

Calculation of cartesian internal molecular coordinates

Arezoo Moradi^{a,*}, Aida Azouji^b

^aDepartment of Statistics, Faculty of Basic Science, University of Kurdistan, Sanandaj, Iran

^bDepartment of Biophysics, North Tehran Branch, Islamic Azad University, Tehran, Iran

(Communicated by Sirous Moradi)

Abstract

For the solution of a variety of problems which concern molecular structures, it is often necessary to calculate the Cartesian coordinates of the atoms from a set of geometric parameters. In order to fully utilize symmetry operations in the calculation of Cartesian coordinates, it is necessary to construct the molecule within a primary coordinate system. This is accomplished by introducing a set of four dummy atoms, by which unit vectors are specified. One atom is placed at the origin and the others at unit distances along each of the three Cartesian axes. In this framework according to the attachment procedures described below. The atoms of a molecule are positioned through the use of a secondary coordinate system. A base coordinate system is defined by the positions of three atoms, numbered 1, 2, and 3. The origin is taken at Atom 1, the negative x-axis passes through Atom 2, and Atom 3 lies in the first or second quadrant of the xy plane. In order to define the angles, the following conventions are adopted: for a general atom j , the atom to which it was attached is called j' , and the atom to which j' was attached is called j'' , and the atom to which j'' was attached is called j''' . Two special cases where these definitions do not suffice are atoms attached to Atom 1 and to Atom 2. Where j is attached to Atom 2, Atom 1 is taken as j'' and Atom 3 as j''' . Where j is attached to Atom 1, Atom 2 is taken as j'' and Atom 3 as j''' . The program described is being used extensively for specifying molecular structures for electron diffraction investigations. Other current uses are in programs for x-ray diffraction of polymers, for estimating magnitudes of energy via postulated force field functions, and in molecular orbital calculations.

Keywords: calculation, cartesian coordinates, internal molecular coordinates
2020 MSC: 74A25, 92E10

1 Introduction

For the solution of a variety of problems which concern molecular structures it is often necessary to calculate the Cartesian coordinates of the atoms from a set of geometric parameters. The type which appears natural to chemists is usually referred to as "valence parameters"; they are bond lengths, valence angles, torsional angles, and dihedral angles. Several schemes have been developed for performing this calculation [2, 4, 8], all derived from a method proposed by Eyring. Generally, these techniques have little flexibility in the choice of parameters which may be employed; it is often not possible to use the most convenient set of valence parameters [10].

The introduction of symmetry operators has substantially enhanced the applicability of the above methods. In addition, the effort required to use the computer program has been reduced by providing the user with a choice of several convenient modes of attachment. These also result in a substantial saving of computer time [1].

*Corresponding author

Email addresses: arezomoradi@rocketmail.com (Arezoo Moradi), aidaazuji99@gmail.com (Aida Azouji)

2 Method of calculation

In order to fully utilize symmetry operations in the calculation of Cartesian coordinates, it is necessary to construct the molecule within a primary coordinate system. This is accomplished by introducing a set of four dummy atoms, by which unit vectors are specified. One atom is placed at the origin and the others at unit distances along each of the three Cartesian axes. In this framework according to the attachment procedures described below. The atoms of a molecule are positioned through the use of a secondary coordinate system [6].

In a wide variety of experimental and theoretical studies involving molecular geometry, a description in terms of internal molecular coordinates (bond lengths, bond and dihedral angles) must be converted to a description in a Cartesian system. Often derivatives of these Cartesian coordinates with respect to certain of the internal coordinates are also desired. While a variety of schemes for conversion from internal to Cartesian descriptions have been described for particular applications/ none appear to combine all the following features: [5] description in terms of an attractive internal coordinate system; [7] applicability to a wide variety of molecular types; [3] analytical calculation of derivatives; [9] efficient organization tailored for machine computation. A method meeting these requirements is described here. During its development over a three-year period, this method, or part thereof, has been applied successfully in such diverse areas as energy minimization in a complex model molecular force field including non-bonded repulsions, the fitting of model structures to electron-diffraction data, and calculation of principal moments of inertia.

3 Calculation of coordinates

The calculation of coordinates as described in this section; the specific definition of internal coordinates used here is given in the following paragraph. The next two paragraphs describe a formulation of this procedure that is readily adapted to subsequent calculation of derivatives. A base coordinate system is defined by the positions of three atoms, numbered 1, 2, and 3. The origin is taken at Atom 1, the negative x axis passes through Atom 2, and Atom 3 lies in the first or second quadrant of the xy plane (see Fig. 1, 2). Coordinates and position vectors without subscripts will refer to this coordinate system. Atom 2 is considered to be bonded to Atom 1, and the bond length must be specified. A coordinate system is defined with origin at Atom 2 and Atom 1 lying on the negative X_2 axis; Atom 3 again lies in the first or second quadrant of the xy plane. From this point, additional atoms are added to this framework by specifying four quantities: the atom already present, to which the new atom is attached, the length of the bond thus formed, a bond angle, and a dihedral angle. In order to define the angles, the following conventions are adopted: for a general atom j , the atom to which it was attached is called j' , the atom to which j' was attached is called j'' , and the atom to which j'' was attached is called j''' . Two special cases where these definitions do not suffice are atoms attached to Atom 1 and to Atom 2. Where j is attached to Atom 2, Atom 1 is taken as j'' and Atom 3 as j''' , where j is attached to Atom 1, Atom 2 is taken as j'' and Atom 3 as j''' .

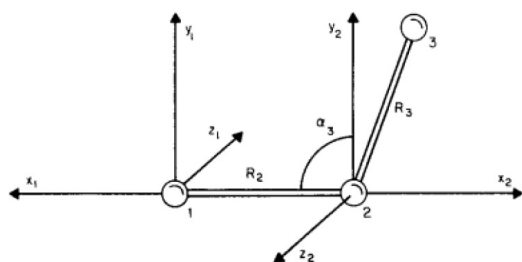


Figure 1: Coordinate systems centered on Atoms 1 and 2

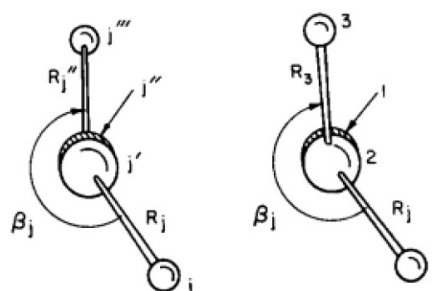


Figure 2: Definition of the angle β . The atom j'' is directly behind j' , and 2 is directly behind 1. The left-hand figure applies to all cases except $j'=2$

On each atom, a Cartesian system $X_j Y_h$ is now defined such that Atom j' lies on the negative X_j axis and atom j'' lies in the third or fourth quadrant of the $X_i Y_j$ plane. The coordinates of a point in this system are related to the coordinates of this same point in the j' system by a rotational transformation and a translation:

$$\begin{bmatrix} x_{j'} \\ y_{j'} \\ z_{j'} \end{bmatrix} = \begin{bmatrix} -\cos \alpha_j & -\sin \alpha_j & 0 \\ \sin \alpha_j \cos \beta_j & -\cos \alpha_j \cos \beta_j & -\sin \beta_j \\ \sin \alpha_j \sin \beta_j & -\cos \alpha_j \sin \beta_j & \cos \beta_j \end{bmatrix} \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} + \begin{bmatrix} -R_j \cos \alpha_j \\ R_j \sin \alpha_j \cos \beta_j \\ R_j \sin \alpha_j \sin \beta_j \end{bmatrix} \quad (3.1)$$

$$\begin{bmatrix} x_{j'} \\ y_{j'} \\ z_{j'} \\ 1 \end{bmatrix} = \begin{bmatrix} -\cos \alpha_j & -\sin \alpha_j & 0 & -R_j \cos \alpha_j \\ \sin \alpha_j \cos \beta_j & -\cos \alpha_j \cos \beta_j & -\sin \beta_j & R_j \sin \alpha_j \cos \beta_j \\ \sin \alpha_j \sin \beta_j & -\cos \alpha_j \sin \beta_j & \cos \beta_j & R_j \sin \alpha_j \sin \beta_j \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} \quad (3.2)$$

Or using q_j for the expanded coordinate vector and B_j for the transformation matrix:

$$q_{j'} = B_j \cdot q_{j*} \quad (3.3)$$

Transformation back to the base system on Atom 1 is then accomplished through a series of matrices B . For example, for an atom attached to a chain that is eventually attached to Atom 3,

$$q_1 = B_2 \cdot B_3 \cdot \dots \cdot B_{j''} \cdot B_{j'} \cdot B_j \cdot q_{j*} \quad (3.4)$$

The advantage of the 4×4 matrix notation is that Eq. (3.4) is a simple matrix product rather than a combination of products and sums. In order to make this process applicable to Atoms 2 and 3 and the definitions of the B matrices consistent. With these provisions, the matrices B suffice to transform the position vectors of all atoms to the base system. For convenience and efficiency it is desirable to calculate and store the 4×4 matrix A_j which transforms a position vector in the j system directly to the base coordinate system:

$$q = A_j \cdot q_j \quad (3.5)$$

$$A_j = B_k \cdot \dots \cdot B_{j''} \cdot B_{j'} \cdot B_j, \quad (3.6)$$

where k is the atom number (perhaps 2) through which the atom chain from j reaches atom 1. Clearly only a single matrix multiplication is now required as each new atom is added:

$$A_j = A_{j'} \cdot B_{j*} \quad (3.7)$$

The matrices A_j resembles the set B ; in that the fourth row contains three zeros and a one in that order. Further, the first three elements of the first column of A_j are the components in the base coordinate system of a unit vector along the X ; axis. Unit vectors along the Y ; and Z ; axes are similarly contained in the second and third columns. The first three elements of the fourth column are the coordinates of the atom j in the base system. These identifications may be useful in computational schemes involving particular molecular properties and will be important to one recipe for calculation of coordinate derivatives.

If only the atomic coordinates are desired, clearly the first three columns of a matrix A_j need not be calculated if j is a "terminal atom," that is, if no atom is to be described as attached to it. However, all the A_j s are needed for the second (and shorter) of the two methods to be described for obtaining coordinate derivatives.

4 Possible applications

The program described is being used extensively for specifying molecular structures for electron diffraction investigations. The memory requirements are relatively small, and hence it can be effectively inserted as a subroutine in larger programs which require the results of this type of calculation. Other current uses are in programs for x-ray diffraction of polymers, for estimating magnitudes of energy via postulated force field functions, and in molecular orbital calculations.

References

- [1] M. Brehm, M. Thomas, S. Gehrke, and B. Kirchner, *TRAVIS—A free analyzer for trajectories from molecular simulation*, J. Chem. Phys. **152** (2020), no. 16, 164105.
- [2] M. Ernst, F. Sittel, and G. Stock, *Contact-and distance-based principal component analysis of protein dynamics*, J. Chem. Phys. **143** (2015), no. 24, 12B640-1.
- [3] M. Ernst, S. Wolf, and G. Stock, *Identification and validation of reaction coordinates describing protein functional motion: Hierarchical dynamics of T4 lysozyme*, J. Chem. Theory Comput. **13** (2017), no. 10, 5076–5088.

-
- [4] S.R. Hare, L.A. Bratholm, D.R. Glowacki, and B.K. Carpenter, *Low dimensional representations along intrinsic reaction coordinates and molecular dynamics trajectories using interatomic distance matrices*, Chem. Sci. **10** (2019), no. 43, 9954–9968.
- [5] A. Hospital, J.R. Goñi, M. Orozco, and J.L. Gelpí, *Molecular dynamics simulations: advances and applications*, Adv. Appl. Bioinf. Chem. **8** (2015), 37–47.
- [6] F. Martín-García, E. Papaleo, P. Gomez-Puertas, W. Boomsma, and K. Lindorff-Larsen, *Comparing molecular dynamics force fields in the essential subspace*, PLoS One **10** (2015), no. 3, e0121114.
- [7] F. Sittel, T. Filk, and G. Stock, *Principal component analysis on a torus: Theory and application to protein dynamics*, J. Chem. Phys. **147** (2017), no. 24, 244101.
- [8] F. Sittel, A. Jain, and G. Stock, *Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates*, J. Chem. Phys. **141** (2014), no. 1, 07B605-1.
- [9] H.B. Thompson, *Calculation of cartesian coordinates and their derivatives from internal molecular coordinates*, J. Chem. Phys. **47** (1967), no. 9, 3407–3410.
- [10] D. Vlachakis, E. Bencurova, N. Papangelopoulos, and S. Kossida, *Current state-of-the-art molecular dynamics methods and applications*, Adv. Protein Chem. Struct. Biol. **94** (2014), 269–313.