

Discrete Differential Geometry and the Structural Study of Protein Complexes

Naoto Morikawa

Genocript, Kanagawa, Japan Email: nmorika@genocript.com

How to cite this paper: Morikawa, N. (2017) Discrete Differential Geometry and the Structural Study of Protein Complexes. *Open Journal of Discrete Mathematics*, **7**, 148-164. https://doi.org/10.4236/ojdm.2017.73014

Received: May 18, 2017 **Accepted:** July 7, 2017 **Published:** July 10, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

Abstract

This paper proposes a novel four-dimensional approach to the structural study of protein complexes. In the approach, the surface of a protein molecule is to be described using the intersection of a pair of four-dimensional triangular cones (with multiple top vertexes). As a mathematical toy model of protein complexes, we consider complexes of closed trajectories of *n*-simplices ($n = 2, 3, 4, \cdots$), where the design problem of protein complexes corresponds to an extended version of the Hamiltonian cycle problem. The problem is to find "a set of" closed trajectories of *n*-simplices which fills the *n*-dimensional region defined by a given pair of n+1-dimensional triangular cones. Here we give a solution to the extended Hamiltonian cycle problem in the case of n=2 using the discrete differential geometry of triangles (*i.e.*, 2-simplices).

Keywords

Discrete Differential Geometry, *n*-Simplices, Hamiltonian Cycle Problem, Protein Complexes, Vector Bundle

1. Introduction

Proteins are called the workhorse molecules of life, playing a crucial role in essentially every activity of living organisms. A protein molecule is made from one or more long chains of amino acids, which normally folds into a well-defined three-dimensional structure. It is the precise shape of the folded structure that determines the function of proteins in a cell.

Most cellular processes are not carried out by random collisions between freely diffusing proteins. Proteins usually interact with other proteins and assemble into complexes to carry out their function [1] [2] [3]. It is therefore crucial to understand and control the formation of protein complexes for understanding biological activity in the cell. In particular, structural characterization of the components of complexes, such as shape complementarity at proteinprotein interfaces, is the key to understanding the function of proteins.

In the last two decade, huge number of protein structures is experimentally determined via high-throughput structural genomics pipelines. However experimental determination of their functions is lagged far behind the pace due to the labor-intensive and time-consuming nature of the process. Urgently needed are improved computational approaches to function prediction of the proteins with known structure [4].

It is however extremely difficult to describe the shape of proteins without visual inspection on a three-dimensional display. The fundamental question is how to describe the geometry of such a highly complicated shape as proteins.

In most of previous studies, the surface of proteins is described using concepts developed in computational geometry and topology, such as the Voronoi diagram, the Delaunay simplices, and the alpha shape representation [5] [6] [7]. As for protein complexes, the topological arrangement of their subunits is usually represented as a graph [8] [9].

The Hamiltonian cycle problem on a regular triangular mesh: a) A region in a regular triangular lattice. b) A Hamiltonian cycle through the region.

An extended version of the Hamiltonian cycle problem on a regular triangular mesh: *de nove design* of complexes of closed trajectories of triangles. Shown are all the three sets of closed trajectories of triangles which cover the specified region. In this case, the region has no Hamiltonian cycle.

In this paper, we propose a novel mathematical toy model which is intended for the structural study of protein complexes. While physics and mathematics have been inspired each other in their long relationship, the relationship between biology and mathematics is still to come. In our case, it is the relation between real protein complexes and the new mathematical toy model. That is, it is critical to justify why such new toy models are indeed relevant and practically useful.

To justify the usefulness of mathematical tools in biology, I'd like to mention the case of the Estrada index introduced by Ernesto Estrada [10] in 2000. The Estrada index was originally proposed as a molecular structure descriptor, and the protein structure has been investigated by using the Estrada index and the normalized Laplacian Estrada index [11] extensively in mathematics in the past decade. The Estrada indices have also found a range of applications in chemistry and complex networks. These days, a dynamic version of the Estrada indices are proposed [12] to study large-scale time-evolving networks which arise naturally in a variety of areas from peer-to-peer telecommunication to online human social behavior to neuroscience.

As for other mathematical approaches to protein structure analysis, most of them are application of known mathematical techniques to the structural study of proteins, such as, distance geometry [13], the knot theory [13], and persistent homology [14]. Differential geometric techniques are also applied to the analysis of the backbone structure of proteins [15].

In our model, instead of open chains of amino acids, we consider closed trajectories of *n*-simplices using the discrete differential geometry of *n*-simplices $(n = 2, 3, \dots)$ [16] [17]. Then, interaction of open chains of amino acids (*i.e.*, proteins) is mimicked with "recombination", such as fusion and fission, of closed chains of *n*-simplices. The advantage of our model lies in the correspondence between the shape of a complex of closed trajectories of *n*-simplices and (a projection image of) the intersection of a pair of n+1-dimensional cones.

Using the mathematical toy model, we will consider the problem of designing protein complexes from scratch (*de novo design* of protein complexes [18] [19] [20]). That is, we will consider the problem of finding a set of closed trajectories of *n*-simplex that forms a specified *n*-dimensional shape: *de nove* design of complexes of closed trajectories of *n*-simplices. For simplicity, we consider the case of n=2 only.

2. Problem

The problem we consider here is an extended version of the Hamiltonian cycle problem on a regular triangular mesh. A Hamiltonian cycle of triangles (*i.e.*, 2-simplices) is a closed trajectory through a given triangular mesh which visits each triangle exactly once, where the trajectory passes triangles through a common edge. As shown in **Figure 1(a)**, meshes are given as a region in a two-dimensional regular triangular lattice. In this case, a Hamiltonian cycle is obtained as shown in **Figure 1(b)**.

To study the formation of a complex of closed trajectory of triangles, we consider not only a single but also multiple closed trajectories of triangles to cover the given region. In the case of **Figure 2**, two closed trajectories are required.

In what follows, we will propose a novel method for finding all the sets of closed trajectories which cover a given region of triangles.

3. Differential Structure on the Mesh

To define a differential structure on a regular triangular mesh, we stack unit cubes diagonally in the three-dimensional Euclidean space E^3 (Figure 3(a)).

By piling up unit cubes orderly in the direction of (-1,-1,-1) in E^3 , a "mountain range-like shape object" consisting of multiple triangular cones is obtained as shown in the upper part of Figure 3(c). If we draw a thick straight



Figure 1. The Hamiltonian cycle problem on a regular triangular mesh. (a) A region in a regular triangular lattice; (b) A Hamiltonian cycle through the region.



line diagonally on the three upper faces of each unit cube, we will obtain a "drawing" on the slope of the mountain range-like shape object (**Figure 3(a)** and **Figure 3(d)**). It is the drawing which specifies a flow of "slant" triangles (along the thick polygonal lines) on the slope.

Then, we define a flow of "flat" triangles on a plane which is perpendicular to the direction of (1,1,1) in E^3 by projecting the flow of "slant" triangles on the plane (the lower part of **Figure 3(c)**)). In the case of **Figure 3(c)**, we obtain a closed trajectory of flat triangles of length 30 and others. In this section we give the precise definition of the differential structure on a regular triangular mesh.

For space saving purposes, we use monomial in indeterminates x_0, x_1 and x_2 to represent the coordinate of points in the three-dimensional Euclidean space. For example, point $p = (l, m, n) \in \mathbb{Z}^3$ is identified with monomial $x_0^l x_1^m x_2^n$, where \mathbb{Z} denotes the set of all integers. Then, points (l+k, m, n), (l, m+k, n) and (l, m, n+k) are represented by monomials px_0^k , px_1^k and px_2^k respectively. (Note that $x_i x_i = x_i x_i$ for all pairs of *i* and *j*.)

3.1. Triangle Tiles

Shown in the upper part of **Figure 3(a)** is a unit cube with a thick straight line drawn diagonally on each of the upper three faces, which is located at the origin



Figure 2. An extended version of the Hamiltonian cycle problem on a regular triangular mesh: *de nove* design of complexes of closed trajectories of triangles. Shown are all the three sets of closed trajectories of triangles which cover the specified region. In this case, the region has no Hamiltonian cycle.



Figure 3. Differential structure on a regular triangular mesh. (a) A unit cube and its projection on a plane perpendicular to the direction of (1,1,1) in E^3 ; (b) The projection of "slant triangles" onto a "flat triangle"; (c) A "mountain range-like shape object" obtained by piling up unit cubes orderly along the diagonal direction, whose peaks are $P_0 = (1,0,0)$, $P_1 = (0,0,1)$, $P_2 = (-2,2,1)$, $P_3 = (0,3,0)$ and $P_4 = (2,2,-1)$; (d) A "drawing" on the slope of the mountain range-like shape object of (c).

O of a three-dimensional Cartesian coordinate system defined by three axes φ_0 , φ_1 and φ_2 . Let $P_a = 1$, $P_b = x_0$, $P_c = x_0 x_1$, and $P_d = x_1 \in \mathbb{Z}^3$. Then, the upper face $P_a P_b P_c P_d$ on the $p_0 p_1$ -plane is divided into two "slant triangles", $P_a P_b P_c$ and $P_a P_d P_c$, by the line segment $P_a P_c$. The other upper faces are also divided into two "slant triangles" similarly.

Shown in the lower part of **Figure 3(a)** is the projected image of the unit cube on a plane which is perpendicular to the direction of (1,1,1) in E^3 . The unit cube at *O* is projected onto a hexagon, which is divided into six "flat triangles" by the image of the three thick line segments on the cube.

The schematic drawing of **Figure 3(b)** shows the projection of slant triangles onto a flat triangle. Using the projection, we will define a discrete differential structure on the set of flat triangles, *i.e.*, a regular triangular mesh.

Let Sym_3 be the symmetric group on a finite set of three symbols. For $a \in \mathbb{Z}^3$ and $\rho \in Sym_3$, let $a \begin{bmatrix} x_{\rho((0)} x_{\rho(1)} \end{bmatrix}$ denote the convex hull of three points a, $ax_{\rho(0)}$ and $ax_{\rho(0)}x_{\rho(1)} \in \mathbb{Z}^3$, *i.e.*,

$$\begin{split} a\Big[x_{\rho(0)}x_{\rho(1)}\Big] &\coloneqq \left\{a^{\lambda_0}\left(ax_{\rho(0)}\right)^{\lambda_1}\left(ax_{\rho(0)}x_{\rho(1)}\right)^{\lambda_2} \mid \lambda_0, \lambda_1, \lambda_2 \in \mathbb{R}, \\ \lambda_0, \lambda_1, \lambda_2 \ge 0, \lambda_0 + \lambda_1 + \lambda_2 = 1\right\} \subset \mathbb{R}^3, \end{split}$$

where \mathbb{R} denotes the set of all real numbers.

For example, the "slant triangle" $P_a P_d P_c$ defined above is denoted by $[x_1 x_0] = a [x_{\rho(0)} x_{\rho(1)}]$, where $a = x_0^0 x_1^0 x_2^0 = 1$ and $\rho = (0,1)$.

Definition 3.1 We define the set S_2 of all slant (triangle) tiles by

$$S_2 \coloneqq \left\{ a \left[x_{\rho(0)} x_{\rho(1)} \right] \mid a \in \mathbb{Z}^3, \rho \in Sym_3 \right\}.$$

The set B_2 of all *flat* (*triangle*) *tiles* is defined as the quotient of S_2 by "shift operator" σ , *i.e.*,

$$B_2 \coloneqq S_2 / \sigma$$

where $\sigma\left(a\left[x_{\rho(0)}x_{\rho(1)}\right]\right) \coloneqq ax_{\rho(0)}\left[x_{\rho(1)}x_{\rho(2)}\right].$

We identify B_2 with the projection image of "slant triangles" on a plane perpendicular to vector (1,1,1) mentioned above. Then, the schematic drawing of **Figure 3(b)** shows the equivalence class of a slant tile $s \in S_2$ and the corresponding flat tile $s \mod \sigma \in B_2$.

3.2. Tangent Space at a Flat Triangle Tile

A tangent bundle-like local structure TB_2 is defined on B_2 by **Definition 3.2** $TB_2 \rightarrow B_2, \pi$

$$\begin{cases} TB_2 \coloneqq S_2/\sigma^3, \\ \pi: TB_2 \to B_2, \pi(s \mod \sigma^3) = s \mod \sigma. \end{cases}$$

Let $s \in S_2$. Then, we obtain

 $\pi^{-1}(s \mod \sigma) = \{s \mod \sigma^3, \sigma(s) \mod \sigma^3, \sigma^2(s) \mod \sigma^3\}.$

Definition 3.3 (Tangent space) For $s \in S_2$, we call $\pi^{-1}(s \mod \sigma)$ the tangent space of B_2 at $s \mod \sigma$.

Definition 3.4 For $s = a \left[x_{\rho(0)} x_{\rho(1)} \right] \in S_2$, the gradient Ds of s is defined by

$$Ds := x_{\rho(0)} x_{\rho(1)} \Big(= x_{\rho(1)} x_{\rho(0)} \Big).$$

Then, we can identify TB_2 with

$$B_2 \times \{x_0 x_1, x_1 x_2, x_0 x_2\}$$

via the one-to-one correspondence

 $s \mod \sigma^3 \sim (s \mod \sigma, Ds).$

Note that the monomial Ds of $s \in S_2$ corresponds to the direction of the thick line on the "slant triangle" s which is described in subsection 3.1 above (Figure 3).

3.3. Vector Field on B₂

Having defined a tangent-bundle like structure (TM_2, B_2, π) on a set of triangles, now we consider the inverse of the projection map π .

Definition 3.5 A section γ of (TB_2, B_2, π) is a map $B_2 \rightarrow TB_2$ such that $\pi(\gamma(t)) = t$ for all $t \in B_2$.

For a section γ of (TB_2, B_2, π) , the value of γ on $t \in B_2$ is given by $\gamma(t) = s \mod \sigma^3 \in TB_2$

for some $s = a \Big[x_{\rho(0)} x_{\rho(1)} \Big] \in S_2$. Let s_D and s_U be two adjacent slant tiles of s in S_2 defined by



Figure 4. Local trajectory. (a) The local trajectory specified by $s = [x_1 x_0] \in S_2$.

 $s_D = x_1[x_0x_1]$ and $s_U = [x_1x_2]$; (b) The smoothness condition on a section γ . Colored gray is $\gamma([x_1x_0] \mod \sigma)$ and colored white is $\gamma([x_1x_2] \mod \sigma)$. Shown above are the gradient of the white tile. The gradient of the gray tile is x_0x_1 ; (c) Smooth sections of (TB_2, B_2, π) on a hexagonal region composed of six flat tiles; (d) Sections of

 (TB_2, B_2, π) which dose not satisfy the smoothness condition. The corresponding singular flat tiles are colored gray in the lower part.

Then, the set of three slant tiles, $\{s_D, s, s_U\}$, makes up a "continuous mountain path" along the thick polygonal line (*i.e.*, along the gradient Ds) at s in S_2 (Figure 4(a)). By projecting these slant tiles on B_2 , we obtain a trajectory of flat tiles of length three at $s \mod \sigma$.

To consider the "smoothness" of the section γ , we firstly define a local trajectory passing through $t \in B_2$ as follows.

Definition 3.6 Let $s \in S_2$. The local trajectory specified by s is the set $\{s_D \mod \sigma, s \mod \sigma, s_U \mod \sigma\} \subset B_2$,

of three consecutive flat tiles passing though $s \mod \sigma$.

Let γ be a section on B_2 . Then, $\gamma(t)$ $(t \in B_2)$ can assume one of the three values of the corresponding tangent space $\pi^{-1}(t)$. For example, $\gamma(s_U \mod \sigma)$ can assume one of the three values of

$$\pi^{-1}(s_U \mod \sigma) = \left\{ x_0^{-1}[x_0 x_1] \mod \sigma^3, [x_1 x_2] \mod \sigma^3, x_1[x_2 x_0] \mod \sigma^3 \right\}$$

where $s = [x_1 x_0] \in S_2$ and $s_U = [x_1 x_2]$.

However some of the slant tiles are not connected smoothly to $\gamma(s \mod \sigma)$ in TB_2 . In this case,

$$x_1[x_2x_0] \mod \sigma^3 \sim (s_U \mod \sigma, x_0x_2)$$

is not connected smoothly to $\gamma(s \mod \sigma) = s \mod \sigma^3$ as shown in **Figure 4(b)**.

To obtain a "smooth" trajectory, we will impose a condition on sections of (TB_2, B_2, π) .

Definition 3.7 (Smoothness condition) Let γ be a section of (TB_2, B_2, π) and $t \in B_2$. Let $\gamma(t) = s \mod \sigma^3$, where $s = a \left[x_{\rho(0)} x_{\rho(1)} \right] \in S_2$. The smoothness condition on γ at t is defined by

$$D(\gamma(s_U \mod \sigma)) = x_{\rho(0)} x_{\rho(1)} \text{ or } x_{\rho(0)} x_{\rho(2)}, D(\gamma(s_D \mod \sigma)) = x_{\rho(0)} x_{\rho(1)} \text{ or } x_{\rho(1)} x_{\rho(2)}.$$

In what follows, we will only consider the sections of (TB_2, B_2, π) which satisfies the smoothness condition at every flat tiles of B_2 .

Remark $x_{\rho(0)}$ corresponds to (the direction of) the contact edge between s and s_U . $x_{\rho(1)}$ corresponds to (the direction of) the contact edge between s and s_D .

Definition 3.8 (Vector field) A vector filed on B_2 is a section of

 (TB_2, B_2, π) which satisfies the smoothness condition at every flat tiles of B_2 .

Shown in **Figure 4(c)** are all the six types of "local" smooth sections of (TB_2, B_2, π) on a hexagonal region composed of six flat tiles of B_2 . By patching these "local" sections together, we will obtain a "global" section of (TB_2, B_2, π) .

Note that some of the "global" sections do not satisfy the smoothness condition as shown in **Figure 4(d)**. The *singular* flat tile of a section γ of (TB_2, B_2, π) is the flat tile where γ dose not satisfy the smoothness condition. A singular flat tile is assigned either no gradient (*i.e.*, without thick edge), two gradients (*i.e.*, two thick edges), or three gradients.

Let $\mu = \{t[i] | i \in I\} \subset B_2$ be a trajectory of a vector field V, where I is a sub-

set of the set \mathbb{N} of natural numbers. Then, we can define the second derivative of the trajectory as follows.

Definition 3.9 The second derivative $D^2V(t[i])$ of V along μ is a binary-valued (U or D) function defined by

$$D^{2}V(t[i+1]) \coloneqq \begin{cases} D^{2}V(t[i]), & \text{if } DV(t[i+1]) = DV(t[i]) \\ -D^{2}V(t[i]), & \text{otherwise} \end{cases}$$

where -D := U and -U := D.

In [16], the conformation of a protein backbone structure is encoded into a 16-valued sequence using the second derivative of trajectories of tetrahedrons (*i.e.*, 3-simplices).

3.4. Vector Fields Induced by Tangent Cones

In the beginning of this section, we constructed a "mountain range-like shape object" by piling up unit cubes diagonally. (Using the terminology defined above, it is a section of (TB_2, B_2, π) .)

Unit cubes are piled up to form a union of triangular cones, which can be identified by its top vertexes. For example, the object shown in the upper part of **Figure 3(c)** is identified by five peaks P_0 , P_1 , P_2 , P_3 , and P_4 .

Definition 3.10 For $A \subset \mathbb{Z}^3$, a tangent cone $ConeA \subset \mathbb{Z}^3$ is defined by

ConeA := {
$$px_0^{l_0}x_1^{l_1}x_2^{l_2} | p \in A \text{ and } l_0, l_1, l_2 \ge 0$$
 }.

We denote the set of all the "top vertexes" of *ConeA* by p(ConeA). Then, the mountain range-like shape object of **Figure 3(c)** is given by

$$Cone\left\{x_{0}, x_{2}, x_{0}^{-2}x_{1}^{2}x_{2}, x_{1}^{3}, x_{0}^{2}x_{1}^{2}x_{2}^{-1}\right\}$$

For a tangent cone c, let dc be the set of all the slant tiles on the surfaces of c, *i.e.*,

$$dc := \left\{ a \left[x_{\rho(0)} x_{\rho(1)} \right] \in S_2 \mid a, a x_{\rho(0)} \text{ and } a x_{\rho(0)} x_{\rho(1)} \text{ are on the surfaces of } c \right\}.$$

For $z \in \mathbb{Z}^3$ and c = ConeA ($A \subset \mathbb{Z}^3$), set

$$l_{c}(z) \coloneqq \max_{z \in \mathcal{L}} \left\{ \min_{z \in \mathcal{L}} \{ l_{0}, l_{1}, l_{2} \} \mid z = x_{0}^{l_{0}} x_{1}^{l_{1}} x_{2}^{l_{2}} p \}.$$

Then, dc is given as follows.

Lemma 3.11 For c = ConeA $(A \subset \mathbb{Z}^3)$,

$$dc = \left\{ a \left[x_{\rho(0)} x_{\rho(1)} \right] \in S_2 \mid l_c(a) = l_c(a x_{\rho(0)}) = l_c(a x_{\rho(0)} x_{\rho(1)}) = 0 \right\}$$

Proof. Let $z, p \in \mathbb{Z}^3$. Then, $z = x_0^{l_0} x_1^{l_1} x_2^{l_2} p$ for some $(l_0, l_1, l_2) \in \mathbb{Z}^3$. (l_0, l_1, l_2) is the coordinate of *z* with respect to "origin" p. In particular,

$$z \in Cone \{p\} \Leftrightarrow l_0, l_1, l_2 \ge 0,$$

$$z \in d (Cone \{p\}) \Leftrightarrow \min \{l_0, l_1, l_2\} = 0.$$

The result follows immediately.

The surfaces of a tangent cone *c* induce a vector field of (TB_2, B_2, π) as follows.

Definition 3.12 For c = ConeA $(A \subset \mathbb{Z}^3)$, a vector field V_c induced by c is defined by

$$V_c(t) \coloneqq s \mod \sigma^3 \quad (t \in B_2),$$

where $s \in dc$ such that $t = s \mod \sigma$. The value $V_c(t)$ is uniquely determined at every flat tile of B_2 .

For example, in Figure 3(c), the thick polygonal lines on the surfaces of the tangent cone $Cone \{x_0, x_2, x_0^{-2} x_1^2 x_2, x_1^3, x_0^2 x_1^2 x_2^{-1}\}$ shows the vector field induced by the tangent cone.

Note that all the smooth sections shown in Figure 4(c) are induced by a tangent cone as indicated in the figure.

Proposition 3.13 For any vector field V, there exists a tangent cone c such that $V = V_c$.

Proof. Let V be a vector filed on B_2 and let $E = \{U_i | i \in \mathbb{N}\}$ be a decomposition of B_2 into hexagonal regions of six flat tiles:

$$B_2 = \bigcup \{ U_i \mid i \in \mathbb{N} \}$$

For $U_i \in E$, we let $V|_{U_i}$ denote the restriction of V on the hexagon U_i .

Because of the smoothness condition, V is locally induced by a tangent cone as shown in **Figure 3(c)**. That is, there exists a tangent cone c_i for each

$$\begin{aligned} U_i \in E \quad \text{such that} \quad V \big|_{U_i} = V_{c_i} \big|_{U_i}, \text{ i.e.,} \\ V(t) = V_{c_i}(t) \quad \text{for } \forall t \in U_i. \end{aligned}$$

Moreover, by considering all combinations, we can assume

$$V\Big|_{U_j \cup U_k} = V_{c_j \cup c_k}\Big|_{U_j \cup U_k} \tag{1}$$

for any pair of adjacent hexagons U_i and U_k , where $c_i \cup c_k$ denote the union of two tangent cones c_j and c_k , *i.e.*, Cone $p(c_j) \cup p(c_k)$.

Suppose that $V \neq V_c$ for $c = \bigcup_{i \in \mathbb{N}} c_i$. Then,

 $\exists U_a \in E \text{ such that } V_{c_a} \Big|_U \neq V_c \Big|_{U_a}$

In particular, $\exists b \in \mathbb{N}$ such that

$$\exists t_0 \in U_a$$
 such that $V_c(t_0) = V_{c_b}(t_0)$.

In other words, tangent cone c_a is (partially) covered by tangent cone c_b on U_a .

Then, there exists a circular loop Θ of hexagons of E around U_b such that

$$\exists U_{e} \in \Theta \text{ such that } V_{c_{e}} \Big|_{U_{e}} \neq V_{c_{e} \cup c_{b}} \Big|_{U_{e}} \text{ and } V_{c_{i}} \Big|_{U_{i}} = V_{c_{i} \cup c_{b}} \Big|_{U_{i}} \text{ for } \forall U_{i} \in \Omega,$$

where Ω is the set of all the hexagons of E contained in the circular region surrounded by Θ . Because of the shape of the tangent cones, $\exists U_f \in \Omega$ such that U_f is adjacent to U_e and c_e is (partially) covered by c_f on U_e , *i.e.*,

$$\exists t_1 \in U_e$$
 such that $V_c(t_1) = V_{c_f}(t_1) \neq V_{c_e}(t_1)$.

In particular,



$$V_{c_e}\Big|_{U_e} \neq V_{c_e \cup c_f}\Big|_{U_e}$$

which is in contradiction to Equation (1).

4. The Boundary of a Closed Trajectory

Now let's go back to our problem described in section 2. Using the terminology given in section 3, the problem is stated as follows.

Problem 4.1. (*De nove* design of complexes of closed trajectories of triangles) Given a region in B_2 , find all the vector fields on B_2 which give a decomposition of the region into closed trajectories.

If there exists such a vector field, we can describe the boundary of the region using a pair of three-dimensional cones as explained in this section.

The cones are defined in another lattice which is associated with \mathbb{Z}^3 . Recall that the three-dimensional lattice \mathbb{Z}^3 is generated by x_0 , x_1 and x_2 . The associated lattice is defined as follows.

Definition 4.2. The conjugate lattice \mathbb{L}_3 is the lattice which is generated by x_0x_1 , x_1x_2 and x_0x_2 .

Note that the gradient of a slant tile corresponds to one of the three coordinate axes of the conjugate lattice \mathbb{L}^3 . In particular, a trajectory of slant tiles corresponds to a zig-zag walk (with gaps) on the grid of \mathbb{L}^3 .

Two types of cones are defined in \mathbb{L}_3 :

Definition 4.3. For $A \subset \mathbb{L}_3$, a cotangent cone $Cone^*A \subset \mathbb{L}^3$ is defined by

$$Cone^*A := \left\{ p(x_1 x_2)^{l_0} (x_0 x_2)^{l_1} (x_0 x_1)^{l_2} \mid p \in A \text{ and } l_0, l_1, l_2 \ge 0 \right\}.$$

For $A \subset \mathbb{L}_3$, a *cotangent roof* $Roof^* A \subset \mathbb{L}^3$ is defined by

$$Roof^*A := \left\{ p \in \mathbb{L}^3 \mid \exists N > 0 \text{ such that } p\left(x_1 x_2\right)^N, p\left(x_0 x_2\right)^N, p\left(x_0 x_1\right)^N \in Cone^*A \right\}.$$

In other words, $Roof^*A$ is obtained by putting as many unit cubes as possible on $Cone^*A$.

For example, shown in Figure 5(c) is



Figure 5. Cotangent roofs associated with a closed trajectory on B_2 . (a) The boundary of the closed trajectory of **Figure 3(c)**; (b) The cotangent roof of the region, where $Q_0 = x_2^{-1}$, $Q_1 = x_0^{-1}$ and $Q_2 = x_0^{-2}x_1$; (c) The inverted cotangent roof of the region, where $Q_3 = x_0^3 x_1^3 x_2$ and $Q_4 = x_0^2 x_1^4 x_2^3$.

$$Roof^* \left\{ x_0, x_2, x_0^{-2} x_1^2 x_2, x_1^3, x_0^2 x_1^2 x_2^{-1} \right\} \left(= Cone^* \left\{ x_2^{-1}, x_0^{-1}, x_0^{-2} x_1 \right\} \right).$$

Inverted cones are also defined similarly:

Definition 4.4. For $A \subset \mathbb{L}_3$, an inverted cotangent cone $ICone^*A \subset \mathbb{L}^3$ is defined by

$$ICone^*A := \left\{ p(x_1 x_2)^{l_0} (x_0 x_2)^{l_1} (x_0 x_1)^{l_2} \mid p \in A \text{ and } l_0, l_1, l_2 \le 0 \right\}.$$

For $A \subset \mathbb{L}_3$, an *inverted cotangent roof* $IRoof^*A \subset \mathbb{L}^3$ is defined by

$$IRoof^*A := \left\{ p \in \mathbb{L}^3 \mid \exists N < 0 \text{ such that } p\left(x_1 x_2\right)^N, p\left(x_0 x_2\right)^N, p\left(x_0 x_1\right)^N \in ICone^*A \right\}.$$

For example, shown in Figure 5(c) is

$$IRoof^* \{ x_0, x_2, x_0^{-2} x_1^2 x_2, x_1^3, x_0^2 x_1^{-1} \} \Big(= ICone^* \{ x_0^3 x_1^3 x_2, x_0^2 x_1^4 x_2^3 \} \Big).$$

Then, the boundary of a closed trajectory of a vector field on B_2 can be described using a pair of a cotangent roof and an inverted cotangent roof as shown below.

Let w be a cotangent cone. We denote by $\partial(w)$ the set of all the lattice points of \mathbb{L}^3 which resides on the surface of the cone w. $\partial(w)$ is called the *boundary lattice points* of w. The boundary lattice points of an inverted cotangent cone is also defined in the same manner.

Proposition 4.5. Let V_c be a vector field of (TB_2, B_2, π) induced by a tangent cone c whose top vertexes p(c) are in \mathbb{L}_3 . Let $\mu = \{t[i] | i \in I\}$ $(I \subset \mathbb{N})$ be a closed trajectory of V_c . Let R_{μ} be the region swept by the trajectory μ .

Then, there exist a cotangent cone w and an inverted cotangent cone iv such that the boundary of R_{μ} is uniquely specified by the intersection of $\partial(w)$ and $\partial(iv)$.

The pair (w, iv) is called a *boundary pair* (of the region R_{μ}) and the specified region is denoted by $R_{(w,iv)}$, *i.e.* $R_{(w,iv)} = R_{\mu}$.

Proof. Let $\Lambda = \{s[i] | i \in I\}$ be a subset of S_2 such that

 $V_c(t[i]) = s[i] \mod \sigma^3 \quad (i \in I).$

Because of the smoothness condition, we may assume slant tiles of Λ are connected "smoothly" in S_2 without any gap. Let A be the set of all vertexes of the slant tiles of Λ . Define cones w and *iv* by

$$\begin{cases} w = Roof^*A, \\ iv = IRoof^*A. \end{cases}$$

Then, the boundary of R_{μ} is obtained by connecting the points of $\pi(\partial(w) \cap \partial(iv))$ on B_2 , where π denotes the projection of the lattice points of \mathbb{Z}^3 on the corresponding vertexes of flat tiles of B_2 .

Remark Roof ${}^*p(c)$ is not defined if $p(c) \notin \mathbb{L}^3$.

For example, in the case of the closed trajectory given in **Figure 3(d)**, the boundary of R_{μ} is uniquely specified by

$$\begin{cases} w = Cone^* \{ Q_0, Q_1, Q_2 \}, \\ iv = ICone^* \{ Q_3, Q_4 \} \end{cases}$$



as shown in Figure 5.

Corollary 4.6. Let *R* be a region in B_2 . Then, *R* has a closed-trajectory decomposition if and only if there exists a pair of a cotangent cone w and an inverted cotangent cone iv such that $R = R_{(w,iv)}$. The pair (w,iv) is also called a boundary pair (of *R*).

Proof. (\rightarrow) Let $\{\mu_i | i \in I\}$ ($I \subset \mathbb{N}$) be a closed-trajectory decomposition of R and let $\{(w_i, iv_i) | i \in I\}$ be their boundary pairs. Set

$$\begin{cases} w = \bigcup_{i \in I} w_i, \\ iv = \bigcup_{i \in I} iv_i. \end{cases}$$

Then, $R = \bigcup_{i \in I} R_{\mu_i} = \bigcup_{i \in I} R_{(w_i, iv_i)}$.

(←) A closed-trajectory decomposition of *R* is induced by *Cone* $\partial(w) \cap \partial(iv)$.

Remark Let \mathbb{T} be the set of all tangent cones. Let \mathbb{C} be the set of all cotangent cones. Let \mathbb{L} be the set of all the regions in B_2 which are defined by boundary pairs. Then, an \mathbb{L} -valued function is defined on $\mathbb{T} \times \mathbb{C}$ by $\langle ConeA, Cone^*B \rangle :=$ "the region in B_2 which is specified by the intersection of ConeA and Cone^{*}B ". In particular,

$$\langle Cone \,\partial(w) \cap \partial(iv), w \rangle = R_{(w,iv)},$$

for a boundary pair (w, iv).

5. Extended Hamiltonian Cycle Problem on B₂

5.1. Problem

By Corollary 4.6, we can paraphrase Problem 4.1 as follows.

Problem 5.1. (*De nove* design of complexes of closed trajectories of triangles) Given a boundary pair (w,iv), find all the tangent cones which induce such a vector field that gives a decomposition of the region $R_{(w,iv)}$ into closed trajectories (Figure 6).



Figure 6. The extended Hamiltonian cycle problem on B_2 (See also **Figure 1**). (a) A pair of a cotangent cone and an inverted cotangent cone which specifies the boundary of a region: $Roof^*A$ and $IRoof^*B$, where $A, B \subset \mathbb{L}^3$; (b) A tangent cone which induces such a vector field whose trajectories don't traverse the specified boundary: Cone C, where $C = \partial (Roof^*A) \cap \partial (IRoof^*B)$.

One of the solutions to the problem is obtained immediately, *i.e.*,

Cone $\partial(w) \cap \partial(iv)$ (Figure 6(b)). In this section, we consider how to find all solutions to the problem.

5.2. Closed-Trajectory Decomposition

Definition 5.2. For $A \subset \mathbb{Z}_3$, a tangent roof $RoofA \subset \mathbb{Z}^3$ is defined by $Roof A := \left\{ p \in \mathbb{Z}^3 \mid \exists N > 0 \text{ such that } p(x_0)^N, p(x_1)^N, p(x_2)^N \in Cone A \right\}$. In other words, RoofA is obtained by putting as many unit cubes as possible on ConeA.

Definition 5.3. For $A \subset \mathbb{Z}_3$, a (tangent) ceiling Ceil $A \subset \mathbb{Z}^3$ is defined by Ceil A := Cone C,

where

$$C = \bigcup \{ B \mid Roof^*B = Roof^*A \text{ and } IRoof^*B = IRoof^*A \} \subset \mathbb{Z}^3.$$

For $A \subset \mathbb{Z}_3$, a (*tangent*) *floor* $FloorA \subset \mathbb{Z}^3$ is defined by *Floor* A := Cone C,

where

$$C = \bigcap \{B \mid Roof^*B = Roof^*A \text{ and } IRoof^*B = IRoof^*A \} \subset \mathbb{Z}^3$$

It follows immediately that

Floor
$$A \subset Ceil A$$
,
Roof $p(Floor A) = Roof A$,
Roof $p(Ceil A) = Roof A$,

where p(c) denotes the set of all the "top vertexes" of a cone c.

For a boundary pair (w, iv), let $W_{(w,iv)}$ be the set of all the tangent cones c such that

Floor
$$C \subset c \subset Ceil C$$
,

where $C = \partial(w) \cap \partial(iv)$.

Now all solutions to Problem 5.1 are obtained as follows:

Proposition 5.4. $W_{(w,iv)}$ induces all decompositions of $R_{(w,iv)}$ into closed trajectories.

Proof. (\rightarrow) Let *VF* be the set of all the vector fields whose trajectories don't traverse the boundary of $R_{(w,w)}$. Then,

$$V_c \in VF$$
 for any $c \in W_{(w,iv)}$

because $(Roof^*p(c), IRoof^*p(c)) = (w, iv)$. In particular, V_c induces a decompositions of $R_{(w,iv)}$ into closed trajectories.

 (\leftarrow) Given a decomposition of $R_{(w,iv)}$ into closed trajectories. Then, it can be extended to a vector field V on B_2 . For example, a flow of triangles on $B_2 \setminus R_{(w,iv)}$ is induced by

$$V_{Cone \ \partial(w) \cap \partial(iv)}$$

Then, \exists a tangent cone *c* such that $V = V_c$ by proposition 3.13. Then

$$(Roof^*p(c), IRoof^*p(c)) = (w, iv)$$

because trajectories of V_c don't traverse the boundary of $R_{(w,w)}$.

For example, **Figure 7** shows all solutions to the problem for the boundary pair of **Figure 5**.

5.3. Fusion and Fission of Closed Trajectories

For a vector field V on B_2 and a region R of B_2 , let Dec(V,R) be the set of all closed trajectories of V in R. Dec(V,R) gives a closed-trajectory decomposition of R if it exists. The number of the closed trajectories of Dec(V,R) is denoted by #Dec(V,R).

For a boundary pair (w, iv), let c0 and c1 $(c0 \neq c1)$ be two tangent cones of $W_{(w,iv)}$. Then, vector fields V_{c0} and V_{c1} induce two different decompositions of $R_{(w,iv)}$: $Dec(V_{c0}, R_{(w,iv)})$ and $Dec(V_{c1}, R_{(w,iv)})$. The correspondence

$$Dec(V_{c0}, R_{(w,iv)}) \leftrightarrow Dec(V_{c1}, R_{(w,iv)})$$

gives a "recombination" of closed trajectories from one to the other. In particular, it gives a "fusion" and "fission" of closed trajectories on $R_{(w,w)}$ if

$$\#Dec(V_{c0}, R_{(w,iv)}) > 1 \text{ and } \#Dec(V_{c1}, R_{(w,iv)}) = 1.$$

In the case of **Figure 7**, the region $R_{(w,iv)}$ has three decompositions but no Hamiltonian cycle:

$$\begin{cases}
\# Dec(V_{Cone C}, R_{(w,iv)}) = 2 & (a) \\
\# Dec(V_{Cone C \cup \{x_1^2\}}, R_{(w,iv)}) = 2 & (b) \\
\# Dec(V_{Cone C \cup \{x_1\}}, R_{(w,iv)}) = 2 & (c)
\end{cases}$$

Moreover, it is not difficult to show the following proposition (in the case of closed trajectories of triangles):



Figure 7. All solutions to the extended Hamiltonian cycle problem for a boundary pair (w, iv), where $\partial(w) \cap \partial(iv) = \{x_0, x_2, x_0^{-2}x_1^2x_2, x_1^3, x_0^2x_1^2x_2^{-1}, x_0x_1^3x_2^{-1}\}$ (See also **Figure 2**). Set $C = \partial(w) \cap \partial(iv)$. (a) The vector field induced by *Floor C*; (b) The vector field induce by *Cone C* $\cup \{x_1^2\}$ which is obtained by putting a cube (with top vertex x_1^2) on *Floor C*; (c) The vector field induced by *Ceil C* = *Cone C* $\cup \{x_1\}$, which is obtained by putting a cube (with top vertex x_1) on the vector field of (b).

Proposition 5.5. When a closed trajectory is merged with a closed trajectory of length 6 (which occupies a hexagonal region), they don't fuse together to form a single closed trajectory.

In other words, closed trajectories always split when they interact with a hexagon.

See **Tabel 1** for the distribution of the length of closed trajectories of *n*-simplices (n = 2, 3, 4).

The distribution of the length of closed trajectories of *n*-simplices (n = 2, 3, 4). Two closed trajectories are identified if and only if their sequences of the second derivative coincide with each other by rotational shift, inversion, or reversion.

6. Conclusions

We have considered an extended version of a two-dimensional Hamiltonian cycle problem in a three-dimensional setting, where the boundary of a two-dimensional region is uniquely specified by a pair of three-dimensional cones, *i.e.*, a boundary pair. Using the discrete differential geometry of triangles, all decompositions of the region into closed trajectories of triangles are obtained immediately from the intersection of the boundary pair.

In the structural study of protein complexes, it is essential to characterize surface features such as bumps (convexity) and dents (concavity) of protein molecules. However mathematical surface characterization has not produced any satisfactory results so far, where the surface of protein molecules is usually studied in a three-dimensional setting.

This paper proposes a novel mathematical approach to the structural study of protein complexes, *i.e.*, an approach from a four-dimensional setting, where the surface of protein molecules is to be described by a pair of four-dimensional cones (with multiple top vertexes) as in the case of complexes of closed trajectories of triangles.

In our approach, protein molecules are to be represented as closed trajectories of tetrahedrons, where shape complementarity is expressed inherently. In particular, we could define fusion and fission of molecules (*i.e.*, closed trajectories) naturally.

As a future research subject, we are considering whether there exist any (algebraic) equations a given boundary pair satisfies. If there exists a set of such equations that specifies the given boundary pair, it is possible to represent the shape

Table 1. The distribution of the length of closed trajectories of *n*-simplices (n = 2, 3, 4). Two closed trajectories are identified if and only if their sequences of the second derivative coincide with each other by rotational shift, inversion, or reversion.

Length	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36
2-simplices	0	0	1	0	1	0	2	0	5	0	11	0	27	0	78	0	234	0
3-simplices	0	0	1	0	0	3	0	0	6	0	0	17	0	0	42	0	0	118
4-simplices	0	0	1	0	0	2	1	1	3	4	3	4	5	13	10	10	12	41



of a protein molecule as a solution for a system of equations. In particular, we would obtain another protein molecule of the same function if a given set of equations has more than one solution.

References

- [1] Alberts, B. (1998) The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell*, 92, 291-294. https://doi.org/10.1016/S0092-8674(00)80922-8
- [2] Ponstingl, H., Kabir, T., Gorse, D. and Thornton, J.M. (2005) Morphological Aspects of Oligomeric Protein Structures. *Progress in Biophysics and Molecular Biology*, 89, 9-35.
- [3] Pereira-Leal, J.B., Levy, E.D. and Teichmann, S.A. (2006) The Origins and Evolution of Functional Modules: Lessons from Protein Complexes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361, 507-517. https://doi.org/10.1098/rstb.2005.1807
- [4] Lee, D., Redfern, O. and Orengo, C. (2007) Predicting Protein Function from Sequence and Structure. *Nature Reviews Molecular Cell Biology*, 8, 995-1005. https://doi.org/10.1038/nrm2281
- [5] Edelsbrunner, H. (2001) Geometry and Topology for Mesh Generation. Cambridge University Press, Cambridge. <u>https://doi.org/10.1017/CBO9780511530067</u>
- [6] Gerstein, M and Richards, F.M. (2001) Protein Geometry: Volumes, Areas, and Distances. In: Rossman, M.G. and Arnold, E., Eds., *The International Tables for Crystallography*, Vol. F, Chap. 22, Kluwer, Dordrecht, 531-539.
- Ban, Y.-E.A., Edelsbrunner, H. and Rudolph, J. (2006) Interface Surfaces for Protein-Protein Complexes. *Journal of the ACM*, 53, 361-378. https://doi.org/10.1145/1147954.1147957
- [8] Levy, E.D., Pereira-Leal, J.B., Chothia, C. and Teichmann, S.A. (2006) 3D Complex: A Structural Classification of Protein Complexes. *PLoS Computational Biology*, 2, e155. <u>https://doi.org/10.1371/journal.pcbi.0020155</u>
- [9] Perica, T., Marsh, J.A., Sousa, F.L., Natan, E., Colwell, L.J., Ahnert, S.E. and Teichmann, S.A. (2012) The Emergence of Protein Complexes: Quaternary Structure, Dynamics and Allostery. *Biochemical Society Transactions*, 40, 475-491. <u>https://doi.org/10.1042/BST20120056</u>
- [10] Estrada, E. (2000) Characterization of 3D Molecular Structure. *Chemical Physics Letters*, **319**, 713-718.
- [11] Li, J., Guo, J. and Shiu, W.C. (2014) The Normalized Laplacian Estrada Index of a Graph. *Filomat*, 28, 365-371. https://doi.org/10.2298/FIL1402365L
- Shang, Y. (2015) Laplacian Estrada and Normalized Laplacian Estrada Indices of Evolving Graphs. *PLoS ONE*, **10**, e0123426. https://doi.org/10.1371/journal.pone.0123426
- [13] Taylor, W.R. and Aszodi, A. (2004) Protein Geometry, Classification, Topology and Symmetry: A Computational Analysis of Structure. Taylor & Francis, UK.
- [14] Edelsbrunner, H., Letscher, D. and Zomorodian, A. (2002) Topological Persistence and Simplification. *Discrete & Computational Geometry*, 28, 511-533. https://doi.org/10.1007/s00454-002-2885-2
- [15] Goriely, A., Hausrath, A. and Neukirch, S. (2008) The Differential Geometry of Proteins and Its Applications to Structure Determination. *Biophysical Reviews and Letters*, 3, 77-101. <u>https://doi.org/10.1142/S1793048008000629</u>

- [16] Morikawa, N. (2014) Discrete Differential Geometry of n-Simplices and Protein Structure Analysis. Applied Mathematics, 5, 2458-2463. https://doi.org/10.4236/am.2014.516237
- [17] Morikawa, N. (2016) Discrete Differential Geometry of Triangles and Escher-Style Trick Art. Open Journal of Discrete Mathematics, 6, 161-166. https://doi.org/10.4236/ojdm.2016.63013
- [18] Woolfson, D.N., Bartlett, G.J., Burton, A.J., Heal, J.W., Niitsu, A., Thomson, A.R. and Wood, C.W. (2015) De Novo Protein Design: How Do We Expand into the Universe of Possible Protein Structures? Current Opinion in Structural Biology, 33, 16-26.
- [19] Huang, P.-S., Boyken, S.E. and Baker, D. (2016) The Coming of Age of de Novo Protein Design. Nature, 537, 320-327. https://doi.org/10.1038/nature19946
- [20] Norn, C.H. and Andre, I. (2016) Computational Design of Protein Self-Assembly. Current Opinion in Structural Biology, 39, 39-45.

🔆 Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc. A wide selection of journals (inclusive of 9 subjects, more than 200 journals) Providing 24-hour high-quality service User-friendly online submission system Fair and swift peer-review system Efficient typesetting and proofreading procedure Display of the result of downloads and visits, as well as the number of cited articles Maximum dissemination of your research work Submit your manuscript at: http://papersubmission.scirp.org/

Or contact ojdm@scirp.org

