Acta Cryst. (1973). A**29**, 201

The Optimum Distribution of Counting Times for Determining Integrated Intensities with a Diffractometer

BY J. K. MACKENZIE
Division of Chemical Physics, CSIRO, Clayton, Victoria, Australia

AND E. J. WILLIAMS
Department of Statistics, University of Melbourne, Parkville, Victoria, Australia

(*Received* 19 *October* 1972; *accepted* 31 *October* 1972)

For a given total counting time in a diffractometer experiment the variance (or square of the standard deviation) of the estimated integrated intensity of a reflexion is least when the counting time at each point is proportional to the square root of the intensity at that point. If the ratio of the peak to background intensity does not exceed 10 and data points are equidistant, a constant counting time at each point gives a variance not more than 1·4 times the minimum possible value. For greater peak/background ratios a further improvement in variance is achieved. The effect of subtracting off the background intensity is equivalent to regarding the line as extended, at background intensity, over a further range equal to the range over which measurements are made. Full advantage of the optimum choice can be obtained by computer control of the diffractometer, and then less time is required for collection of data of the same accuracy.

1. Introduction

A diffractometer is normally used for the accurate experimental determination of X-ray line profiles. In such an experiment the X-ray intensities η_r are estimated at a series of points x_r across the line profile by measuring the number of counts N_r obtained in time T_r at each of these points. This paper is concerned with the problem of choosing the individual counting times T_r so that, for a given total counting time $T=\sum T_r$, the greatest possible accuracy (minimum variance due to counting statistics) is obtained for the estimated integrated intensity of the reflexion. Furthermore, a simple but practical approximation to the optimum choice is indicated for the case of a manually controlled diffractometer. With full computer control there is every reason to take full advantage of the optimum choice.

When the estimates $y_r=N_r/T_r$ of the intensities η_r have been determined the integrated intensity is estimated by means of some approximate integration formula of the type

$$I=\int y\mathrm{d}x \simeq \sum \alpha_r y_r\,, \tag{1.1}$$

where the α_r are fixed constants determined only by the x_r and the nature of the approximation. The choice of this integration formula will not be considered here but it will be assumed that it has been so chosen that if the values of η_r were known exactly the formula would approximate the true value of the integral with sufficient accuracy. Within the limitations of this accuracy the final results obtained in this paper for the error in the estimated value of the integrated intensity are independent of the actual integration formula used; although, in practice, a trapezoidal formula is often the most convenient.

It has been assumed implicitly above that the intensity η_r decreases to zero as one moves away from the centre of the reflexion. However, in practice, this does not occur because of the presence of a background intensity; this background intensity will be assumed to be constant. Thus, the background intensity, integrated over the range of the experimental observations, must be subtracted from the integrated intensity obtained from the actual measurements. Since the estimate of the background intensity is subject to error, it is clear that the accuracy of the final result will depend, for a given total counting time, on a suitable balance between the times spent counting on the reflexion and that spent counting the background. The optimum balance is derived below.

In principle, the presence of the constant background has another effect on the accuracy of the final results for reflexions with very extended tails such as arise from heavily worked metals, for the final result is obtained by subtracting the integral of the (constant) background from the integral of the measured intensities. Thus, as the range of the measurements increases the final result is obtained as the difference between two quantities which are getting larger and larger and so the accuracy will ultimately decrease. Therefore, there is an optimum point at which measurements should be stopped even though the true intensity of the reflexion may not be quite zero at this stopping point. This problem is also considered below and it is shown that the optimum stopping point is usually experimentally inaccessible.

There are two other common procedures which will affect the final accuracy of the results. These are the removal from the actual measurements of the effects of one of the components of a doublet in the incident radiation and the removal of the effect of instrumental broadening. The problems raised by these two procedures are not discussed in detail in the present paper.

Instrumental broadening is sometimes removed by the use of the Fourier transform of the profile. Now this transform is obtained essentially by multiplying the intensity by a factor sin ux or cos ux and integrating over all values of x for which the intensity is not zero. Since sin ux and cos ux are numerically less than or equal to unity, it is clear that the error in the Fourier transform is always numerically less than or equal to the error in the integrated intensity. Thus, any experimental design which tends to minimize the error in the integrated intensity will equally tend to minimize this simple upper bound for the error in the Fourier transform apart, of course, from any error due to the inadequacy of the numerical integration formula which may be used.

2. Minimum variance of an integral

At each point of observation x_r a count N_r is obtained in time T_r and the square of the standard deviation or variance of such a count is equal to its expected value

$$V(N_r)=E(N_r)=\eta_r T_r\,. \tag{2.1}$$

Then, the estimated intensity is

$$y_r=N_r/T_r\,, \tag{2.2}$$

and its variance

$$V(y_r)=\eta_r/T_r\,. \tag{2.3}$$

Hence, since the observations at each point are independent, the variance of the sum S in (1.1) is given by

$$V(S)=V(\sum \alpha_r y_r)=\sum \alpha_r^2\eta_r/T_r\,. \tag{2.4}$$

The problem is to minimize $V(S)$ subject to the condition that the total time of observation T is constant. This can be done by the following device. Since

$$T=\sum T_r\,, \tag{2.5}$$

equation (2.4) can be written

$$V(S)=\sum \alpha_r^2\eta_r/T_r\,.\,\sum T_r/T\,. \tag{2.6}$$

Now applying Cauchy's inequality* (Hardy, Littlewood & Pólya, 1934; Theorem 7, p. 16) to the product of sums on the right of (2.6) it follows that

$$V(S)\geq(\sum \alpha_r\eta_r^{1/2})^2/T\,, \tag{2.7}$$

with equality, if and only if

$$T_r=\lambda\alpha_r\eta_r^{1/2}\,. \tag{2.8}$$

Here, λ is a constant of proportionality determined by the condition that $\sum T_r=T$; from which it follows that

$$\lambda=T/\sum \alpha_r\eta_r^{1/2}\,. \tag{2.9}$$

* The result follows from the identity

$$\sum a_r^2\sum b_r^2-(\sum a_r b_r)^2=\tfrac{1}{2}\sum\sum(a_r b_s-a_s b_r)^2.$$

Thus, the choice of T_r proportional to $\alpha_r\eta_r^{1/2}$ gives $V(S)$ a minimum value which cannot be improved since it is independent of T_r.

Note that since the sum appearing in (2.7) and (2.9) approximates to $\int \eta^{1/2}\,dx$ the results are independent, to this order of approximation, of the integration formula used.

3. The effect of the background

For a reflexion with a finite range X_0 *i.e.* having zero true intensity outside a range of length X_0 the background intensity β can be estimated as $b=N_b/T_b$ from a count N_b obtained in time T_b at some point outside this range. In this case the above treatment can be generalized to account for the effect of subtracting the background intensity. The estimate of the integrated intensity will be

$$I=\sum_{r=1}^{n}\alpha_r y_r-bX\,, \tag{3.1}$$

where X ($\geq X_0$) is the range over which observations are made.

Then,

$$V(I)=\sum_{r=1}^{n}\alpha_r^2\eta_r/T_r+\beta X^2/T_b\,, \tag{3.2}$$

and it is clear that for any given choice of α_r, T_r and T_b $V(I)$ increases with X. Thus, X should be chosen as small as possible, *i.e.* $X=X_0$.

Writing $\eta_{n+1}=\beta$, $\alpha_{n+1}=X$, $T_{n+1}=T_b$, equation (3.2) becomes identical in form with equation (2.4) so that the previous results apply. In particular, the minimum possible variance (2.7) can be written, on replacing the sum by an integral, as

$$V_{\min}=\text{Min }V(I)=\left(\int_0^X\eta^{1/2}\,dx+\beta^{1/2}X\right)^2/T\,. \tag{3.3}$$

For a fixed X this variance is inversely proportional to T and it is also proportional to the incident intensity of the X-ray beam I_0, say, because η is proportional to I_0. However, the integrated intensity, as here defined, is also proportional to I_0 so that the fractional accuracy (coefficient of variation) that is achieved is proportional to $(1/I_0T)^{1/2}$. This last statement is true whatever the choice of the T_r but by choosing T_r proportional to $\alpha_r\,\eta_r^{1/2}$ the constant of proportionality is made a minimum.

Since $\sum\alpha_s=X(s=1,n)$, the correction for background can be written $-\sum\alpha_s b_s$ with variance $\beta\sum\alpha_s^2/T_s$ appropriate to a series of fictitious observations b_s of the background β taken over a range X with counting times T_s. Now the optimum choice for T_s given by (2.8) is $T_s=\mu\alpha_s$ so that $T_b=\sum T_s=\mu\sum\alpha_s$ and this variance becomes $\beta(\sum\alpha_s)^2/T_b=\beta X^2/T_b$ in agreement with (3.2). Thus, provided T_s is chosen proportional to α_s in the background region, the variance (3.2) of the final result is *the same as though the tail of the profile were extended a distance X at height β* and the integrated intensity of this extended line were being estimated.

4. Practical considerations

To see what advantage may be obtained from an optimum choice, the variance $V_{\min}$ will be compared with the (greater) variance V_1 obtained by choosing T_r proportional to α_r. For equally spaced data with a trapezoidal integration formula the latter choice implies T_r constant (with the exception of the end points) and that half the total time is spent counting the background alone ($\alpha_{n+1}=\sum\alpha_r=X$). It will be shown that for plausible line shapes and when no correction for background is made the ratio $V_1/V_{\min}$ is less than about 1·4 provided the ratio of the peak intensity to the background is less than 25; if the latter ratio is less than 10 the stated limit for $V_1/V_{\min}$ holds for any profile whatsoever. Thus, for most profiles and a manually controlled diffractometer a two-stage choice of T_r with changes in T_r whenever the intensity drops by a factor of about 25 will give a suitable compromise between simplicity of experimental design and a reasonable approximation to the greatest possible accuracy. On the other hand, with computer control it is quite easy to change T_r at every point and so collect data of the same accuracy in up to 40% less time.

The origin of x will be taken at the peak of the profile and the units of x will be taken such that the half width of the profile at half true-peak height is unity. In these units the accessible range of x will not usually be greater than ± 15. In accordance with the remark at the end of the last section the effect of the background correction is equivalent to considering an extended line. Thus, no further attention need be explicity paid to this matter.

It follows from (2.4) and (2.7), on replacing sums by integrals, that for half the profile

$$V_1/V_{\min}=X\int_0^X \eta \, dx\Big/\left[\int_0^X \eta^{1/2}\, dx\right]^2. \qquad (4.1)$$

Now if ζ is the true line intensity so that $\eta=\zeta+\beta$, it follows on differentiating with respect to β and applying Tchebychef's inequality (Hardy, Littlewood & Pólya, 1934; Theorem 43, p. 43) that $V_1/V_{\min}$ decreases as the background increases. Thus, $V_1/V_{\min}$ will be overestimated if, as in the comparisons in Table 1, η is replaced by ζ.

Table 1. *Comparison of $V_1/V_{\min}$ for various profiles*

True profile	X	$V_1/V_{\min}$
Gaussian exp $-\frac{1}{2}x^2$	3	1·20
Cauchy or $1/(1+x^2)$ Lorentzian	3	1·13
	5	1·28
	10	1·63
	20	2·23
Linear $1-x/X$		1·125
Inverse square $1/x^2$ Tail region $1\le x\le X$	5	1·24
	10	1·53

The results in Table 1 show that provided T_r is constant in regions where the intensity drops by less than a factor 25 then $V_1/V_{\min}$ for the corresponding region is less than about 1·3. Thus, with a manually controlled diffractometer, a two-stage choice for the T_r will cope with profiles having a peak/background ratio of at least 500 and probably 1000. The point at which the change in counting time should be reduced is at $X_1=5$ and thereafter the inverse square law result shows that the new value of T_r can be used up to at least $X=25$ which should be sufficient to cope with an extended profile.

There remains the problem of determining the factor by which the T_r used in the peak region should be reduced for the tail region and also the question of when to use a two-stage design as opposed to a single counting time.

Denoting the integral of η over the relevant range by I_r where the subscripts 1, 2 denote values in the peak and tail regions respectively, it follows from (2·4) and (2.8) that since $T_1X_1+T_2X_2$ is constant, the best choice of T_1/T_2 is

$$T_1/T_2=(I_1/X_1)^{1/2}/(I_2/X_2)^{1/2}, \qquad (4.2)$$

where $X=X_1+X_2$ and a trapezoidal integration formula has been assumed. The corresponding generalization of (4.1) is

$$(V_2/V_{\min})^{1/2}=[(I_1X_1)^{1/2}+(I_2X_2)^{1/2}]\Big/\int_0^X \eta^{1/2}\, dx\,. \qquad (4.3)$$

and this is always less than the greater of the two ratios $(I_1X_1)^{1/2}/\int_0^{X_1}\eta^{1/2}\,dx$ and $(I_2X_2)^{1/2}/\int_{X_2}^{X_1}\eta^{1/2}\,dx$. Note that the value of X_2 must include the addition due to the extension of the profile.

Now for *any* profile for which the intensity always lies between the peak intensity η_p and the background intensity β the ratio

$$V_1/V_{\min}\le[1+(\eta_p/\beta)^{1/2}]^2/4(\eta_p/\beta)^{1/2}. \qquad (4.4)$$

(Hardy, Littlewood & Pólya, 1934; theorem 71, p. 62). Thus, for $\eta_p/\beta=10$, $V_1/V_{\min}=1·37$; the limiting value of $V_1/V_{\min}$ is achieved for a profile which extends at a constant intensity η_p out to $x=1$ and thereafter has a constant tail intensity β for a further distance $(\eta_p/\beta)^{1/2}$. For $\eta_p/\beta=25$ the above limit is 1·80 while for the same value of the ratio η_p/β a Cauchy line gives a maximum of 1·34 when X is about 15 and a Gaussian line gives a maximum of 1·56 when X is about 10. It would seem then that changing from a one to a two-stage design for $\eta_p/\beta>10$ should, in practice, give $V_1/V_{\min}<1·3$.

5. The effect of extended tails

For a profile with extended tails, and possibly an infinite range, measurements can of necessity be made only over a finite range X. If X is too small there will be a large error arising from the neglect of the residual tail area outside this range. On the other hand, if X is too large (3.2) indicates that there will again be a large error arising from the subtraction of the background. Thus, there will be an optimum stopping point. Extended profiles also present another difficulty; the determination of the background in the presence of a small but not necessarily negligible intensity of the true profile at the point of measurement. For the moment it will be assumed that the background can be independently determined.

A rough way of estimating the optimum stopping point is to minimize, for variation of X, the mean-square error, *i.e.* the sum of the error variance (3.3) and the square of the neglected residual tail area (the bias). Writing $\eta=\zeta+\beta$ and assuming that $\zeta/\beta\leq 1$ in the residual-tail region and also that the residual-tail area is small compared with the integral of the background over the range of measurement, it follows that the expression to be minimized is

$$V\simeq(A+2\beta^{1/2}X)^2/T+[\int_X^\infty \zeta\,\mathrm{d}x]^2, \tag{5.1}$$

where

$$A=\int_0^\infty[(\zeta+\beta)^{1/2}-\beta^{1/2}]\,\mathrm{d}x\,. \tag{5.2}$$

Note that A increases as the background decreases relative to the peak intensity of the line.

For a Cauchy line profile

$$\zeta=k/(1+x^2)\,, \tag{5.3}$$

$$A=[k/(k+\beta)^{1/2}]D[k^{1/2}/(k+\beta)^{1/2}]\,, \tag{5.4}$$

where D is the complete elliptic integral tabulated by Jahnke & Emde (1938).* Inspection of these tables shows that for $k/\beta<1000$, A is within a factor of 3 of the value $A=2k^{1/2}$. Now if, as will usually be the case, $2\beta^{1/2}X_{\text{opt}}\gg A$, the optimum stopping point, found by equating the derivative of (5.1) with respect to X to zero, is given approximately by

$$(X_{\text{opt}})^4=(k/\beta)^2\beta T/4\,. \tag{5.5}$$

Thus, the condition $2\beta^{1/2}X_{\text{opt}}\gg A$ becomes $4\beta k^2T\gg A^4$ or, putting $A=2k^{1/2}$,

$$\beta T\gg 4\,. \tag{5.6}$$

However, βT is the expected number of counts which would be obtained if the background were counted for the whole time T and will usually be greater than 10^4. Thus for $k/\beta>2$, $X_{\text{opt}}>10$.

The conclusion is that, except possibly for lines with very low peak intensities relative to the background, the optimum stopping point is experimentally inaccessible and so counting should continue as far out into the tails as possible.

When the line has extended tails and the background has to be determined in the presence of a non-negligible intensity from the line the problem is more complex. If it can be assumed that in the tail region the true intensity is given by

$$\eta=\beta+\kappa f(x)\,, \tag{5.7}$$

where $f(x)$ is of known functional form, then the values of β and κ could be estimated from observations made in the tail region by, say, the method of least-squares. Then, the estimate of the integrated intensity of the line would be

$$I=\sum\alpha_r y_r-bX+k\int_x^\infty f(x)\,\mathrm{d}x\,. \tag{5.8}$$

Since the estimated values of b, k and the y_r used in determining them will be correlated the exact conditions for the optimum experimental design will not be very simple and the matter will not be considered further here.

This paper is a revision of a draft manuscript nearly 15 years old and written when the authors were located respectively in the C.S.I.R.O. Divisions of Tribophysics and of Mathematical Statistics. It was not published then because it did not seem that real advantages for data collection would accrue without automatic control of the diffractometer. We wish to thank Dr R. C. G. Killean for drawing our attention to the relevance of our results to data collection with modern instrumentation. We realise, of course, that a great deal of work has been carried out on these problems in the intervening years, and our omission of references to it does not imply that we think it unimportant.

* The parameter denoted by k^2 in these tables is equal to $k/(k+\beta)$ in the present notation.

References

Hardy, G. H., Littlewood, J. E. & Pólya, G. (1934). *Inequalities.* Cambridge Univ. Press.

Jahnke, P. R. E. & Emde, F. (1938). *Tables of Functions,* 3rd ed., p. 83. New York: Stechert.