Least-Squares Refinement of Molecular Structures from Gaseous Electron-Diffraction Sector-Microphotometer Intensity Data. I. Method*

By Kenneth Hedberg† and Machio Iwasaki‡

Department of Chemistry, Oregon State University, Corvallis, Oregon, U.S.A.

An adaptation of the method of least squares to the refinement of molecular structures, based on gaseous electron-diffraction sector-microphotometer intensity data, is presented. The method has the advantages of permitting simultaneous refinement of many more parameters than can be handled by trial-and-error methods and of opening the possibility of obtaining realistic standard errors and error correlations in place of the more subjective 'error limits'. Although the problem of weighting of the observations is not solved, it is felt that the method provides a more straightforward way to take account of factors known to introduce uncertainties than does the alternative approach based on radial distribution functions.

A brief account of general experience with the method is given.

Introduction

For many years molecular structure refinement based on gaseous electron-diffraction data has been carried out largely by trial-and-error methods, for example by comparing theoretical intensity curves corresponding to various, slightly different combinations of structural parameters with the experimental intensity, or by fitting Gaussian curves to the peaks of the radial distribution obtained by Fourier transformation of the experimental intensity. Such methods are effective, though laborious, when the number of structural parameters is small and when the peaks of radial distribution curve are well resolved; however, in more complicated cases little can be done other than to make plausible assumptions about the less important parameters and to refine the more important ones, reporting the results with error limits sufficiently large to include the effect of uncertainty in the assumptions. In recent years this problem has been made more acute by the rapid development of the experimental techniques of gaseous electron diffraction: a well-executed experiment today yields a scattered electron distribution extending to approximately twice the scattering angle obtainable ten years ago and of more accurately known intensity. Analysis of such data by trial-and-error methods is even more laborious than before, partly because one is interested in realizing the more accurate parameter values inherent in the richer data, but also because one wishes to press the possibility of determining values for the less important parameters.

An obvious solution to the problem of refining

molecular structures determined from gaseous electron-diffraction data is through the method of least squares, so successfully used in X-ray crystallography (Hughes, 1941). One possible approach, to which some attention has been given (*e.g.* Bonham & Bartell, 1959) involves fitting theoretical peaks to the observed radial distribution curve. A second, and to us more appealing, approach is based on intensity distributions. Based as it is on the individual observations of scattered intensity (or on directly related measurements of photographic density as is most common practice), this approach permits a more natural assignment of weights and more easily understood error estimates. It has been discussed briefly by Cruickshank & Viervoll (1949) and, particularly with respect to use of visual intensity data, in detail by Hamilton & Schomaker (1954).

Some years ago one of us (K. H.) proposed a method for applying least squares to sector-microphotometer electron-diffraction intensity data (Bastiansen, Hedberg & Hedberg, 1956), and in that investigation (of 1,3,5,7-cyclo-octatetraene) fifteen parameters were refined. This gratifying success (the refinement of four parameters by the correlation technique is a practical limit) has encouraged us to develop the method, and after many applications it has now become a standard part of electron-diffraction structure investigation in this laboratory. Its major advantages have proven to be the large number of parameters which may be handled, the statistically significant error estimates and error correlations which come naturally from the calculations, and the speed of refinement.

It is the purpose of this article to describe the main features of the method and to summarize some of our general experience with its application. The article following describes an adaptation of the method to an automatic computer. In the third article of the

---

series the method is illustrated in application to a typical problem, the refinement of cyclopropane.

Matrix formulation of least squares

It is convenient to begin by summarizing the principles of least-squares curve-fitting in matrix form (see *e.g.* Arley & Buch, 1950). We suppose

$$f = f(x_1, x_2, \ldots, x_m) \tag{1}$$

is a non-linear function of m parameters x_q by which we seek to represent N $(N > m)$ experimental points. If approximate values x^0 are known for the m parameters, we have, neglecting higher order terms,

$$f = f(x_1^0, x_2^0, \ldots, x_m^0) + (\partial f / \partial x_1)_0 \Delta x_1$$
$$+ (\partial f / \partial x_2)_0 \Delta x_2 + \ldots + (\partial f / \partial x_m)_0 \Delta x_m \tag{2}$$

or, in matrix notation

$$\mathbf{F}_N = \mathbf{F}_N^0 + \mathbf{A}_{Nm} \mathbf{X}_m . \tag{3}$$

Equations (3) are the conditional equations. $\mathbf{F}_N$ and $\mathbf{F}_N^{\text{obs}}$ differ by $\mathbf{V}_N$, so that use of the observed values of the function gives, after dropping subscripts,

$$\mathbf{F}^{\text{obs}} + \mathbf{V} = \mathbf{F}^0 + \mathbf{A}\mathbf{X} \tag{4}$$

$$\mathbf{V} = \mathbf{F}^0 - F^{\text{obs}} + \mathbf{A}\mathbf{X} \tag{5}$$

$$= -\mathbf{N} + \mathbf{A}\mathbf{X} . \tag{6}$$

Introducing the diagonal weight matrix $\mathbf{P}_{NN} = \mathbf{P}$ and applying the least-squares criterion $\mathbf{V}'\mathbf{P}\mathbf{V} \to$ minimum ($\mathbf{V}'$ is the transpose of $\mathbf{V}$) we have

$$(\partial / \partial \mathbf{X})(\mathbf{V}'\mathbf{P}\mathbf{V}) = \mathbf{A}'\mathbf{P}\mathbf{V} = 0 . \tag{7}$$

Combination of (6) and (7) gives the normal equations

$$\mathbf{B}\mathbf{X} = \mathbf{Y} \tag{8}$$

where

$$\mathbf{B} = \mathbf{A}'\mathbf{P}\mathbf{A} \tag{9}$$

and

$$\mathbf{Y} = \mathbf{A}'\mathbf{P}\mathbf{N} . \tag{10}$$

The solutions are

$$\mathbf{X} = \mathbf{B}^{-1}\mathbf{Y} \tag{11}$$

and the standard errors $\sigma(x_q)$ are

$$\sigma(x_q) = [(B_{qq}^{-1})(\mathbf{V}'\mathbf{P}\mathbf{V})/(N - m)]^{\frac{1}{2}} \tag{12}$$

$$= [(B_{qq}^{-1})(\mathbf{N}'\mathbf{P}\mathbf{N} - \mathbf{X}_1'\mathbf{Y})/(N - m)]^{\frac{1}{2}}. \tag{13}$$

Application to intensity data

The problem of refinement by least squares based on intensity data is the problem of fitting a theoretical curve, which is a function of the molecular parameters, to an experimental curve derived from the measured intensity distribution. The particular type of experimental curve to be fitted depends upon the method of data reduction practised, and might well be different from case to case, although usually not. Our refinement method currently makes use of the following well-known intensity expression, which corresponds to a Gaussian probability distribution of distance displacements during molecular vibration:

$$kI_t(s) = k \sum_n A_n r_n^{-1} \exp\left(-\tfrac{1}{2} l_n^2 s^2\right) \sin s r_n . \tag{14}$$

The parameters whose best values are sought are k, part or all of the r's, and part or all of the l's; the amplitudes A_n, which may be constant or vary with s, are assumed known. The parameter k is an amplitude scale factor (the I^{obs} and I^{calc} are usually on different, arbitrary scales) and is of no structural interest. As a consequence of molecular symmetry, about which a good deal is often known, it is usual to regard only part of the r's as independent. In principle connections among the l's in the exponential vibration factors exist also, but it is convenient in practice to regard them as all independent. Refinement of some r's or l's corresponding to very weakly contributing distances may be unsuccessful or converge to unreasonable values: in such event their values may be assumed and the refinement carried out on the remaining parameters.

The trial structure and the N matrix

The process of refinement begins with selection of a suitable trial structure, which is usually obtained by analysis of the radial distribution curve calculated by Fourier inversion of the experimental intensity curve. Since this procedure is adequately described or referenced in almost any account of a gaseous electron-diffraction investigation, it need not be discussed here. The N matrix, a column matrix of N elements, may be obtained from the observed intensity curve and the theoretical curve calculated for the trial structure from (14) by the formula

$$\mathbf{N} = I^{\text{obs}}(s) - kI_t^{\text{calc}}(s) . \tag{15}$$

Here $I^{\text{obs}}(s)$ has been identified with $\mathbf{F}^{\text{obs}}$ and $kI_t^{\text{calc}}(s)$ with $\mathbf{F}^0$. In our work the interval $\Delta s = \tfrac{1}{4}$ Å^{-1} is most often used in forming these differences, although larger intervals are also common.* A trial value of k is readily obtained by rough comparison of the scales of I^{obs} and I^{calc} and, of course, is chosen so as to make N small. In further cycles of refinement N is calculated each time from the improved structure derived from the preceding cycle.

---

* The value $\tfrac{1}{4}$ derives from our usual practice of taking data at that interval from the continuous microphotometer traces of the photographic plates.

The matrix of derivatives A

The rectangular matrix $\mathbf{A} = \{A_{Nm}\}$ has elements of the type $(\partial k I_t^{\text{calc}}(s)/\partial k)$, $(\partial k I_t^{\text{calc}}(s)/\partial l)$, and $(\partial k I_t^{\text{calc}}(s)/\partial r)$. The first of these is obvious from (14), and since the l's are regarded as independent, the second is given by $-k l_n s^2 I_n(s)$, where $I_n(s)$ is a single term of (14) associated with l_n. The derivatives with respect to distance parameters are somewhat more difficult to calculate because cognizance must be taken of changes in the d dependent distances (d_d) as a function of the i distances chosen as parameters (r_i). In terms of these two types of distances the total intensity defined by (14) may be written

$$kI_t(s) = k[\sum_i I_i(r_i) + \sum_d I_d(d_d)] \qquad (16)$$

and the derivatives expressed by

$$\partial k I_t^{\text{calc}}/\partial r_i = k[\partial I_i^{\text{calc}}/\partial r_i + \sum_d (\partial I_d^{\text{calc}}/\partial d_d)(\partial d_d/\partial r_i)] \qquad (17)$$

where, of course,

$$\partial I_n^{\text{calc}}/\partial r_n$$
$$= A_n r_n^{-1} \exp(-\tfrac{1}{2}l_n^2 s^2)(s \cos sr_n - r_n^{-1} \sin sr_n) . \qquad (18)$$

Thus, in order to form the column vectors of the $\mathbf{A}$ matrix corresponding to the parameters r_i one must first calculate the quantities $\{\partial d_d/\partial r_i\} = \{J_{ai}\} = \mathbf{J}$ and then use (17). The direct calculation of $\mathbf{J}$, which requires keeping all but one distance parameter constant, is often laborious. We have found it more convenient to describe the structure in terms of bond angles and bond distances, which constitute a more natural set of parameters, and to calculate $\mathbf{J}$ from variation in these by transformations to the set of distance parameters. Details of this calculation are found in the Appendix.

The weight matrix P

In theory the weight matrix should reflect the effect of correlation, which surely exists, among the data. Because the problem of correlation is a complicated one, which needs thorough study, we have so far chosen to regard the measurements as independent, and to employ a diagonal weight matrix $\mathbf{P} = \{P_{NN}\}$; a rough estimate of the effect of correlation is made following the refinement. Weight functions of unity (except for ranges of zero corresponding to small or large scattering angles), and ones calculated according to

$$P = \text{const} \times s \exp(-bs^2) \qquad (19)$$

where b is an arbitrary constant, have been most common. Equation (19) has no special virtue, other than that of fitting our general experience concerning quality of the data.

Results and recycling

The use of the three matrices described above is indicated by equations (8) to (13). The quantities $\mathbf{X}$, of course, are the corrections to the distance, mean amplitude, and scale factor parameters, and the refined structure is obtained by adding them to the starting values. From this result a new $\mathbf{F}^0$ may be obtained and a new cycle begun. When recycling, particularly in later stages when the changes are small, it is often possible to use the $\mathbf{J}$ and $\mathbf{B}$ (hence $\mathbf{B}^{-1}$) matrices from the proceeding refinement cycle. We have found however, that it is usually best to recalculate $\mathbf{A}$ each time.

The error matrix M

The standard errors are given by (12) or (13), but the most complete statement of the refinement results includes as well the correlations among errors. This is expressed in the off-diagonal elements of $\mathbf{B}^{-1}$ by the equations

$$\varrho_{qp}\sigma(x_q)\sigma(x_p) = B_{qp}^{-1}(\mathbf{V'PV})/(N-m) \qquad (20)$$

$$= B_{qp}^{-1}S^2. \qquad (21)$$

Since the parameters actually adjusted are not necessarily the most natural set, it is of interest to compute from (12) or (13) and (20) or (21) the errors and correlations to be associated with such a set. The desired result is the error matrix $\mathbf{M}$:

$$\mathbf{M} = \{\mu_{qp}\} = S^2\mathbf{CB}^{-1}\mathbf{C'}. \qquad (22)$$

Here $\mathbf{C}$ is the matrix of coefficients transforming the parameters adjusted to the set of interest.

Some general experience

We have found the method generally easy to use, and very much faster than conventional refinement methods. However, it must be applied with care, and convergence is not always obtained. Such difficulties may be anticipated, of course, by the appearance of the radial distribution function from which the trial structure is taken. Indeed, some useful information is contained in non-convergence, which tends to indicate which parameters are marginally determinable — a circumstance otherwise often difficult to ascertain. The following paragraphs summarize some experience, much of which is drawn from work carried out in this laboratory by Prof. Otto Bastiansen.

In simple cases, characterized by well resolved peaks in the radial distribution curve, the method converges nicely even from a relatively poor trial structure (distance and mean amplitude parameters in error by several hundredths of 1 Å). The results

for the more important distances and associated mean amplitudes are usually within 0·001 Å of the values obtained by a careful analysis of the radial distribution curve. The main advantage of the least-squares method is speed, and the error information obtained.

In complicated cases, where the radial distribution curve is poorly resolved, good trial structures are usually needed. Convergence is sometimes obtained from a poor trial structure, but many cycles of refinement may be needed. More often, some of the parameters diverge and one faces the choice of looking for a better trial structure, of applying some sort of reduced-shift correction or of simplifying the problem by fixing some of the less important parameters. Even so, in such cases the more important parameters usually refine satisfactorily.

Results for the mean amplitudes of important distances are often quite different for refinements based on different ranges of the intensity curve: in general their values are found to be smaller using inner data. The reason is unclear, but one is reminded that systematic uncertainties in factors used in data reduction (for example, the form factors and exposure correction) are greatest in the inner region. If a similar effect for distance parameters is present, it is much less marked.

The standard errors obtained from a successful refinement are surprisingly small (*i.e.* 0·001–0·002 Å for bond distances) for the more important parameters. Since our weight matrices have not taken account of correlations among the data, these errors connected to parameter values are unrealistic. Nevertheless, their small magnitudes are an indication of high internal consistency among the data, and one is tempted to hope that once the major sources of correlation such as wave length of electrons, apparatus constants, *etc.* are understood and properly handled, such magnitudes will not be unreasonable.

The magnitudes of the errors corresponding to a distance parameter and to the mean-amplitude parameter associated with that distance are usually very similar. This is in contrast to the magnitudes of the errors corresponding to different distances or mean amplitudes, which may differ widely.

As would be expected, errors associated with well-resolved peaks in the radial distribution curve are relatively little correlated with others, while with poorly resolved or unresolved peaks they are highly correlated.

Possible improvements

A most important improvement could be made through the use of a non-diagonal weight matrix, in which proper account could be taken of correlations among the observations; such a weight matrix would lead to a more realistic assessment of errors. The nature of the correlations is not yet clear, but the sources are well known: uncertainties in experimental constants (wave length, camera distance, and sector function), in the photographic process (density-exposure relationship), and in the process of drawing in the experimental background.

Our description of the method has been based upon the simple equation (14), which is unsatisfactory when the probability distribution of displacements is not Gaussian. The least-squares refinement of parameters strongly affected by internal rotation or by out-of-plane vibrations, as well as investigation of second-order effects such as the anharmonicity of normal vibrations therefore requires a more general intensity function in place of (14). Further, one might anticipate the wish to adjust the phase factors $\cos \Delta \eta_{ij}$. It is possible to incorporate all such modifications.

APPENDIX

It is desired to calculate the elements of a matrix

$$\mathbf{J} = \{\mathbf{J}_{di}\} = \{\partial d_d / \partial r_i\} \tag{23}$$

which expresses the dependence of one type of distance ('dependent') on another type ('independent') used as parameters, starting from a model defined in terms of bond angles and bond distances. We define the following matrices:

$$\mathbf{X} = \{x_{i1}\} = \{r_1, r_2, \ldots, r_b, \alpha_1, \alpha_2, \ldots, \alpha_a\} \tag{24}$$

$$\mathbf{Y} = \{y_{h1}\} = \{r_{b+1}, r_{b+2}, \ldots, r_i, d_1, d_2, \ldots, d_d\} \tag{25}$$

$$\xi = \{\xi_{i1}\} = \{r_1, r_2, \ldots, r_b, r_{b+1}, r_{b+2}, \ldots, r_i\} \tag{26}$$

$$\eta = \{\eta_{k1}\} = \{d_1, d_2, \ldots, d_d, \alpha_1, \alpha_2, \ldots, \alpha_a\}. \tag{27}$$

$\mathbf{X}$ may be identified with the original set of a bond-angle and b bond-distance parameters, $\mathbf{Y}$ with the set of distances dependent on $\mathbf{X}$, ξ with a set of i distance parameters specifying the model, and η with a set of d distance and a angle parameters dependent on ξ. Clearly $i = a + b$ and $h = a + d$. Adopting the notation

$$\mathbf{D}^u_v = \{(D^u_v)_{hi}\} = \begin{pmatrix} \dfrac{\partial u_1}{\partial v_1} & \dfrac{\partial u_1}{\partial v_2} & \cdots & \dfrac{\partial u_1}{\partial v_i} \\ \vdots & & & \\ \dfrac{\partial u_h}{\partial v_1} & \dfrac{\partial u_h}{\partial v_2} & \cdots & \dfrac{\partial u_h}{\partial v_i} \end{pmatrix} \tag{28}$$

it is seen that the desired $\mathbf{J}$ matrix comprises the first d rows of $\mathbf{D}^\eta_\xi$, which is readily calculated from $\mathbf{D}^\eta_x$ and $\mathbf{D}^\xi_x$:

$$\mathbf{D}^\eta_\xi = \mathbf{D}^\eta_x \mathbf{D}^x_\xi \tag{29}$$

$$= \mathbf{D}^\eta_x (\mathbf{D}^\xi_x)^{-1}. \tag{30}$$

The matrices $\mathbf{D}^\eta_x$ and $\mathbf{D}^\xi_x$ are easily written down; they have the form

$$\mathbf{D}^{\eta}_{x} = \begin{pmatrix} \mathbf{D}^{d}_{r} & \mathbf{D}^{d}_{\alpha} \\ {}_{d,b} & {}_{d,a} \\ \hline \mathbf{O} & \mathbf{E} \\ {}_{a,b} & {}_{a,a} \end{pmatrix} \quad r = r_1, r_2, \ldots, r_b \qquad (31)$$

$$\mathbf{D}^{\xi}_{x} = \begin{pmatrix} \mathbf{E} & \mathbf{O} \\ {}_{b,b} & {}_{b,a} \\ \hline \mathbf{D}^{r'}_{r} & \mathbf{D}^{r'}_{\alpha} \\ {}_{a,b} & {}_{a,a} \end{pmatrix} \quad \begin{matrix} r = r_1, r_2, \ldots, r_b \\ r' = r_{b+1}, r_{b+2}, \ldots r_i \end{matrix} \qquad (32)$$

where $\mathbf{E}$ and $\mathbf{O}$ are the indentity and null matrices.

We wish to thank Prof. Otto Bastiansen and Mr Fred N. Fritsch for helpful discussion, and for significant contribution to the $\mathbf{J}$ matrix calculation.

References

ARLEY, N. & BUCH, K. R. (1950). *Introduction to the Theory of Probability and Statistics*. Chapter 12. New York: Wiley.

BASTIANSEN, O., HEDBERG, L. & HEDBERG, K. (1956). *J. Chem. Phys.* **27**, 1311.

BONHAM, R. A. & BARTELL, L. S. (1959). *J. Chem. Phys.* **31**, 702.

CRUICKSHANK, D. W. J. & VIERVOLL, H. (1949). *Acta Chem. Scand.* **3**, 560.

HAMILTON, W. A. & SCHOMAKER, V. (1954). Unpublished. But see Hamilton, W. A., Ph D. Thesis, California Institute of Technology, 1955.

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

Least-Squares Refinement of Molecular Structures from Gaseous Electron-Diffraction Sector-Microphotometer Intensity Data. II. Adaptation to Automatic Computation*

BY MACHIO IWASAKI,† F. N. FRITSCH AND KENNETH HEDBERG‡

Department of Chemistry, Oregon State University, Corvallis, Oregon, U.S.A.

A description is given of a computer program representing an adaptation of a least-squares method of molecular structure refinement based on gaseous electron-diffraction sector-microphotometer data.

Introduction

The principles of a least-squares refinement method, based on gaseous electron-diffraction sector-microphotometer data, were described in the preceding article (Hedberg & Iwasaki, 1964; hereafter called H & I). In a typical application of the method some 200 observations and 10–15 parameters may be handled, and the calculations involved are extensive enough to merit use of an automatic computer. Because the power of the method should attract increasing attention from workers in gaseous electron diffraction, and because the computational problems posed are quite different from those encountered in X-ray diffraction, a description of a computer program seems appropriate.

In the account given here we seek to present only the main features of the program and in a way to make clear the sequence of operations. The actual program, which will be of little interest to others

since it was written for a non-standard ALWAC III-E, differs in unimportant details from the flow diagrams. The complexity of the program is in part dictated by the nature of the computational problem and in part by the properties of the rather slow computer superposed on our objectives of making the refinement as automatic as possible and as versatile as the basic intensity equation H & I (14) permits. Thus, it has been necessary to build in special features such as optional approximations which allow savings of time, or checking calculations with optional recycling to protect a large investment of time. Use of one of the widely available high-speed computers, where a refinement cycle for the typical problem mentioned above would require seconds or minutes instead of 1–1½ hours as it does with ALWAC III-E, would make these features unnecessary and lead to simplifications.

List of symbols

For further explanation of many of the following quantities see H & I.

A Matrix of derivatives.

$$\mathbf{A} = \mathbf{A}_{Nm} = \{ \partial k I_t^{calc}(s) / \partial x_i \}_{Nm}$$

B $\mathbf{B} = \mathbf{A}'\mathbf{PA}$.