BioGeometry NEWS
February 2003
biogeometry.cs.duke.edu

Research

The weighted volume derivative of a space filling diagram

This work is a combined effort of Herbert Edelsbrunner (Computer Science, Duke University) and Patrice Koehl (Structural Biology, Stanford University).

Motivation. It is widely believed that characterizing the geometry of a protein molecule is essential for understanding its folding process as well as its interactions with other biomolecules and small ligands. Among all geometric measures, volume is probably the most fundamental property to study. Atomic volumes have been used for quantifying packing interactions in proteins, as well as in energy functions for implicit solvent models. In the latter, the effects of water on the protein of interest are accounted for using continuum models for electrostatics, and a volume or surface term for hydrophobic effects. Inclusion of such potentials in a biomolecular simulation require fast algorithms for computing the volume occupied by the protein, and its derivatives with respect to the atomic positions.

Computing molecular volume using Alpha Shapes. Computational methods that evaluate the solvent-excluded volume of a molecule can be divided into approximate and exact methods. Most of the approximate methods rely on numerical integrations. The first analytical method to compute the volume of a molecule was based on the inclusion-exclusion principle, with the approximation that only intersections of up to three balls need to be considered. Including all possible orders of intersection remains computational difficult and expensive. The Alpha Shape theory solves this problem exactly by using Delaunay triangulations and their filtrations. An overview of this approach is given in figure 1.

Volume derivatives. The distinction between approximate and exact computation also applies to existing methods for computing the derivatives of the solvent-excluded volume with respect to atomic coordinates. All these methods have to take a large number of singularities into account, where approximations are usually required. The Alpha Shape theory proposes a robust solution to this problem, by implementing arbitrary precision arithmetic to avoid numerical problems and systematically resolving all singularities without explicitly perturbing the positions of the ball centers. The latter method is referred to as Simulation of Simplicity. We have derived a theorem that provides the Weighted Volume Derivative of a space filling diagram, and implemented this theorem within the Alpha Shapes theory, to provide an efficient, robust, exact computation of the derivatives of volumes. Figure 2 illustrates this theorem for three balls.

There is an inherent difficulty in using a potential based on volume for energy minimization or molecular dynamics. Although the volume is continuous in the position of the atoms, its derivatives may not be. We examine this issue within the framework of the Alpha Shape method and relate discontinuities to combinatorial changes in the subcomplex of the Delaunay triangulation that is dual to the space-filling diagram of the molecule.

Implementation and software availability. We have written a new version of the Alpha Shape software, AlphaVol, specific to molecular simulation applications. AlphaVol includes implementations of the volume and weighted volume derivative formulas. Computation of the volume and its derivatives of a 250-residue proteins using AlphaVol requires about 0.37 seconds on a 1000 Mhz Pentium processor. The 0.37 seconds roughly breaks down to 0.11 seconds for computing the regular triangulation, 0.03 seconds for generating the dual complex, 0.17 seconds to compute the weighted volume, and 0.06 seconds to compute the weighted volume derivatives. AlphaVol is available online at http://biogeometry.duke.edu/software/proshape.

A paper based on this work will soon appear in the Proceedings of the National Academy of Sciences (USA).

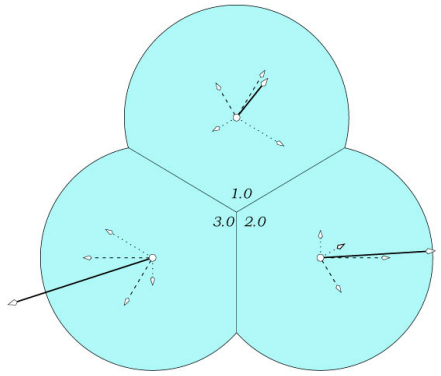


Figure 2: Three equal-sized balls with different weights as indicated. For each pair we get components of the gradient normal to (dashed) and parallel to (dotted) the separating facet. The solid vectors are the sums of the components as well as the portions of the gradient that correspond to the different balls

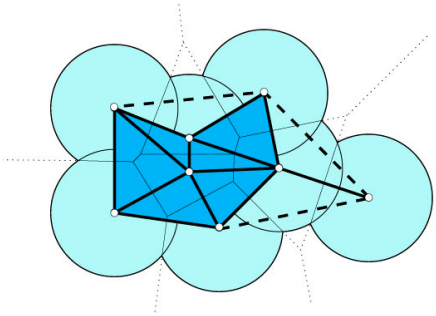


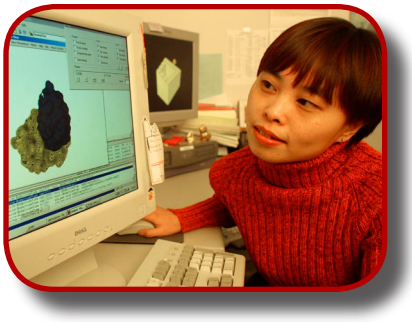
Figure 1: Computing geometric properties of a molecule. As common in biology, atoms are treated as intersecting balls, whose union forms the space-filling diagram. Richards and others proposed to study this union using the Voronoi diagram (or more exactly the power diagram since the balls have different radii). That diagram divides the space into convex cells, one per atom. In the two-dimensional example shown here, the edges that separate these cells are shown as solid inside and dotted outside the union of disks. Note that the power cells of some surface atoms extend to infinity. The superimposed Delaunay triangulation (thick solid and dashed lines) is the dual of the power diagram obtained by drawing a line segment between two ball centers if their convex cells share a common edge. Despite their different appearance, the Delaunay triangulation and the power diagram contain exactly the same information. The key to connect the power diagram to the molecule is to consider the intersection of the two: this defines a convex decomposition of the space-filling diagram (i.e the light shaded area inside the disks, divided into convex regions by the solid part of the Voronoi edges). The dual of this decomposition is the dual complex (thick solid lines and shaded triangles). The dual complex is a subset of the Delaunay triangulation, and contains all simplices (tetrahedra, triangles and edges) that correspond to overlapping atoms. The dual complex contains all information about a molecule required to compute its surface and volume. In this paper, we show that the same dual complex can be used to compute the derivative of the volume.

Student Profile: Yusu Wang

Yusu Wang, a fifth-year PhD student in the Department of Computer Science at Duke University, is working with PIs Agarwal and Edelsbrunner on shape representation and analysis. After completing her BS from Tsinghua University, she joined Duke as a graduate student in 1998. Initially, she worked in the area of computer graphics, but later her research interests shifted to developing algorithms for representing and analyzing 3-dimensional shapes, especially the shapes of proteins. Over the last three years she has developed algorithms for computing the writhing number of a polygonal curve in 3-space, simplifying a curve in two or higher dimensions, computing a small, robust approximation of a point set that preserves the underlying shape, and measuring similarity between two shapes.

The writhing number is used to capture the physical phenomenon that a chord tends to coil when it is twisted. The writhing number and its variants play an important role in understanding various geometric conformations of circular DNAs in solutions and in classifying folding families of proteins. Yusu proved a relationship between the writhing number of a polygonal curve and the average winding number of its Gauss map [1]. Using this relationship she developed the first subquadratic algorithm for computing the writhing number of a polygonal curve. She also proposed a simpler algorithm that runs in linear time on inputs that are typical in practice, including protein backbones.

While working on the writhing-number project, Yusu got interested in simplifying 3-dimensional curves because she needed to remove local noise from a protein backbone before computing its writhing number. Curve simplification is also used to save memory space needed to store a curve and to expedite the processing of a curve. She developed a simple, greedy approximation algorithm for curve simplification under the so-called Frèchet error measure [2]. In many situations this error-measure is better than the more widely used Hausdorff error measure if one wants



to preserve global features of a curve. She tested her algorithm on a variety of curves. Her experimental results showed that her algorithm is faster than the commonly used algorithms, such as the Douglas-Peucker heuristic, and produces small-size approximations of the input curves.

Recently, Yusu developed an algorithm for approximating a set of points by a small subset that preserves certain geometric properties of the input points. For example, if a simple shape (such as a line or circle) fits well through the subset, then it also fits well through the original point set. Although an algorithm was known for this problem, it was incapable of handling outliers. Her algorithm is robust and can handle outliers. Her work suggests promising ideas for a high-dimensional clustering algorithm that can handle outliers.

Currently, Yusu participates in a project at Duke focused on investigating geometric approaches of docking two proteins. She investigates methods from Morse theory to find promising initial relative placements of the proteins, which will then be refined using a local-improvement heuristic. She represents the surface of a molecule as a 2-manifold, defines a Morse function on this surface, and computes the critical points of this Morse function. The goal is to find appropriate functions so that its critical points capture geometric information, such as pockets and protrusions, as well as biochemical information, such as the electrostatic potential. She is developing a matching algorithm to output promising initial placements of proteins. Her matching algorithm takes advantage of both geometric information of a given shape and topological information based on the Morse function. Experimental results in two dimensions suggest that the new method outputs reasonably good matchings while inspecting many fewer conformations than the exhaustive search or traditional geometric-hashing approaches.

In parallel to the above approach to docking, she investigates the computational complexity of minimizing the Hausdorff distance between two sets of spheres under a group of motions. She developed an efficient exact algorithm for the two-dimensional case allowing translations only and an approximation algorithm for the 3-dimensional case. She also studied the problem of minimizing the Hausdorff distance under the constraint of having no collisions, a problem that has received little attention in the past, while it is essential in docking proteins [3].

Yusu is also very active in the department. She has served on a number of committees and is currently a student representative in the department. She and a few other students started a weekly student research seminar.

Publications

[1] P. K. Agarwal, H. Edelsbrunner, Y. Wang. Computing the writhing number of a polygonal knot. *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*. 791-799, 2002.

[2] P. K. Agarwal, S. Har-Peled, N. H. Mustafa, Y. Wang. Near-Linear Time Approximation Algorithms for Curve Simplification. Proceedings of the *10th European Symposium on Algorithms*. 29-41, 2002. (Invited to the special issue of *Algorithmica*.)

[3] P. K. Agarwal, S. Har-Peled, M. Sharir, Y. Wang. Hausdorff distance between points, disks, and balls. Proceedings of the *19th Annual Symposium on Computational Geometry*. To appear. 2003.

[4] S. Har-Peled, Y. Wang. Shape fitting with outliers. Proceedings of the *19th Annual Symposium on Computational Geometry*. To appear. 2003.

- Profile by Pankaj Agarwal

The BioGeometry project is funded by the National Science Foundation under grant CCR-00-86013. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.