Topology of Cyclo-Octane Energy Landscape

S. Martin\textsuperscript{1}, W. M. Brown\textsuperscript{2}, S. Pollock\textsuperscript{3}, A. P. Thompson\textsuperscript{1}, E. A. Coutsias\textsuperscript{3}, and J.-P. Watson\textsuperscript{1}

4/5/2011

\textsuperscript{1}\textsuperscript{Sandia National Laboratories, \textsuperscript{2}Oak Ridge National Laboratories, \textsuperscript{3}University of New Mexico
Overview of Work

• Cyclo-Octane has been studied as a model problem in computational chemistry for over 40 years.
  – Multiple conformations of similar energy.
  – Complex energy landscape.

• We used/developed new methods from dimension reduction, computational algebraic geometry, and computational topology to better understand cyclo-octane.

• As a result, we have completely characterized the conformation space of cyclo-octane:
  – The energy landscape, together with the topology of the conformation space, explains experimental observations that boat-chair is the dominant conformation of cyclo-octane.
Cyclo-Octane

- Eight membered ring $C_8H_{16}$ studied as a model problem for over 40 years in computational chemistry.

- “Cyclo-octane is unquestionably the conformationally most complex cycloalkane owing to the existence of so many forms of comparable energy.” (Hendrickson, 1967).

- Three stable conformations: boat, boat-chair, and crown.

- Conformation space thought to be 2D due to ring closure constraint.
Enumerating Cyclo-Octane Conformations

- Cyclo-octane conformation can be described analytically using kinematic loop closure (Coutsias et al., 2005) or distance geometry (Portas et al. 2007).
  - Bond lengths and angles are fixed, while torsions are varied.
  - Algebraic equations (degree 16 for cyclo-octane) are solved to enumerate solutions.
  - At least 6 torsions are required, otherwise we have $s - 6$ degrees of freedom ($s = 8$ for cyclo-octane).
Enumerating Cyclo-Octane Conformations
(Example using Distance Constraints)

Define:

\[ D(1, 2, 3, \ldots, k) = \begin{bmatrix}
0 & r_{1,2} & r_{1,3} & \cdots & r_{1,k} & 1 \\
r_{2,1} & 0 & r_{2,3} & \cdots & r_{2,k} & 1 \\
r_{3,1} & r_{3,2} & 0 & \cdots & r_{3,k} & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & 1 \\
r_{k,1} & r_{k,2} & r_{k,3} & \cdots & 0 & 1 \\
1 & 1 & 1 & \cdots & 1 & 0
\end{bmatrix} \]

where \( r_{i,j} = \| \mathbf{p}_i - \mathbf{p}_j \|^2 \). The constraints are given by

\[
D(1, 2) > 0 \quad D(\mathbf{R}, i) = 0 \\
D(1, 2, 3) < 0 \quad D(\mathbf{R}, j) = 0 \\
D(1, 2, 3, 4) > 0 \quad D(\mathbf{R}, i, j) = 0
\]

where \( \mathbf{R} = 1, \ldots, 4 \).
Dimension Reduction of Cyclo-Octane*

We applied nonlinear dimension reduction methods to the space of cyclo-octane conformations.

Dimension Reduction of Cyclo-Octane

- In (Brown et al., 2008), we applied Principal Component Analysis (PCA), IsoMap, Locally Linear Embedding, and an Autoencoder (neural network).
  - Best results were obtained using Isomap (Tenenbaum et al., 2000).
  - Embedding dimension of conformation space was estimated to be 5.
  - Intrinsic dimension was estimated to be 2.
Questions Raised by Dimension Reduction of Cyclo-Octane

- This is a 3D visualization of a 5D object – what is in the other 2 dimensions?
- Are apparent intersections actually intersections (or just “singularities of projection”)?
- What is the topology of this object?
- What does this mean (in terms of molecular conformation)?
Triangulation of Cyclo-Octane

• Need triangulation to compute topological invariants such as homology (more later).

• Problems:
  – Existing surface reconstruction methods are limited to 3D, except incremental projection algorithm (Freedman, 2007).
  – (Freedman, 2007) assumes manifold surface.
  – We have non-manifold surface in 24 (ring atoms) or 72 (all atoms) dimensions.
Triangulation of Cyclo-Octane*

- Solution:
  - Model non-manifold neighborhoods as two intersecting planes in 3D.
  - Decompose non-manifold neighborhood into an intersecting line and two isolated planes
  - Triangulate intersections then use (Freedman, 2007) to triangulate surface.

Fitting Two Intersecting Planes

• Overview:
  – We use PCA to project a non-manifold neighborhood into xyz coordinates.
  – We fit a quadratic polynomial to the data using a least squares fit subject to the constraint that it must factor.
  – We factor the polynomial, find the intersection of the two planes, and split the neighborhood accordingly.

• Details:
  – Denote a polynomial by
    \[ f(x, y, z) = a_{11}x^2 + 2a_{12}xy + 2a_{13}xz + 2a_{14}x + a_{22}y^2 + 2a_{23}yz + 2a_{24}y + a_{33}z^2 + 2a_{34}z + a_{44} = 0. \]
  – Write \( A = (a_{ij}) \), denote by \( A_3 \) the 3x3 upper left submatrix of \( A \), and let
    \[ T_2 = (a_{11}a_{22} - a_{12}^2) + (a_{11}a_{33} - a_{13}^2) + (a_{22}a_{33} - a_{23}^2). \]
  – Let \( m_j = [x_j^2 \ 2x_jy_j \ 2x_jz_j \ 2x_j \ y_j^2 \ 2y_jz_j \ 2y_j \ z_j^2 \ 2z_j \ 1] \) and \( M = [m_j] \).
  – We must solve
    \[
    \begin{array}{l}
    \min_a \quad a^T M^T M a \\
    \text{s.t.} \\
    \quad \text{rank}(A) = \text{rank}(A_3) = 2, \\
    \quad T_2 \leq 0, \quad \|A_3\|_F = 1.
    \end{array}
    \]
Fitting Two Intersecting Planes
(More Details)

We re-write the minimization

$$\min_{a} \ a^T M^T M a$$

s.t. \quad \text{rank}(A) = \text{rank}(A_3) = 2,$$

$$T_2 \leq 0, \quad \|A_3\|_F = 1.$$

By observing that

$$\begin{bmatrix}
\text{rank}(A) = 2,
\text{rank}(A_3) = 2,
\|A_3\|_F = 1
\end{bmatrix} = \begin{bmatrix}
\lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T
\end{bmatrix}$$

so that we can minimize the following equivalent problem

$$\min_{q_1, q_2, \lambda_1, \lambda_2} \sum_j \left( \lambda_1 q_1^T X_j q_1 + \lambda_2 q_2^T X_j q_2 \right)^2$$

s.t. \quad \|q_1\| = \|q_2\| = 1, \quad q_1^T q_2 = 0,$$

$$\lambda_1, \lambda_2 \neq 0, \quad \lambda_1^2 + \lambda_2^2 = 1,$$

$$T_2 = \sum_{j<k\leq3} \det([\lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T]_{jk}) \leq 0,$$

where $X_j$ is a matrix of quadratic monomials in $xyz$ coordinates for data point $j$. 
Fitting Two Intersecting Planes (Still More Details)

- We can obtain a very good initial estimate for the minimization using the following procedure:
  - Get unconstrained solution by setting $a^*$ to be the right singular vector of $M$.
  - Form matrix $A^* = (a_{ij}^*)$ and normalize such that $\|A_3^*\|_F = 1$.
  - Decompose $A_3^* = \lambda_1 \tilde{q}_1 \tilde{q}_1^T + \lambda_2 \tilde{q}_2 \tilde{q}_2^T$
  - Solve for $q_1 = [\tilde{q}_1 \ r]^T$, $q_2 = [\tilde{q}_2 \ s]^T$

\[
\begin{bmatrix}
\lambda_1 \tilde{q}_1 & \lambda_2 \tilde{q}_2
\end{bmatrix}
\begin{bmatrix}
r \\
s
\end{bmatrix}
= 
\begin{bmatrix}
a_{14}^* \\
a_{24}^* \\
a_{34}^*
\end{bmatrix}
\]

- Now re-set $a^* = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T$

- This estimate works without further optimization for cyclo-octane!

- Not shown: how to factor $a^*$.
Triangulation of Cyclo-Octane

- Performed surface reconstruction 5 times using randomly selected subsamples (at least $\varepsilon$ distance apart) of 24 dimensional (ring atom only) cyclo-octane data.
  - 6,040 samples ($\varepsilon = .12$)
  - 7,114 samples ($\varepsilon = .11$)
  - 8,577 samples ($\varepsilon = .10$)
  - 10,503 samples ($\varepsilon = .9$)
  - 13,144 samples ($\varepsilon = .8$)

- Verified local topology in each case as homotopic to a point.
Recent development in computational algebraic topology allow us to investigate the topology of triangulated/point set data.

- Available tools such as Plex (comptop.stanford.edu) and Chomp (chomp.rutgers.edu) can compute topological invariants known as Betti numbers.
- Betti numbers count number of connected components ($\beta_0$), number of loops ($\beta_1$), number of voids ($\beta_2$), etc. using algebraic homology.

Torus Betti Numbers: 1,2,1
Sphere Betti Numbers: 1,0,1
Topology of Cyclo-Octane*

• We used computational topology tools to compute Betti numbers of cyclo-octane conformations.
  – Plex (comptop.stanford.edu) to compute boundary maps and Laplacian operators.
  – Linbox (www.linalg.org) to compute ranks of Laplacians.
  – Afra Zomordian’s persistence codes (not publicly available).

• We computed the Betti numbers for each of the 5 triangulations.
  – $\beta_0 = 1, \beta_1 = 1, \beta_2 = 2$.

Decomposition of Cyclo-Octane

- Betti numbers (1,1,2) are uninformative, due to the non-manifold nature of the conformation space.
- However, we can use our triangulation to decompose the space into two components via the self-intersections.
Cyclo-Octane Canonical Basis

• To resolve the identity of the hourglass, we derived a canonical basis from the distance constraints which gives an analytical version of the Isomap coordinates.

– The canonical basis in torsion coordinates is given by

\[
\begin{align*}
\mathbf{c}_r &= (u, -u, u, -u, u, -u, u, -u)^T \\
\mathbf{b}_1 &= (0, v, 0, -v, 0, v, 0, -v)^T \\
\mathbf{b}_2 &= (v, 0, -v, 0, v, 0, -v, 0)^T \\
\mathbf{c}_1 &= (v, 0, -v, w, -v, 0, v, -w)^T \\
\mathbf{c}_2 &= (v, -w, v, 0, -v, w, -v, 0)^T,
\end{align*}
\]

where

\[
\begin{align*}
\cos u &= \frac{(1 - \sqrt{2} - \cos \theta_b)}{(1 + \cos \theta_b)} \\
\cos v &= \frac{\cos^2 \theta_b}{\sin^2 \theta_b} \\
\cos w &= \frac{(3 \cos^2 \theta_b - 1)}{\sin^2 \theta_b}.
\end{align*}
\]
Analytic Cyclo-Octane Decomposition

- Triangulation allowed decomposition of conformation space into a sphere and a Klein bottle intersecting in two rings.
- Klein bottle is a unique discovery in the field of molecular conformation.
Understanding Cyclo-Octane Conformations

• Structure of conformation space can be explained by the geometry of the cyclooctane molecule.
• There are ten “canonical conformations” which can be grouped into three families: Crown (Cr), Boat (B), and Boat-Chair (BC).
• There are twice as many BC conformations as there are Cr conformations, which causes a twist in the conformation space.
• The twist forms a Mobius strip.
• The Klein bottle forms because there are two Mobius strips (due to symmetry by reflection) joined at their edges.
• The energy landscape, together with the topology of the conformation space, explains experimental observations that BC is the dominant conformation of cyclooctane.
Conclusions

• Our work on cyclo-octane has revealed novel discoveries in field of molecular motion.
  – We have used new tools from computational algebraic geometry/topology to complete characterize a 40 year old model problem.
  – Algebraic singularities are evidence of previously unsuspected mathematical complexity.
  – Klein bottle evidence of previously unassumed non-orientable structure.

• In addition, the cyclo-octane data has motivated the need for better data analysis algorithms.
  – Previous algorithms assume a manifold structure. For certain datasets this assumption is inadequate, and such algorithms will fail.
  – We need a new class of algorithms that can handle algebraic structure, including singularities.
  – The algorithms that we developed to analyze cyclo-octane are just one small step towards this new class of algorithms.
  – Future challenges include data set size, dimension of structure, different types of singularities, etc.