3.3. MOLECULAR MODELLING AND GRAPHICS

Hardware systems which use true floating-point representations have less need of homogeneous coordinates and for these N and W may normally be set to unity.

3.3.1.1.3. *Notation*

In this chapter the conventions of matrix algebra will be adhered to except where it is convenient to show operations on elements of vectors, matrices and tensors, where a subscript notation will be used with a modified summation convention in which summation is over lower-case subscripts *only*. Thus the equation

$$x_I = A_{Ij}X_j$$

is to be read 'For any I, x_I is $A_{Ij}X_j$ summed over j'.

Subscripts using the letter i or later in the alphabet will relate to the usual three dimensions and imply a three-term summation. Subscripts a to h are not necessarily so limited, and, in particular, the subscript a is used to imply summation over atoms of which there may be an arbitrary number.

We shall use the superscript T to denote a transpose, and also use the Kronecker delta, δ_{IJ}, which is 1 if $I = J$ and zero otherwise, and the tensor ε_{IJK} which is 1 if I, J and K are a cyclic permutation of 1, 2, 3, -1 if an anticyclic permutation, and zero otherwise.

$$\varepsilon_{IJK} = (I-J)(J-K)(K-I)/2 \quad 1 \le I, J, K \le 3.$$

A useful identity is then

$$\varepsilon_{iJK}\varepsilon_{iLM} = \delta_{JL}\delta_{KM} - \delta_{JM}\delta_{KL}.$$

Single modulus signs surrounding the symbol for a square matrix denote its determinant, and around a vector denote its length.

The symbol $\simeq$ is defined in the previous section.

3.3.1.1.4. *Standards*

The sections of this chapter concerned with graphics are primarily concerned with the mathematical aspects of graphics programming as they confront the applications programmer. The implementations outlined in the final section have all, so far as the author is aware, been developed *ab initio* by their inventors to deal with these aspects using their own and unrelated techniques and protocols. It is clear, however, that standards are now emerging, and it is to be hoped that future developments in applications software will handle the graphics aspects through one or other of these standards.

First among these standards is the Graphical Kernel System, GKS, defined in *American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Graphical Kernel System (GKS) Functional Description* (1985) and described and illustrated by Hopgood *et al.* (1986) and Enderle *et al.* (1984). GKS became a full International Standards Organization (ISO) standard in 1985, and its purpose is to standardize the interface between application software and the graphics system, thus enhancing portability of software. Specifications for Fortran, Pascal and Ada formulations are at an advanced stage of development. Its value to crystallographers is limited by the fact that it is only two-dimensional. A three-dimensional extension known as GKS-3D, defined in *International Standards Organisation, International Standard Information Processing Systems – Computer Graphics – Graphical Kernel System for Three Dimensions (GKS-3D), Functional Description* (1988) became an ISO standard in 1988. Perhaps of greatest interest to crystallographers, however, is the Programmers' Hierarchical Interactive Graphics System (PHIGS) (Brown, 1985; Abi-Ezzi & Bunshaft, 1986) since this allows hierarchical segmentation of picture content to exist in both the applications software and the graphics device in a related manner, which GKS does not. Some graphics devices now

available support this type of working and its exploitation indicates the choice of PHIGS. Furthermore, Fortran implementations of GKS and GKS-3D require points to be stored in arrays dimensioned as $X(N)$, $Y(N)$, $Z(N)$ which may be equivalenced (in the Fortran sense) to $XYZ(N, 3)$ but not to $XYZ(3, N)$, which may not be convenient. PHIGS also became an International Standard in 1988: *American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Programmer's Hierarchical Graphics System (PHIGS) Functional Description, Archive File Format, Clear-Text Encoding of Archive File* (1988). PHIGS has also been extended to support the capability of raster-graphics machines to represent reflections, shadows, see-through effects *etc.* in a version known as PHIGS+ (van Dam, 1988).

Increasingly, manufacturers of graphics equipment are orienting their products towards one or other of these standards. While these standards are not the subject of this chapter it is recommended that they be studied before investing in equipment.

In addition to these standards, related standards are evolving under the auspices of the ISO for defining images in a file-storage, or metafile, form, and for the interface between the device-independent and device-dependent parts of a graphics package. Arnold & Bono (1988) describe the ANSI and ISO Computer Graphics Metafile standard which provides for the definition of (two-dimensional) images. The definition of three-dimensional scenes requires the use of (PHIGS) archive files.

3.3.1.2. *Orthogonal (or rotation) matrices*

It is a basic requirement for any graphics or molecular-modelling system to be able to control and manipulate the orientation of the structures involved and this is achieved using orthogonal matrices which are the subject of these sections.

3.3.1.2.1. *General form*

If a vector $\mathbf{v}$ is expressed in terms of its components resolved onto an axial set of vectors $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ which are of unit length and mutually perpendicular and right handed in the sense that $(\mathbf{X} \times \mathbf{Y}) \cdot \mathbf{Z} = +1$, and if these components are v_I, and if a second set of axes $\mathbf{X}'$, $\mathbf{Y}'$, $\mathbf{Z}'$ is similarly established, with the same origin and chirality, and if $\mathbf{v}$ has components v'_I on these axes then

$$v'_I = a_{Ij}v_j,$$

in which a_{IJ} is the cosine of the angle between the ith primed axis and the jth unprimed axis. Evidently the elements a_{IJ} comprise a matrix $\mathbf{R}$, such that any row represents one of the primed axial vectors, such as $\mathbf{X}'$, expressed as components on the unprimed axes, and each column represents one of the unprimed axial vectors expressed as components on the primed axes. It follows that $\mathbf{R}^T = \mathbf{R}^{-1}$ since elements of the product $\mathbf{R}^T\mathbf{R}$ are scalar products among perpendicular unit vectors.

A real matrix whose transpose equals its inverse is said to be *orthogonal*.

Since $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ can simultaneously be superimposed on $\mathbf{X}'$, $\mathbf{Y}'$ and $\mathbf{Z}'$ without deformation or change of scale the relationship is one of rotation, and orthogonal matrices are often referred to as rotation matrices. The operation of replacing the vector $\mathbf{v}$ by $\mathbf{R}\mathbf{v}$ corresponds to rotating the axes from the unprimed to the primed set with $\mathbf{v}$ itself unchanged. Equally, the same operation corresponds to retaining fixed axes and rotating the vector in the opposite sense. The second interpretation is the one more frequently helpful since conceptually it corresponds more closely to rotational operations on objects, and it is primarily in this sense that the following is written.

If three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ form the edges of a parallelepiped, then its volume V is

361

$$V = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \varepsilon_{ijk} u_i v_j w_k$$

and if these vectors are transformed by the matrix $\mathbf{R}$ as above, then the transformed volume V' is

$$V' = \varepsilon_{lmn} u'_l v'_m w'_n = \varepsilon_{lmn} a_{li} a_{mj} a_{nk} u_i v_j w_k.$$

But the determinant of $\mathbf{R}$ is given by

$$|\mathbf{R}|\varepsilon_{IJK} = \varepsilon_{lmn} a_{lI} a_{mJ} a_{nK}$$

so that

$$V' = |\mathbf{R}| V$$

and the determinant of $\mathbf{R}$ must therefore be +1 for a transformation which is a pure rotation. Nevertheless orthogonal matrices with determinant -1 exist though these do not describe a pure rotation. They may always be described as the product of a pure rotation and inversion through the origin and are referred to here as improper rotations. In what follows all references to orthogonal matrices refer to those with positive determinant only, unless stated otherwise.

An important general form of an orthogonal matrix in three dimensions was derived as equation (1.1.4.32) and is

$$\mathbf{R} = \begin{pmatrix} l^2 + (m^2 + n^2)\cos\theta & lm(1 - \cos\theta) - n\sin\theta & nl(1 - \cos\theta) + m\sin\theta \\ lm(1 - \cos\theta) + n\sin\theta & m^2 + (n^2 + l^2)\cos\theta & mn(1 - \cos\theta) - l\sin\theta \\ nl(1 - \cos\theta) - m\sin\theta & mn(1 - \cos\theta) + l\sin\theta & n^2 + (l^2 + m^2)\cos\theta \end{pmatrix}$$

or

$$R_{IJ} = (1 - \cos\theta) l_I l_J + \delta_{IJ} \cos\theta - \varepsilon_{IJk} l_k \sin\theta,$$

in which l, m and n are the direction cosines of the axis of rotation (which are the same when referred to either set of axes under either interpretation) and θ is the angle of rotation. In this form, and with $\mathbf{R}$ operating on column vectors on the right, the sign of θ is such that, when viewed along the rotation axis *from* the origin *towards* the point *lmn,* the object is rotated clockwise for positive θ with a fixed right-handed axial system. If, under the same viewing conditions, the axes are to be rotated clockwise through θ with the object fixed then the components of vectors in the object, on the new axes, are given by $\mathbf{R}$ with the same *lmn* and with θ negated. This is the transpose of $\mathbf{R}$, and if $\mathbf{R}$ is constructed from a product, as below, then each factor matrix in the product must be transposed and their order reversed to achieve this. Note that if, for a given rotation, the viewing direction from the origin is reversed, l, m, n and θ are all reversed and the matrix is unchanged.

Any rotation about a reference axis such that two of the direction cosines are zero is termed a *primitive rotation*, and it is frequently a requirement to generate or to interpret a general rotation as a product of primitive rotations.

A second important general form is based on Eulerian angles and is the product of three such primitives. It is

$$\mathbf{R} = \begin{pmatrix} \cos\varphi_3 & -\sin\varphi_3 & 0 \\ \sin\varphi_3 & \cos\varphi_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\varphi_2 & 0 & \sin\varphi_2 \\ 0 & 1 & 0 \\ -\sin\varphi_2 & 0 & \cos\varphi_2 \end{pmatrix} \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 & 0 \\ \sin\varphi_1 & \cos\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} (\cos\varphi_3\cos\varphi_2\cos\varphi_1 & -(\cos\varphi_3\cos\varphi_2\sin\varphi_1 & \cos\varphi_3\sin\varphi_2 \\ -\sin\varphi_3\sin\varphi_1) & +\sin\varphi_3\cos\varphi_1) & \\ (\sin\varphi_3\cos\varphi_2\cos\varphi_1 & (-\sin\varphi_3\cos\varphi_2\sin\varphi_1 & \sin\varphi_3\sin\varphi_2 \\ +\cos\varphi_3\sin\varphi_1) & +\cos\varphi_3\cos\varphi_1) & \\ -\sin\varphi_2\cos\varphi_1 & \sin\varphi_2\sin\varphi_1 & \cos\varphi_2 \end{pmatrix}$$

which is commonly employed in four-circle diffractometers for which $\varphi = -\varphi_1$, $\chi = \varphi_2$ and $\omega = -\varphi_3$. In terms of the fixed-axes–moving-object conceptualization this corresponds to a rotation φ_1 about Z followed by φ_2 about Y followed by φ_3 about Z. In the familiar diffractometer example, when $\chi = 0$ the φ and ω axes are both vertical and equivalent. If φ is altered first, then the χ axis is

still in the direction of a fixed Y axis, but if ω is altered first it is not. Since all angles are to be rotations about fixed axes to describe a rotating object it follows that it is φ rather than ω which corresponds to φ_1. In general, when rotating parts are mounted on rotating parts the rotation closest to the moved object must be applied first, forming the right-most factor in any multiple transformation, with the rotation closest to the fixed part as the left-most factor, assuming data supplied as column vectors on the right.

Given an orthogonal matrix, in either numerical or analytical form, it may be required to discover θ and the axis of rotation, or to factorize it as a product of primitives. From the first form we see that the vector

$$v_I = \varepsilon_{Ijk} a_{jk},$$

consisting of the antisymmetric part of $\mathbf{R}$, has elements $-2\sin\theta$ times the direction cosines l, m, n, which establishes the direction immediately, and normalization using $l^2 + m^2 + n^2 = 1$ determines $\sin\theta$. Furthermore, the trace is $1 + 2\cos\theta$ so that the quadrant of θ is also fixed. This method fails, however, if the matrix is symmetrical, which occurs if $\theta = \pi$. In this case only the direction of the axis is required, which is given by

$$l : m : n = (a_{23})^{-1} : (a_{31})^{-1} : (a_{12})^{-1}$$

for non-zero elements, or $l = \sqrt{\frac{1}{2}(a_{11} + 1)}$ *etc.*, with the signs chosen to satisfy $a_{12} = 2lm$ *etc.*

The Eulerian form may be factorized by noting that $\tan\varphi_1 = -a_{32}/a_{31}$, $\tan\varphi_3 = a_{23}/a_{13}$, $\cos\varphi_2 = a_{33}$. There is then freedom to choose the sign of $\sin\varphi_2$, but the choice then fixes the quadrants of φ_1 and φ_3 through the elements in the last row and column, and the primitives may then be constructed. These expressions for φ_1 and φ_3 fail if $\sin\varphi_2 = 0$, in which case the rotation reduces to a primitive rotation about Z with angle $(\varphi_1 + \varphi_3)$, $\varphi_2 = 0$, or $(\varphi_3 - \varphi_1)$, $\varphi_2 = \pi$.

Eulerian angles are unlikely to be the best choice of primitive angles unless they are directly related to the parameters of a system, as with the diffractometer. It is often more important that the changes to primitive angles should be quasi-linearly related to θ for any small rotations, which is not the case with Eulerian angles when the required rotation axis is close to the X axis. In such a case linearized techniques for solving for the primitive angles will fail. Furthermore, if the required rotation is about Z only $(\varphi_1 + \varphi_3)$ is determinate.

Quasi-linear relationships between θ and the primitive rotations arise if the primitives are one each about X, Y and Z. Any order of the three factors may be chosen, but the choice must then be adhered to since these factors do not commute. For sufficiently small rotations the primitive rotations are then $l\theta$, $m\theta$ and $n\theta$, whilst for larger θ linearized iterative techniques for finding the primitive rotations are likely to be convergent and well conditioned.

The three-dimensional space of the angles φ_1, φ_2 and φ_3 in either case is non-linearly related to θ. In the Eulerian case the worst non-linearities occur at the origin of φ-space. Equally severe non-linearities occur in the quasi-linear case also but are 90° away from the origin and less likely to be troublesome.

Neither of the foregoing general forms of orthogonal matrix has ideally convenient properties. The first is inconvenient because it uses four non-equivalent variables l, m, n and θ, with a linking equation involving l, m and n, so that they cannot be treated as independent variables for analytical purposes. The second form (the product of primitives) is not ideal because the three angles, though independent, are not equivalent, the non-equivalence arising from the non-commutation of the primitive factors. In the remainder of this section we give two further forms of orthogonal matrix which each use three variables which are independent and strictly equivalent, and a third form using four whose squares sum to unity.

The first of these is based on the diagonal and uses the three independent variables p, q, r, from which we construct the auxiliary variables

$$P = \pm\sqrt{1 + p - q - r}, \; Q = \pm\sqrt{1 - p + q - r},$$
$$R = \pm\sqrt{1 - p - q + r}, \; S = \pm\sqrt{1 + p + q + r},$$

then

$$\boldsymbol{R} = \begin{pmatrix} p & \frac{1}{2}[PQ - RS] & \frac{1}{2}[PR + QS] \\ \frac{1}{2}[PQ + RS] & q & \frac{1}{2}[QR - PS] \\ \frac{1}{2}[PR - QS] & \frac{1}{2}[QR + PS] & r \end{pmatrix}$$

is orthogonal with positive determinant for any of the sixteen sign combinations. The signs of P, Q, R and S are, respectively, the signs of the direction cosines of the rotation axis and of $\sin\theta$. Using also $T = \sqrt{4 - S^2}$, which may be deemed positive without loss of generality,

$$l = P/T, m = Q/T, n = R/T, \sin\theta = ST/2,$$
$$\cos\theta = 1 - T^2/2 = S^2/2 - 1.$$

Although p, q and r are independent, the point $[pqr]$ is bound, by the requirement that P, Q, R and S be real, to lie within a tetrahedron whose vertices are the points $[111]$, $[1\bar{1}\bar{1}]$, $[\bar{1}1\bar{1}]$ and $[\bar{1}\bar{1}1]$, corresponding to the identity and to 180° rotations about each of the axes. The facts that the identity occurs at a vertex of the feasible region and that $(1 - \cos\theta)$, rather than $\sin\theta$, is linear on p, q and r in this vicinity make this form suitable only for substantial rotations.

The second form consists in defining a rotation vector $\mathbf{r}$ with components u, v, w such that $u = lt$, $v = mt$, $w = nt$ with $t = \tan(\theta/2)$ and $\mathbf{r} \cdot \mathbf{r} = t^2$. Then the matrix

$$\boldsymbol{R} = \begin{pmatrix} \dfrac{1 + u^2 - v^2 - w^2}{1 + t^2} & \dfrac{2(uv - w)}{1 + t^2} & \dfrac{2(uw + v)}{1 + t^2} \\ \dfrac{2(uv + w)}{1 + t^2} & \dfrac{1 - u^2 + v^2 - w^2}{1 + t^2} & \dfrac{2(vw - u)}{1 + t^2} \\ \dfrac{2(uw - v)}{1 + t^2} & \dfrac{2(vw + u)}{1 + t^2} & \dfrac{1 - u^2 - v^2 + w^2}{1 + t^2} \end{pmatrix}$$
$$R_{IJ} = (1 + t^2)^{-1}[\delta_{IJ}(1 - u_k u_k) + 2(u_I u_J - \varepsilon_{IJl} u_l)]$$

is orthogonal and the variables u, v, w are independent, equivalent and unbounded, and, unlike the previous form, small rotations are quasi-linear on these variables. As examples, $\mathbf{r} = [100]$ gives 90° about X, $\mathbf{r} = [111]$ gives 120° about [111].

$\boldsymbol{R}$ then transforms a vector $\mathbf{d}$ according to

$$\boldsymbol{R}\mathbf{d} = \mathbf{d} + \frac{2}{1 + t^2}\{(\mathbf{r} \times \mathbf{d}) + [\mathbf{r} \times (\mathbf{r} \times \mathbf{d})]\}.$$

Multiplying two such matrices together allows us to establish the manner in which the rotation vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ combine.

$$\mathbf{r} = \frac{\mathbf{r}_2 + \mathbf{r}_1 + \mathbf{r}_2 \times \mathbf{r}_1}{1 - \mathbf{r}_2 \cdot \mathbf{r}_1}$$

for a rotation $\mathbf{r}_1$ followed by $\mathbf{r}_2$, so that rotations expressed in terms of rotation angles and axes may be compounded into a single such rotation without the need to form and decompose a product matrix.

Note that if $\mathbf{r}_1$ and $\mathbf{r}_2$ are parallel this reduces to the formula for the tangent of the sum of two angles, and that if $\mathbf{r}_1 \cdot \mathbf{r}_2 = 1$ the combined rotation is always 180°. Note, too, that reversing the order of application of the rotations reverses only the vector product.

If three rotations $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are applied successively, $\mathbf{r}_1$ first, then their combined rotation is

$$\mathbf{r} = [\mathbf{r}_3(1 - \mathbf{r}_1 \cdot \mathbf{r}_2) + \mathbf{r}_2(1 + \mathbf{r}_3 \cdot \mathbf{r}_1) + \mathbf{r}_1(1 - \mathbf{r}_3 \cdot \mathbf{r}_2)$$
$$+ \mathbf{r}_3 \times \mathbf{r}_2 + \mathbf{r}_3 \times \mathbf{r}_1 + \mathbf{r}_2 \times \mathbf{r}_1]$$
$$\times [1 - \mathbf{r}_1 \cdot \mathbf{r}_2 - \mathbf{r}_2 \cdot \mathbf{r}_3 - \mathbf{r}_3 \cdot \mathbf{r}_1 - \mathbf{r}_3 \cdot (\mathbf{r}_2 \times \mathbf{r}_1)]^{-1}.$$

Note the irregular pattern of signs in the numerator.

Similar ideas, using a vector of magnitude $\sin(\theta/2)$, are developed in Aharonov *et al.* (1977).

The third form of orthogonal matrix uses four variables, λ, μ, ν and σ, which comprise a four-dimensional vector $\boldsymbol{\rho}$, such that $\lambda = ls$, $\mu = ms$, $\nu = ns$ with $s = \sin(\theta/2)$ and $\sigma = \cos(\theta/2)$. In terms of these variables

$$\boldsymbol{R} = \begin{pmatrix} (\lambda^2 - \mu^2 - \nu^2 + \sigma^2) & 2(\lambda\mu - \nu\sigma) & 2(\lambda\nu + \mu\sigma) \\ 2(\mu\lambda + \nu\sigma) & (-\lambda^2 + \mu^2 - \nu^2 + \sigma^2) & 2(\mu\nu - \lambda\sigma) \\ 2(\lambda\nu - \mu\sigma) & 2(\mu\nu + \lambda\sigma) & (-\lambda^2 - \mu^2 + \nu^2 + \sigma^2) \end{pmatrix}.$$

Two further matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ may be defined (Diamond, 1988),

$$\boldsymbol{S} = \begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix} \text{ and } \boldsymbol{T} = \begin{pmatrix} \sigma & -\nu & \mu & \lambda \\ \nu & \sigma & -\lambda & \mu \\ -\mu & \lambda & \sigma & \nu \\ -\lambda & -\mu & -\nu & \sigma \end{pmatrix},$$

which are themselves orthogonal (though $\boldsymbol{S}$ has determinant -1) and which have the property that

$$\boldsymbol{S}^2 = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{pmatrix}$$

so that, for example, if homogeneous coordinates are being employed (Section 3.3.1.1.2)

$$\begin{pmatrix} x' \\ y' \\ z' \\ w \end{pmatrix} = \begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix} \begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$$

is a rotation of (x, y, z, w) through the angle θ about the axis (l, m, n). With suitably pipelined hardware this forms an efficient means of applying rotations since the 'overhead' of establishing $\boldsymbol{S}$ is so trivial.

$\boldsymbol{T}$ has the property that the rotation vector $\boldsymbol{\rho}$ arising from a concatenation of n rotations is

$$\boldsymbol{\rho} = \boldsymbol{T}_n \boldsymbol{T}_{n-1} \ldots \boldsymbol{T}_1 \boldsymbol{\rho}_0,$$

in which $\boldsymbol{\rho}_0^T$ is the vector $(0, 0, 0, 1)$ which defines a null rotation. This equation may be used as a basis for factorizing a given rotation into a concatenation of rotations about designated axes (Diamond, 1990*a*).

Finally, an exact rotation of the vector $\mathbf{d}$ may be obtained without using matrices at all by writing

$$\mathbf{d} = \sum_{0}^{\infty} \mathbf{d}_n$$

in which

$$\mathbf{d}_n = \frac{1}{n}(\boldsymbol{\theta} \times \mathbf{d}_{n-1})$$

and $\mathbf{d}_0$ is the initial position which is to be rotated. Here $\boldsymbol{\theta}$ is a vector with direction cosines l, m and n, and magnitude equal to the required rotation angle in radians (Diamond, 1966). This method is particularly efficient when $|\boldsymbol{\theta}| \ll 1$ or when the number of vectors to be transformed is small since the overhead of establishing $\boldsymbol{R}$ is eliminated and the process is simple to program. It is the three-dimensional analogue of the power series for $\sin\theta$ and $\cos\theta$ and has the same convergence properties.

3.3.1.2.2. *Measurement of rotations and strains from coordinates*

Given the coordinates of a molecular fragment it is often a requirement to relate the fragment to its image in some standard orientation by a transformation which may be required to be a pure rotation, or may be required to be a combination of rotation and strain. Of the methods reviewed in this section all except (iv) are concerned with pure rotation, ignoring any strain that may be present, and give the best rigid-body superposition. In all these methods, unless inhomogeneous strain is being considered, the best possible superposition is obtained if the centroids of the two images are first brought into coincidence by translation and treated as the origin.

Methods (i) to (v) seek transformations which perform the superposition and impose on these, in various ways, the requirements of orthogonality for the rotational part. All these methods therefore need some defence against indeterminacy that arises in the general transformation if one or both of the fragments is planar, and, if improper rotations are to be excluded, need a defence against these also if the fragment and its image are of opposite chirality. Methods (vi) and (vii) pay no attention to the general transformation and work with variables which are intrinsically rotational in character, and always produce an orthogonal transformation with positive determinant, with no degeneracy arising from planar fragments which need no special attention. Even collinear atoms cause no problem, the superposition being performed correctly but with an arbitrary rotation about the length of the line being present in the result. These methods are therefore to be preferred over the earlier ones unless the purpose of the operation is to detect differences of chirality, although this, too, can be detected with a simple test.

In this review we adopt the same notation for all the methods which, unavoidably, means that symbols are used in ways which differ from the original publications. We use the symbol x for the vector set which is to be rotated and X for the vector set whose orientation is not to be altered, and write the residuals as

$$e_{IA} = D_{Ij}x_{jA} - X_{IA}$$

and, by choice of origin,

$$W_a x_{Ia} = W_a X_{Ia} = 0_I$$

for weights W. The quadratic residual to be minimized is

$$E = W_a e_{ia} e_{ia}$$

and we define the matrix $M_{IJ} = W_a x_{Ia} X_{Ja}$ and use l for the direction cosines of the rotation axis.

(i) McLachlan's first method (McLachlan, 1972, 1982) is iterative and conceptually the simplest. It sets

$$D_{IJ} = A_{Ik}R_{kJ}$$

in which A and R are both orthogonal with R being a current estimate of D and A being an adjustment which, at the beginning of each cycle, has a zero angle associated with it. One iterative cycle estimates a non-trivial A, after which the product AR replaces R.

$$A_{IJ} = (1 - \cos\theta)l_I l_J + \delta_{IJ}\cos\theta - \varepsilon_{IJk}l_k \sin\theta$$

and

$$\left(\frac{\partial A_{IJ}}{\partial \theta}\right)_{\theta=0} = -\varepsilon_{IJk}l_k,$$

therefore

$$\left(\frac{\partial E}{\partial \theta}\right)_{\theta=0} = 2W_a\left(\frac{\partial A_{ij}}{\partial \theta}\right)_{\theta=0} R_{jk}x_{ka}(A_{il}R_{lm}x_{ma} - X_{ia})$$
$$= 2\varepsilon_{ijl}R_{jk}M_{ki}l_l.$$

For this to vanish for all possible rotation axes l the vector

$$g_L = \varepsilon_{ijL}R_{jk}M_{ki}$$

must vanish, *i.e.* at the end of the iteration R must be such that the matrix

$$N_{JI} = R_{Jk}M_{kI}$$

is symmetrical. The vector g represents the couple exerted on the rotating body by forces $2W_A(R_{Ij}x_{jA} - X_{IA})$ acting at the atoms. Choosing

$$l_L = g_L/|g|$$

gives the greatest $|\partial E/\partial \theta|_{\theta=0}$ and $(\partial E/\partial \theta)$ vanishes when

$$\tan\theta = \frac{\varepsilon_{ijk}N_{ji}l_k}{N_{pq}(l_p l_q - \delta_{pq})}$$

in which N is constructed from the current R matrix. A is then constructed from l and this θ and AR replaces R. The process is iterative because a couple about some new axis can appear when rotation about g eliminates the couple about g.

Note that for each rotation axis l there are two values of θ, differing by π, which reduce $|g|$ to zero, corresponding to maximum and minimum values of E. The minimum is that which makes

$$\frac{\partial^2 E}{\partial \theta^2} = 2(\text{tr } N - l_i N_{ij}l_j)$$

positive. Adding π to θ alters R and N and negates this quantity.

Note, too, that the process is essentially characterized as that which makes the product RM symmetrical with R orthogonal. We return to this point in (iii).

(ii) Kabsch's method (Kabsch, 1976, 1978) minimizes E with respect to the nine elements of D, subject to the six constraints

$$D_{kI}D_{kJ} - \delta_{IJ} = 0_{IJ},$$

by using an auxiliary function

$$F = L_{ij}(D_{ki}D_{kj} - \delta_{ij})$$

in which L is symmetric containing six Lagrange multipliers. The Lagrangian function

$$G = E + F$$

then has minima with respect to the elements of D at locations which are dependent, *inter alia*, on the elements of L. By suitably choosing L a minimum of G may be brought into coincidence with the constrained minimum of E. A minimum of G occurs where

$$\frac{\partial G}{\partial D_{IJ}} = 2D_{Ik}(S_{Jk} + L_{Jk}) - 2M_{JI} = 0_{IJ}$$

and the 9×9 matrix

$$\frac{\partial^2 G}{\partial D_{MK}\partial D_{IJ}} = 2\delta_{MI}(S_{JK} + L_{JK})$$

is positive definite, block diagonal, and has

$$S_{JK} = W_a x_{Ja} x_{Ka}$$

which is symmetrical. Thus L must be chosen so as to make the symmetric matrix $(S + L)$ such that

$$D(S + L)^T = M^T$$

with $\boldsymbol{D}$ orthogonal, or $\boldsymbol{RN} = \boldsymbol{M}^T$ with $\boldsymbol{R}$ replacing $\boldsymbol{D}$ since we are now confined to the orthogonal case, and $\boldsymbol{N}$ is symmetric and positive definite.

(iii) Comparison of the Kabsch and McLachlan methods. Using the initials of these authors as subscripts, we have seen that the Kabsch solution involves solving

$$\boldsymbol{R}_{\text{WK}}\boldsymbol{N}_{\text{WK}} = \boldsymbol{M}^T$$

for an orthogonal matrix $\boldsymbol{R}_{\text{WK}}$ given that $\boldsymbol{N}_{\text{WK}}$ is symmetrical and positive definite. Thus

$$\boldsymbol{MM}^T = \boldsymbol{N}_{\text{WK}}^T\boldsymbol{R}_{\text{WK}}^T\boldsymbol{R}_{\text{WK}}\boldsymbol{N}_{\text{WK}} = \boldsymbol{N}_{\text{WK}}^2$$

and

$$\boldsymbol{R}_{\text{WK}} = \boldsymbol{M}^T(\boldsymbol{MM}^T)^{-1/2}.$$

By comparison, the McLachlan treatment leads to an orthogonal $\boldsymbol{R}$ matrix satisfying

$$\boldsymbol{R}_{\text{ADM}} = \boldsymbol{N}_{\text{ADM}}\boldsymbol{M}^{-1}$$

in which $\boldsymbol{N}_{\text{ADM}}$ is also symmetric and positive definite, which similarly leads to

$$\boldsymbol{R}_{\text{ADM}} = (\boldsymbol{M}^T\boldsymbol{M})^{1/2}\boldsymbol{M}^{-1}.$$

These seemingly different expressions for $\boldsymbol{R}_{\text{WK}}$ and $\boldsymbol{R}_{\text{ADM}}$ are, in fact, equal, as the following shows

$$\boldsymbol{R}_{\text{WK}} = \boldsymbol{R}_{\text{ADM}}\boldsymbol{R}_{\text{ADM}}^{-1}\boldsymbol{R}_{\text{WK}} = \boldsymbol{R}_{\text{ADM}}\boldsymbol{M}\boldsymbol{N}_{\text{ADM}}^{-1}\boldsymbol{M}^T\boldsymbol{N}_{\text{WK}}^{-1},$$

therefore

$$\boldsymbol{R}_{\text{WK}}^T\boldsymbol{R}_{\text{WK}} = \boldsymbol{I}$$
$$= \boldsymbol{N}_{\text{WK}}^{-1}\boldsymbol{M}\boldsymbol{N}_{\text{ADM}}^{-1}\boldsymbol{M}^T\boldsymbol{R}_{\text{ADM}}^T\boldsymbol{R}_{\text{ADM}}\boldsymbol{M}\boldsymbol{N}_{\text{ADM}}^{-1}\boldsymbol{M}^T\boldsymbol{N}_{\text{WK}}^{-1}.$$

Multiplying on both sides by $\boldsymbol{N}_{\text{WK}}$ gives

$$\boldsymbol{N}_{\text{WK}}^2 = (\boldsymbol{M}\boldsymbol{N}_{\text{ADM}}^{-1}\boldsymbol{M}^T)^2,$$

and since both $\boldsymbol{N}$ matrices are positive definite

$$\boldsymbol{N}_{\text{WK}} = \boldsymbol{M}\boldsymbol{N}_{\text{ADM}}^{-1}\boldsymbol{M}^T$$

and conversely

$$\boldsymbol{N}_{\text{ADM}} = \boldsymbol{M}^T\boldsymbol{N}_{\text{WK}}^{-1}\boldsymbol{M},$$

therefore

$$\boldsymbol{R}_{\text{WK}} = \boldsymbol{M}^T\boldsymbol{M}^{T-1}\boldsymbol{N}_{\text{ADM}}\boldsymbol{M}^{-1} = \boldsymbol{R}_{\text{ADM}}.$$

(iv) Diamond's first method. This method (Diamond, 1976a) differs from the previous ones in that the transformation $\boldsymbol{D}$ is allowed to be a general transformation which is then factorized into the product of an orthogonal matrix $\boldsymbol{R}$ and a symmetrical matrix $\boldsymbol{T}$. The transformation of $\boldsymbol{x}$ to fit $\boldsymbol{X}$ is thus interpreted as the combination of homogeneous strain and pure rotation in which $\boldsymbol{x}$ is subjected to strain and the result is rotated.

$$\boldsymbol{D} = \boldsymbol{RT}$$
$$\boldsymbol{T}^2 = \boldsymbol{D}^T\boldsymbol{D}$$
$$\boldsymbol{T} = (\boldsymbol{D}^T\boldsymbol{D})^{1/2}$$
$$\boldsymbol{R} = \boldsymbol{D}(\boldsymbol{D}^T\boldsymbol{D})^{-1/2}.$$

Furthermore, the solution for $\boldsymbol{D}$ is

$$\boldsymbol{D} = \boldsymbol{M}^T\boldsymbol{S}^{-1}$$

(in the notation of Kabsch), so that

$$\boldsymbol{R} = \boldsymbol{M}^T\boldsymbol{S}^{-1}(\boldsymbol{S}^{-1}\boldsymbol{MM}^T\boldsymbol{S}^{-1})^{-1/2}$$

which may be compared with the results of the previous paragraph.

Although this $\boldsymbol{R}$ matrix by itself (*i.e.* applied without $\boldsymbol{T}$) does not produce the best rotational superposition (*i.e.* smallest E), it is the one which *exactly* superposes the only three vectors in $\boldsymbol{x}$ whose mutual dispositions are conserved, on their equivalents in $\boldsymbol{X}$, so that the rotation so found is arguably the best defined one.

Alternatives based on $\boldsymbol{D} = \boldsymbol{TR}$, $\boldsymbol{D}^{-1} = \boldsymbol{RT}$, $\boldsymbol{D}^{-1} = \boldsymbol{TR}$ are all easily developed, and these ideas are developed by Diamond (1976a) to include non-homogeneous strains also.

(v) McLachlan's second method. This method (McLachlan, 1979) is based on the properties of the 6×6 matrix

$$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{M} \\ \boldsymbol{M}^T & \boldsymbol{0} \end{pmatrix}$$

and is immune to singularity of $\boldsymbol{M}$. If $\mathbf{p}$ and $\mathbf{q}$ are three-dimensional vectors such that $(\mathbf{p}^T, \mathbf{q}^T)$ is an eigenvector of this matrix then

$$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{M} \\ \boldsymbol{M}^T & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M}\mathbf{q} \\ \boldsymbol{M}^T\mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{p}\lambda \\ \mathbf{q}\lambda \end{pmatrix}.$$

If $\mathbf{q}$ is negated the second equality is maintained provided λ is also negated. Therefore an orthogonal 6×6 matrix

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{H} \\ \boldsymbol{K} & -\boldsymbol{K} \end{pmatrix}$$

(consisting of 3×3 partitions) exists for which

$$\begin{pmatrix} \boldsymbol{H}^T & \boldsymbol{K}^T \\ \boldsymbol{H}^T & -\boldsymbol{K}^T \end{pmatrix}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{M} \\ \boldsymbol{M}^T & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{H} & \boldsymbol{H} \\ \boldsymbol{K} & -\boldsymbol{K} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Lambda} \end{pmatrix}$$

in which $\boldsymbol{\Lambda}$ is diagonal and contains non-negative eigenvalues. The reverse transformation shows that

$$\boldsymbol{M} = 2\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{K}^T$$

and multiplying the eigenvectors together gives

$$\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{K}^T\boldsymbol{K} = \tfrac{1}{2}\boldsymbol{I} = \boldsymbol{HH}^T = \boldsymbol{KK}^T.$$

Therefore

$$2\boldsymbol{KH}^T\boldsymbol{M} = 4\boldsymbol{KH}^T\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{K}^T = 2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{K}^T,$$

but $2\boldsymbol{KH}^T$ is orthogonal and $2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{K}^T$ is symmetrical, therefore [by paragraphs (i) and (iii) above] $2\boldsymbol{KH}^T$ is the required rotation. Similarly, forming

$$\boldsymbol{M}^T = 2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{H}^T$$
$$2\boldsymbol{M}^T\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T = 4\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T = 2\boldsymbol{KH}^T$$

corresponds to the Kabsch formulation [paragraphs (ii) and (iii)] since $2\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T$ is symmetrical and the same rotation, $2\boldsymbol{KH}^T$, appears.

Note that the determinant of the orthogonal matrix so found is twice the product of the determinants of $\boldsymbol{H}$ and of $\boldsymbol{K}$, and since the positive eigenvalues are collected into $\boldsymbol{\Lambda}$ it follows that the sign of the determinant of $\boldsymbol{M}$ is the same as the sign of the determinant of the resulting orthogonal matrix. If this is negative it means that the best superposition is obtained if one vector set is inverted and that $\boldsymbol{x}$ and $\boldsymbol{X}$ are of opposite chirality.

Expanding the expression for E, the weighted sum of squares of errors, for an orthogonal transformation shows that this is least when the trace of the product $\boldsymbol{RM}$ is greatest. In this treatment

$$\text{tr}(\boldsymbol{RM}) = \text{tr}(2\boldsymbol{KH}^T \cdot 2\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{K}^T) = \text{tr}(2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{K}^T) = \text{tr}(\boldsymbol{\Lambda}).$$

Hence, if the eigenvalues in $\boldsymbol{\Lambda}$ and $-\boldsymbol{\Lambda}$ are arranged in decreasing

order of modulus, and if the determinant of M is negative, then exchanging the third and sixth columns of

$$\begin{pmatrix} H & H \\ K & -K \end{pmatrix}$$

produces a product $2KH^T$ with positive determinant (*i.e.* a proper rotation) at minimum cost in residual. Similarly, if M is singular and one or more eigenvalues in Λ vanishes it is necessary only to complete an orthonormal set of eigenvectors such that the determinants of H and K have the same sign.

(vi) MacKay's method. MacKay (1984) was the first to consider the rotational superposition problem in terms of the vector $\mathbf{r}$ of Section 3.3.1.2.1. Using quaternion algebra he showed that if a vector $\mathbf{x}$ is rotated to $\mathbf{X} = R\mathbf{x}$ then

$$(\mathbf{X} - \mathbf{x}) = \mathbf{r} \times (\mathbf{X} + \mathbf{x}),$$

where $|\mathbf{r}| = \tan(\theta/2)$ and the direction of $\mathbf{r}$ is the axis of rotation, as may also be shown from elementary considerations. MacKay then solves this for the vector $\mathbf{r}$ by least squares given the vector pairs $\mathbf{X}$ and $\mathbf{x}$. The individual errors are

$$e_{IA} = \varepsilon_{Ijk} r_j (X_{kA} + x_{kA}) - (X_{IA} - x_{IA})$$

and

$$E = W_a e_{ia} e_{ia}.$$

Setting $\partial E/\partial r_P = 0_P$ gives

$$W_a \varepsilon_{iPk} \varepsilon_{ilm} r_l (X_{ka} + x_{ka})(X_{ma} + x_{ma})$$
$$= W_a \varepsilon_{iPk}(X_{ka} + x_{ka})(X_{ia} - x_{ia})$$

which reduces to

$$2\mathbf{V} = -(Q + Q_0)\mathbf{r}$$

in which

$$Q = M + M^T - 2I \operatorname{tr} M$$
$$Q_0 = S + S' - I(\operatorname{tr} S + \operatorname{tr} S')$$
$$V_I = \varepsilon_{Ijk} M_{jk}$$
$$S_{IJ} = W_a x_{Ia} x_{Ja}$$
$$S'_{IJ} = W_a X_{Ia} X_{Ja}.$$

Thus a direct solution for $\mathbf{r}$ is obtained,

$$\mathbf{r} = -2(Q_0 + Q)^{-1}\mathbf{V},$$

the elements of which are u, v and w, and may be used to construct the orthogonal matrix as in Section 3.3.1.2.1. $Q + Q_0$ may be obtained directly from $X + x$.

If the requisite rotation is 180°, $(Q_0 + Q)$ is singular and cannot be inverted. In this case any row or column of the adjoint of $(Q_0 + Q)$ is a vector in the direction of the axis. Normalizing this vector to unity, giving l, gives the requisite orthogonal matrix as

$$R = 2ll^T - I.$$

Note that MacKay's residual E is quadratic in $\mathbf{r}$. E therefore has one minimum and no maximum, and the minimum is reached on the first cycle of least squares. By contrast, the objective function E that is minimized in methods (i), (ii), (v) and (vii) has one minimum, one maximum and two saddle points in the space of the vector $\mathbf{r}$, as shown in (vii).

It may be shown (Diamond, 1989) that if MacKay's solution vector $\mathbf{r}$ is denoted by $\mathbf{r}_M$ and that given by the other methods [except (iv)] by $\mathbf{r}_O$ then

$$\mathbf{r}_M = \mathbf{r}_O - A^{-1}B\mathbf{r}_O$$

in which A and B are real symmetric, positive semi-definite. A is positive definite unless all the individual vector sums $(\mathbf{X} + \mathbf{x})$ are parallel, as can happen when the best rotation is 180°. Thus the MacKay method only gives the same result as the other methods if:

(*a*) the initial orientation is optimal, for then $\mathbf{r}_O = \mathbf{0}$, or

(*b*) perfect fitting is possible, for then $B = \mathbf{0}$, or

(*c*) all the residual vectors (after fitting by $\mathbf{r}_O$) are parallel to $\mathbf{r}_O$, for then B is singular such that $B\mathbf{r}_O = \mathbf{0}$. In general, $|\mathbf{r}_M| \le |\mathbf{r}_O|$. $\mathbf{r}_O$ may be found by iterating $\mathbf{r}_M$, but $\mathbf{x}$ must be replaced by $R\mathbf{x}$ on each iteration.

(vii) Diamond's second method. This is closely related to MacKay's method, but uses a four-dimensional vector $\boldsymbol{\rho}$ with components λ, μ, ν and σ in which λ, μ and ν are the direction cosines of the rotation axis multiplied by $\sin(\theta/2)$ and σ is $\cos(\theta/2)$. In terms of such a vector Diamond (1988) showed that

$$E = E_0 - 2\boldsymbol{\rho}^T P \boldsymbol{\rho}$$

in which E is the weighted sum of squares of coordinate differences, as before, E_0 is its value before any rotation is applied and P is the matrix

$$P = \begin{pmatrix} Q & \mathbf{V} \\ \mathbf{V}^T & 0 \end{pmatrix}.$$

The rotation matrix R corresponding to the vector $\boldsymbol{\rho}$ is then the last of the forms for R given in Section 3.3.1.2.1.

The minimum E is therefore E_0 minus twice the largest eigenvalue of P since $\boldsymbol{\rho}^T \boldsymbol{\rho} = 1$, and a stationary value of E occurs when $\boldsymbol{\rho}$ is any of the four eigenvectors of P. E thus has a maximum, a minimum and two saddle points, in general, and its value may be determined before any coordinates are transformed. Diamond also showed that the orientations giving these stationary values are related by the operations of 222 symmetry. Equivalent results have also been obtained by Kearsley (1989).

As an alternative to solving a 4×4 eigenproblem, Diamond also showed that the vector $\mathbf{r}$, as in MacKay's solution, may be obtained by iterating

$$\alpha_0 = E_0/2$$
$$\mathbf{r}_n = (\alpha_n I - Q)^{-1}\mathbf{V}$$
$$\alpha_{n+1} = \frac{\mathbf{V} \cdot \mathbf{r}_n + \alpha_n r_n^2}{1 + r_n^2}$$

which has the property that if X and x are exactly superposable then α_0 is the exact solution and no iteration is necessary. If X and x are similar but not exactly superposable then a small number of iterations may be required to reach a stable $\mathbf{r}$ vector, though the matrix Q_0 is not required. As in MacKay's solution, $(\alpha I - Q)$ is singular at the end of the iteration if the required rotation is 180°, but the MacKay and Diamond methods both have the advantage that improper rotations are never generated by these means, and methods based on P and $\boldsymbol{\rho}$ rather than Q and $\mathbf{r}$ are trouble-free for 180° rotations. The iterative loop in this method does not require $R\mathbf{x}$ to be redetermined on each cycle.

Finally, it may be shown that if p_1, p_2, p_3, p_4 are the eigenvalues of P arranged in descending order and

$$p_1 - p_2 - p_3 + p_4$$

is negative, then a closer superposition may be obtained by reversing the chirality of one of the vector sets, and the R matrix constructed from $\boldsymbol{\rho}_4$ optimally superimposes $R\mathbf{x}$ onto $-X$, the enantiomer of X (Diamond, 1990*b*).

3.3.1.2.3. *Orthogonalization of impure rotations*

There are several ways of deriving a strictly orthogonal matrix from a given approximately orthogonal matrix, among them the following.

(i) The Gram–Schmidt process. This is probably the simplest and the easiest to compute. If the given matrix consists of three column vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ (later referred to as primers) which are to be replaced by three column vectors $\mathbf{u}_1, \mathbf{u}_2$ and $\mathbf{u}_3$ then the process is

$$\mathbf{u}_1 = \mathbf{v}_1/|\mathbf{v}_1|$$
$$\mathbf{u}_2 = \mathbf{v}_2 - (\mathbf{u}_1 \cdot \mathbf{v}_2)\mathbf{u}_1$$
$$\mathbf{u}_2 = \mathbf{u}_2/|\mathbf{u}_2|$$
$$\mathbf{u}_3 = \mathbf{v}_3 - (\mathbf{u}_1 \cdot \mathbf{v}_3)\mathbf{u}_1 - (\mathbf{u}_2 \cdot \mathbf{v}_3)\mathbf{u}_2$$
$$\mathbf{u}_3 = \mathbf{u}_3/|\mathbf{u}_3|.$$

As successive vectors are established, each vector $\mathbf{v}$ has subtracted from it its components in the directions of established vectors, and the remainder is normalized. The method will fail at the normalization step if the vectors $\mathbf{v}$ are not linearly independent. Otherwise, the process may be extended to any number of dimensions.

The weakness of the method is that, though $\mathbf{u}_1$ differs from $\mathbf{v}_1$ only in scale, $\mathbf{u}_N$ may differ grossly from $\mathbf{v}_N$ as the various columns are not treated equivalently.

(ii) A preferable method which treats all vectors equivalently is to iteratively replace the matrix M by $\frac{1}{2}(M + M^{T-1})$.

Defining the residual matrix E as

$$E = MM^T - I,$$

then on each iteration E is replaced by

$$E^2(MM^T)^{-1}/4$$

and convergence necessarily ensues.

(iii) A third method resolves M into its symmetric and antisymmetric parts

$$S = \tfrac{1}{2}(M + M^T), \quad A = \tfrac{1}{2}(M - M^T), \quad M = S + A$$

and constructs an orthogonal matrix for which only S is altered. A determines l, m, n and θ as shown in Section 3.3.1.2.1, and from these a new S may be constructed.

(iv) A fourth method is to treat the general matrix M as a combination of pure strain and pure rotation. Setting

$$M = RT$$

with R orthogonal and T symmetrical gives

$$T = (M^TM)^{1/2}, \quad R = M(M^TM)^{-1/2}.$$

The rotation so found is the one which exactly superposes those three mutually perpendicular directions which remain mutually perpendicular under the transformation M.

$T - I$ is then the strain tensor of an unrotated body.

Writing $M = TR$, $T = (MM^T)^{1/2}$, $R = (MM^T)^{-1/2}M$ may also be useful, in which $T - I$ is the strain tensor of a rotated body. See also Section 3.3.1.2.2 (iv).

3.3.1.2.4. *Eigenvalues and eigenvectors of orthogonal matrices*

If R is the orthogonal matrix given in Section 3.3.1.2.1 in terms of the direction cosines l, m and n of the axis of rotation, then it is clear that (l, m, n) is an eigenvector of R with eigenvalue unity because

$$R\begin{pmatrix} l \\ m \\ n \end{pmatrix} = \begin{pmatrix} l \\ m \\ n \end{pmatrix}.$$

Consideration of the determinant $|R - \lambda I| = 0$ shows that the sum of the three eigenvalues is $1 + 2\cos\theta$ and that their product is unity. Hence the three eigenvalues are 1, $e^{i\theta}$ and $e^{-i\theta}$. Since R is real, its product with any real vector is also real, yet its product with an eigenvector must, in general, be complex. Thus the eigenvectors must themselves be complex.

The remaining two eigenvectors $\mathbf{u}$ may be found using the results of Section 3.3.1.2.1 (*q.v.*) according to

$$R\mathbf{u} = \mathbf{u} + \frac{2}{1 + t^2}\{(\mathbf{r} \times \mathbf{u}) + [\mathbf{r} \times (\mathbf{r} \times \mathbf{u})]\} = \mathbf{u}e^{\pm i\theta} = \mathbf{u}\frac{1 \pm it}{1 \mp it},$$

which is solved by any vector of the form

$$\mathbf{u} = \mathbf{l} \times \mathbf{v} \mp i\mathbf{l} \times (\mathbf{l} \times \mathbf{v})$$

for any real vector $\mathbf{v}$, where $\mathbf{l}$ is the normalized axis vector, $\mathbf{l}t = \mathbf{r}$, $|\mathbf{l}| = 1$, $t = \tan(\theta/2)$. Eigenvectors for the two eigenvalues may have unrelated $\mathbf{v}$ vectors though the sign choices are coupled. If the vector $\mathbf{v}$ is rotated about $\mathbf{l}$ through an angle φ the corresponding vector $\mathbf{u}$ is multiplied by $e^{-i\varphi}$ and remains an eigenvector. Using superscript signs to denote the sign of θ in the eigenvalue with which each vector is associated, the matrix

$$U = (\mathbf{l}, \ \mathbf{u}^+, \ \mathbf{u}^-)$$

has the properties that

$$RU = U\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\theta} & 0 \\ 0 & 0 & e^{-i\theta} \end{pmatrix}$$

and

$$U^{*T}U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2|\mathbf{l} \times \mathbf{v}^+|^2 & 0 \\ 0 & 0 & 2|\mathbf{l} \times \mathbf{v}^-|^2 \end{pmatrix}$$

which places restrictions on $\mathbf{v}$ if this is to be the identity. Note that the 23 element vanishes even in the absence of any relationship between $\mathbf{v}^+$ and $\mathbf{v}^-$.

A convenient form for U, symmetrical in the elements of $\mathbf{l}$, is obtained by setting $\mathbf{v}^+ = \mathbf{v}^- = [111]$ and is

$$U = \begin{pmatrix} l & \{(m-n) - i[l(l+m+n) - 1]\}/d & \{(m-n) + i[l(l+m+n) - 1]\}/d \\ m & \{(n-l) - i[m(l+m+n) - 1]\}/d & \{(n-l) + i[m(l+m+n) - 1]\}/d \\ n & \{(l-m) - i[n(l+m+n) - 1]\}/d & \{(l-m) + i[n(l+m+n) - 1]\}/d \end{pmatrix}$$

in which the normalizing denominator is given by

$$d = 2\sqrt{1 - lm - mn - nl}.$$

3.3.1.3. *Projection transformations and spaces*

In the following section we address the question of the relationship between the coordinates of a molecular model and the corresponding coordinates on the screen of the graphics device. A good introduction to this topic is given by Newman & Sproull (1973), and Foley *et al.* (1990) give a comprehensive account of the field, including recent developments, especially those arising from the development of raster-graphics technologies.

3.3.1.3.1. *Definitions*

Typically, the coordinates, $\mathbf{X}$, of points in an object to be drawn are held in homogeneous Cartesian form as described in Section 3.3.1.1.2. Such coordinates are said to be in *data space* or world

references