short communications

Rapid calculation of RMSDs using a quaternion-based characteristic polynomial

Douglas L. Theobald

Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO 80309-0215, USA.
Correspondence e-mail: theobal@colorado.edu

A common measure of conformational similarity in structural bioinformatics is the minimum root mean square deviation (RMSD) between the coordinates of two macromolecules. In many applications, the rotations relating the structures are not needed. Several common algorithms for calculating RMSDs require the computationally costly procedures of determining either the eigen decomposition or matrix inversion of a 3×3 or 4×4 matrix. Using a quaternion-based method, here a simple algorithm is developed that rapidly and stably determines RMSDs by circumventing the decomposition and inversion problems.

1. Introduction

Orthogonal rotations are commonly used for comparing macromolecular structures, and the root mean square deviation (RMSD) is a natural metric for quantitating the similarity of two optimally rotated structures (Flower, 1999). Several least-squares algorithms have been proposed for finding the rotation that minimizes the squared distances between corresponding atoms in the structures. The most efficient of these methods requires the eigen decomposition (or inversion) of a key matrix constructed from the sums and squares of the atomic coordinates in the structures being compared. The popular method of Kabsch (1978) involves a 3×3 key matrix, whereas the methods of Diamond (1988), Horn (1987) and Kearsley (1989) use quaternion representations of the rotations with 4×4 key matrices. In each of these methods, the minimizing rotation is formed from the eigenvectors of the key matrices and the minimizing RMSD is found from the eigenvalues of the key matrices.

In Kabsch's method, improper rotation matrices (rotoinversions) may arise and one must verify that the determinant of the rotation matrix is positive (Kabsch, 1978). Quaternion-based methods lack this complication since quaternion rotations are always proper. Thus, with quaternion-based methods the minimum RMSD can be calculated using the eigenvalues of the key matrix alone, without knowing the eigenvectors. If only RMSDs are desired, the quaternion-based methods can save a considerable amount of computation by avoiding a complete eigen decomposition of the key matrix.

Diamond (1988) proposed the fastest method known for determining minimum RMSDs using an iterative algorithm. Diamond's method requires one 3×3 matrix inversion per cycle in order to find the largest eigenvalue of a quaternion-based 4×4 matrix. While matrix inversions are generally expensive, the inversion of a 3×3 matrix can be calculated analytically and relatively quickly. Diamond's method, however, is unstable when the minimizing rotation is near $180°$. This problem is especially grave when the coordinate differences are of magnitude similar to the coordinates themselves, such as is common when superimposing relatively small fragments of proteins or nucleic acids.

Based on Horn's quaternion superposition method (Horn, 1987), which is relatively unknown in structural biology, an extremely fast, numerically stable and accurate algorithm is presented here for

determining minimum RMSDs. This algorithm does not require separate consideration of special problematic cases and it bypasses the computationally costly diagonalization and inversions commonly used to find the needed eigenvalues. Rather, a quick and simple Newton–Raphson root-finding method is used to determine the appropriate eigenvalue from the characteristic polynomial of the key matrix. This quaternion-based characteristic polynomial (QCP) algorithm requires significantly less computation than the alternate methods, including Diamond's.

2. The superposition problem

A molecular structure consisting of N atoms can be represented as an $N \times 3$ matrix, where the three columns correspond to the x, y and z coordinates. The superposition objective is to find the orthogonal rotation and translation that minimizes the squared Euclidean distance between the rows of two matrices corresponding to the two structures being compared. The translational component of the problem can be removed from the outset by translating each structure so that its respective centroid is at the origin. When superimposing structure **B** onto structure **A**, in matrix notation the problem is to minimize the sum of squared errors E with respect to the orthogonal rotation **R**:

$$E = \|\mathbf{BR} - \mathbf{A}\|_F^2, \qquad (1)$$

where $\|\mathbf{X}\|_F$ denotes the Frobenius (or Euclidean) norm of the matrix **X**,

$$\|\mathbf{X}\|_F = [\text{tr}(\mathbf{X}'\mathbf{X})]^{1/2}, \qquad (2)$$

and $\mathbf{X}'$ denotes the transpose of **X** and $\text{tr}\,\mathbf{X}$ denotes the matrix trace of **X**. Upon expansion of equation (1), it can be shown that

$$E = G_A + G_B - 2\text{tr}(\mathbf{MR}) = (N)\text{RMSD}^2, \qquad (3)$$

where G_A is the inner product of structure **A**,

$$G_A = \text{tr}(\mathbf{A}'\mathbf{A}) = \sum_i^N (x_{A,i}^2 + y_{A,i}^2 + z_{A,i}^2),$$

and **M** is the inner product of the matrices **A** and **B**. The inner product matrix **M** is given by

$$\mathbf{M} = \mathbf{B}'\mathbf{A} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix}, \qquad (4)$$

where

$$S_{xy} = \sum_i^N x_{B,i} y_{A,i}. \qquad (5)$$

3. A quaternion-based characteristic quartic polynomial for calculating RMSDs

Horn (1987) has shown that the sum of squared errors E can be minimized by solving for the largest positive eigenvalue λ_{max} of a 4×4 symmetric key matrix $\mathbf{K}$:

$$\begin{bmatrix} (S_{xx} + S_{yy} + S_{zz}) & S_{yz} - S_{zy} & S_{zx} - S_{xz} & S_{xy} - S_{yx} \\ S_{yz} - S_{zy} & (S_{xx} - S_{yy} - S_{zz}) & S_{xy} + S_{yx} & S_{zx} + S_{xz} \\ S_{zx} - S_{xz} & S_{xy} + S_{yx} & (-S_{xx} + S_{yy} - S_{zz}) & S_{yz} + S_{zy} \\ S_{xy} - S_{yx} & S_{zx} + S_{xz} & S_{yz} + S_{zy} & (-S_{xx} - S_{yy} + S_{zz}) \end{bmatrix}.$$

The eigenvector corresponding to the largest eigenvalue of $\mathbf{K}$ is a quaternion equivalent to the optimal rotation, where

$$\mathrm{RMSD} = [(G_A + G_B - 2\lambda_{\mathrm{max}})/N]^{1/2}. \qquad (6)$$

Eigenvalues are usually determined by diagonalization of the matrix. Alternatively, one can find the eigenvalues of a symmetric matrix by locating the roots of the characteristic polynomial of the matrix. With $|\mathbf{X}|$ denoting the determinant of matrix $\mathbf{X}$, the characteristic equation for the key matrix $\mathbf{K}$ is a quartic given by

$$C_4\lambda^4 + C_3\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0 = 0,$$

where

$$C_4 = 1$$
$$C_3 = -\mathrm{tr}\,\mathbf{K} = 0$$
$$C_2 = -2\,\mathrm{tr}\,\mathbf{M}'\mathbf{M}$$
$$\quad = -2(S_{xx}^2 + S_{xy}^2 + S_{xz}^2 + S_{yx}^2 + S_{yy}^2 + S_{yz}^2 + S_{zx}^2 + S_{zy}^2 + S_{zz}^2)$$
$$C_1 = -8|\mathbf{M}| = 8(S_{xx}S_{yz}S_{zy} + S_{yy}S_{zx}S_{xz} + S_{zz}S_{xy}S_{yx})$$
$$\quad - 8(S_{xx}S_{yy}S_{zz} + S_{yz}S_{zx}S_{xy} + S_{zy}S_{yx}S_{xz})$$
$$C_0 = |\mathbf{K}| = D + E + F + G + H + I,$$

where

$$D = (S_{xy}^2 + S_{xz}^2 - S_{yx}^2 - S_{zx}^2)^2$$
$$E = [-S_{xx}^2 + S_{yy}^2 + S_{zz}^2 + S_{yz}^2 + S_{zy}^2 - 2(S_{yy}S_{zz} - S_{yz}S_{zy})]$$
$$\quad \times [-S_{xx}^2 + S_{yy}^2 + S_{zz}^2 + S_{yz}^2 + S_{zy}^2 + 2(S_{yy}S_{zz} - S_{yz}S_{zy})]$$
$$F = [-(S_{xz} + S_{zx})(S_{yz} - S_{zy}) + (S_{xy} - S_{yx})(S_{xx} - S_{yy} - S_{zz})]$$
$$\quad \times [-(S_{xz} - S_{zx})(S_{yz} + S_{zy}) + (S_{xy} - S_{yx})(S_{xx} - S_{yy} + S_{zz})]$$
$$G = [-(S_{xz} + S_{zx})(S_{yz} + S_{zy}) - (S_{xy} + S_{yx})(S_{xx} + S_{yy} - S_{zz})]$$
$$\quad \times [-(S_{xz} - S_{zx})(S_{yz} - S_{zy}) - (S_{xy} + S_{yx})(S_{xx} + S_{yy} + S_{zz})]$$
$$H = [(S_{xy} + S_{yx})(S_{yz} + S_{zy}) + (S_{xz} + S_{zx})(S_{xx} - S_{yy} + S_{zz})]$$
$$\quad \times [-(S_{xy} - S_{yx})(S_{yz} - S_{zy}) + (S_{xz} + S_{zx})(S_{xx} + S_{yy} + S_{zz})]$$
$$I = [(S_{xy} + S_{yx})(S_{yz} - S_{zy}) + (S_{xz} - S_{zx})(S_{xx} - S_{yy} - S_{zz})]$$
$$\quad \times [-(S_{xy} - S_{yx})(S_{yz} + S_{zy}) + (S_{xz} - S_{zx})(S_{xx} + S_{yy} - S_{zz})].$$

Using these coefficients, the characteristic equation can be reduced to the following simple form:

$$P(\lambda) = \lambda^4 + C_2\lambda^2 + C_1\lambda + C_0 = 0. \qquad (8)$$

Although the coefficient equations may appear somewhat horrendous, coding them is straightforward. With careful programming, by avoiding duplicate evaluations of several terms (such as $S_{yz} - S_{zy}$ which appears four times), calculating the coefficients from the sums found in equation (4) involves at most only 66 floating point operations (FLOPs).

Horn (1987) proposed solving this characteristic equation using Ferrari's method (Abramowitz & Stegun, 1972) for quartics. While analytically exact, Ferrari's method has seen little use as it is computationally slow and notoriously numerically unstable. Additionally, Horn (1987) omits expressions for C_0 and an equation given for C_2 contains an important typographical error.

Fortunately in this case, the largest positive root of the characteristic polynomial given in equation (8) can be found very quickly and with excellent numerical stability using the Newton–Raphson algorithm. Successful use of the Newton–Raphson method requires a good initial guess for the desired root. From equation (3) and noting that the largest positive eigenvalue of the key matrix $\mathbf{K}$ is equal to $\mathrm{tr}(\mathbf{MR})$ with the optimal rotation, the largest positive root of the characteristic polynomial attains its maximum possible value when the two structures are in identical conformations (i.e. when RMSD = 0). Thus, the maximum possible value of the largest positive root is constrained to be less than or equal to the average of the inner products of the the two structures, $(G_A + G_B)/2$. Using this value as an initial guess in the Newton–Raphson algorithm is essentially fool-proof, since the algorithm will converge first on the largest root of the polynomial. The Newton–Raphson QCP algorithm can be described by the following simple pseudocode:

while($|\lambda - \lambda_{\mathrm{old}}| < \mathrm{precision}$)
  $\lambda_{\mathrm{old}} = \lambda$;
  $\lambda = \lambda - P(\lambda)/[\mathrm{d}P(\lambda)/\mathrm{d}\lambda]$;

where the first derivative of the polynomial is given by $\mathrm{d}P(\lambda)/\mathrm{d}\lambda = 4\lambda^3 + 2C_2\lambda + C_1$.

The Newton–Raphson QCP algorithm given above is much faster than other common methods that require determination of eigenvectors, and it is also faster than methods using eigen decomposition for only eigenvalues. The efficiency and speed of the QCP Newton–Raphson algorithm were compared to the popular method using Householder reduction to tridiagonal form followed by QL decomposition with implicit shifts (H–QL method; Golub & Van Loan, 1996). Each iteration of the QCP algorithm requires merely 11 FLOPS, of which none are square roots. In contrast, the H–QL algorithm requires a comparable number of total FLOPs but over two dozen costly square roots. Diamond's (1988) method is likewise much more efficient than the eigen decomposition methods and also requires no square roots. Each cycle of the Diamond iteration uses ≈ 70 FLOPS (accounting for symmetry considerations), plus 15 FLOPS for set-up. Since both the QCP and Diamond's methods take an average of about five iterations to converge to a relative precision of 10^{-6}, the QCP method uses only about one third of the operations of Diamond's method for eigenvalue determination. Thus, essentially all of the computational effort in the QCP method is expended in calculating the summations of the key matrix $\mathbf{M}$ given in equation (4), a step that is common to all methods and dependent on the number of atoms compared.

The QCP algorithm has been benchmarked on several computing platforms (OSX, Linux and Irix) for $>10^8$ independent superposition calculations with protein fragments ranging from 5 to 500 residues. In all cases, the same RMSD solution was found, within error, as the Kabsch and Kearsley methods. In contrast, especially with smaller

short communications

fragments, Diamond's method converged on an incorrect solution (corresponding to the second largest eigenvalue) in a significant fraction of cases. In the author's experience, if the construction of the key matrices is ignored, the QCP algorithm is over four times as fast as Diamond's method and 30 to 70 times faster than the eigen decomposition methods. An ANSI C library implementing the QCP method is available at http://monkshood.colorado.edu/QCP/ or by request from the author.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed., pp. 17–18. New York: Dover.

Diamond, R. (1988). *Acta Cryst.*, A**44**, 211–216.

Flower, D. R. (1999). *J. Mol. Graph. Model.* **17**, 238–244.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press.

Horn, B. K. P. (1987). *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **4**, 629–642.

Kabsch, W. (1978). *Acta Cryst.* A**34**, 827–828.

Kearsley, S. K. (1989). *Acta Cryst.* A**45**, 208–210.