remains to be seen if similar regularities will also hold for graphs with $n = 12$ and more vertices.

The purpose of this work has been to demonstrate an analytical route to generation of graphs of prescribed form. In the series of papers that will follow the subject will be elaborated and applied to problems of interest in chemistry, such as construction of all molecular skeletons of prescribed valency distribution, construction of all polyhedra of a given size, construction of cluster forms of interest in chemistry of heavy elements, and construction of two-dimensional and three-dimensional connected networks of interest in crystal chemistry. The application can be extended to structures having multiple bonds and loops, as well as to directed graphs when appropriate modifications are introduced which take into account the new constraints on the adjacency matrices. Clearly, some ramifications might be of more interest for their mathematical or chemical, rather than their structural, aspect (for example, enumerating polyhedra or various isomers); as these arise they will be briefly mentioned, and the interested reader will be referred to more complete accounts of such studies presented elsewhere.

The author wishes to thank Dr D. H. Rouvray (Mülheim, Germany) for comments which led to an improvement of the presentation of the material. This work was supported by the US Department of Energy, Division of Basic Energy Sciences.

References

BALABAN, A. T. (1966). *Rev. Roum. Chim.* **11**, 1097–1116.
BALABAN, A. T. (1967). *Rev. Roum. Chim.* No. 1 (last, unnumbered, page).
BALABAN, A. T. (1970). *Rev. Roum. Chim.* **15**, 463–485.
BALABAN, A. T. (1972). *Rev. Roum. Chim.* **17**, 865–881; 883–896.
BALABAN, A. T. (1973). *Rev. Roum. Chim.* **18**, 635–653.
BALABAN, A. T. (1974). *Rev. Roum. Chim.* **19**, 1185–1196; 1611–1619.
BRÜCKNER, M. (1900). *Vielecke and Vielfläche.* Leipzig: Teubner.
DUNITZ, J. D. & PRELOG, V. (1968). *Angew. Chem.* **80**, 700.
GRACE, D. W. (1965). *Computer Search for Non-isomorphic Convex Polyhedra.* Stanford Computational Center Technical Report No. CS15.
HARARY, F. (1968). *Graph Theory.* Reading, Mass.: Addison-Wesley.
KURATOWSKI, K. (1930). *Fundam. Math.* **15**, 271–283.
LEDERBERG, J. (1965). *A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs.* NASA Report CR-68898.
LEDERBERG, J. (1966). *Systematics of Organic Molecules, Graph Topology and Hamilton Circuits.* NASA Report CR-68899.
MACKAY, A. L. (1975). *J. Chem.* Phys. **62**, 308–309.
RANDIĆ, M. (1974). *J. Chem. Phys.* **60**, 3920–3928.
RANDIĆ, M. (1976). *Chem. Phys. Lett.* **42**, 283–287.
RANDIĆ, M. (1977). *J. Chem. Inf. Comput. Sci.* **17**, 171–180.
STEINITZ, E. (1922). In *Encyklopädie der Mathematischen Wissenschaften*, IIIAB12, p. 1.

Acta Cryst. (1978). A**34**, 282–284

Stable Calculation of Coordinates from Distance Information

BY GORDON M. CRIPPEN AND TIMOTHY F. HAVEL

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco 94143, *USA*

(*Received* 12 *August* 1977; *accepted* 18 *October* 1977)

A new method is described for the calculation of Cartesian coordinates for n points given the $n \times n$ matrix of interpoint distances. The algorithm is faster than some earlier methods, and it is remarkably stable with respect to both numerical roundoff errors and errors in the given distance matrix. The resultant coordinates have their origin near the center of mass and axes approximately along the three principal rotational axes. The calculation is described of distances to the center of mass directly from the distance matrix. Results of computer trials of the algorithm are given.

Introduction

Our recent work (Crippen, 1977*a*,*b*; Kuntz, Crippen & Kollman, 1977) on the calculation of the conformation of proteins by the 'distance geometry' approach has as an important step the computation of atomic coordinates given a trial matrix of interatomic distances. The $n \times n$ trial matrix **D** is chosen to be elementwise bounded by a matrix of upper bound distances **U** and one of lower bound distances **L**, but even so, it is usually

impossible to find coordinates ($\mathbf{v}_i$, $i = 1, \ldots, n$) in three dimensions which match **D** exactly. In other words

$$||\mathbf{v}_i - \mathbf{v}_j|| = d_{ij} \text{ for all } i \text{ and } j, \quad (1)$$

has usually no solution for arbitrary **D**. What we desire are trial coordinates that approximately satisfy (1), which we can subsequently refine until

$$l_{ij} \leq ||\mathbf{v}_i - \mathbf{v}_j|| \leq u_{ij} \quad (2)$$

holds for all i and j. This problem is not confined to conformational calculations, but arises whenever incomplete or inaccurate geometric information on interpoint distances must be converted into Cartesian coordinates. The errors in the distances may stem from experimental inaccuracies or simply from numerical roundoff. A number of numerical solutions have been proposed (Crippen, 1977*a*,*b*; Mackay, 1974), but they are either time-consuming, or the trial coordinates are ill-conditioned (*i.e.* small changes in **D** give rise to large changes in the **v**'s), or the trial coordinates can be very far from satisfying (1) even when **D** is only slightly in error. The method given in the next section has proven in our experience to be a considerable improvement on all three counts.

The algorithm

I. Given the trial distances matrix $\mathbf{D} = (d_{ij})$ where i and $j = 1, \ldots, n$, calculate the distance of each point i to the center of mass, denoted as point O.

$$d_{iO}^2 = n^{-1} \sum_{j=1}^{n} d_{ij}^2 - n^{-2} \sum_{j=2}^{n} \sum_{k=1}^{j-1} d_{jk}^2. \quad (3)$$

See the Appendix for proof of (3). If **D** corresponds to distances between points located in Euclidean space of any number of dimensions, then (3) holds because the proof depends only on the most general properties of vectors.

II. Following Mackay (1974) we next calculate the $n \times n$ metric matrix $\mathbf{G} = (g_{ij})$ using the center-of-mass distances

$$g_{ij} = \tfrac{1}{2}(d_{iO}^2 + d_{jO}^2 - d_{ij}^2). \quad (4)$$

Equation (4) is simply the law of cosines, because g_{ij} is the dot product of the two vectors from the center of mass to points i and j respectively. Depending on the nature of the **D** matrix, the points should in general be viewed as still being located in $n-1$-dimensional Euclidean space, where $n - 1 > 3$.

III. Find the (largest) three eigenvalues, λ_1, λ_2 and λ_3, of **G**, such that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$. Similarly determine their corresponding eigenvectors, given as the three columns of the $n \times 3$ matrix **W**. Since **G** is real and symmetric (and usually positive semi-definite), a convenient algorithm for this step is the method of 'exhaustion' as described by Faddeev & Faddeeva (1963).

IV. Finally, the three coordinates of each of the n points, v_{ij}, $j = 1, \ldots, 3$ and $i = 1, \ldots, n$ are calculated by

$$v_{ij} = \lambda_j^{1/2} w_{ij}. \quad (5)$$

In the cases where some $\lambda_j < 0$ due to errors in **D**, we arbitrarily use $|\lambda_j|$ in place of λ_j. It is easy to show that (5) is correct by noting that

$$g_{ij} = \sum_{k=1}^{n} w_{ik} w_{ij} \lambda_k$$

when *all* eigenvalues and eigenvectors have been calculated. But if we knew the coordinates of the points in n-dimensional space, then

$$g_{ij} = \sum_{k=1}^{n} v_{ik} v_{jk}$$

by the definition of the metric matrix. Equating corresponding terms in the two sums yields (5).

Discussion

The stability of the algorithm is readily apparent. Calculating G with respect to the center of mass, as opposed to some possibly distant point, ensures that the g_{ij}'s are not uniform in magnitude and sign, and that the resultant coordinates are all comparable in magnitude. Choosing the first three largest eigenvalues amounts to choosing the coordinate axes along the largest three principal axes of the collection of points, thus attributing the least possible scatter to the fourth and higher dimensions. Therefore the **v**'s constitute an optimal fit to **D** [*i.e.* the least deviation from (1)] in this sense.

It has been our experience that the use of (3) gives non-negative distances to the center of mass for almost all reasonable choices of **D**. It can still happen that an eigenvalue will be negative, but at least some sort of useful coordinates will result, nevertheless. It should be noted that the coordinate generation of Mackay (1974) involves a Choleski decomposition of **G**, which also requires **G** to be positive definite.

In timed trials, the present algorithm runs some 20% faster than our previous method (Crippen, 1977*b*) when placing 80 points. The resultant trial coordinates give a consistently better match to **D** initially, and the success of the subsequent refinement to satisfy (2) is so assured that failure to converge in refinement can be taken as an indication of a geometric impossibility built into the boundary matrices, **U** and **L**. The method has been successful in conformational calculations on pancreatic

trypsin inhibitor, tobacco mosaic virus coat protein, rubredoxin, and lysozyme. Mackay's (1974) method is potentially much more rapid, since only a triangularization of **G** is involved, and no eigenvalues. However, in our hands, the resultant coordinates are very sensitive to small errors in **D**, apparently as a result of the inherent instability of Choleski decomposition of nearly singular matrices.

This work was supported by a grant from the Academic Senate of the University of California. The authors thank Dr I. D. Kuntz, for stimulating and helpful discussions.

APPENDIX

Proof of equation (3)

Let $\mathbf{r}_{kl}$ denote the vector from point l to point k; n = total number of points; and O denotes the center-of-mass point. From the definition of the center of mass of an array of points, each of unit mass:

$$\sum_{j=1}^{n} \mathbf{r}_{jO} = \mathbf{0} = \sum_{j=1}^{n} (\mathbf{r}_{1O} + \mathbf{r}_{1j})$$

so that

$$\mathbf{r}_{1O} = -n^{-1} \sum_{j=2}^{n} \mathbf{r}_{1j}$$

and

$$d_{1O}^2 = \mathbf{r}_{1O} \cdot \mathbf{r}_{1O} = n^{-2} \sum_{j=2}^{n} \sum_{k=2}^{n} \mathbf{r}_{1j} \cdot \mathbf{r}_{1k}.$$

By the law of cosines,

$$\begin{aligned} d_{1O}^2 &= (2n^2)^{-1} \sum_{j=2}^{n} \sum_{k=2}^{n} (d_{1j}^2 + d_{1k}^2 - d_{jk}^2) \\ &= (2n^2)^{-1} \left[2(n-1) \sum_{j=2}^{n} d_{1j}^2 - 2 \sum_{2=j<k}^{n} \sum d_{jk}^2 \right] \\ &= (n-1)/(n^2) \sum_{j=2}^{n} d_{j1}^2 - n^{-2} \sum_{2=j<k}^{n} \sum d_{jk}^2 \\ &= n^{-1} \sum_{j=1}^{n} d_{j1}^2 - n^{-2} \sum_{1=j<k}^{n} \sum d_{jk}^2. \end{aligned}$$

Since the labelling of the points is arbitrary, then we have the general formula

$$d_{iO}^2 = n^{-1} \sum_{j=1}^{n} d_{ij}^2 - n^{-2} \sum_{j=2}^{n} \sum_{k=1}^{j-1} d_{jk}^2.$$

References

CRIPPEN, G. M. (1977*a*). *J. Comput. Phys.* **24**(1), 96–107.
CRIPPEN, G. M. (1977*b*). *J. Comput. Phys.* In the press.
KUNTZ, I. D., CRIPPEN, G. M. & KOLLMAN, P. A. (1977). To be published.
MACKAY, A. L. (1974). *Acta Cryst.* A**30**, 440–447.
FADDEEV, D. K. & FADDEEVA, V. N. (1963). *Computational Methods of Linear Algebra*, pp. 307, 328–330. San Francisco: W. H. Freeman.

Acta Cryst. (1978). A**34**, 284–288

Diffraction from Dislocations and Grain Boundaries – An Optical Analogue Study

BY B. K. MATHUR, B. K. SAMANTARAY AND G. B. MITRA

Department of Physics, Indian Institute of Technology, Kharagpur 721302, *India*

(*Received* 22 *March* 1977; *accepted* 26 *July* 1977)

Optical transforms of the models of atomic configurations around edge and screw dislocations in f.c.c. and b.c.c. lattices as well as grain boundaries have been obtained with the help of a laser diffractometer. The models used were based on computer simulation studies of other workers. It has been observed that the intensity at the reciprocal lattice points splits into annular haloes or takes the 'figure of eight' shape in some cases. The directional dependence of the splitting has been compared with the existing theories. It has also been observed that with ordering of the dislocations at the grain boundaries, the diffraction pattern resembles that of a single dislocation.

Introduction

As a result of elastic strain around the lattice defects there is a displacement of atoms from their normal lattice sites. Huang (1947), considering a random distribution of defects, each producing a spherically symmetric displacement field $\mathbf{U} = \mathbf{r}/|\mathbf{r}|^3$, has shown that crystals containing such defects would give rise to