long tails which may not be distinguishable from background scattering due to other sources.

Our estimates of error from this source are admittedly very rough and subjective and are intended only as indications of the possible magnitudes of the errors and their variation with the relevant parameters. The actual errors of measurement will depend on a number of factors. The curves of Fig. 2 will be convoluted in practice with an instrumental measuring function and a mosaic spread function which will vary in width and form with the distance of the reciprocal-lattice point from the origin and also with the angle between the **h** vector and the repetition vector **R**. The estimates of background will depend on the measuring routine used.

In the reciprocal-space representation of Fig. 1(*c*) it is seen that for our simple model the value of $\mathbf{h}.\boldsymbol{\Delta}$ is zero on one line through the origin and varies systematically with distance from this line. Since from Fig. 4 it is seen that there is only a weak dependence of the error on $\mathbf{h}.\boldsymbol{\Delta}$, the effect will be to produce only a small systematic variation of measured intensity. However, the variation of the form and size of the measuring volume in reciprocal space (Fig. 3) will tend to give a larger systematic error with position of the reciprocal-lattice point since the varying range of measurement in terms of $2\pi\varepsilon R$ will give an error ranging from almost zero to the maximum indicated in Fig. 4.

In general, the errors in relative intensities will be less if F_1 and F_2 are of the same sign and differ by less than 50%. It is for the occasional pairs of reflections for which F_1 and F_2 have opposite sign that the relatively large errors may be produced.

It is suggested that when microtwinning is indicated by streaking in X-ray diffraction or electron diffraction patterns or by contrast effects in high-resolution electron micrographs (Wenk, 1976) estimates may be made of the possible errors on the basis of the above considerations in relation to the techniques used for measurement of X-ray diffraction intensities. In this way improved accuracy may be obtained in structure analyses when microtwinning is present.

The authors are grateful to Drs Gabrielle and J. D. H. Donnay for stimulating discussions in the initiation of this project and in review of the manuscript, and also to Dr R. Von Dreele for advice on diffractometer measuring techniques. The work was supported by NSF grant DMR 76-06108.

References

Cowley, J. M. (1976*a*). *Acta Cryst.* A**32**, 83–87.
Cowley, J. M. (1976*b*). *Acta Cryst.* A**32**, 88–91.
Grainger, C. T. (1969). *Acta Cryst.* A**25**, 427–434.
Iijima, S. & Buseck, P. R. (1976). *Electron Microscopy in Mineralogy*, edited by H.-R. Wenk, pp. 319–323. Heidelberg: Springer.
McLaren, A. C. & Marshall, D. B. (1974). *Contrib. Mineral. Petrol.* **44**, 237–249.
Prince, E., Donnay, G. & Martin, R. F. (1973). *Am. Mineral.* **58**, 500–507.
Wenk, H.-R. (1976). *Electron Microscopy in Mineralogy.* Heidelberg, New York: Springer.
Zachariasen, W. H. & Plettinger, H. A. (1965). *Acta Cryst.* **18**, 710–716.

Acta Cryst. (1978). A**34**, 743–746

Minimizing the Variance in Integrals and Derivatives of the Electron Density

By C. L. Davis and E. N. Maslen

Department of Physics, University of Western Australia, Nedlands, Western Australia

(*Received* 21 *November* 1977; *accepted* 28 *March* 1978)

A method for minimizing the variance in integrals and derivatives of electron densities by filtering the scattering amplitudes is established. The optimum filters for the density and its integrals and derivatives are shown to be the same. Calculations with experimental data show that the variance of integrated densities is not very sensitive to the shape of the region of integration, indicating that good estimates may be made using simple shapes. A very convenient expression is given for the variance of the integral over a sphere. Coppens & Hamilton [*Acta Cryst.* (1968), B**24**, 925–929] have shown that accurate estimates of integrated densities are possible. Filtering can further improve these estimates. By contrast, good estimates of derivatives of the electron density remain unlikely, even using filtered diffraction data.

Introduction

In quantitative studies of the structure of materials by the analysis of diffraction data it is often of interest to calculate gradients and volume integrals of the electron density. When the density is evaluated by Fourier synthesis of the structure factors, $\mathbf{F}(\mathbf{S})$ ($\mathbf{S}$ is the reciprocal-lattice vector), difficulties arise in the treatment of the experimental errors. The variance in the density is the sum over all reflections of the variance in the structure factor for each reflection (see, for example, Cruickshank, 1949). Ideally, the Fourier and variance summations extend over the infinite number of reciprocal-lattice vectors. The smoothness of the time-average electron density guarantees that $\mathbf{F}(\mathbf{S}) \rightarrow 0$ as $|\mathbf{S}| \rightarrow \infty$, so that the Fourier summation converges in principle.

In practice there is always a non-zero uncertainty in the structure factor measurement due to background, even when the true value of $\mathbf{F}(\mathbf{S})$ approaches zero. Even in this limit there will be errors in the measured structure factors which produce 'noise' in the calculated density distribution. It is apparent also that the variance summation may diverge. Estimates of the variance made by terminating the summation at the limit of the experimental data set will depend on the extent of the data set, which is somewhat arbitrary.

Essentially the same problem arises in estimating the variance in integrals and derivatives of the density. The situation is less serious for integrated densities, but worse for derivatives. Coppens & Hamilton (1968) have shown that each term in the variance summation for an integral over a parallelepiped region is multiplied by a function of **S**. This function is the square of

$$D(\mathbf{S}) = \textstyle\int_V \exp(-i\mathbf{S}.\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{1}$$

where V is the volume of the parallelepiped. $D(\mathbf{S})$ is the Fourier transform of a function which has a value of one inside the parallelpiped and zero elsewhere. $D(\mathbf{S})$ is peaked around the origin of reciprocal space. The larger the volume V, the more rapidly $D(\mathbf{S})$ approaches zero away from the origin. The effect of this factor is to reduce the contribution of high-order terms to the variance. Normally, the predominantly weak high-order reflections are relatively less accurately measured than the strong reflections which occur at low angles. Coppens & Hamilton (1968) have pointed out that it is therefore possible to measure an integrated density relatively more accurately than the density itself. For a large enough volume the series-termination effect will be negligible. If the volume of integration is small enough for the value of $D(\mathbf{S})$ to be appreciable over the range of the data, however, the termination effect may still be serious.

Minimum-variance analysis

A technique which in principle removes the dependence of the variance of the electron density of the data cut-off has been proposed by Davis, Maslen & Varghese (1978). They define a modulated experimental density

$$\rho_M(\mathbf{r}) = \frac{1}{V}\sum M\mathbf{F}_e \exp(-i\mathbf{S}.\mathbf{r}) \tag{2}$$

where M is a function to be optimized, $\mathbf{F}_e$ is an estimated structure factor (complex) and V is the unit-cell volume. The summation extends over all observed reflections.

A quantity Q_M derived from $\rho_M(\mathbf{r})$ by integration or differentiation may be written as

$$Q_M = \frac{1}{V}\sum M\mathbf{F}_e D(\mathbf{S}). \tag{3}$$

For an integral, $D(\mathbf{S})$ is given by (1) where V may now be a general volume in the unit cell, not necessarily a parallelepiped. If Q_M is a derivative of the density, $D(\mathbf{S})$ will be obtained by applying the differential operator to $\exp(-i\mathbf{S}.\mathbf{r})$.

We seek the function $M = \boldsymbol{M}$ which minimizes a residual with the form of an estimate of the mean variance of Q_M in the unit cell

$$\langle \sigma^2(Q_M)\rangle_V = \frac{1}{V}\int \langle [Q_M - Q]^2\rangle\,\mathrm{d}\mathbf{r}. \tag{4}$$

Here Q is the true value of the quantity being considered and the expectation brackets in the right-hand side of the equation indicate an average over an ensemble of observations of Q.

Following the argument of Davis, Maslen & Varghese (1978) we find that the optimum form for M is

$$\boldsymbol{M} = \frac{\langle \mathbf{F}_e.\mathbf{F}\,D(\mathbf{S})\,D(-\mathbf{S})\rangle}{\langle \mathbf{F}_e.\mathbf{F}_e D(\mathbf{S})\,D(-\mathbf{S})\rangle} = \frac{\langle \mathbf{F}_e.\mathbf{F}\rangle}{\langle \mathbf{F}_e.\mathbf{F}_e\rangle}, \tag{5}$$

where $\mathbf{F}$ is the true value of the structure factor estimated by $\mathbf{F}_e$.

This is the same as the optimum filter for the density itself. The useful properties of $\boldsymbol{M}$ may be readily appreciated if the observations $\mathbf{F}_e$ are assumed to be symmetrically distributed around $\mathbf{F}$. For a strong reflection which is measured with small relative error, $\langle \mathbf{F}_e.\mathbf{F}\rangle \simeq \langle \mathbf{F}_e.\mathbf{F}_e\rangle$ so $\boldsymbol{M} \simeq 1$. For very weak reflections ($|\mathbf{F}| \rightarrow 0$) which predominate at high scattering angles we expect $\langle \mathbf{F}_e.\mathbf{F}\rangle \rightarrow 0$ while $\langle \mathbf{F}_e.\mathbf{F}_e\rangle$ approaches the variance in $\mathbf{F}_e$, so that $\boldsymbol{M} \rightarrow 0$. In this region of the data the measured structure factors are essentially all noise.

In practice, where only one measurement of each reflection is available, it may be necessary to replace the ensemble averages in (5) with averages over regions of reciprocal space, such as small intervals of $S = |\mathbf{S}|$ (Davis, Maslen & Varghese, 1978).

Estimate of the variance

The essential features of the variance may be illustrated by considering the special case of a centrosymmetric structure. In this case the structure factors are real numbers. With the assumption that F_e is symmetrically distributed around F and with the optimum filter function we obtain

$$\begin{aligned}\sigma^2(Q_M) = {} & \frac{1}{V^2}\sum MD(\mathbf{S})\,D(-\mathbf{S})\,\sigma^2(F_e)\\ & + \frac{1}{V^2}\sum MD(\mathbf{S})\,D(\mathbf{S})\,\sigma^2(F_e)\\ & + \frac{1}{V^2}\sum\sum{}'\,(1-M)(1-M')\langle F_e\rangle\langle F'_e\rangle\\ & \times D(\mathbf{S})\,D(\mathbf{S}'). \qquad (6)\end{aligned}$$

$\sum'$ is a restricted sum over all $\mathbf{S}' \neq \mathbf{S}$ or $-\mathbf{S}$ and $\sigma^2(F_e)$ is the variance of the observation F_e.

Significance of the terms in the variance

In general, for both integrals and derivatives, $D(\mathbf{S})$ contains an imaginary exponential factor which is periodic in $\mathbf{S}$ and position within the unit cell. The first term in (6), therefore, is independent of position while the other terms are position dependent.

The second term is negligible except when the derivative or integral is evaluated near a special position in the unit cell determined by the crystal symmetry (Cruickshank & Rollett, 1953). At these positions it may be of similar magnitude to the first term.

The third term in the variance is closely related to the 'mean square noise', $\langle[Q_M - Q_1]^2\rangle$, where Q_1 is evaluated from (3) by putting $M = 1$ for all reflections. The importance of this term will depend strongly on the form of $D(\mathbf{S})$.

It should be kept in mind that any error in the scale factor for the structure factors is likely to dominate the variance near nuclear positions.

Variance in integrated densities

Because $(1 - M)$ is small for most low-order reflections while $D(\mathbf{S})$ is small for the higher-order reflections, the third term in (6) will be negligible for an integrated density. At a general position in the unit cell the variance will be given by the first term only. Thus, for the centrosymmetric case

$$\sigma^2(Q_M) = \frac{1}{V^2}\sum MD(\mathbf{S})\,D(-\mathbf{S})\,\sigma^2(F_e). \qquad (7)$$

Example

A set of accurate X-ray data has recently been collected for copper sulphate pentahydrate $CuSO_4.5H_2O$, for which neutron diffraction parameters have been determined by Bacon & Titterton (1975). The optimum filter function is given in Davis, Maslen & Varghese (1978), and reproduced here in Fig. 2(*b*). M drops to a minimum of 0·317 at the data limit, $|\mathbf{S}|/4\pi = 1{\cdot}07$ Å^{-1}. Variances for integrated difference densities have been calculated for spherical regions of various radii. The appropriate form of $D(\mathbf{S})$ for an integral over a sphere was given by Kobayashi, Marumo & Saito (1972). For a sphere of radius R, with $S = |\mathbf{S}|$, equation (7) becomes

$$\sigma^2[Q_M\,(\text{sphere})] = \frac{9V^2}{V^2}\sum M\left[\frac{\sin(RS) - RS\cos(RS)}{(RS)^3}\right]^2\sigma^2(F_e). \qquad (8)$$

The results of this calculation, which are plotted in Fig. 1, show that the integrals are accurately determined.

As expected, the value of $\sigma(Q_M)/V$ approaches $\sigma[\Delta\rho_M] = 0{\cdot}066$ e Å^{-3} as the volume V approaches zero. When the radius is increased, $\sigma(Q_M)/V$ decreases as the contribution of high-order reflections is reduced.

Fig. 2(*a*) is a plot of the weighting function $MD(\mathbf{S})$ $D(-\mathbf{S})$, averaged over small intervals of S, for a sphere of 0·25 Å radius.

Beyond the data limit the weighting function is small, so extending the data set will make little difference to the variance of the integral.

Sensitivity to the shape of the region

For many interesting volumes of integration it may be difficult or impossible to express $D(\mathbf{S})$ in analytic form. To test the sensitivity of the variance to the shape of the region of integration, variances were also calculated from (7) for integrals over parallelepiped regions. The

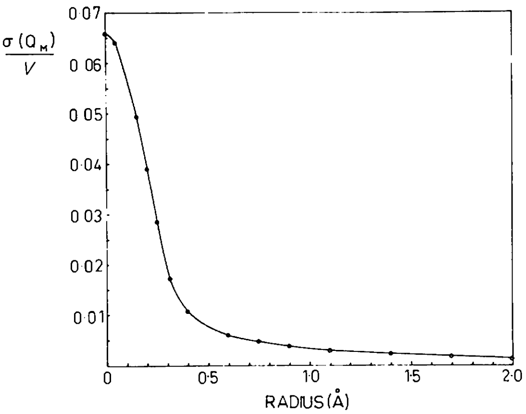


Fig. 1. Standard deviation (e Å^{-3}) of the average difference density of copper sulphate pentahydrate for averages taken over spheres of various radii.

Table 1. *Standard deviations for integrals of the difference density of copper sulphate pentahydrate*

Region of integration	$\sigma[Q_M]/V$
Sphere of radius = 0·25 Å	0·0286
Parallelepipeds of same volume as sphere with slides in ratio 1 : 1 : 1	0·0265
1 : 1 : 3	0·0267
1 : 1 : 6	0·0245

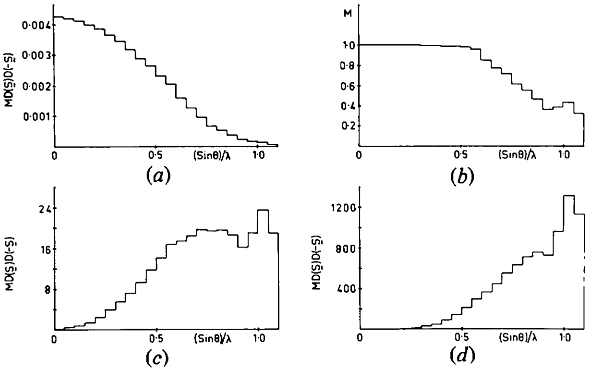
sides were taken parallel to the crystal axes and the expression of Coppens & Hamilton (1968) was used for $D(\mathbf{S})$. The parallelepipeds considered all had the same volume as a sphere of radius 0·25 Å, but the sides were in differing ratios. The values of the standard deviations are listed in Table 1.

The figures show that the standard deviation is not very sensitive to the shape of the region. For compact shapes, a good upper limit for the standard deviation of the integral is given by the spherical result, equation (8). This expression is considerably simpler to evaluate than the parallelepiped expression.

Variance in derivatives of the density

The difficulty in obtaining accurate estimates of derivatives of electron densities has been pointed out by Price & Maslen (1978). This can be appreciated in terms of the minimum-variance analysis by considering the special case of the nth partial derivative of $\rho_M(\mathbf{r})$ with respect to x, say, where $\mathbf{r} = (x,y,z)$ and $\mathbf{S} = (S_x,S_y,S_z)$ in some coordinate frame. For this case

$$D(\mathbf{S}) = (-iS_x)^n \exp(-i\mathbf{S}.\mathbf{r}). \tag{9}$$



(*a*) (*b*) (*c*) (*d*)

Fig. 2. The weighting function $MD(\mathbf{S})\,D(-\mathbf{S})$ as a function of S for the variance given by equation (7) using copper sulphate data for (*a*) an integral over a sphere of 0·25 Å radius, (*b*) the difference density, (*c*) the first derivative, and (*d*) the second derivative. Note the different vertical scales.

The weighting function $MD(\mathbf{S})\,D(-\mathbf{S})$ in the first term of (6) can be written as

$$MD(\mathbf{S})\,D(-\mathbf{S}) \simeq \frac{1}{3^n} M\langle S^{2n}\rangle_S \tag{10}$$

where S_x^{2n} has been replaced by its average value $1/3^n \times \langle S^{2n}\rangle_S$ in a small interval of S. This weighting function is plotted for the first derivative ($n = 1$) and the second derivative ($n = 2$) of the copper sulphate difference density in Fig. 2(*c*) and (*d*).

It is apparent from the size of the weighting function that the estimated variance of the derivatives will be much larger than the variance for the density, and its integrals. What is more, even though M decreases at high angles, the S^{2n} factor causes the high-order reflections to dominate the variance. Termination of the data causes the variance to be seriously underestimated.

In real-space terms, the derivatives are critically dependent on the fine structure of the Fourier map of the density. These features are heavily affected by random errors in the high-angle data. For the copper sulphate data no reasonable estimate of even the first derivative could be made.

Conclusion

Integrated electron densities may be accurately estimated from filtered diffraction data. The variance of the integrated density is not very sensitive to the shape of the region of integration. Approximating the region by a sphere of the same volume gives a good upper estimate of the variance which is easy to evaluate.

Reasonable estimates of derivatives of the difference density and their variances could not be obtained for copper sulphate pentahydrate, even using filtering techniques.

Computing costs were covered by a grant from the University of Western Australia Research Committee. One of us (CLD) acknowledges the receipt of a Commonwealth Postgraduate Research Award.

References

Bacon, G. E. & Titterton, D. H. (1975). *Z. Kristallogr.* **141**, 330–341.

Coppens, P. & Hamilton, W. C. (1968). *Acta Cryst.* B**24**, 925–929.

Cruickshank, D. W. J. (1949). *Acta Cryst.* **2**, 65–84.

Cruickshank, D. W. J. & Rollett, J. S. (1953). *Acta Cryst.* **6**, 705–707.

Davis, C. L., Maslen, E. N. & Varghese, J. N. (1978). *Acta Cryst.* A**34**, 371–377.

Kobayashi, A., Marumo, F. & Saito, Y. (1972). *Acta Cryst.* B**28**, 2709–2715.

Price, P. F. & Maslen, E. N. (1978). *Acta Cryst.* A**34**, 173–183.