

3차원에서 단백질의 보로노이 다이어그램 계산을 위한 모서리추적 알고리즘 및 응용

Edge Tracing Algorithm for Computation of Euclidean Voronoi Diagram for Protein and its Applications

조영송¹, 김동욱¹, 조철형², 김덕수³

¹ 한양대학교, Voronoi diagram 연구단, {ycho, donguk}@voronoi.hanyang.ac.kr

² 한양대학교, 산업공학과, murick@voronoi.hanyang.ac.kr

³ 한양대학교, 산업공학과, dskim@hanyang.ac.kr

Abstract

Although there are many applications of Euclidean Voronoi diagram for spheres in a 3D space in various disciplines from sciences and engineering, it has not been studied as much as it deserves. In this paper, we present an edge-tracing algorithm to compute the Euclidean Voronoi diagram of 3-dimensional spheres in $O(mn)$ in the worst-case, where m is the number of edges of Voronoi diagram and n is the number of spheres. After building blocks for the algorithm, we show an example of Voronoi diagram for atoms using actual protein data and discuss its applications for protein analysis.

Key word: Euclidean Voronoi diagram, edge-tracing, rational quadratic Bézier curve.

1. 서론

보로노이 다이어그램은 계산 기하학 분야에서 가장 중요한 연구 분야 중 하나이며, 다양한 과학 및 공학 분야에서 응용 가능한 방법으로 알려져 있다. 2 차원 및 고차원 점집합에 대한 보로노이 다이어그램은 이미 많은 연구가 수행되어 그 특성들이 잘 알려져 있다. 그러나, 2 차원에서의

원집합 및 3 차원에서의 구집합 등에 대한 유클리디안 보로노이 다이어그램은 응용분야에서의 중요성에도 불구하고, 많은 연구가 이루어지지 않은 것이 사실이다.

1967 년 Bernal 과 Finney [5]가 처음 소개한 이후로, 단백질 구조를 위한 보로노이 다이어그램(Voronoi diagram)의 연구는 원자들의 중심점 집합 P 에 대한 보로노이 다이어그램 $VD(P)$ 와 그것의 dual 구조인 Delaunay 삼각망이 주로 사용되었다. Baumgart [4]가 winged-edge 구조를 발표한 시기가 1974 년이기 때문에, 이때는 계산기하 분야에서조차도 2 차원 $VD(P)$ 의 효율적인 계산이 쉽지 않았을 때였다.

3 차원 $VD(P)$ 는 주어진 점 집합에 대하여 공간을 각 점에 가까운 영역으로 분할 하는 공간 tessellation 이다. 이런 $VD(P)$ 는 수학적으로 잘 정의되어 있으며, 효과적이고 안정적인 코드들이 나와있기 때문에, 생물학 분야의 많은 연구들이 $VD(P)$ 를 이용하여 왔다 [1,9,16,19,20].

$VD(P)$ 는 원자 종류에 따른 반지름의 차이를 고려할 수 없기 때문에, Richards 는 1974 년 두 개의 이웃한 원자들의 반지름 비율에 따라 bisector 를 반지름이 작은 원자 쪽으로 이동시키는 스키마(scheme)을 제시하였다 [19]. 이렇게 되면 $VD(P)$ 가 타당한 tessellation 을 한다고 볼 수

없으며, 보로노이 꼭지점은 수학적으로 정확하게 정의될 수 없다. Richards 는 이것을 “vertex error”로 이름 지었다.

이런 vertex error 를 인식하여, Gellatly 와 Finney 는 1982 년 보로노이 모서리를 이동하는 대신에 radical plane 을 사용하는 방법을 소개하였다 [8]. 이런 radical plane 접근법은 1987 년 Aurenhammer [2]에 의해서 명명된 원집합 B 의 파워 다이어그램(Power diagram) PD(B)와 사실 똑같다고 볼 수 있으며, 계산기하 분야에서는 그 당시 알려지지 않은 방법이었다. 그 후로 PD(B)는 어느 수준까지는 원자들의 반지름을 고려하기 때문에 생물학적 문제들을 해결하는데 종종 사용되어왔다 [3]. PD(B)의 이론도 역시 수학적으로 잘 정의되어있으며, 빠르고 안정적인 코드들이 공개되어 있다 [21]. 그러나 PD(B)는 두 개의 원자 사이의 거리를 접평면을 이용하여 계산하기 때문에, 유클리디안 공간을 충분히 반영했다고 보기 어렵다.

유클리디안 거리(Euclidean distance)에 대한 원집합의 보로노이 다이어그램의 경우, 최근에 간단하지만 매우 빠르고 강건한 알고리즘이 개발되었다 [11,12]. 3 차원 구집합에 대한 유클리디안 보로노이 다이어그램의 경우, 관련된 연구가 진행되고 있으나[13,14], 아직 구체적인 알고리즘이나 구현결과는 보고된 바 없다. 본 논문에서는 3 차원 구집합의 유클리디안 보로노이 다이어그램에 대한 실질적인 알고리즘을 제안한다. 제안하는 알고리즘의 Luchinikov [15]등에 의해 제시된 방법에 기반하고 있다.

2. 용어 및 정의

$B = \{b_1, b_2, \dots, b_n\}$ 을 보로노이 다이어그램의 제너레이터라 하자. 이때, n 은 3 차원 구로서, 중심좌표와 반경은 각각 $c_i = (x_i, y_i, z_i)$ 와 r_i 로 정의한다. B 와 관련하여, 각 b_i 에 대해 보로노이 영역(Voronoi region) VR_i 이 존재하며, 이는 $VR_i = \{p \mid \text{dist}(p, b_i) < \text{dist}(p, b_j), \quad i \neq j\}$ 로

정의된다. 이때 $\text{dist}(p, b_i)$ 는 한 점에서 구까지의 유클리디안 거리이다.

각 보로노이 영역은 두 이웃한 구에 의해 정의된 다수의 보로노이 면(Voronoi face)에 의해 한정되며, 이 면은 항상 쌍곡면의 일부이다. 하나의 보로노이 면이 다른 면과 만나 보로노이 모서리(Voronoi edge)를 형성하고, 보로노이 모서리들은 보로노이 꼭지점(Voronoi Vertex)에서 교차한다.

본 논문에서 대상으로 하는 구들은 하나의 구가 다른 한 구를 완벽히 포함하는 경우는 없다고 가정한다. 그리고 보로노이 모서리는 항상 3 개의 구에 의해 정의되고, 꼭지점은 항상 4 개의 구에 의해 정의된다고 가정한다.

3. 보로노이 다이어그램의 기하요소 계산

보로노이 다이어그램의 정의로부터, 보로노이 꼭지점은 4 개의 인접한 구에 외접하는 구의 중심이다. 이때 외접하는 구는 주어진 어떤 구들과도 교차하지 않는다. 4 개의 구가 주어졌을 때, 외접하는 구의 계산은 주어진 네 구에 대해 등거리에 있는 점들은 표현하는 2 차 방정식을 계산함으로써 얻을 수 있다 [7].

보로노이 모서리는 항상 인접한 세 구에 대해 등거리에 놓인 점들의 궤적으로 정의된다. 이때, 가장 작은 반경을 갖는 구를 원점에 위치시키고, 모든 세 구의 반경을 가장 작은 반경만큼 축소하였을 때, 보로노이 모서리는 3 개의 방정식의 해이며, 이는 평면 곡선으로서 코닉 섹션(conic section)이다. 보로노이 모서리가 코닉 섹션이므로, 이는 2 차 유리 베이지어 곡선(rational quadratic Bézier curve)의 형태로 정확하게 표현될 수 있다 [6].

위상적으로 인접한 두 개의 구에 의해 정의된 보로노이 면은 두 구에 대해 등거리에 있는 점들의 집합이다. 그러므로 보로노이 면은 항상 쌍곡면의 일부이며, 이에 대한 음함수식은 쉽게 얻을 수 있다.

Figure 1. 은 3 차원 보로노이 다이어그램의 예로서, (a)에 있는 3 종류의 반지름을 갖는 15 개의 구들 중, 중앙에 위치한 구의 보로노이 영역을 (b)에서 보여준다. 이 예제는 Luchinikov 등의 논문에서 제시된 모델과 유사하다 [15].

4. 모서리 추적 알고리즘에 의한 보로노이 다이어그램의 위상요소 구축

보로노이 다이어그램의 위상요소는 모서리 추적 알고리즘에 의해 얻어진다. 모서리 추적 알고리즘의 기본적인 개념은 다음과 같이 매우 간단하다.

먼저 알고리즘은 이웃한 네 개의 적합한 구로부터 정의된 외접하는 구를 계산함으로써, 초기 보로노이 꼭지점, v_0 을 찾는다. 이는 네

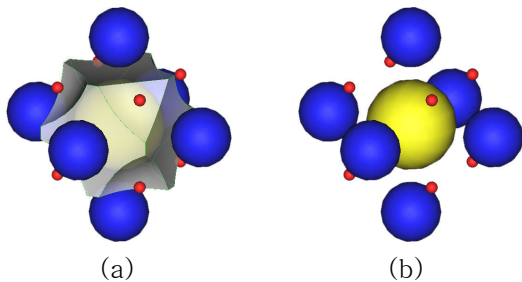


Figure 1. Voronoi region, (a) 15 spheres with 3 types of radius, (b) Voronoi region of the centered sphere.

개의 구의 조합에 대해 계산한 외접구가 주어진 어떠한 구와도 교차하지 않는지 시간이 필요하나, 전 처리과정에서 적합한 버킷을 이용하여 분류하며, 평균 $O(1)$ 시간으로 v_0 를 계산할 수 있다.

일단 v_0 가 구해졌다면, v_0 로부터 나가는 네 개의 보로노이 모서리 e_0, e_1, e_2, e_3 , 이 쉽게 얻어지며, 이를 모서리 스택(edge-stack)에 넣는다. 이때 각 모서리는 그들의 시작점을 v_0 로 정의한다. 이 스택으로부터 하나의 모서리를 꺼낸 후, 꺼낸 모서리를 정의하는 세 구와 나머지 $n-3$ 개의 구들로부터 외접하는 구를 계산한다. 이때 각 외접하는 구는 주어진 어떠한 구와도 교차하지 않아야 하므로 이를 위한 검사가 필요하다. 이

방법은 실질적으로 최악의 경우 $O(n^2)$ 수행시간이 요구된다. 이는 알고리즘의 설계를 좀더 주의 깊게 하여 다음과 같이 향상시킬 수 있다.

스택에서 꺼낸 모서리를 정의하는 세 개의 구가 주어졌을 때, 먼저 후보 집합에 있는 $n-3$ 개의 구들 중 하나의 구 b_i 로부터 외접구 T_i 를 계산한다. T_i 는 주어진 모서리의 끝점을 정의하는 외접구의 후보가 된다. 이후 후보 집합의 다른 하나의 구 b_j 를 이용하여 외접구 T_j 를 계산한다. 이때, 만일 b_j 가 T_i 와 교차한다면, 주어진 모서리의 끝점을 정의하는 외접구의 후보는 T_j 가 되고, 교차하지 않는다면, T_i 와 T_j 중 현재 모서리의 시작점과 모서리를 정의하는 구들 중 하나의 중심과 T_i 와 T_j 각각의 중심이 이루는 각이 더 작은 외접구를 후보로 선택한다. 이 방법은 최악의 경우와 평균 모두 $O(n)$ 의 수행시간을 요구한다. 다음은 알고리즘을 요약한 내용이다.

Algorithm 1. Find End-Vertex

Step 1. Compute a tangent sphere S_1 from 3 defining balls and b_1 . $S \leftarrow S_1$

Step 2. For $k = 2$ to $n - 3$,

Step 2.1. Compute S_k from the 3 balls and b_k .

Step 2.2 If S_k has the smaller angle distance, then $S \leftarrow S_k$.

End-for

Step 3. Output end vertex as the center of S .

만약 주어진 어떠한 구와도 교차하지 않는 외접하는 구를 찾았다면, 이 외접하는 구의 중심이 스택으로부터 꺼낸 현재 모서리의 끝점이 된다. 이 점 역시 보로노이 꼭지점이므로, 이 꼭지점으로부터 나가는 3 개의 보로노이 모서리가 정의되어 다시 스택에 저장되고, 이들 세 모서리의 시작점은 현재 찾은 외접구의 중심이 된다. 이 과정을 통해 스택에 있는 하나의 모서리에 대한 계산이 완료되고,

이러한 과정은 스택이 완전히 비워질 때까지 수행된다.

이 방법은 보로노이 모서리의 수(m)만큼 수행되어야 한다. 모서리의 수는 최악의 경우 $O(n^2)$ 이지만 평균 $O(n)$ 이다. 각 모서리에 대해 $O(n)$ 의 수행시간을 요구하므로, 전체 알고리즘이 갖는 최악의 경우 수행시간은 $O(mn)$ 이다. 그러나 Halperin 은 단백질 안에 있는 원자들을 대상으로 할 경우 각 보로노이 꼭지점과 모서리, 면의 개수가 최악의 경우에도 $O(n)$ 을 따른다는 것을 발견하였다 [10]. 이런 중요한 성질은 단백질의 원자들의 분포가 가지는 두 가지 성질에 기인한다. Pauli 의 exclusion principle 에 따르면, 하나의 원자는 또 하나의 다른 원자를 완전히 포함하지 않는다 [17]. 여기에 원자들의 반지름 차이는 항상 일정 범위 안에 항상 놓이게 된다는 성질이 추가가 된다. 이것은 앞에서도 언급한 단백질을 구성하는 원자가 항상 6 개이기 때문에 가능하다. 결국 단백질 원자들의 보로노이 다이어그램에서 모든 꼭지점, 모서리, 면의 개수들은 최악의 경우에도 선형적이라고 말할 수 있다.

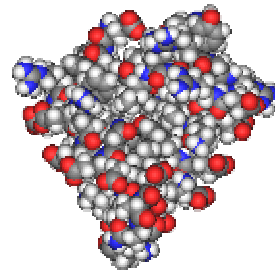
5. 구현 결과 및 응용

Figure 2. (a)는 1,016 개의 원자들로 구성된 1a1u 이라는 코드네임을 가진 단백질 이다 [22]. 일단 보로노이 다이어그램이 계산되면, (b)와 (c)와 같은 원자들의 컨벡스 헐(convex hull)이나, 임의의 육면체 내에 존재하는 구형의 빈 공간을 찾는 문제들은 쉽게 계산할 수 있다.

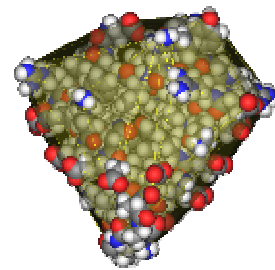
사실 단백질 1a1u 는 두 개의 작은 단백질이 결합된 것으로 이들 둘 사이의 경계면들은 Figure 2. (b)에서처럼, 보로노이 면의 일부이다. 이는 생물학적으로 중요한 의미를 갖는 것으로서 역시 보로노이 다이어그램을 이용하여 쉽게 얻어낼 수 있다.

6. 결론

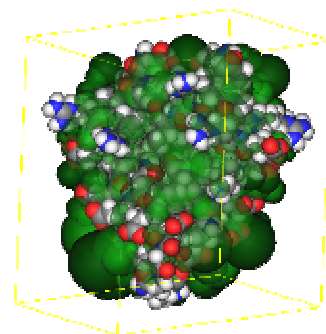
3 차원 공간상에서 구들에 대한 유클리디안 보로노이 다이어그램은 다양한 과학 및 공학분야에서 응용될 수 있는 중요한



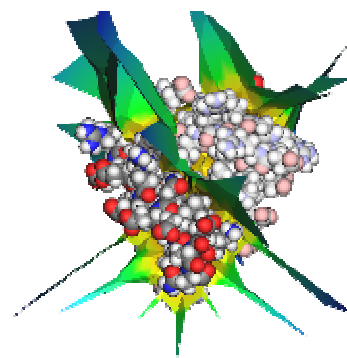
(a)



(b)



(c)



(d)

Figure 2. PDB code 1a1u consist of 1016 atoms; (a) atomic structure, (b) convex hull, (c) internal

voids, and (d) interaction faces

생물학적으로 매우 중요한 의미를 갖는 것으로서 역시 보로노이 다이어그램을 이용하여 쉽게 얻어낼 수 있다. 방법임에도 불구하고, 많은 연구가 이루어지지 않은 것이 사실이다.

본 논문에서는 3 차원 구에 대한 유클리디안 보로노이 다이어그램을 계산하는 모서리 추적 알고리즘을 제시한다. 이 알고리즘은 최악의 경우 수행시간은 $O(mn)$ 이다. m 과 n 은 각각 보로노이 다이어그램의 모서리의 수와 주어진 구의 수이다. 그리고 본 논문에서는 보로노이 모서리가 2 차 유클리드 베이지어 곡선으로 표현될 수 있는 코닉 섹션이며, 보로노이 면의 적합한 표현법을 제시하였다.

그리고 현재 모서리 추적 알고리즘이 정확한 계산(exact computation) 및 구현 속도 측면에서 좀더 향상될 수 있으며, 이를 추후 연구과제로 고려하고 있다.

후 기

본 연구는 과학기술부 창의적연구진흥사업의 지원으로 수행되었습니다.

참고문헌

[1] Angelov, B., Sadoc, J.-F., Jullien, R., Soyer, A., Mornon, J.-P., and Chomilier, J. "Nonatomic solvent-driven Voronoi tessellation of proteins: an open tool to analyze protein folds", *Proteins: Structure, Function, and Genetics*, Vol.49(2002), pp.446-456.

[2] Aurenhammer, F. Power diagrams: properties, algorithms and applications. *SIAM Journal of Computing*, Vol.16(1987), pp.78-96.

[3] Bajaj, C.L., Pascucci, V., Shamir, A., Holt, R.J., and Netravali, A.N. "Dynamic maintenance and visualization of molecular surfaces". *Discrete Applied Mathematics*, Vol.127(2003), pp.23-51.

[4] Baumgart B. "Geometric Modelling for Computer Vision". Ph.D. Thesis, Stanford, CA, USA, 1974.

[5] Bernal, J.D. and Finney, J.L. "Random close-packed hard-sphere model II. Geometry of random packing of hard spheres". *Discussions of the Faraday Society*, Vol.43(1967), pp.62-69.

[6] Farin, G. *Curves and Surfaces for Computer-Aided Geometric Design: A Practical Guide*, fourth ed. Academic Press, 1996.

[7] Gavrilova, M. and Rokne, J. "Updating the topology of the dynamic Voronoi diagram for spheres in Euclidean d-dimensional space", *Computer Aided Geometric Design*, Vol.20, No.4(2003), pp.231-242.

[8] Gellatly, B.J. and Finney, J.L. "Calculation of protein volumes: an alternative to the Voronoi procedure". *Journal of Molecular Biology*, Vol. 161, No.2(1982), pp. 305-322.

[9] Goede, A., Preissner, R., and Frömmel, C. "Voronoi Cell: New Method for Allocation of Space among Atoms: Elimination of Avoidable Errors in Calculation of Atomic Volume and Density", *Journal of Computational Chemistry*, Vol.18, No.9(1997), pp.1113-1123.

[10] Halperin, D., Overmars, M.H. "Spheres, Molecules, and Hidden Surface Removal". *Proc. 10th ACM Symposium on Computational Geometry*, pp. 113-122, 1994.

[11] Kim, D.-S., Kim, D., and Sugihara, K. "Voronoi diagram of a circle set from Voronoi diagram of a point set: I. Topology. *Computer Aided Geometric Design*, Vol.18, No.6(2001), pp.541-562.

[12] Kim, D.-S., Kim, D., and Sugihara, K. "Voronoi diagram of a circle set from Voronoi diagram of a point set: II. Geometry. *Computer Aided Geometric Design*, Vol.18, No.6(2001), pp.563-585.

[13] Kim, D.-S., Cho, Y., and Kim, D. "Edge-tracing algorithm for Euclidean Voronoi diagram of 3D spheres". *Proc. 16th Canadian Conference on Computational Geometry*, pp. 176-179, 2004.

- [14] Kim, D.-S., Cho, Y., and Kim, D. "Euclidean Voronoi diagram of 3D balls and its computation via tracing edges". *Computer-Aided Design*(in printing).
- [15] Luchnikov, V.A., Medvedev, N.M., Oger, L., and Troadec, J.-P. "Voronoi-Delaunay analysis of voids in systems of nonspherical particles", *Physical review E*, Vol.59, No.6(1999), pp.7205-7212.
- [16] Montoro, J.C.G. and Abascal, J.L.F. "The Voronoi Polyhedra as Tools for Structure Determination in Simple Disordered Systems", *The Journal of Physical Chemistry*, Vol.97, No.16(1993), pp.4211-4215.
- [17] Noggle, J.H. "Physical Chemistry". 3rd ed. Freedom Academy Publishing Co., 1996.
- [18] Okabe, A., Boots, B., and Sugihara, K. *Spatial Tessellations: Concepts and Applications of Voronoi Diagram*, John Wiley & Sons, 1992.
- [19] Richards, F.M. "The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density", *Journal of Molecular Biology*, Vol.82(1974), pp.1-14.
- [20] Voloshin, V.P., Beaufils, S., and Medvedev, N.M. "Void space analysis of the structure of liquids", *Journal of Molecular Liquids*, Vol.96-97(2002), pp.101-112.
- [21] Computational Geometry Algorithms Library (CGAL) Homepage. <http://www.cgal.org/>, 2005.
- [22] RCSB Protein Data Bank. <http://www.rcsb.org/pdb/>, 2005.