GUIDE TO COMPUTING RMSD OF 2 VECTOR SETS

— ARUN KONAGURTHU

The goal is to compute RMSD (root-mean-square-deviation) of 2 vector sets:

$$U = \{\vec{u}_1, \vec{u}_2, ---, \vec{u}_n\}$$

and $$V = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$$

Each of the vectors is in 3D and is of the form (x, y, z).
It is also assumed that any $\vec{u}_i$ corresponds to any $\vec{v}_i$ (for all $i \in \{1, \ldots, n\}$).

RMSD of these two vector sets (i.e, U and V} can be computed by following these steps:

1. TRANSLATE the sets U and V, (separately) such that the geometric centers of vectors in U and vectors in V are moved to origin. [NOTE: The geometric center of a set of vectors is simply its centroid. That is, if $\{\vec{s}_1, \vec{s}_2, \ldots \vec{s}_n\}$ are vectors in a set, centroid is given by $\frac{\sum_{i=1}^{n} \vec{s}_i}{n}$.]

This gives a new translated set of vectors U' and V'.
Let the vectors in $U' = \{\vec{u}_1', \vec{u}_2', \ldots \vec{u}_n'\}$ be of the form $\vec{u}_i' \equiv (x_i, y_i, z_i)$.
Similarly, let the vectors in $V' = \{\vec{v}_1', \vec{v}_2', \ldots \vec{v}_n'\}$ be of the form $\vec{v}_i' = (x_i', y_i', z_i')$.

(2) Given the vectors in sets U' and V', RMSD Computati
involves finding eigen values to the following
4x4 square symmetric matrix:

$$
M = \begin{bmatrix}
\sum_{i=1}^{n}(x_m^2+y_m^2+z_m^2) & \sum_{i=1}^{n}(y_p z_m - y_m z_p) & \sum_{i=1}^{n}(x_m z_p - x_p z_m) & \sum_{i=1}^{n}(x_p y_m - x_m y_p) \\
\sum_{i=1}^{n}(y_p z_m - y_m z_p) & \sum_{i=1}^{n}(y_p^2+z_p^2+x_m^2) & \sum_{i=1}^{n}(x_m y_m - x_p y_p) & \sum_{i=1}^{n}(x_m z_m - x_p z_p) \\
\sum_{i=1}^{n}(x_m z_p - x_p z_m) & \sum_{i=1}^{n}(x_m y_m - x_p y_p) & \sum_{i=1}^{n}(x_p^2+z_p^2+y_m^2) & \sum_{i=1}^{n}(y_m z_m - y_p z_p) \\
\sum_{i=1}^{n}(x_p y_m - x_m y_p) & \sum_{i=1}^{n}(x_m z_m - x_p z_p) & \sum_{i=1}^{n}(y_m z_m - y_p z_p) & \sum_{i=1}^{m}(x_p^2+y_p^2+z_m^2)
\end{bmatrix}
$$

(REFER THIS MATRIX IN THE PAPER BY KEARSLEY THAT IS INCLUD
IN THE SUPPORTING MATERIAL FOLDER (rmsdpaper.pdf))

In the above matrix 'x_m' refers to $(x_i' - x_i)$,
'y_m' refers to $(y_i' - y_i)$
and 'z_m' refers to $(z_i' - z_i)$

Similarly 'x_p' refers to $(x_i' + x_i)$
'y_p' refers to $(y_i' + y_i)$
and 'z_p' refers to $(z_i' + z_i)$

Finding eigenvalues of M (use MATLAB routine eig())
gives 4 values all of which are real and ≥ 0
(This is ~~your~~ a sanity check for your script). Call them
$\lambda_1, \ldots, \lambda_4$.

RMSD of superposition of U and V is given by
$\sqrt{\frac{\lambda}{n}}$ where λ is $\min\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$.

END