weighted syntheses, attention will still have to be directed towards reducing the ΔF's, possibly more by improving the molecular models used in calculation (which will improve the series-termination correction) than by improving the accuracy of intensity measurement.

The chief possibility of mis-estimation of error either in the least-squares or the Fourier process is if assumption (1) is not fulfilled. This was discussed in an earlier paper (Cruickshank, 1949, § 14), and it was shown that the estimated errors could be either too large or too small. An important feature is the accuracy of the finite-series correction, which requires that the peaks and ripples of the observed and calculated densities should be nearly the same. This may be checked by examining $\Delta\rho = \rho_o - \rho_c$; if the slopes of $\Delta\rho$ are zero in regions around the atomic centres, and if no periodic diffraction ripples are present, then the finite-series corrections have been accurately made and the errors may be overestimated by using $\Delta F = F_o - F_c$; on the other hand, if substantial diffraction ripples are present in $\Delta\rho$, the errors may be overestimated. Brill *et al.* (1939) have suggested that $\Delta\rho$ maps may also be useful in locating hydrogen atoms; because if ρ_c does not contain any hydrogen atoms, the H atoms, if detectable, will appear as maxima in $\Delta\rho$ which may be easier to locate than the slight bulges in the ordinary contour lines.

I should like to thank Dr W. Cochran for suggesting some improvements in the presentation of this paper, and Mr A. P. Robertson for discussions on statistical problems.

References

Brill, R., Grimm, H., Hermann, C. & Peters, C. (1939). *Ann. Phys., Lpz.*, **34**, 393.

Booth, A. D. (1946). *Proc. Roy. Soc.* A, **188**, 77.

Booth, A. D. (1947). *Nature, Lond.*, **160**, 196.

Cochran, W. (1948*a*). *Acta Cryst.* **1**, 138.

Cochran, W. (1948*b*). *Nature, Lond.*, **161**, 765.

Cox, E. G. & Cruickshank, D. W. J. (1948). *Acta Cryst.* **1**, 92.

Cruickshank, D. W. J. (1949). *Acta Cryst.* **2**, 65.

Hughes, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.

Jeffrey, G. A. (1947). *Proc. Roy. Soc.* A, **188**, 322.

Kendall, M. G. (1946). *The Advanced Theory of Statistics*, **2**. London: Griffin.

Weatherburn, C. E. (1947). *Mathematical Statistics.* Cambridge: University Press.

Whittaker, E. T. & Robinson, G. (1940). *The Calculus of Observations*, 3rd ed. London: Blackie.