

SPHERES METHOD INTERSECTION USED TO CALCULATE PROTEIN STRUCTURES

Inajara da Silva Freitas

Prof. Dr. Luiz Carlos Matioli

Universidade Federal do Paraná – UFPR

inajarafreitas@gmail.com

June de 2013



Protein Data Bank (PDB)

Protein Structure

Health and Diseases

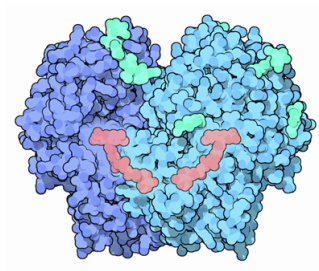
Biotechnology and Nanotechnology



Glucose Oxidase

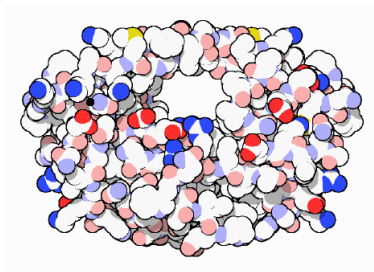
Diabetes - detect glucose levels

Preserver - honey natural preserver



HIV-1 Protease

Enzyme used in the treatment of AIDS



How to find the shape of these structures?

X-ray Crystallography

Nuclear Magnetic Resonance (NMR Spectroscopy)

Electron Microscopy



Cases

- ▶ Sparse Matrix
- ▶ Full Matrix



Given a set of C atom pairs (i, j) of a set of n atoms with the distances d_{ij} defined over C , find the positions $x_1, \dots, x_n \in \mathbb{R}^3$ atoms in the molecule structure so that

$$\|x_i - x_j\| = d_{ij} \quad \forall (i, j) \in C$$



Exact Distance

Given 4 points in \mathbb{R}^3

$$x_1 = (u_1, v_1, w_1)^T$$

$$x_2 = (u_2, v_2, w_2)^T$$

$$x_3 = (u_3, v_3, w_3)^T$$

$$x_4 = (u_4, v_4, w_4)^T$$

we can build a system to find the remaining points.



Exact Distance

We denote the distance between atom i and j by $d_{i,j}$ where $j = 1, \dots, 4$. We set the system according to the problem that we are looking for.

$$\|x_i - x_1\| = d_{i,1}$$

$$\|x_i - x_2\| = d_{i,2}$$

$$\|x_i - x_3\| = d_{i,3}$$

$$\|x_i - x_4\| = d_{i,4}$$



Exact Distance

Raise squared on both sides of each equality

$$\|x_i - x_1\|^2 = \|x_i\|^2 - 2x_i^T x_1 + \|x_1\|^2 = d_{i,1}^2$$

$$\|x_i - x_2\|^2 = \|x_i\|^2 - 2x_i^T x_2 + \|x_2\|^2 = d_{i,2}^2$$

$$\|x_i - x_3\|^2 = \|x_i\|^2 - 2x_i^T x_3 + \|x_3\|^2 = d_{i,3}^2$$

$$\|x_i - x_4\|^2 = \|x_i\|^2 - 2x_i^T x_4 + \|x_4\|^2 = d_{i,4}^2$$



Exact Distance

Replacing $x_j^T x_j$, where $j = 1, \dots, 4$, by $x_j = (u_j, v_j, w_j)^T$

$$\|x_i - x_1\|^2 = \|x_i\|^2 - 2u_i u_1 - 2v_i v_1 - 2w_i w_1 + \|x_1\|^2 = d_{i,1}^2$$

$$\|x_i - x_2\|^2 = \|x_i\|^2 - 2u_i u_2 - 2v_i v_2 - 2w_i w_2 + \|x_2\|^2 = d_{i,2}^2$$

$$\|x_i - x_3\|^2 = \|x_i\|^2 - 2u_i u_3 - 2v_i v_3 - 2w_i w_3 + \|x_3\|^2 = d_{i,3}^2$$

$$\|x_i - x_4\|^2 = \|x_i\|^2 - 2u_i u_4 - 2v_i v_4 - 2w_i w_4 + \|x_4\|^2 = d_{i,4}^2$$



Exact Distance

Subtracting the other equations from the first, we get the following system

$$A = 2 \begin{pmatrix} u_1 - u_2 & v_1 - v_2 & w_1 - w_2 \\ u_1 - u_3 & v_1 - v_3 & w_1 - w_3 \\ u_1 - u_4 & v_1 - v_4 & w_1 - w_4 \end{pmatrix}$$

$$b_i = \begin{pmatrix} (\|x_1\|^2 - \|x_2\|^2) - (d_{i,1}^2 - d_{i,2}^2) \\ (\|x_1\|^2 - \|x_3\|^2) - (d_{i,1}^2 - d_{i,3}^2) \\ (\|x_1\|^2 - \|x_4\|^2) - (d_{i,1}^2 - d_{i,4}^2) \end{pmatrix}$$

In matricial form: $Ax_i = b_i$



Computational Results

We have $x_i = (u_i, v_i, w_i)^T$.

We will set

- ▶ $u_1 = 0, v_1 = 0$ e $w_1 = 0,$
- ▶ $u_2 = d_{2,1}, v_2 = 0$ e $w_2 = 0,$
- ▶ $w_3 = 0.$

With this we have

- ▶ $x_1 = (0, 0, 0)^T$
- ▶ $x_2 = (u_2, 0, 0)^T$
- ▶ $x_3 = (u_3, v_3, 0)^T$
- ▶ $x_4 = (u_4, v_4, w_4)^T$



Computational Results

To find the values of x_3 we solve the following system

$$d_{3,1}^2 = \|x_3 - x_1\|^2$$

$$d_{3,2}^2 = \|x_3 - x_2\|^2$$

we find

$$v_3 = \pm (d_{3,1}^2 - u_3^2)^{1/2}$$

$$u_3 = (d_{3,1}^2 - d_{3,2}^2)/(2u_2) + u_2/2$$



Computational Results

To find the values of x_4 we solve the following system

$$d_{4,1}^2 = \|x_4 - x_1\|^2$$

$$d_{4,2}^2 = \|x_4 - x_2\|^2$$

$$d_{4,3}^2 = \|x_4 - x_3\|^2$$

we find

$$u_4 = (d_{4,1}^2 - d_{4,2}^2)/(2u_2) + u_2/2$$

$$v_4 = (d_{4,2}^2 - d_{4,3}^2 - (u_4 - u_2)^2 + (u_4 - u_3)^2)/2v_3 + v_3/2$$

$$w_4 = \pm (d_{4,1}^2 - u_4^2 - v_4^2)^{1/2}$$



Computational Results

With the coordinates of the first four atoms, we can determine the remaining coordinates solving the system as $Ax_i = b_i$

Protein	Time/second	N° of atoms
1HAA	0.018870	1310
1AJV	0.010216	1516
1PTQ	0.007361	402
1HOE	0.007767	558
1HIV	0.010275	1502

Tabela: Test Results - Linear Method



Intersection between spheres

The problems of intersection between spheres, is to find $x \in \mathbb{R}^n$ such

$$\|x - x_i\|^2 = d_i^2 \quad i = 1, \dots, n.$$

We will set d_i the distance between points x and x_i , ie the radius of the sphere. Solve the problem using Gaussian Elimination.



Gaussian Elimination

We have

$$\|x - x_i\|^2 = d_i^2 \quad i = 1, \dots, n$$

which is equivalent to

$$x^T x - 2x^T x_i + x_i^T x_i = d_i^2$$

isolating $x^T x_i$ we have

$$x^T x_i = \frac{x^T x + x_i^T x_i - d_i^2}{2}$$

for $r = x^T x$ and $b_i = x_i^T x_i - d_i^2$ equation will be of the form

$$x^T x_i = \frac{r + b_i}{2} \quad i = 1, \dots, n$$



Gaussian Elimination

Represented in matrix form

$$A^T x = \frac{re + b}{2}$$

where $e = [1, 1, 1, \dots, 1, 1]^T \in \mathbb{R}^n$. We defined A invertible, there is A^{-1} such that

$$A^{-T} A^T x = A^{-T} \frac{(re + b)}{2}$$

$$\text{ie } x = \frac{A^{-T} re + A^{-T} b}{2}.$$



Gaussian Elimination

For $u = A^{-T} e$ e $v = A^{-T} b$

$$x = \frac{ru + v}{2}$$

We have

$$r = x^T x = \frac{(ru + v)^T}{2} \frac{(ru + v)}{2}$$

or else

$$4r = r^2 u^T u + ru^T v + rv^T u + v^T v$$



Gaussian Elimination

We can see that $u^T v, v^T u \in \mathbb{R}$ e $u^T v = v^T u$. Then the equation can be written as:

$$u^T u r^2 + (2u^T v - 4)r + v^T v = 0$$

which leads us to seek a solution to r , determining the roots of the last equation, ie:

$$r = \frac{2 - u^T v \pm \sqrt{(u^T v - 2)^2 - (u^T u)(v^T v)}}{u^T u}$$



Algorithm

Data Input: $A = [x_1, x_2, \dots, x_n]$ array of the centers of spheres and $d = (d_1, d_2, \dots, d_n)$ radii of spheres.

operations:

For $i = 1: n$

$$b_i = x_i^T x_i - d_i^2$$

end

$$A^{-1} = \text{gaus_jordan}(A)$$

$$u = A^{-T} e$$

$$v = A^{-T} b$$

$$\text{root} = (u^T v - 2)^2 - (u^T u)(v^T v)$$



Algorithm

- If $root = 0$:

$$r = \frac{2 - (u'v)}{u'u};$$

$$x = \frac{(ru) + v}{2};$$

- If $root > 0$:

$$r_1 = \frac{2 - u'v + \sqrt{root}}{u'u};$$

$$r_2 = \frac{2 - u'v - \sqrt{root}}{u'u};$$

$$x = \left[\frac{r_1 u + v}{2}, \frac{r_2 u + v}{2} \right];$$

- If $root < 0$:

There is no solution;

end



Aplication

We will use the points x_2 , x_3 , x_4 as the centers of spheres and maintain x_1 as the starting point so the structure will start at the origin.

The matrix A is given by:

$$A = [x_2, x_3, x_4] = \begin{pmatrix} u_2 & u_3 & u_4 \\ 0 & v_3 & v_4 \\ 0 & 0 & w_4 \end{pmatrix}$$

The distances d_{ij} for $j = 2, 3, 4$ are the radii of the spheres center x_j .



Aplication

By using the method of spheres intersection we obtain:

- ▶ 1 point x_i - in this case x_i ; this is the point we are looking for.
- ▶ 2 points x_{i1} and x_{i2} - we can find out the correct point checking which one is closest to the distance d_{i1} .
- ▶ 0 points - this case shold not occur since all distances exist.

To ensure this result just put a $\varepsilon > 0$ and check

$(u^T v - 2)^2 - (u^T u)(v^T v) \leq \varepsilon$ instead of the condition

$(u^T v - 2)^2 - (u^T u)(v^T v) = 0$.



New Algorithm

Data Input: $A = [x_1, x_2, \dots, x_n]$ array of sphere centers and $d = (d_1, d_2, \dots, d_n)$ radii of spheres.

Operations:

For $i = 1:n$

$$b_i = x_i^T x_i - d_i^2$$

end

$$A^{-1} = \text{gaus_jordan}(A)$$

$$u = A^{-T} e$$

$$v = A^{-T} b$$

$$\text{root} = (u^T v - 2)^2 - (u^T u)(v^T v)$$



New Algorithm

- If $root = 0$ or ($root < \varepsilon$ and $root > -\varepsilon$):

$$r = \frac{2 - (u'v)}{u'u};$$

$$x = \frac{(ru) + v}{2};$$

- If $root > 0$:

$$r_1 = \frac{2 - u'v + \sqrt{root}}{u'u};$$

$$r_2 = \frac{2 - u'v - \sqrt{root}}{u'u};$$

$$x_1 = \frac{r_1 u + v}{2};$$

$$x_2 = \frac{r_2 u + v}{2};$$

If ($\|x_1\| - \|x_2\| \leq \|x_1\| - \|x_2\|$)

$$x = x_1;$$

Else

$$x = x_2;$$

end



Computational Results

Table with the tests results:

Protein	Time/seconds	N ^o of atoms
1HAA	0.146791	1310
1AJV	0.168065	1516
1PTQ	0.065965	402
1HOE	0.078101	558
1HIV	0.165347	1502

Tabela: Test Results - Quadratic Method



Geometric Center

To put the structures in its geometric center do

$$x_c(j) = \frac{1}{n} \sum_{i=1}^n X(i,j)$$

$$y_c(j) = \frac{1}{n} \sum_{i=1}^n Y(i,j)$$

for $j = 1, 2, 3$.



To translate the structures

$$X_1(:, 1) = X(:, 1) - xc(1)$$

$$Y_1(:, 1) = Y(:, 1) - yc(1)$$

$$X_1(:, 2) = X(:, 2) - xc(2)$$

$$Y_1(:, 2) = Y(:, 2) - yc(2)$$

$$X_1(:, 3) = X(:, 3) - xc(3)$$

$$Y_1(:, 3) = Y(:, 3) - yc(3)$$



Rotation

Root-Mean-Square Deviation (RMSD) is used to measure the degree of similarity of two structures X_1 and Y_1 in a way that they have the same geometric center, we define::

$$RMSD(X_1, Y_1) = \frac{\min_Q \|X_1 - Y_1 Q\|_F}{\sqrt{n}}$$

Where $Q_{3 \times 3}$ is the rotation matrix, and $Q Q^T = I$.



Algorithm

Given X Y

Calculate, x_c , y_c , X_1 , Y_1

$$C = Y_1^T X_1$$

$$C = U \Sigma V^T$$

$$Q = UV^T$$

$$RMSD(X_1, Y_1) = \frac{\min_Q \|X_1 - Y_1 Q\|_F}{\sqrt{n}}$$

end



Results

Protein	N° of atoms	Linear	Quadratic
1HAA	1310	4.97177553078e-04	4.97177553078e-04
1AJV	1516	4.99166258766e-04	4.99166258766e-04
1PTQ	402	5.02486001361e-04	5.02486001361e-04
1HOE	558	4.89971795038e-04	4.89971795038e-04
1HIV	1502	5.03796641867e-04	5.03796641867e-04

Tabela: Test Results - RMSD



Conclusions

Does it work for real problems?

PDB access via MATLAB



Next steps

Analise data in a Sparse Matrix






Can i fix this Matrix?

Convergence of these Methods

Comparison of methods



Bibliography

-  DAVIS, R. T.; ERNST, C.; WU, D. - *Protein structure determination via an efficient geometric build-up algorithm*. BMC Structural Biology, 2010.
-  Dong e Wu - *A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances*. Journal of Global Optimization, pg 365 - 375, 2002.
-  I.D. Coope - *Reliable computation of the points of intersection of n spheres in \mathbb{R}^n* . Journal of Global Optimization, pg 365 - 375, 2002.
-  RCSB PDB - *RCSB Protein Data Bank*: www.rcsb.org/.
-  Michael F. Souza - *SUAVIZAÇÃO HIPERBÓLICA APLICADA À OTIMIZAÇÃO DE GEOMETRIA MOLECULAR*. Tese de Doutorado, UFRJ, 2010.

