Polynomial representation of DNA and proteins

Naoto Morikawa

GENOCRIPT, 27-22-1015, Sagami-ga-oka 1-chome, Zama-shi, Kanagawa 228-0001 Japan
(e-mail: nmorika@f3.dion.ne.jp)

March 3, 2003.

Abstract. A new number system is developed to measure shapes of proteins. In the system the structure of a protein is represented by polynomials whose coefficients are all 1. Biochemical reactions of proteins correspond to arithmetic equations and protein folding corresponds to integration. The system also gives another example of quantization of space via quantization of orientation, i.e., via directional nature of covalent bonds.

Key words. hetero number – polynomial – polyhedron – DNA – protein – coding – discrete mathematics – geometry – moduli

Contents

1. Introduction

This paper gives a way to "measure" or to "express as a number" shapes of proteins, where the structure of a protein is represented by polynomials whose coefficients are all 1. Unlike the active field of molecular dynamics simulation, mathematical approach has been rarely taken to study (static) geometry of proteins.

One of the reasons is lack of mathematical theory suited to protein structure analysis. At best one can counts on optimal packing or knot theory. For example, [6] computes optimal packing of DNA and [7] searches knots in proteins. This paper will pave a new way for protein structure analysis and hopefully enlarge the field of mathematics.

Recall that natural numbers are represented as a sum of 1s. For example $6 = 1+1+1+1+1+1$. Factoring 6 into $2*3$, we obtain $(1+1+1+1+1+1) = (1+1)*(1+1+1)$. What would happen if we replace 1 with indeterminates $x_0, x_1, \ldots, x_{N-1}$? In this paper we define a number system as a set of subsets of the free group $\langle x_0, x_1, \ldots, x_{N-1} \rangle$ generated by $x_0, x_1, \ldots, x_{N-1}$. And we give the condition when a subset is a number and show how it is factored into prime numbers.

In the case of natural numbers, the number system is "homo" in the sense that $x_0 = x_1 = \cdots = x_{N-1} = 1$. In contrast, we call the new system "hetero numbers". Because of its "heterogeneity", every hetero number is associated with a geometrical object. In particular, a prime number is associated with a closed chain of polyhedrons.

We can also interpret arithmetic operations geometrically. In particular, various biochemical reactions of proteins can be associated with arithmetic equations. Moreover the polynomial representation of a protein is determined by the second derivative of its shape. By differentiation all the geometrical information is encoded into a $\{+1, -1\}$ sequence. That is, "protein folding" corresponds to a process of integration and "genetic coding" corresponds to differentiation.

1.1. Biological background

Here we give a brief introduction to *PROTEOMICS*. For further reference, for example, see [2] or [15]. Proteomics is the word coined to express the principal theme of the life science community in the post-genome era. That is, the study of proteins that are encoded by the genes of an organism. It includes characterization of the biological functions of proteins (functional genomics) and analysis of the corresponding 3-dimensional ($3D$) structures (structural genomics). One can retrieve actual structure data determined by X-ray crystallography or NMR spectroscopy from the Protein Data Bank (PDB, [16]), where more than $14,000$ structures are deposited.

1.1.1. What is protein Protein is a polymer of amino-acids linked by peptide bonds. The order of amino-acids are encoded by a gene in a genome.

In nature proteins are folded into a well-defined $3D$ structure (the native state). This process is called protein folding. Free energy of the native state is significantly lower than that of alternative structures. Protein folding is reproducible and a protein always folds into the particular naive state rapidly.

The functional properties of proteins depend upon their $3D$ structure. For example, the $3D$ structure of a protein determines the active site of an enzyme, the binding site of a drug or the binding site for another protein. These active sites or binding sites are key to understand the function of a protein in the cell, or to understand how particular molecular targets interact with drugs. Therefore the knowledge of the $3D$ structures of proteins is crucial to study full potential of genomic information.

Amino-acids. There are 20 naturally occurring amino-acids. All amino-acids have a common structure, i.e., a tetrahedral (sp^3) carbon atom (C_α) to which four asymmetric groups are connected: an amino group (NH_2), a carboxyl group ($COOH$), a hydrogen atom (H) and another chemical group (denoted by R) which varies from one amino-acid to another.

Note that an atom with sp^3 covalent bond has four orbitals pointing towards the corners of an imaginary tetrahedron. In an ideal conformation the angles between any two orbitals would be 109.5 degrees. That is, sp^3 orbitals are tetrahedral in shape. The peptide unit together with the C_α are termed back-bone and the R is termed side-chain.

Amino-acid sequence. Amino-acids are linked by peptide bonds to form a polypeptide chain during protein synthesis. The peptide bond is formed by a condensation reaction between the amino and carboxyl group, which releases a water molecule and forms a covalent bond between them:

$$\cdots\text{-}COOH + H_2N\text{-}\cdots \longrightarrow \cdots\text{-}CO\text{-}NH\text{-}\cdots + H_2O.$$

The remains of each amino-acid, that is R, is called an (amino-acid) residue. The distance between one C_α atom to the next C_α atom is about $3.8\mathring{A}$ (angstrom). In eukaryotes the average length of (domains of) proteins is 153 ± 87 residues ([9]).

The conformation space of an amino-acid sequence. Because of the rigidity of covalent bonds, freedom of motion is restricted to rotations around these bonds. Each amino-acid can rotate around two such bonds. By convention, the angle of rotation around the N-C_α bond is denoted by Φ and the angle around the C_α-C by Ψ. Thus the conformation of whole polypeptide chain is determined by these angles for each amino-acids.

$$\cdots\text{-}N\text{-}C_\alpha\text{-}C\text{-}N\text{-}C_\alpha\text{-}C\text{-}N\text{-}C_\alpha\text{-}C\text{-}\cdots$$

How dose a protein finds its naive state in a finite time? (Levinthal paradox) Let's consider the case of a small protein made up of a polypeptide chain of 100 residues. Suppose that there are only two possible configurations for each residue. Then there are about 10^{30} possible conformations. If it takes 10^{-11} seconds to convert one configuration into another, a random search of all conformations would require 10^{11} years. On the other hand, doubling time of bacteria can be less than 30 minutes ([4]). The protein-folding is one of the major unsolved problems ([1]).

Genetic code. The order of amino-acids in a polypeptide chain is determined by its corresponding genetic code. A genetic code is a sequence of nucleotides, coded in triplets (codons) along the DNA or mRNA. (The DNA sequence of a gene can be used to predict the mRNA sequence.) The evidence gathered over years indicates that it would be possible to deduct the $3D$ structure of any protein from its amino-acid sequence, i.e., its genetic code. That is, all of the information needed to specify the $3D$ structure is contained within its genetic code.

Notable is the following fact. On one hand, proteins with similar amino-acids sequence (the degree of sequence identity is more than 30%) have similar structure and function. On the other hand, a replacement of a single amino-acid in the sequence may unfold or alter the whole structure, as is the case with sickle cell anemia. The deduction/prediction of the structure from a code is another major unsolved problem ([12]).

1.1.2. Protein structure The structure of proteins are organized in a structural hierarchy.

Primary structures. Primary structure is the amino-acid sequence. Two amino-acid sequence are said to be homologous if they are very similar. The homology relationship among sequences indicate that they are evolutionary related.

Secondary structures. Secondary structure is the conformational arrangement of the backbone segments of a polypeptide chain. There are two main types of them: α-helices and β-sheets, where a β-sheet is a “pleated” sheet of β-strands. Both types are characterized by hydrogen-bonding between the NH and CO groups in the chain and formed when a number of consecutive residues have the same rotation angles Φ and Ψ. Typically the length of a α-helix is about 12 residues and the depth of a pleats of a β-sheet is about 6 residues. α-helices and β-sheets are connected by loop regions of various lengths and irregular shape.

Certain arrangements of a few consecutive secondary structure elements with a specific geometric arrangement are present in many different protein structures. They are called either supersecondary structures or motifs. Motifs are formed by packing side chains from adjacent α-helices or β-strands close to each other.

Tertiary structures. Tertiary structure is the spatial organization of secondary structures within a single polypeptide chain. Combinations of secondary structures (α-helices, β-sheets, loop regions, and motifs) which pack together to form folded shapes are called domains. Small proteins usually consist of only one domain. But larger proteins may contain a number of domains connected by segments of polypeptide chain which lack regular secondary structure.

Among the about $14,000$ proteins in PDB, more than 25,000 domains are reported. The core $3D$ structure of a domain is called a fold. All data available support the hypothesis that there are only a few thousand possible folds in all human proteins ($\sim 35,000$ protein genes), despite the requirement for a vast variety of different functions.

Quaternary structures. Some proteins consists of several polypeptide chains (subunits) that associate into a molecule in a specific way and called multimeric proteins. These subunits are held together by hydrogen bonds, van der Waals and coulombic forces. Quaternary structure is the organization of two or more polypeptide chains with tertiary structure. These subunits can function either independently of each other or cooperatively among subunits. Proteins that have only one chain are called monomeric.

1.1.3. Protein taxonomy The structures of proteins show regularities and many proteins have the same or similar folds even in cases when they have no obvious sequence similarity, i.e., evolutionary relation. With the growing database of known protein conformations, the classification of the $3D$ structures now plays a central role in understanding the principles of protein structure and function. Many classification databases are available on the web, for example, see [13] or [14]. The goal of structural genomics involves generating a set of structures representative of most of the possible folds for specific protein domains and then solving the structures for new proteins by homology (sequence similarity) ([11]).

Classification schemes are to some extent subjective and there are a number of different classifications of folds, derived from manual or automatic structure comparison. Here we give an example of the classifications. Items are arranged from top to bottom increasing structural similarity.

Because of its subjectivity, the classifications are not robust nor amenable to automation. To avoid the problem [8] proposes a “periodic table” for proteins. This paper also gives a new approach for protein classification.

Class. One common way of classification is to divide all proteins or protein domains into 5 major groups or classes,: (i) all α (or α/α), (ii) all β (or β/β), (iii) α/β, (iv) $\alpha+\beta$ and (v) a fifth group of proteins which do not fit in any of the other categories. The α/β and $\alpha+\beta$ differ in the way that the a/b proteins mainly are built up of parallel β-strands connected by α-helices, while the $\alpha+\beta$ proteins have β-strands and α-helices connected in a more irregular fashion. Note that class is based on the structure of the protein core and small elements of secondary structures are usually ignored.

Fold. The number, type, connectivity, and arrangement of secondary structures defines the fold of a protein. Fold recognition attempt to detect similarities between $3D$ structures that are not accompanied by any significant sequence similarity. There are many approaches to try and find folds that are compatible with a particular sequence. Frequently fold similarities are recognized by eye-following structure determination.

Family. Protein families are often arranged in a hierarchy. Superfamily refer to a group of structurally or functionally related proteins not necessarily of common evolutionary origin. If proteins with the same fold share a common ancestor, they are referred to as homologs. If they do not share a common ancestor, they are analogs. Closely related proteins with a recent common ancestor comprise families.

Superfolds and supersites. Many superfamilies share certain protein folds. It suggests that the fold has arisen many times by convergent evolution. Such folds are termed superfolds. The same superfolds from different superfamilies may perform different functions. Though some superfolds from different superfamilies show a tendency to bind molecules (ligands) in a common location (binding site), that is, perform the same function. These locations are termed supersites. Supersites are thought to be a property of the fold, such as the alignment of nonhydrogen-bonded main-chain atoms.

1.1.4. Protein function The ultimate goal of genome projects is to determine the function and biological role for all genes in a genome. And one should remember the vagueness of the term “function”. The function of a protein is defined by its molecular activity and mechanism of action (what it dose and how it acts), by its expression characteristics (when and where it acts) and more. Usually structural data carry only the information of biochemical function of the protein such as specific bindings to ligands (or drugs) of a protein.

Proteins are stable mechanical constructs that allow certain internal motions to enable their biological function. The folded domains can serve as module for building up large assemblies such as virus particles or muscle fibers. Or they can provide specific catalytic or binding sites as found in enzymes or proteins that carry oxygen or that regulate the function of DNA. For example, enzyme active sites are usually in distinct clefts on the surface of the protein. In contrast, protein-protein interaction sites are highly exposed and vary in character. These binding sites or active sites are key to understand the function of a protein in the cell. In particular, to understand how proteins interact with drugs.

1.1.5. Beyond protein A large molecule made of repeating subunits (monomer) linked by covalent bonds is called polymer. (A molecule of just a few monomers, generally from 3 to 10, is called oligomer.) For example, protein is a linear polymer of amino-acids and DNA is a linear polymer of nuclei acids. Small proteins give simple examples of self-assemble linear polymers. That is, proteins form complex architectures without external intervention and perform highly sophisticated functions.

The ability of this self-assemble process will have wide application in the field of nanotechnology. It will be more efficient to construct molecular-scale machines of diverse functions by folding up essentially linear polymer (bottom-up approach) than by placing an atom at a time into a $3D$ assembly (top-down approach) ([10]). Chemists have already begun to synthesize polymers with properties that are similar to those of proteins. It is now possible to assemble specific sequence of diverse monomer sets into chain lengths that are nearly sufficient for tertiary structure formation ([5]).

1.2. Basic idea

The basic idea behind this paper is the following observation: let's pile up N-cubes (unit cubes in $\mathbf{R}^N$) in a diagonal direction and project the resulting surface onto an $N-1$-dimensional hypersurface along the diagonal axis. By printing a pattern on the surface of each N-cube, one obtains a drawing made up of the patterns, which defines a flow of tiles. In this section we illustrate it in the case of $N=3$.

Consider a unit cube in $\mathbf{R}^3$ whose vertices are, say, given by $v_1 = (0,0,0)$, $v_x = (1,0,0)$, $v_y = (0,1,0)$, $v_z = (0,0,1)$, $v_{xy} = (1,1,0)$, $v_{xz} = (1,0,1)$, $v_{yz} = (0,1,1)$ and $v_{xyz} = (1,1,1)$. And draw lines $\overline{v_1 v_{xy}}$, $\overline{v_1 v_{yz}}$ and $\overline{v_1 v_{xz}}$. Each face is divided into two slant triangle-tiles by the lines. For example, triangles $v_1 v_x v_{xy}$ and $v_1 v_y v_{xy}$ for the face $v_1 v_x v_{xy} v_y$.

By piling up unit cubes in the direction from v_{xyz} to v_1, we obtain an object of peaks-and-valleys (Fig.1(a)). The lines form a flow of slant tiles along the slopes. In particular, there are three closed orbits of tiles.

By projecting all closed orbits on the hypersurface $\{(\lambda_0, \lambda_1, \lambda_2) : \lambda_0 + \lambda_1 + \lambda_2 = 0\} \subset \mathbf{R}^3$, we obtain a 2-dimensional shape V. The shape V is uniquely determined by its ten peaks:

$$\begin{aligned}&\{(0,-2,-2),(1,-3,-2),(2,-3,-3),(3,-4,-3),(3,-5,-2),\\&\quad(4,-6,0),(2,-4,0),(1,-4,1),(0,-4,2),(-3,-3,0)\} \subset \mathbf{Z}^3.\end{aligned}$$

And we represent the shape by a polynomial

$$\begin{aligned}f =& x^0y^{-2}z^{-2} + x^1y^{-3}z^{-2} + x^2y^{-3}z^{-3} + x^3y^{-4}z^{-3} + x^3y^{-5}z^{-2}\\&+ x^4y^{-6}z^0 + x^2y^{-4}z^0 + x^1y^{-4}z^1 + x^0y^{-4}z^2 + x^{-3}y^{-3}z^0,\end{aligned}$$

where a term $x^l y^m z^n$ corresponds to a peak $(l,m,n) \in \mathbf{Z}^3$.

In the same way, each closed orbit defines the subshapes: V_0, V_1 and V_2. Their representations (f_i for V_i) are

$$\begin{aligned}f_0 &= x^0y^{-2}z^{-2} + x^1y^{-3}z^{-2} + x^2y^{-4}z^0\\&\quad + x^1y^{-4}z^1 + x^0y^{-4}z^2 + x^{-3}y^{-3}z^0,\\f_1 &= x^1y^{-3}z^{-2} + x^2y^{-3}z^{-3} + x^3y^{-4}z^{-3} + x^3y^{-5}z^{-2},\\f_2 &= x^3y^{-5}z^{-2} + x^4y^{-6}z^0 + x^2y^{-4}z^0.\end{aligned}$$

We denote the relation among f, f_0, f_1 and f_2 by an equation

$$f = f_0 \oplus f_1 \oplus f_2,$$

which is the "prime factoring" of f.

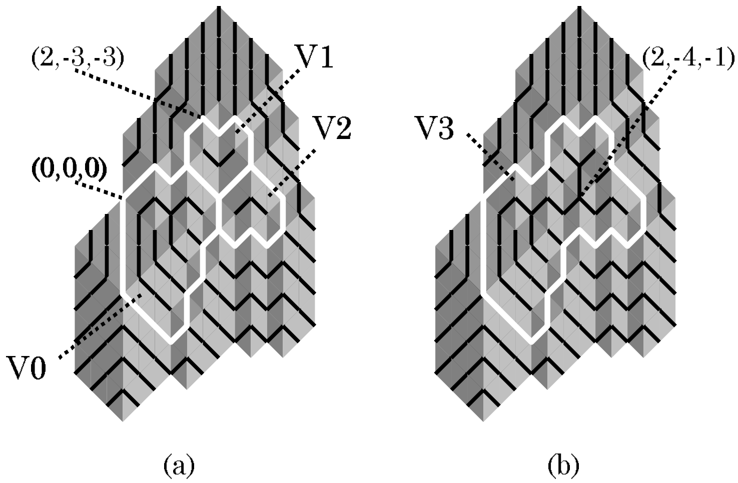


Fig. 1. Peaks-and-valleys

Note that the shape is determined by the gradient of the slant tiles along an orbit: "up (U)" and "down (D)" . Therefore we can encode the shape by assigning "U/D" to each tile. For example, the code of V_1 is given by

$$D-D-U-U-D-U-D-D-D-U-U-U-D-U.$$

This is a mathematical version of "genetic coding".

On the other hand, by putting one more cube on the surface, the shape V fuses into a single orbit V_3 (Fig.1(b)). We denote the relation between the two decompositions by

$$(f_0 \oplus f_1 \oplus f_2) * x^2 y^{-4} z^{-1} = f_3,$$

where f_3 is the representation of V_3.

Lastly we give a look-and-feel in the case of $N = 4$. By piling up 4-dimensional cubes, we can decompose a rhombic dodecahedron into a chain of 24 tetrahedron-tiles (Fig.2). Its polynomial representation is

$$f' = xyw + xz + yz + wz.$$

The "U/D"-code of the chain is

$$\begin{aligned} U-D-D-U-D-U-U-D& \\ -U-D-D-U-D-U-U-D& \\ -U-D-D-U-D-U-U-D&. \end{aligned}$$

This is the "genetic code" of a rhombic dodecahedron. Another representation

$$f'' = xyz + xyw + xzw + yzw.$$

defines a dodecahedron divided into four equal parts (see example 5).

The chain of tetrahedrons satisfies the following conditions:

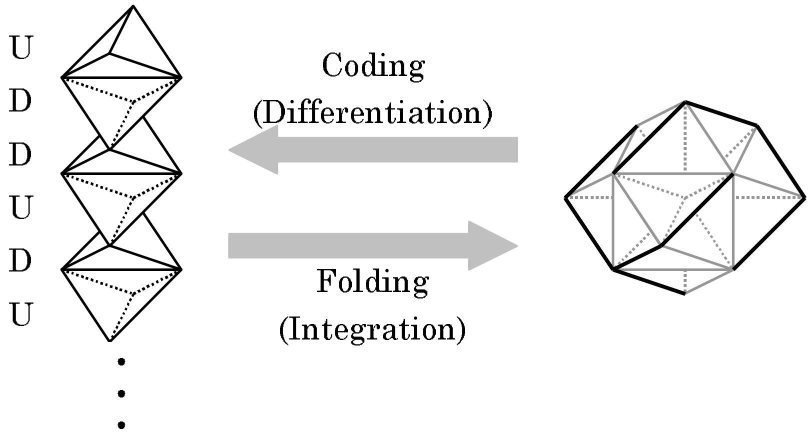


Fig. 2. The "genetic code" of a rhombic dodecahedron

(1) Each tetrahedron consists of 4 short edges and 2 long edges, where the ratio of the length is $\sqrt{3}/2$.
(2) Tetrahedrons are connected via long edges and rotate around the edges.

Any shape represented by a polynomial is obtained by folding some chains of adequate length according to their "U/D" codes and holding them together. (The chain structure itself was already reported in [3].)

2. Differential geometry of polyhedrons

2.1. Preliminary

2.1.1. Lattice L with two orders Let $N > 2$ and L be a free Abelian group generated by indeterminates $x_0, x_1, x_2, \ldots, x_{N-1}$. We denote the multiplication in L simply by ab for $a, b \in L$ and set $1 := \prod_{0 \le i < N} x_i^0$. Let H be a cyclic subgroup of L generated by $e := x_0 x_1 x_2 \cdots x_{N-1}$:

$$L := \left\{ \prod_{0 \le i < N} {x_i}^{l_i} : l_i \in \mathbf{Z}\ (0 \le i < N) \right\},$$
$$H := \left\{ e^k \in L : k \in \mathbf{Z} \right\} \subset L.$$

The canonical projection $L \to L/H$ is denoted by π.

For $a = \prod_{0 \le i < N} {x_i}^{l_i} \in L$, we set $|a|_L := \sum_{0 \le i < N} l_i$. The function $|-|_L$ on L induces a function on L/H and we denote it by $|\pi(a)|_{L/H}$, i.e., $|\pi(a)|_{L/H} \equiv |a|_L \bmod N$.

Set

$$L_0 := \{a \in L : |a|_L = 0\} \subset L,$$
$$L^* := \left\{ \prod_{0 \le i < N} (e/x_i)^{l_i} : l_i \in \mathbf{Z}\ (0 \le i < N) \right\} \subset L.$$

We also consider a scalar extension of L:

$$L_{\mathbf{R}} := \left\{ \prod_{0 \le i < N} {x_i}^{\lambda_i} : \lambda_i \in \mathbf{R}\ (0 \le i < N) \right\},$$
$$(L/H)_{\mathbf{R}} := \left\{ \prod_{0 \le i < N} {x_i}^{\lambda_i} \in L_{\mathbf{R}} : \sum_{0 \le i < N} \lambda_i = 0 \right\} \subset L_{\mathbf{R}}.$$

A projection $\pi_{\mathbf{R}}$ from $L_{\mathbf{R}}$ onto $(L/H)_{\mathbf{R}}$ is given by

$$\pi_{\mathbf{R}} \left(\prod_{0 \le i < N} {x_i}^{\lambda_i} \right) := \left(\prod_{0 \le i < N} {x_i}^{\lambda_i} \right) / e^{(\lambda_0 + \lambda_1 + \cdots + \lambda_{N-1})/N}.$$

L is a lattice of $L_{\mathbf{R}}$ and L_0 is a lattice of $(L/H)_{\mathbf{R}}$. Occasionally we identify L/H with $\pi_{\mathbf{R}}(L) \subset (L/H)_{\mathbf{R}}$.

L is partially ordered in two ways and each order is associated with a $\mathbf{Z}$-valued function as follows.

Definition 1. *Let* $a_0 = \prod_{0 \le i < N} {x_i}^{l_i}$, $a_1 = \prod_{0 \le i < N} {x_i}^{m_i} \in L$.

(1) Order T *and its height function* $h_T : L \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$a_0 \ge_T a_1 \overset{\text{def}}{\Longleftrightarrow} l_i \le m_i\ (0 \le i < N),$$
$$h_T[a, b] := -N \min \left\{ h \in \mathbf{Q} : a \ge_T b e^h \right\}.$$

(2) Order C *and its height function* $h_C : L \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$a_0 \ge_C a_1 \overset{\text{def}}{\Longleftrightarrow} \sum_{0 \le i < N} l_i/(N-1) - l_i \le \sum_{0 \le i < N} m_i/(N-1) - m_i$$
$$(0 \le i < N),$$
$$h_C[a, b] := -N \min \left\{ h \in \mathbf{Q} : a \ge_C b e^h \right\}.$$

In particular,

$$\prod_{0 \le i < N} (e/x_i)^{L_i} \ge_C \prod_{0 \le i < N} (e/x_i)^{M_i} \Longleftrightarrow L_i \le M_i\ (0 \le i < N).$$

By elementary calculation we obtain the following.

Lemma 1. *Let* $a_0 = \prod_{0\le i<N} {x_i}^{l_i}$, $a_1 = \prod_{0\le i<N} {x_i}^{m_i} \in L$.

(1) $h_T[a_0, \pi_{\mathbf{R}}(a_1)] = -\max_{0\le k<N} \{N(l_k - m_k) + \sum_i m_i\} \in \mathbf{Z}$.

(2) $h_C[a_0, \pi_{\mathbf{R}}(a_1)] = -\max_{0\le k<N} \Big\{N\Big(\big(\sum_i l_i - (N-1)l_k\big) - \big(\sum_i m_i - (N-1)m_k\big)\Big) + \sum_i m_i\Big\} \in \mathbf{Z}$.

(3) $|a_0|_L = -h_T[a_0, \pi_{\mathbf{R}}(a_0)] = -h_C[a_0, \pi_{\mathbf{R}}(a_0)]$.

2.1.2. Powerset PL of L Let PL be the set of all finite subsets of L. We denote an element $f = \{a_0, a_1, \ldots, a_{n-1}\} \subset L$ by $f = \sum_{0\le i<n} a_i$. And we write set theoretic operations additively, i.e., $0 := \emptyset$, $f + g := f \cup g$ and $f - g := \{a \in f : a \notin g\}$. That is, we treat f as if it were a polynomial whose coefficients are all 1. In fact f is a finite subset of L.

The orders in L are extended to orders in PL.

Definition 2. *Let* $f, g \in PL$.

(1) Order T and its height function $h_T : PL \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$f \ge_T g \overset{\text{def}}{\Longleftrightarrow} \forall a' \in g, \exists a \in f \text{ s.t. } a \ge_T a',$$
$$h_T[f, b] := \max\{h_T[a, b] : a \in f\}.$$

(2) Order C and its height function $h_C : PL \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$f \ge_C g \overset{\text{def}}{\Longleftrightarrow} \forall a' \in g, \exists a \in f \text{ s.t. } a \ge_C a',$$
$$h_C[f, b] := \max\{h_C[a, b] : a \in f\}.$$

For $U \subset PL$, we define

$$\max_T\{f : f \in U\} := \text{the smallest subset } g \text{ of } \sum_{f\in U} f \text{ s.t. } g \ge_T \sum_{f\in U} f,$$

$$\max_C\{f : f \in U\} := \text{the smallest subset } g \text{ of } \sum_{f\in U} f \text{ s.t. } g \ge_C \sum_{f\in U} f.$$

2.2. Tile space and its bundles

2.2.1. Tile space First we define the space B^{N-1} of $N-1$-dimensional tiles where we will consider differential geometry. The space B^{N-1} is given by a partition of $(L/H)_{\mathbf{R}}$ into N-hedrons.

Let $a_i \in L$ $(0 \le i < N)$ and set

$$[a_0, a_1, \ldots, a_{N-1}] := \left\{ \prod_{0\le i<N} {a_i}^{\lambda_i} : \sum_{0\le i<N} \lambda_i = 1, \quad 0 \le \lambda_i \in \mathbf{R}(0 \le i < N) \right\}.$$

We call $[a_0, a_1, \ldots, a_{N-1}]$ an N-*hedron* in $L_{\mathbf{R}}$. An N-hedron in $(L/H)_{\mathbf{R}}$ is also defined similarly and set

$$\pi([a_0, a_1, \ldots, a_{N-1}]) := [\pi_{\mathbf{R}}(a_0), \pi_{\mathbf{R}}(a_1), \ldots, \pi_{\mathbf{R}}(a_{N-1})].$$

Let S_N be the N-th symmetric group and denote $g \in S_N$ by $(x_{g(0)}x_{g(1)} \cdots x_{g(N-1)})$. Let id be the unit element of S_N, i.e., $id := (x_0x_1 \cdots x_{N-1})$.

For $(a, g) \in L_{\mathbf{R}} \times S_N$ and $(b, g) \in (L/H)_{\mathbf{R}} \times S_N$, set

$$\begin{aligned}
[a, g]_T &:= \big[a, ax_{g(0)}, ax_{g(0)}x_{g(1)}, \\
&\qquad \ldots, ax_{g(0)}x_{g(1)} \ldots x_{g(N-2)}\big] \subset L_{\mathbf{R}}, \\
[a, g]_C &:= \big[a, a(e/x_{g(N-1)}), a(e/x_{g(N-1)})(e/x_{g(N-2)}), \\
&\qquad \ldots, a(e/x_{g(N-1)})(e/x_{g(N-2)}) \ldots (e/x_{g(1)})\big] \subset L_{\mathbf{R}}, \\
[b, g]_B &:= \big[b, \pi_{\mathbf{R}}(bx_{g(0)}), \pi_{\mathbf{R}}(bx_{g(0)}x_{g(1)}), \\
&\qquad \ldots, \pi_{\mathbf{R}}(bx_{g(0)}x_{g(1)} \ldots x_{g(N-2)})\big] \subset (L/H)_{\mathbf{R}}.
\end{aligned}$$

We call $[a, g]_T$ and $[a, g]_C$ *slant tiles* and $[a, g]_B$ a *(base) tile*.

Let $a \in L$ s.t. $|a|_L \equiv -m \bmod N$ $(0 \le m < N)$ and $g = (y_0y_1 \cdots y_{N-1})$. Then $\pi([a, g]_T) = [\pi_{\mathbf{R}}(ay_0y_1 \cdots y_{m-1}), (y_my_{m+1} \cdots y_{m-1})]_B$.

Definition 3 (Tile space). *The* tile space B^{N-1} *is defined by*

$$B^{N-1} := \{[b, g]_B : (b, g) \in L_0 \times S_N\}.$$

B^{N-1} gives an N-hedron tiling of $(L/H)_{\mathbf{R}}$. It follows immediately that $B^{N-1} = \{\pi([a, g]_T) : (a, g) \in L \times S_N\} = \{\pi([a, g]_C) : (a, g) \in L \times S_N\}$.

Example 1 (Shape of a tile).

(1) The case of $N = 3$. $[1, id]_B = \pi([1, x_0, x_0x_1]) = [1, x_0/e^{1/3}, x_0x_1/e^{2/3}]$. The set $\{[1, g]_B : g \in S_3\} \subset B^2$ gives a division of a hexagon.

(2) The case of $N = 4$. $[1, id]_B = \pi([1, x_0, x_0x_1, x_0x_1x_2]) = [1, x_0/e^{1/4}, x_0x_1/e^{2/4}, x_0x_1x_2/e^{3/4}]$. The tetrahedron consists of 4 short edges and 2 long edges, where the ratio of the length is $\sqrt{3}/2$. The set $\{[1, g]_B : g \in S_4\} \subset B^3$ gives a division of a rhombic dodecahedron.

Note that rhombic dodecahedrons can be fitted together to fill $\mathbf{R}^3$ with no overlaps and no gaps.

2.2.2. Tangent bundle and D^2 bundle

Definition 4 (Tangent bundle). *Let* $(a, g) \in L \times S_N$. *We set* $D[a, g]_T := [\pi_{\mathbf{R}}(a), g]_T \subset L_{\mathbf{R}}$ *and call it the* gradient *or the* first derivative *of* $[a, g]_T$. *We also define* $D[a, g]_C := [\pi_{\mathbf{R}}(a), g]_C$. *Set*

$$\begin{aligned}
T(B^{N-1}) &:= \{D[a, g]_T : (a, g) \in L \times S_N\}, \\
T^*(B^{N-1}) &:= \{D[a, g]_C : (a, g) \in L \times S_N\}.
\end{aligned}$$

We define projections onto B^{N-1} *by* $\pi(D[a,g]_T) := \pi([a,g]_T)$ *and* $\pi(D[a,g]_C) := \pi([a,g]_C)$. $T(B^{N-1})$ *is the* tangent bundle *over* B^{N-1} *and* $T^*(B^{N-1})$ *is the* cotangent bundle *over* B^{N-1}. *We denote the set of all sections of* $T(B^{N-1})$ *over* $U \subset B^{N-1}$ *by* $\mathcal{O}_T(U)$. *Elements of* $\mathcal{O}_T(U)$ *are called* vector fields *over* U. $\mathcal{O}_{T^*}(U)$ *are also defined similarly.*

Let $t = \pi([a,g]_T) \in B^{N-1}$, where $g = (y_0 y_1 \cdots y_{N-1}) \in S_N$. Then the fiber $\mathcal{O}_T(t)$ of $T(B^{N-1})$ over t is given by

$$\begin{aligned}\mathcal{O}_T(t) = \{&D[a,g]_T,\\ &D[ay_0,(y_1\cdots y_{N-1}y_0)]_T,\\ &D[ay_0y_1,(y_2\cdots y_{N-1}y_0y_1)]_T,\\ &\dots, D[ay_0y_1\cdots y_{N-2},(y_{N-1}y_0\cdots y_{N-2})]_T\}.\end{aligned}$$

The fiber $\mathcal{O}_{T^*}(t)$ of $T^*(B^{N-1})$ over t is also given in the same way.

For $t = [b,g]_B \in B^{N-1}$ and $k \in \mathbf{Z}$, set

$$\begin{aligned} a(t,k) &:= \begin{cases} be^M & \text{if } k \equiv 0 \bmod N,\\ bx_{g(0)}x_{g(1)}\dots x_{g(m-1)}e^M & \text{else.}\end{cases}\\ g(t,k) &:= (x_{g(m)}x_{g(m+1)}\cdots x_{g(N-1)}x_{g(0)}\dots x_{g(m-1)}),\end{aligned}$$

where $m, M \in \mathbf{Z}$ s.t. $k = -(m+MN)$ and $0 \le m < N$.

Proposition 1.

(1) $T(B^{N-1}) \simeq B^{N-1} \times \mathbf{Z}/(N)$ *by the isomorphism of sets*

$$\begin{aligned} D[a,g]_T &\mapsto (\pi([a,g]_T), -|\pi(a)|_{L/H}),\\ (t, k \bmod N) &\mapsto D[a(t,k), g(t,k)]_T.\end{aligned}$$

(2) $T^*(B^{N-1}) \simeq B^{N-1} \times \mathbf{Z}/(N)$ *by the isomorphism of sets*

$$\begin{aligned} D[a,g]_C &\mapsto (\pi([a,g]_C), -|\pi(a)|_{L/H}),\\ (t, k \bmod N) &\mapsto D[a(t,k), g(t,k)]_C.\end{aligned}$$

Remark 1. We denote a one-to-one correspondence between two sets A and B by $A \simeq B$.

Now we consider the orbit defined by $d = D[a,g]_T \in T(B^{N-1})$. Let $a_0 = a$ and $a_i = ax_{g(0)}x_{g(1)}\cdots x_{g(i-1)}$ $(0 < i < N)$. Set

$$\begin{aligned} a'_0 &:= (x_{g(0)}/x_{g(N-1)})a_0 \in L,\\ a'_{N-1} &:= (x_{g(N-1)}/x_{g(N-2)})a_{N-1} \in L.\end{aligned}$$

$\pi_{\mathbf{R}}(a'_0)$ is the reflection of $\pi_{\mathbf{R}}(a_0)$ in the face $\pi([a_1,a_2,\dots,a_{N-1}]) \subset (L/H)_{\mathbf{R}}$ and $\pi_{\mathbf{R}}(a'_{N-1})$ is the reflection of $\pi_{\mathbf{R}}(a_{N-1})$ in the face $\pi([a_0,a_1,\dots,a_{N-2}]) \subset (L/H)_{\mathbf{R}}$.

Definition 5 (Local orbit and orbit). *Let $t \in B^{N-1}$. The* local orbit *defined by $d = D[a, g]_T \in T(B^{N-1})$ is the set of tiles $\{t_{-1}, t_0, t_1\} \subset B^{N-1}$ s.t. $t_0 = \pi(d)$ and $\{t_{-1}, t_1\} = \{\pi(d_U), \pi(d_D)\}$, where*

$$d_U := D[a_0, a_1, \ldots, a_{N-2}, a'_{N-1}] \in T(B^{N-1}),$$
$$d_D := D[a'_0, a_1, \ldots, a_{N-2}, a_{N-1}] \in T(B^{N-1}).$$

Note that $d \simeq (t_0, -|a|_L \bmod N)$ and the local orbit is uniquely determined by $|a|_L \bmod N$. An orbit *is a sequence of tiles $\{t[i]; i \in [l_0, l_1]\} \subset B^{N-1}$ s.t. each consecutive triplet is a local orbit, i.e., $t[i] = \pi(d)$ and $\{t[i-1], t[i+1]\} = \{\pi(d_U), \pi(d_D)\}$ for some $d \in \mathcal{O}_T(t[i])$.*

An infinite orbit $\{t[i] : i \in \mathbf{Z}\}$ is called closed *if there exists $k \in \mathbf{Z}$ s.t. $t[i+k] = t[i]$ for $\forall i \in \mathbf{Z}$. Let $Obt = \{t[i] : i \in \mathbf{Z}\}$ be a closed orbit. We call $\min\{k : k > 0, t[i+k] = t[i]\ \forall i \in \mathbf{Z}\}$ the* period *of Obt.*

We identify a closed orbit Obt with the finite orbit $\{t[i]; i \in [0, k-1]\}$, where $k \in \mathbf{Z}$ is the period of Obt.

Proposition 2 (Differentiation of an orbit). *Let $\{t[i]\} \subset B^{N-1}$ be an orbit. Then its first derivative $\{Dt[i]\} \subset T(B^{N-1})$ is computed as follows:*

$$Dt[i] \simeq \begin{cases} (t[i], -|Da[i-1]|_{L/H} + 1) & \text{if } (t[i+1], t[i]) = (t[i]_U, t[i-1]_U), \\ (t[i], -|Da[i-1]|_{L/H}) & \text{if } (t[i+1], t[i]) = (t[i]_D, t[i-1]_U), \\ (t[i], -|Da[i-1]|_{L/H}) & \text{if } (t[i+1], t[i]) = (t[i]_U, t[i-1]_D), \\ (t[i], -|Da[i-1]|_{L/H} - 1) & \text{if } (t[i+1], t[i]) = (t[i]_D, t[i-1]_D). \end{cases}$$

where $Dt[i] = [Da[i], g[i]]_T$, $t[i]_U = \pi((Dt[i])_U)$, $t[i]_D = \pi((Dt[i])_D)$.

Proposition 3 (Integration of D). *Let $\{d[i]\} \subset T(B^{N-1})$ be a sequence of the first derivatives. Then the orbit $\{t[i]\} \subset B^{N-1}$ s.t. $Dt[i] = d[i]$ is computed as follows:*

(Case1) If $t[i] = \pi(d[i-1]_U)$, then

$$t[i+1] = \begin{cases} \pi(d[i]_U) & \text{if } |Da[i]|_{L/H} \equiv |Da[i-1]|_{L/H} + 1, \\ \pi(d[i]_D) & \text{if } |Da[i]|_{L/H} \equiv |Da[i-1]|_{L/H}, \\ \text{undefined} & \text{else}. \end{cases}$$

(Case2) If $t[i] = \pi(d[i-1]_D)$, then

$$t[i+1] = \begin{cases} \pi(d[i]_D) & \text{if } |Da[i]|_{L/H} \equiv |Da[i-1]|_{L/H} - 1, \\ \pi(d[i]_U) & \text{if } |Da[i]|_{L/H} \equiv |Da[i-1]|_{L/H}, \\ \text{undefined} & \text{else}. \end{cases}$$

where $d[i] = [Da[i], g[i]]_T$ and $\equiv$ denotes congruence modulo N.

Lastly we define the second derivative of an orbit $\{t[i]\} \subset B^{N-1}$. It is defined only if the gradient $Dt[i]$ of consecutive tiles share $N-1$ vertices. There are only two cases for each direction:

(1) Gradients of $\pi(d_U)$ which share $N-1$ vertices with $d \in T(B^{N-1})$.

$$D[a_0, \ldots, a_{N-2}, a'_{N-1}] \simeq (\pi(d_U), -|\pi(a)|_{L/H}),$$
$$D[a'_{N-1}/e, a_0, \ldots, a_{N-2}] \simeq (\pi(d_U), -|\pi(a)|_{L/H} + 1).$$

(2) Gradients of $\pi(d_D)$ which share $N-1$ vertices with $d \in T(B^{N-1})$.

$$D[a'_0, a_1, \ldots, a_{N-1}] \simeq (\pi(d_D), -|\pi(a)|_{L/H}),$$
$$D[a_1, \ldots, a_{N-1}, a'_0 e] \simeq (\pi(d_D), -|\pi(a)|_{L/H} - 1).$$

Definition 6 (D^2 bundle). *For an orbit $\{t[i]\} \subset B^{N-1}$, the* second derivative $\{D^2 t[i]\}$ *of the orbit is defined by*

$$D^2 t[i] := +1 \text{ or } -1,$$
$$D^2 t[i+1] := \begin{cases} -D^2 t[i] & \text{if } |Da[i+1]|_{L/H} \equiv |Da[i]|_{L/H}, \\ D^2 t[i] & \text{if } |Da[i+1]|_{L/H} \equiv |Da[i]|_{L/H} \pm 1, \\ \text{undefined} & \text{else}. \end{cases}$$
$$D^2 t[i-1] := \begin{cases} -D^2 t[i] & \text{if } |Da[i-1]|_{L/H} \equiv |Da[i]|_{L/H}, \\ D^2 t[i] & \text{if } |Da[i-1]|_{L/H} \equiv |Da[i]|_{L/H} \pm 1, \\ \text{undefined} & \text{else}. \end{cases}$$

where $Dt[i] = [Da[i], g[i]]_T$ and $\equiv$ denotes congruence modulo N.

Moreover we suppose that

$$t[i+1] = \begin{cases} \pi((Dt[i])_U) & \text{if } D^2 t[i] > 0, \\ \pi((Dt[i])_D) & \text{if } D^2 t[i] < 0. \end{cases}$$

The D^2 bundle *over B^{N-1} is defined by*

$$T^2(B^{N-1}) := B^{N-1} \times \{+1, -1\}.$$

We denote the set of all sections of $T^2(B^{N-1})$ over $U \subset B^{N-1}$ by $\mathcal{O}_{T^2}(U)$.

Proposition 4 (Integration of D^2). *Let $\{c[i]\} \subset T^2(B^{N-1})$ be a sequence of the second derivatives. Then the orbit $\{t[i]\} \subset B^{N-1}$ s.t. $D^2 t[i] = c[i]$ is computed as follows:*

$$Dt[i+1] \simeq \begin{cases} (t[i]_U, -|Da[i]|_{L/H} + 1) & \text{if } c[i+1] > 0, c[i] > 0, \\ (t[i]_U, -|Da[i]|_{L/H}) & \text{if } c[i+1] < 0, c[i] > 0, \\ (t[i]_D, -|Da[i]|_{L/H}) & \text{if } c[i+1] > 0, c[i] < 0, \\ (t[i]_D, -|Da[i]|_{L/H} - 1) & \text{if } c[i+1] < 0, c[i] < 0. \end{cases}$$
$$t[i+1] = \pi(Dt[i+1]),$$

where $Dt[i] = [Da[i], g[i]]_T$, $t[i]_U = \pi((Dt[i])_U)$, $t[i]_D = \pi((Dt[i])_D)$.

2.3. Affine vector field

Here we define two line bundles over B^{N-1} and consider their sections induced by an element of PL.

2.3.1. Space of slant tiles

Definition 7 (Space of slant tiles). *Set*

$$S(B^{N-1}) := \{[a,g]_T : (a,g) \in L \times S_N\},$$
$$S^*(B^{N-1}) := \{[a,g]_C : (a,g) \in L \times S_N\}.$$

Note that $\pi(S(B^{N-1})) = \pi(S^*(B^{N-1})) = B^{N-1}$. We denote the set of all sections of $S(B^{N-1})$ over $U \subset B^{N-1}$ by $\mathcal{O}_S(U)$. We also define $\mathcal{O}_{S*}(U)$ similarly.

Proposition 5.
(1) $S(B^{N-1}) \simeq B^{N-1} \times \mathbf{Z}$ by the isomorphism of sets

$$[a,g]_T \mapsto (\pi([a,g]_T), -|a|_L),$$
$$(t,k) \mapsto [a(t,k), g(t,k)]_T.$$

(2) $S^*(B^{N-1}) \simeq B^{N-1} \times \mathbf{Z}$ by the isomorphism of sets

$$[a,g]_C \mapsto (\pi([a,g]_C), -|a|_L),$$
$$(t,k) \mapsto [a(t,k), g(t,k)]_C.$$

Thus $T(B^{N-1}) \simeq S(B^{N-1})/(N)$ and $T^*(B^{N-1}) \simeq S^*(B^{N-1})/(N)$ by the natural projection from $\mathbf{Z}$ to $\mathbf{Z}/(N)$.

2.3.2. Polynomial sections A term $a \in L$ defines slopes

$$\left\{ a \prod_{0 \le i < N} {x_i}^{\lambda_i} : \lambda_i \in \mathbf{R}, \min_{0 \le i < N} \{\lambda_i\} = 0 \right\} \subset L_{\mathbf{R}},$$

which induce an element of $\mathcal{O}_S(B^{N-1})$. We define sections as follows.

Definition 8. *Let $f \in PL$ and $t = [b_0, b_1, \ldots, b_{N-1}] \in B^{N-1}$.*
(1) Polynomial section of $S(B^{N-1})$.

$$|Sf|(t) := \max\{h_T[f, b_j] : 0 \le j < N\} \in \mathbf{Z},$$
$$Sf(t) := [a(t, |Sf|(t)), g(t, |Sf|(t))]_T \in \mathcal{O}_S(t).$$

(2) Polynomial section of $S^*(B^{N-1})$.

$$|S^*f|(t) := \max\{h_C[f, b_j] : 0 \le j < N\} \in \mathbf{Z},$$
$$S^*f(t) := [a(t, |S^*f|(t)), g(t, |S^*f|(t))]_C \in \mathcal{O}_{S^*}(t).$$

A polynomial section induces a vector field over B^{N-1} as follows.

Definition 9 (Affine vector field). *Let* $t \in B^{N-1}$ *and* $f \in PL$.

$$Df(t) := D\left[a(t, |Sf|(t)), g(t, |Sf|(t))\right]_T \in \mathcal{O}_T(t).$$

The vector field induced by $f \in PL$ *is called an* $N-1$*-dimensional* affine vector field *over* B^{N-1} *and denoted by* $X(f)$*:* $X(f) \in \mathcal{O}_T(B^{N-1})$.

A closed orbit of an $N-1$*-dimensional affine vector field is called an* $N-1$*-dimensional* affine loop. *We denote the set of all* $N-1$*-dimensional affine loops by* ALP^{N-1}.

Let $f \in PL$. Since the value of $|Sf|$ varies "continuously", we can define the second derivative of f.

Definition 10 (The 2nd derivative of f). *Let* $d = Df(t) \in \mathcal{O}_T(t)$ *and set*

$$D^2 f(\pi(d_U)) := 2\left(|Sf|(\pi(d_U)) - |Sf|(t) - 1/2\right) D^2 f(t),$$
$$D^2 f(\pi(d_D)) := 2\left(|Sf|(t) - |Sf|(\pi(d_D)) - 1/2\right) D^2 f(t).$$

We call $D^2 f \in \mathcal{O}_{T^2}(B^{N-1})$ *the* second derivative *of* f.

Orbits of $X(f)$ $(f \in PL)$ are calculated from $D^2 f$ as follows.

Proposition 6. *The orbit* $\{t[i]\} \subset B^{N-1}$ *of* $X(f)$ *which contains* $t \in B^{N-1}$ *is obtained by the following steps:*

(Step 0) $t[0] = t$ *and* $D^2 f(t[0]) = 1$.
(Step $k+1$*)*

$$t[k+1] = \begin{cases} \pi(Df(t[k])_U) & \textit{If } D^2 f(t[k]) > 0, \\ \pi(Df(t[k])_D) & \textit{If } D^2 f(t[k]) < 0. \end{cases}$$

$$t[-(k+1)] = \begin{cases} \pi(Df(t[-k])_D) & \textit{If } D^2 f(t[-k]) > 0, \\ \pi(Df(t[-k])_U) & \textit{If } D^2 f(t[-k]) < 0. \end{cases}$$

Note that orbits of $X(f)$ has no branch nor endpoint, i.e., they are infinite sequences or loops. In other words, $X(f)$ has no "singular points".

Proposition 7 (Integration of affine vector field). *Let* $\{c[i]\} \subset T^2(B^{N-1})$ *be the second derivative of some* $f \in PL$. *Then the orbit* $\{t[i]\} \subset B^{N-1}$ *s.t.* $D^2 t[i] = c[i]$ *is computed as follows:*

$$Sf(t[i+1]) \simeq \begin{cases} (t[i]_U, -|a[i]|_L + 1) & \textit{if } c[i+1] > 0, c[i] > 0, \\ (t[i]_U, -|a[i]|_L) & \textit{if } c[i+1] < 0, c[i] > 0, \\ (t[i]_D, -|a[i]|_L) & \textit{if } c[i+1] > 0, c[i] < 0, \\ (t[i]_D, -|a[i]|_L - 1) & \textit{if } c[i+1] < 0, c[i] < 0. \end{cases}$$

$$t[i+1] = \pi(Sf(t[i+1])),$$

where $Sf(t[i]) = [a[i], g[i]]_T$, $t[i]_U = \pi(Df(t[i])_U)$ *and* $t[i]_D = \pi(Df(t[i])_D)$. *Moreover we obtain* $f = \max_T\{\sum_i a[i]\}$.

2.3.3. Polynomial representation of affine loops Lastly we show that an affine loop is uniquely represented by an element of PL.

Definition 11 (Defining polynomial of an orbit). *Let $Obt = \{t[i] : i \in [l_0, l_1]\} \subset B^{N-1}$ be an orbit of $X(f)$ ($f \in PL$). We set*

$$p(Obt) := \max_T \left\{ \sum_{l_0 \le i \le l_1} a(t[i], |Sf|(t[i])) \right\} \in PL$$

and call it the defining polynomial *of the orbit Obt.*

Theorem 1 (Polynomial representation of affine loops). *Let $Obt, Obt' \in ALP^{N-1}$. Then $p(Obt)/p(Obt') \in H$ if and only if $Obt = Obt'$.*

We denote the unique affine loop corresponds to $f \in PL$ by $obt(f)$.

We need some preparation to prove the theorem.

Definition 12 (Frame of an orbit). *Let $Obt = \{t[i] : i \in [l_0, l_1]\} \subset B^{N-1}$ be an orbit of $X(f)$ ($f \in PL$) and set $Sf(t[i]) = [a[i], g[i]]_T$. Let $\{i_0, i_1, \ldots, i_{m-1}\}$ be a maximal subset of $[l_0, l_1]$ s.t.*

$$\{a[i_j] : 0 \le j < m\} \subset p(Obt),$$
$$a[i_j] \ne a[i_{j+1 \bmod m}] \ (0 \le j < m).$$

We set $fr(Obt) := \{a[i_j] : 0 \le j < m\} \subset L$ and call it the frame *of the orbit Obt. The frame $fr(Obt)$ gives the order of terms of $p(Obt)$ without repeat, in which the orbit goes by.*

Definition 13 (Connectable pair). *Let $c, c' \in L$. We call c and c' are* connectable *if $\exists Obt \in ALP^{N-1}$ s.t. $fr(Obt)$ contains the consecutive pair of c and c'. By definition, any consecutive pair $(c[k], c[k+1])$ in $fr(Obt)$ is connectable for $\forall Obt \in ALP^{N-1}$.*

A sequence $\{c[i] : 0 \le i < m\} \subset L$ is a connectable loop *if every consecutive pair and $(c[m-1], c[0])$ are connectable. Let $lp = \{c[i] : 0 \le i < m\} \subset L$ be a connectable loop. We call lp* minimal *if $c[i]$ and $c[j]$ are not connectable for $i \ne j \pm 1 \bmod m$.*

For $\forall Obt \in ALP^{N-1}$, $fr(Obt)$ is a connectable loop but it may not be minimal.

It follows immediately that, for every consecutive triplet in a connectable loop, there exists a unique affine loop whose frame includes the triplet. That is, every consecutive triplet determines an affine loop uniquely.

Remark 2. Let $(a, g) \in L \times S_N$, where $g = (y_0 y_1 \cdots y_{N-1})$. Let $k = m + M(N-1) \in \mathbf{Z}$, where $m, M \in \mathbf{Z}$ s.t. $0 \le m < N$. Set

$$UP^k(a, g) := (a/(y_{N-2} y_{N-3} \cdots y_{N-m-1}))\,(y_{N-1}/e)^M.$$

Then the definition of connectable pair is explicitly given by

$$\begin{aligned}&\exists[a,g]_T \in S \text{ and } k,k' > 0 \text{ s.t.}\\ &\qquad c = UP^k(a,g),\\ &\qquad c' = UP^{k'}(a(y_0/y_{N-1}),(y_{N-1}y_1\cdots y_{N-2}y_0)).\end{aligned}$$

Let $Obt = \{t[i]\} \subset B^{N-1}$ be an orbit of $X(f)$ $(f \in PL)$ and $Sf(t[0])$ $= [a,g]_T$, where $g = (y_0y_1\cdots y_{N-1})$. Then we have

(Case1) If $D^2f(t[0]) = +1, D^2f(t[1]) = -1$, then

$$Sf(t[1]) = [a,(y_0y_1\cdots y_{N-3}y_{N-1}y_{N-2})]_T\,.$$

(Case2) If $D^2f(t[0]) = -1, D^2f(t[1]) = +1$, then

$$Sf(t[1]) = [a(y_0/y_{N-1}),(y_{N-1}y_1\cdots y_{N-3}y_{N-2}y_0)]_T\,.$$

(Case3) If $D^2f(t[i]) = -1$ $(0 \le i < k)$ for $k > 0$, then

$$\begin{aligned}Sf(t[k-1]) = \big[&ay_0y_1\cdots y_{m-1}(e/y_{N-1})^M,\\ &(y_my_{m+1}\cdots y_{m-1}y_{N-1})\big]_T\,.\end{aligned}$$

(Case4) If $D^2f(t[i]) = +1$ $(0 \le i < k)$ for $k > 0$, then

$$\begin{aligned}Sf(t[k-1]) = \big[&(a/(y_{N-2}y_{N-3}\cdots y_{N-m-1}))(y_{N-1}/e)^M,\\ &(y_{N-m-1}y_{N-1}\cdots y_{N-m-2}y_{N-1})\big]_T\,.\end{aligned}$$

In the case of Case3 (resp. Case4) the orbit flows from $t[0]$ to $t[k-1]$ along the vector $\pi_{\mathbf{R}}(e/y_{N-1})$ (resp. $\pi_{\mathbf{R}}(y_{N-1}/e)$).

Lemma 2. *For any minimal connectable loop $lp \subset L$, there exists a unique affine loop $Obt \in ALP^{N-1}$ s.t. $fr(Obt) = lp$.*

Proof. Let $lp = \{c[i] : 0 \le i < m\} \subset L$. As noted above, there exist N vectors $\pi_{\mathbf{R}}(e/x_i)$ $(0 \le i < N)$ of $(L/H)_{\mathbf{R}}$, along one of which any orbit flows. The vectors induce a partition of $(L/H)_{\mathbf{R}}$ at $\forall b \in (L/H)_{\mathbf{R}}$:

$$\begin{aligned}&U\left(b;\pi_{\mathbf{R}}(e/x_i),\pi_{\mathbf{R}}(e/x_j)\right)\\ &:= \left\{\pi_{\mathbf{R}}\left(b\prod_{0\le k<N} {x_k}^{s_k}\right) : s_k \in \mathbf{R} \text{ s.t. } s_k \ge 0\ (\forall k \ne i,j), s_i, s_j \le 0\right\}.\end{aligned}$$

By an elementary calculation, we obtain

$$(L/H)_{\mathbf{R}} = \sum_{i\ne j} U\left(b;\pi_{\mathbf{R}}(e/x_i),\pi_{\mathbf{R}}(e/x_j)\right) \quad \text{(disjoint union)}.$$

Let $k \in \mathbf{Z}$ s.t. $0 \le k < m$. We denote the direction of the orbit from $c[k]$ to $c[k{+}1 \bmod m]$ by $dir[k]_+$ and the direction from $c[k]$ to $c[k{-}1 \bmod m]$

by $dir[k]_-$: $\exists x_j, x_{j'}$ s.t. $dir[k]_+ = \pi_{\mathbf{R}}(e/x_j)$ and $dir[k]_- = \pi_{\mathbf{R}}(e/x_{j'})$. In the following we abbreviate $k \pm 1 \bmod m$ to $k \pm 1$.

First we show that

$$\{\pi_{\mathbf{R}}(c[i]) : 0 \leq i < m\} \subset U(\pi_{\mathbf{R}}(c[k]); dir[k]_+, dir[k]_-).$$

Suppose that there exists $i \in [0, m)$ s.t. $\pi_{\mathbf{R}}(c[i]) \notin U(\pi_{\mathbf{R}}(c[k]); dir[k]_+, dir[k]_-)$. We may assume $\pi_{\mathbf{R}}(c[i]) \in U(\pi_{\mathbf{R}}(c[k]); dir[k]_+, \pi_{\mathbf{R}}(e/x_{j''}))$. Since e/x_j, $e/x_{j'}$ and $e/x_{j''}$ are independent as vectors of $\mathbf{R}^N$, it implies that $\exists i' \in [0, m)$ s.t. $i' \not\equiv k \pm 1 \bmod m$ and $c[i']$ and $c[k]$ are connectable. Since the loop is minimal, it is a contradiction.

Let $Obt \subset B^{N-1}$ be the unique orbit determined by the triplet $\{c[k-1], c[k], c[k+1]\}$. We show that $fr(Obt) = lp$. Let $fr(Obt) = \{c'[i]\} \subset L$ s.t. $c'[0] = c[k-1]$, $c'[1] = c[k]$ and $c'[2] = c[k+1]$. By the above discussion,

$$\{\pi_{\mathbf{R}}(c'[i]) : i = 0, 1, 2\} \subset U(\pi_{\mathbf{R}}(c'[2]); dir[k+1]_+, dir[k+1]_-).$$

It implies that Obt flows into some tile $\pi([c'[2], g]_T)$ along $1/dir[k+1]_-$ and leaves the tile along $dir[k+1]_+$, which leads to another tile $\pi([c[k+2], g']_T)$. That is, $c'[3] = c[k+2]$. Continuing the process, we obtain the claim. The lemma follows immediately. □

Now we give the proof of theorem 1.

Proof (of theorem 1). It is enough to show the 'only if' part. Let $fr(Obt) = \{c[i] : 0 \leq i < m\}$ and $fr(Obt') = \{c'[i] : 0 \leq i < m'\}$. For any pair $(c'[k], c'[k+1 \bmod m'])$ of $fr(Obt')$, $\exists k_0, k_1 \in \mathbf{Z}$ s.t.

$$k_0 < k_1, c[k_0] = c'[k] \text{ and } c[k_1] = c'[k+1 \bmod m'].$$

Therefore $\{c[i] : k_0 \leq i \leq k_1\}$ is a connectable loop. In particular, there exists a minimal connectable loop $lp = \{c[i] : i \in I\}$, where $I \subset [k_0, k_1]$. We may suppose that $lp \neq fr(Obt)$.

By lemma 2 there exists $Obt'' \in ALP^{N-1}$ s.t. $fr(Obt'') = lp$. On the other hand, $fr(Obt)$ and $fr(Obt'')$ share a consecutive triplet. Therefore $Obt'' = Obt$, which is a contradiction. □

Example 2. V_0, V_1, V_2 and V_3 given in section 1.2 are 2-dimensional affine loops. Their defining polynomials are $p(V_i) = f_i$ $(0 \leq i < 4)$.

3. Heterological algebra

3.1. Affine hetero numbers and their algebra

By theorem 1 an affine loop is identified with its defining polynomial. On the other hand, an affine vector field may have several affine loops. For

$f \in PL$ and a finite subset V of ALP^{N-1}, set

$$V(f) := \{Obt \in ALP^{N-1}; p(Obt) \subset f\} \subset ALP^{N-1},$$
$$p(V) := \max_T \left\{ \sum_{Obt \in V} p(Obt) \right\} \in PL.$$

$V(f)$ consists of finite loops and $p(V(0)) = 0$.

Definition 14 (Affine hetero numbers). *$f \in PL$ is an $N-1$-dimensional* affine hetero number *if $p(V(f)) = f$. We denote the set of all $N-1$-dimensional affine hetero numbers by $\mathbf{AHN}^{N-1}$. Let $f \in \mathbf{AHN}^{N-1}$. Then $V(f) \neq \emptyset$ if and only if $f \neq 0$.*

An affine hetero number is prime *if $V(f)$ consists of a single loop. The set of all primes is identified with ALP^{N-1}.*

Since $V(f)$ ($f \in PL$) is decomposed into affine loops uniquely, we can introduce an algebraic structure in $\mathbf{AHN}^{N-1}$ as follows.

Definition 15 (Algebra of $\mathbf{AHN}^{N-1}$). *Let $f_0, f_1, f = \sum_i a_i \in \mathbf{AHN}^{N-1}$ and $f' = \sum_i a'_i \in PL$.*

(1) Addition $\oplus$:

$$f_0 \oplus f_1 := \begin{cases} f & \text{if } V(f) = V(f_0) + V(f_1), \\ \text{undefined} & \text{else}. \end{cases}$$

(2) (Scalar) multiplication $\otimes$:

$$f' \otimes f := \max_T \left\{ \sum_i \sum_j a'_i a_j \right\}.$$

Example 3 (Embedding of $\mathbf{N}$ into $\mathbf{AHN}^{N-1}$). Let $c = \sum_i e/x_i \in \mathbf{AHN}^{N-1}$. If $g \in \mathbf{AHN}^{N-1}$, then $g \otimes c \in \mathbf{AHN}^{N-1}$ and we call it the *successor* of g. We define an embedding $\Delta : \mathbf{N} \to \mathbf{AHN}^{N-1}$ by

$$\Delta(0) := 1,$$
$$\Delta(n+1) := \Delta(n) \otimes c.$$

Then we have $\Delta(n+m) = \Delta(n) \otimes \Delta(m)$ and

$$\Delta(n) = a_0 c \oplus a_1 c \oplus \cdots \oplus a_{k-1} c,$$

where $\sum_i a_i = \Delta(n-1)$ $(a_i \in L)$. $\Delta(n)$ consists of k copies of c and every instance of c is distinguished from each other by its “label” a_i.(heterogeneity) If we consider the case of $N = 2$, $\Delta(n)$ will consist of n copies of c and

$$\Delta(nm) = \oplus_{0 \leq i < n} x_0^{mi} x_1^{m(n-i-1)} \Delta(m).$$

The next proposition follows immediately.

Proposition 8 (Prime factoring of an affine hetero number). *Let* $f \in \mathbf{AHN}^{N-1}$ *and* $V(f) = \{Obt_0, Obt_1, \ldots, Obt_{m-1}\}$*. Then*

$$f = p(Obt_0) \oplus p(Obt_1) \oplus \cdots \oplus p(Obt_{m-1}),$$

which we call the $\mathrm{prime\ factoring}$ *of* f*.*

3.2. Prehetero numbers

An affine hetero number is defined as a set of closed orbits of an affine vector field. Here we define a cluster of affine hetero numbers "algebraically". In the next section we will use the cluster structure to define "interaction" among affine hetero numbers. Note that it is possible to define affine hetero numbers as factors of the clusters.

3.2.1. Join operator We give a "multi-dimensional version" of the least common multiple of $f \in PL$, which we call the *join* of f.

Definition 16 (Join on PL). *Let* $f \in PL$*.*

(1) $\vee_T f := \max_T \left\{ g \in PL : \exists K \geq 0 \text{ s.t. } f \geq_T \left(\sum_{0 \leq i < N} x_i\right)^K g \right\}$.

(2) $\vee_C f := \max_C \left\{ g \in PL : \exists K \geq 0 \text{ s.t. } f \geq_C \left(\sum_{0 \leq i < N} (e/x_i)\right)^K g \right\}$.

For a finite subset $\{l_0, l_1, \ldots, l_{n-1}\} \subset$ **Z**, let $\max^{(2)}_{0 \leq i < n}\{l_i\}$ be the second element in decreasing order, i.e., $\max^{(2)}_{0 \leq i < n}\{l_i\} := l_{\sigma(1)}$, where $\sigma \in S_N$ s.t. $l_{\sigma(0)} \geq l_{\sigma(1)} \geq \ldots \geq l_{\sigma(n-1)}$.

Lemma 3. *Let* $f = \sum_{0 \leq i < n} a_i \in PL$*, where* $a_i = \prod_{0 \leq j < N} {x_j}^{l_{ij}}$ $(0 \leq i < n)$*. Set* $m_{ij} = \sum_{0 \leq k < N} l_{ik} - l_{ij}$ $(0 \leq i, j < n)$*.*

(1) $\vee_T f$ *is explicitly given by*

$$\vee_T f = \begin{cases} \max_T\{f\} & \text{if } n < N, \\ \max_T \left\{ f + \prod_{0 \leq j < n} {x_j}^{\max^{(2)}_{0 \leq i < n}\{l_{ij}\}} \right\} & \text{if } n = N, \\ \max_T \left\{ f + \sum_\rho \vee_T \left(\sum_{0 \leq i < N} a_{\rho(i)} \right) \right\} & \text{if } n > N. \end{cases}$$

(2) $\vee_C f$ *is explicitly given by*

$$\vee_C f = \begin{cases} \max_C\{f\} & \text{if } n < N, \\ \max_C \left\{ f + \prod_{0 \leq j < n} (e/x_j)^{\max^{(2)}_{0 \leq i < n}\{m_{ij}\}} \right\} & \text{if } n = N, \\ \max_C \left\{ f + \sum_\rho \vee_C \left(\sum_{0 \leq i < N} a_{\rho(i)} \right) \right\} & \text{if } n > N. \end{cases}$$

$\sum_\rho$ *is the sum over the combination* ρ *of* n *objects taken* N *at a time.*

Definition 17 (Base of $f \in PL$). *For* $f \in PL$*, we define*

$$f_b := \max_T \{g \in PL : g \subset L^* \text{ and } g \leq_T f\} \in PL.$$

Note that $f_b \subset L^*$.

Lemma 4. *Let* $f \in PL$*. Then*

$$f_b = \max_T \left\{ f_{(0)} + \left(\sum_{0 \leq i < N} x_i \right)^{N-2} f_{(1)} + \left(\sum_{0 \leq i < N} x_i \right)^{N-3} f_{(2)} + \cdots + \left(\sum_{0 \leq i < N} x_i \right) f_{(N-2)} \right\},$$

where $f_{(k)} := \{a \in f : |a|_L \equiv k \bmod N-1\}$ $(0 \leq k < N-1)$.

3.2.2. Prehetero numbers First we give the definition of a decomposition of B^{N-1} along an affine vector field. A prehetero number is defined as an element of PL which induces an affine vector field with no “boundary”.

Definition 18 (Decomposition of B^{N-1}). *For* $f, f' \in PL$*, set*

$$\begin{aligned}
Out(f,f') &:= \left\{ [b,g]_B \in B^{N-1} : h_C[f', b_i] \leq h_T[f, b_i] \ (0 \leq i < N) \right\}, \\
Bd(f,f') &:= \left\{ [b,g]_B \in B^{N-1} : \begin{cases} h_C[f', b_0] > h_T[f, b_0], \\ h_C[f', b_{N-1}] < h_T[f, b_{N-1}]. \end{cases} \right. \\
&\qquad\qquad \left. \text{or} \quad \begin{cases} h_C[f', b_0] < h_T[f, b_0], \\ h_C[f', b_{N-1}] > h_T[f, b_{N-1}]. \end{cases} \right\}, \\
In(f,f') &:= B^{N-1} - Out(f,f') - Bd(f,f'),
\end{aligned}$$

where $b_0 = b$ *and* $b_i = \pi_{\mathbf{R}}(b x_{g(0)} x_{g(1)} \dots x_{g(i-1)})$ $(0 < i < N)$*. For simplicity, we denote* $Out(f, \vee_C f_b)$*,* $Bd(f, \vee_C f_b)$ *and* $In(f, \vee_C f_b)$ *by* $Out(f)$*,* $Bd(f)$ *and* $In(f)$ *respectively. We call*

$$B^{N-1} = In(f) + Bd(f) + Out(f) \quad \text{(disjoint union)}$$

the decomposition *of* B^{N-1} *along the affine vector field induced by* f*.*

Definition 19 (Prehetero numbers). $f \in PL$ *is an* $N-1$*-dimensional* prehetero number *if* $Bd(f) = \emptyset$*. The set of all* $N-1$*-dimensional prehetero numbers is denoted by* $\mathbf{PHN}^{N-1}$.

Note that $V(f) \subset In(f)$ *for* $\forall f \in \mathbf{PHN}^{N-1}$.

Example 4.

(1) Polynomials given in section 1.2 are all prehetero numbers.
(2) Let $N = 3$ and set $x = x_0,\ y = x_1, z = x_2$. Let

$$f = y^3z + x^3y^2z^{-1} + x^4 + x^2y^2z^2 \in PL.$$

Then $f_b = f$, $\vee_C f_b = x^2yz^{-1} + xy^2z$ and

$$Bd(f) = \{[x^{-1}y, (xyz)]_B, [x^2y^{-1}z^{-1}, (yxz)]_B, [x^2z^{-2}, (zxy)]_B, [yz^{-1}, (zyx)]_B\} \subset B^2.$$

Therefore f is not any prehetero number.

3.2.3. Loop decomposition of a prehetero number Now we show that $\mathbf{PHN}^{N-1} \subset \mathbf{AHN}^{N-1}$. In particular, any prehetero number is factored into prime affine hetero numbers. For example, a rhombic dodecahedron is divided into parts as we will see in example 5 and 7.

Theorem 2 (Loop decomposition of a prehetero number). $\mathbf{PHN}^{N-1} \subset \mathbf{AHN}^{N-1}$. *Moreover, for* $f \in \mathbf{PHN}^{N-1}$*, there exists a unique set of primes* $f_i \in \mathbf{AHN}^{N-1}$ $(0 \le i < m)$ *s.t.*

$$f = f_0 \oplus f_1 \oplus \cdots f_{m-1}.$$

Remark 3. $\mathbf{PHN}^{N-1} = \mathbf{AHN}^{N-1}$ if and only if $N = 3$.

First we show a lemma.

Lemma 5. *For* $f \in PL$*, the number* $|In(f)|$ *of elements of* $In(f)$ *is finite.*

Proof. For $k \in \mathbf{R}$ s.t. $k > 0$, set

$$L_k = \left\{ \prod_{0 \le j < N} (e/x_j)^{l_j} : \max_{0 \le j < N}\{l_j\} < k, \min_{0 \le j < N}\{l_j\} = 0 \right\} \subset L,$$
$$B_k = \{t = [b_0, b_1, \ldots, b_{N-1}] : b_j \in \pi(L_k)\ (0 \le j < N)\} \subset B^{N-1}.$$

By definition, $\exists K > 0$ s.t. $f_b \ge_C \left(\sum_{0 \le i < N}(e/x_i)\right)^K \vee_C f_b$. Choose sufficiently large $k \in \mathbf{R}$ s.t. $\pi\left(\left(\sum_{0 \le i < N}(e/x_i)\right)^K \vee_C f_b\right) \subset \pi(L_k)$.

First we show that, for $b \in (L/H)\backslash\pi(L_k)$,

$$h_C\left[\left(\sum_{0 \le i < N}(e/x_i)\right)^K \vee_C f_b, b\right] = h_C[\vee_C f_b, b]. \qquad (1)$$

Let $b \in L/H$ and $a \in \vee_C f_b$ s.t. $h_C[\vee_C f_b, b] = h_C[a, b]$. Set $m = -h_C[a, b]/N$. It follows immediately that

$$\exists d_j \geq 0 \text{ s.t. } \begin{cases} be^m = a \prod_{0 \leq j < N} (e/x_j)^{d_j}, \\ \min_{0 \leq j < N}\{d_j\} = 0. \end{cases}$$

Suppose that $b \in (L/H) \backslash \pi(L_k)$. Since $\pi\left(a * (\sum_j (e/x_j))^K\right) \subset \pi(L_k)$,

$$\exists d'_j, d''_j \geq 0 \text{ s.t. } \begin{cases} be^m = a \prod_{0 \leq j < N} (e/x_j)^{d'_j} \prod_{0 \leq j < N} (e/x_j)^{d''_j}, \\ \sum_{0 \leq j < N}\{d'_j\} = K, \\ d_j = d'_j + d''_j \quad (0 \leq j < N). \end{cases}$$

Note that $\min_{0 \leq j < N}\{d''_j\} \leq \min_{0 \leq j < N}\{d_j\} = 0$. Therefore

$$h_C\left[a \prod_{0 \leq j < N} (e/x_j)^{d'_j}, b\right] = h_C[\vee_C f_b, b].$$

Since $\left(\sum_{0 \leq i < N} (e/x_i)\right)^K \vee_C f_b \geq_C a \prod_{0 \leq j < N} (e/x_j)^{d'_j}$,

$$h_C\left[\left(\sum_{0 \leq i < N} (e/x_i)\right)^K \vee_C f_b, b\right] \geq h_C[\vee_C f_b, b].$$

On the other hand, $h_C\left[\left(\sum_{0 \leq i < N} (e/x_i)\right)^K \vee_C f_b, b\right] \leq h_C[\vee_C f_b, b]$ for $b \in L/H$. Thus (1) follows.

Next we show that $B^{N-1} \backslash B_k \subset Out(f)$. Note that

$$f_b \geq_T \left(\sum_{0 \leq i < N} (e/x_i)\right)^K \vee_C f_b.$$

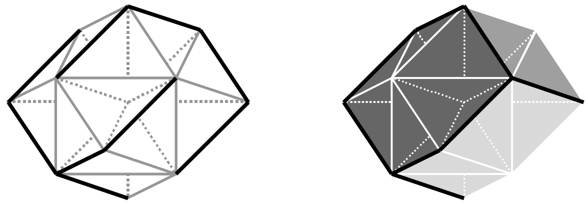
Since $f \geq_T f_b$, it follows that

$$h_T[f, b] \geq h_C\left[\left(\sum_{0 \leq i < N} (e/x_i)\right)^K \vee_C f_b, b\right]$$

Therefore $h_T[f, b] \geq h_C[\vee_C f_b, b]$ for $b \in (L/H) \backslash \pi(L_k)$ by (1). In particular, for $t = [b_0, b_1, \ldots, b_{N-1}] \in B^{N-1} \backslash B_k$,

$$h_T[f, b_j] \geq h_C[\vee_C(f_b), b_j] \; (0 \leq j < N).$$

Hence $B^{N-1} \backslash B_k \subset Out(f)$. That is, $In(f) \subset B_k$. It implies that $|In(f)|$ is finite. □



(a) (b)

Fig. 3. Factoring of a dodecahedron (1)

Proof (of theorem 2). Let $f \in \mathbf{PHN}^{N-1}$ and $Obt \subset B^{N-1}$ be an orbit of $X(f)$. It follows readily that $Obt \subset V(f)$ if $\exists t \in Obt$ s.t. $t \in V(f)$. And Obt is closed if $Obt \subset V(f)$, since $V(f)$ is a finite set by lemma 5.

Let $t \in V(f)$. Then there exists an orbit Obt of $X(f)$ s.t. $t \in Obt$. As noted above, $Obt \subset V(f)$ and $Obt \in ALP^{N-1}$. Since $V(f)$ is covered by a finite set of closed orbits, the theorem follows. □

Example 5. Let $N = 4$. Set $x = x_0$, $y = x_1$, $z = x_2$ and $w = x_3$.

(1) $f = xyw + xz + yz + wz \in \mathbf{PHN}^3$. f is prime and its orbit $obt(f)$ gives a coding of a rhombic dodecahedron (Fig.3(a)).

(2) $f = xyz+xyw+xzw+yzw \in \mathbf{PHN}^3$. f is factored into 4 primes: $f = g_0 \oplus g_1 \oplus g_2 \oplus g_3$, where $g_0 = xyz+xyw+xzw$, $g_1 = xyz+xyw+yzw$, $g_2 = xyz + xzw + yzw$ and $g_3 = xyw + xzw + yzw$ (Fig.3(b)).

3.3. Algebra of prehetero numbers

In section 3.1 we have defined an algebraic structure in $\mathbf{AHN}^{N-1}$. Here we define "interaction" between affine hetero numbers via action of PL. The "interaction" induces the notion of "catalyst" naturally.

3.3.1. Decomposition at infinity First we define orders "at infinity".

Definition 20 (Orders at infinity). *Let $a_0, a_1 \in L$.*

(1) Order T_∞ and its height function $h_T^\infty : L \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$a_0 \geq_T^\infty a_1 \overset{\text{def}}{\Longleftrightarrow} 1/a_0 \geq_T 1/a_1,$$
$$h_T^\infty[a,b] := -N \min\left\{h \in \mathbf{Q} : a \geq_T^\infty b(1/e)^h\right\}.$$

(2) Order C_∞ and its height function $h_C^\infty : L \times (L/H)_{\mathbf{R}} \to \mathbf{Z}$.

$$a_0 \geq_C^\infty a_1 \overset{\text{def}}{\Longleftrightarrow} 1/a_0 \geq_C 1/a_1,$$
$$h_C^\infty[a,b] := -N \min\left\{h \in \mathbf{Q} : a \geq_C^\infty b(1/e)^h\right\}.$$

The orders are extended to orders in PL. We also define $\max_{T^\infty}\{U\}$ and $\max_{C^\infty}\{U\}$ $(U \subset PL)$ as in the case of T and C.

Let $\hat{f} := \sum_{0\le i<n} 1/a_i$ for $f = \sum_{0\le i<n} a_i \in PL$. Then $f \ge_A^\infty g \iff \hat{f} \ge_A \hat{g}$ and $h_A^\infty[f,b] = h_A[\hat{f}, 1/b]$, where A stands for T or C.

Definition 21. *Let $f \in PL$ and $t = [b_0, b_1, \ldots, b_{N-1}] \in B^{N-1}$.*

(1) Polynomial section of $S(B^{N-1})$ at infinity.

$$|S_\infty f|(t) := \max\{h_T^\infty[f, b_j] : 0 \le j < N\} \in \mathbf{Z},$$
$$S_\infty f(t) := [a(t, |S_\infty f|(t)), g(t, |S_\infty f|(t))]_T \in \mathcal{O}_S(t).$$

(2) Polynomial section of $S^*(B^{N-1})$ at infinity.

$$|S_\infty^* f|(t) := \max\{h_C^\infty[f, b_j] : 0 \le j < N\} \in \mathbf{Z},$$
$$S_\infty^* f(t) := [a(t, |S_\infty^* f|(t)), g(t, |S_\infty^* f|(t))]_C \in \mathcal{O}_{S^*}(t).$$

"At infinity" we define a "multi-dimensional version" of the greatest common divisor of $f \in PL$. We call it the *meet* of f.

Definition 22 (Meet on PL). *Let $f \in PL$.*

(1) $\wedge_T f := \max_{T_\infty}\left\{g \in PL : \exists K \ge 0 \; f \ge_T^\infty \left(\sum_{0\le i<N}(1/x_i)\right)^K g\right\}.$

(2) $\wedge_C f := \max_{C_\infty}\left\{g \in PL : \exists K \ge 0 \; f \ge_C^\infty \left(\sum_{0\le i<N}(x_i/e)\right)^K g\right\}.$

Lemma 6. *Let $f = \sum_{0\le i<n} a_i$, where $a_i = \prod_{0\le j<N}(1/x_j)^{l_{ij}}$ $(0 \le i < n)$. Set $m_{ij} = \sum_{0\le k<N} l_{ik} - l_{ij}$ $(0 \le i, j < n)$.*

(1) $\wedge_T f$ is explicitly given by

$$\wedge_T f = \begin{cases} \max_{T^\infty}\{f\} & \text{if } n < N, \\ \max_{T^\infty}\left\{f + \prod_{0\le j<n}(1/x_j)^{\max^{(2)}_{0\le i<n}\{l_{ij}\}}\right\} & \text{if } n = N, \\ \max_{T^\infty}\left\{f + \sum_\rho \wedge_T\left(\sum_{0\le i<N} a_{\rho(i)}\right)\right\} & \text{if } n > N. \end{cases}$$

(2) $\wedge_C f$ is explicitly given by

$$\wedge_C f = \begin{cases} \max_{C^\infty}\{f\} & \text{if } n < N, \\ \max_{C^\infty}\left\{f + \prod_{0\le j<n}(x_j/e)^{\max^{(2)}_{0\le i<n}\{m_{ij}\}}\right\} & \text{if } n = N, \\ \max_{C^\infty}\left\{f + \sum_\rho \wedge_C\left(\sum_{0\le i<N} a_{\rho(i)}\right)\right\} & \text{if } n > N. \end{cases}$$

$\sum_\rho$ is the sum over the combination ρ of n objects taken N at a time.

Let $f \in PL$. We define subsets of L by

$$in(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] < h_C[\vee_C f_b, b] \right\},$$

$$out(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] > h_C[\vee_C f_b, b] \right\},$$

$$bd(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] = h_C[\vee_C f_b, b] \right\}.$$

Since $|In(f)|$ is finite, $in(f)$ is also a finite set.

Definition 23 (Base of $f \in PL$ at infinity). *For $f \in PL$, we define*

$$f^b := \max_{C^\infty} \left\{ g \in U : g \leq_C^\infty g' \textit{ for } \forall g' \in U \right\} \in PL$$

where $U = \{g \in PL : g \subset L^*, g \geq_C^\infty in(f) + bd(f)\}$. Note that $f^b \subset L^*$.

Let $t \in B^{N-1}$ and $Sf(t) = [a_0, a_1, a_2, \ldots, a_{N-1}] \in \mathcal{O}_S(t)$. Set

$$k_{max}(t) := \max \left\{ k : \begin{cases} [a(t,k), g(t,k)]_C = [a'_0, a'_1, \ldots, a'_{N-1}], \\ a(t,k) \in L^*,\ a'_i \geq_C^\infty a_i \text{ for } a_i \in in(f), \\ \pi_{\mathbf{R}}(a'_i) = \pi_{\mathbf{R}}(a_i)\ (0 \leq i < N). \end{cases} \right\}$$

If $in(f) \neq \emptyset$, then f^b is explicitly given by the following.

Lemma 7. *Let $f \in PL$. Then*

$$f^b = \max_{C^\infty} \left\{ \sum_{t \in In(f) + Bd(f)} a(t, k_{max}(t) - (N-1)(N-1)) \right\}.$$

Finally we give the definition of "decomposition of B^{N-1} at infinity" along an affine vector field.

Definition 24 (Decomposition of B^{N-1} at infinity). *For $f, f' \in PL$, set*

$$In_\infty(f,f') := \left\{ [b,g]_B \in B^{N-1} : h_C^\infty[f', b_i] \leq h_T[f, b_i]\ (0 \leq i < N) \right\},$$

$$Bd_\infty(f,f') := \left\{ [b,g]_B \in B^{N-1} : \begin{cases} h_C^\infty[f', b_0] > h_T[f, b_0], \\ h_C^\infty[f', b_{N-1}] < h_T[f, b_{N-1}]. \end{cases} \textit{or} \quad \begin{cases} h_C^\infty[f', b_0] < h_T[f, b_0], \\ h_C^\infty[f', b_{N-1}] > h_T[f, b_{N-1}]. \end{cases} \right\},$$

$$Out_\infty(f,f') := B^{N-1} - In_\infty(f,f') - Bd_\infty(f,f'),$$

where $b_0 = b$ and $b_i = \pi_{\mathbf{R}}(bx_{g(0)}x_{g(1)} \cdots x_{g(i-1)})$ $(0 < i < N)$. We denote $Out_\infty\left(f, \wedge_C f^b\right)$, $Bd_\infty\left(f, \wedge_C f^b\right)$ and $In_\infty\left(f, \wedge_C f^b\right)$ by $Out_\infty(f)$, $Bd_\infty(f)$ and $In_\infty(f)$ respectively. We call

$$B^{N-1} = In_\infty(f) + Bd_\infty(f) + Out_\infty(f) \quad \textit{(disjoint union)}$$

the decomposition *of* B^{N-1} at infinity *along* $X(f)$.

3.3.2. Relation of the decompositions The purpose of this section is to show the following proposition.

Proposition 9. *Let $f \in PL$.*

(1) $In_\infty(f) = In(f) + Bd(f)$,
$Out_\infty(f) + Bd_\infty(f) = Out(f)$.
(2) $Bd(f) = \emptyset$ *if and only if* $Bd_\infty(f) = \emptyset$.

We obtain two lemmas immediately from definitions.

Lemma 8. *Let $t \in B^{N-1}$, $f \in PL$ and $(a, g) \in L \times S_N$. Set $a_0 = a$ and $a_j = ax_{g(0)}x_{g(1)} \dots x_{g(j-1)}$ $(0 < j < N)$.*

(1) Let $Sf(t) = [a, g]_T$. Then

$$h_T[f, \pi_{\mathbf{R}}(a_k)] = |Sf|(t) - k \quad (0 \le k < N).$$

(2) Let $S^*f(t) = [a, g]_C$ and $f \subset L^*$. Then

$$h_C[f, \pi_{\mathbf{R}}(a_{N-k})] = |S^*f|(t) - (N-1)k \quad (0 \le k < N).$$

(3) Let $S_\infty f(t) = [a, g]_T$. Then

$$h_T^\infty[f, \pi_{\mathbf{R}}(a_k)] = |S_\infty f|(t) - k \quad (0 \le k < N).$$

(4) Let $S_\infty^* f(t) = [a, g]_C$ and $f \subset L^*$. Then

$$h_C^\infty[f, \pi_{\mathbf{R}}(a_{N-k})] = |S_\infty^* f|(t) - (N-1)k \quad (0 \le k < N).$$

Lemma 9. *Let $t \in B^{N-1}$ and $f, g \in PL$. Let $s = [a_0, a_1, \dots, a_{N-1}]_T \in \mathcal{O}_S(t)$ and $s' = [a'_0, a'_1, \dots, a'_{N-1}]_C \in \mathcal{O}_{S^*}(t)$. If there exist a_i, a'_j s.t. $h_T[f, \pi_{\mathbf{R}}(a_i)] = h_C[g, \pi_{\mathbf{R}}(a'_j)]$, then $\pi_{\mathbf{R}}(a_i) = \pi_{\mathbf{R}}(a'_j)$.*

Using the lemmas, we have the followings.

Proposition 10. *Let $f, f' \in PL$ and $f' \subset L^*$.*

(1) $Out(f, f') = \{t : |S^*f'|(t) \le |Sf|(t)\}$.
(2) $$Bd(f, f') = \left\{ [b, g]_B : \begin{cases} h_C[f', b] < h_T[f, b], \\ h_C[f', \pi_{\mathbf{R}}(b(e/x_{g(N-1)}))] \\ \quad > h_T[f, \pi_{\mathbf{R}}(b(e/x_{g(N-1)}))]. \end{cases} \right\}.$$

Proof. Let $t \in B^{N-1}$ and set $Sf(t) = [a, g]_T \in \mathcal{O}_S(t)$, $S^*f(t) = [a', g']_C \in \mathcal{O}_{S^*}(t)$. Let $b_0 = \pi_{\mathbf{R}}(a)$, $b_i = \pi_{\mathbf{R}}(ax_{g(0)}x_{g(1)} \dots x_{g(i-1)})$, $b'_0 = \pi_{\mathbf{R}}(a')$ and $b'_i = \pi_{\mathbf{R}}(a'x_{g'(0)}x_{g'(1)} \dots x_{g'(i-1)})$ $(0 < i < N)$.

(1) Since $\pi([a,g]_T) = \pi([a',g']_C)$, there exists $m > 0$ s.t. $b_i = b'_{m+i \mod N}$ $(0 \le i < N)$. By lemma 8, for $i \in [0,N)$,

$$h_C[f', b_i] \le h_T[f, b_i] \Leftrightarrow |Sf|(t) - |S^* f'|(t) \ge i - k_i(N-1),$$

where $k_i \equiv N - m - i \bmod N$ and $0 \le k_i < N$. Note that

$$\max_{0 \le i < N} \{i - k_i(N-1)\} = m * (N-1).$$

Therefore

$$t \in Out(f, f') \Leftrightarrow |Sf|(t) - |S^* f'|(t) \ge m * (N-1).$$

In particular, $|Sf|(t) - |S^* f'|(t) \ge 0$ for $t \in Out(f, f')$.
Next suppose that $|Sf|(t) - |S^* f'|(t) \ge 0$. Since $h_T[f, b_0] \ge h_C[f', b'_0]$, lemma 8 implies that

$$h_T[f, b'_i] \ge h_T[f, b_{N-1}] \ge h_C[f', b'_i] \quad (0 < i < N).$$

Thus it is enough to show that $h_T[f, b'_0] \ge h_C[f', b'_0]$.
If $h_T[f, b_{N-1}] \ge h_C[f', b'_0]$, then $h_T[f, b'_0] \ge h_C[f', b'_0]$.
If $h_T[f, b_{N-1}]\} < h_C[f', b'_0]$, then

$$h_T[f, b_0] \ge h_C[f', b'_0] > h_T[f, b_{N-1}].$$

Since $h_T[f, b_i]$ ranges over N consecutive integers, $\exists b_k$ s.t. $h_T[f, b_k] = h_C[f', b'_0]$. By lemma 9, $b_k = b'_0$, i.e., $h_T[f, b'_0] = h_C[f', b'_0]$.

(2) Suppose that $t \in Bd(f, f')$. By definition, $\exists b'_j$ s.t.

$$h_T[f, b_0] > h_C[f', b'_j] > h_T[f, b_{N-1}].$$

Since $h_T[f, b_i]$ ranges over N consecutive integers, $\exists b_i$ $(0 < i < N-1)$ s.t. $h_T[f, b_i] = h_C[f', b'_j]$. By lemma 9, $b_i = b'_j$. Therefore

$$h_T[f, b_0] > \cdots > h_T[f, b_i] > \cdots > h_T[f, b_{N-1}],$$
$$h_C[f', b_0] < \cdots < h_T[f, b_i] < \cdots < h_C[f', b_{N-1}].$$

Thus $h_C[f', b_0] < h_T[f, b_0]$ and $h_C[f', b_{N-1}] > h_T[f, b_{N-1}]$.
□

Proposition 11. *Let $f, f' \in PL$ and $f' \subset L^*$.*

(1) $In_\infty(f, f') = \{t : |S^*_\infty f'|(t) \le |Sf|(t)\}$.

(2) $$Bd_\infty(f, f') = \left\{ [b,g]_T : \begin{cases} h^\infty_C[f', b] < h_T[f, b], \\ h^\infty_C[f', \pi_{\mathbf{R}}(b(e/x_{g(N-1)}))] \\ \quad > h_T[f, \pi_{\mathbf{R}}(b(e/x_{g(N-1)}))]. \end{cases} \right\}.$$

Proof. The proof is analogous to the proof of proposition 10. □

Lemma 10. *Let* $t \in In(f) + Bd(f)$ *and* $Sf(t) = [a_0, a_1, a_2, \ldots, a_{N-1}]$. *Then* $\exists a_k$ *s.t.* $S^*_\infty f^b(t) = S^* a_k(t)$. *In particular,* $h^\infty_C[f^b, \pi_{\mathbf{R}}(a_i)] \le h_T[f, \pi_{\mathbf{R}}(a_i)]$ $(0 \le i < N)$.

Proof. Since $\{|a_i|_L : 0 \le i < N\}$ ranges over N consecutive integers,

$$\exists a_k \text{ s.t. } |a_k|_L \equiv 0 \bmod (N-1) \text{ and } a_k \neq a_{N-1}.$$

In particular, $a_k \in L^*$. Let $[a(t, k_{max}), g(t, k_{max})]_C = [a'_0, a'_1, \ldots, a'_{N-1}]$. Suppose that $\exists a'_j$ s.t. $|a'_j|_L < |a_k|_L$. Then $|a'_j|_L \le |a_k|_L - (N-1)$ because $a'_j \in L^*$. Since $|a_k|_L - (N-1) < |a_i|_L$ $(0 \le i < N)$, $|a'_j|_L < |a_i|_L$ $(0 \le i < N)$. In particular, $a'_j \not\ge^\infty_C a_i$ $(0 \le i < N)$. By the definition of k_{max}, $\pi_{\mathbf{R}}(a'_j) \notin in(f)$. Therefore $|S^* \vee_C f_b|(t) < -|a'_j|_L$. It implies that $|S^* \vee_C f_b|(t) \le -|a_k|_L$. That is, $t \in Out(f)$, which is a contradiction. Thus $-|a_k|_L \ge -|a'_i|_L$ $(0 \le i < N)$, i.e., $-|a_k|_L \ge k_{max}$.

Let $[a_k, g(t, -|a_k|_L)]_C = [a'_0, a'_1, \ldots, a'_{N-1}]$, where $\pi_{\mathbf{R}}(a'_i) = \pi_{\mathbf{R}}(a_i)$ $(0 \le i < N)$. Then $a'_i \ge^\infty_C a_i$ $(0 \le i < N)$. Therefore $-|a_k|_L \le k_{max}$. That is, $-|a_k|_L = k_{max}$ and $a(t, k_{max}) = a_k$. Since $S^* a_k(t) = [a_k, g(t, -|a_k|_L)]_C$, the result follows. □

Lemma 11. *Let* $f \in PL$ *and* $a \in L^*$ *s.t.* $\vee_C f_b \not\ge_C a$. *Then* $\wedge_C f^b \ge^\infty_C a$ *if and only if* $f^b \ge^\infty_C a$.

Proof. It is enough to show that, if $\wedge_C f^b \ge^\infty_C a$ and $f^b \not\ge^\infty_C a$, then $\vee_C f_b \ge_C a$. By the condition, $\exists K > 0$ s.t. $f^b \ge^\infty_C \left(\sum_{0 \le j < N} (x_j/e)\right)^K a$. Let $k \in [0, N)$. Then $\exists a'_k \in f^b$ and $d'_j(k) \ge 0$ $(0 \le j < N)$ s.t.

$$a'_k \prod_{0 \le j < N} (x_j/e)^{d'_j(k)} = a(x_k/e)^K.$$

That is, $a = a'_k \left(\prod_{0 \le j < N, j \neq k} (x_j/e)^{d'_j(k)}\right) (x_k/e)^{-K + d'_k(k)}$. Since $f^b \not\ge^\infty_C a$, $K - d'_k(k) > 0$ $(0 \le k < N)$. Set

$$c := a \prod_{0 \le j < N} (x_j/e)^{K - d'_j(j)} \in L^*,$$

$$c_k := a \left(\prod_{0 \le j < N, j \neq k} (x_j/e)^{K - d'_j(j)} \right) (x_k/e)^{-D_k} \in L^*,$$

where $D_k = \min\{d'_k(j) : 0 \le j < N, j \neq k\}$ $(0 \le k < N)$. Since $c(e/x_k)^{D_k + K - d'_k(k)} = c_k$ $(0 \le k < N)$, we obtain

$$\sum_{0 \le k < N} c_k \ge_C \left(\sum_{0 \le j < N} (e/x_j) \right)^{K'} c,$$

where $K' := \max\{D_k + K - d'_k(k) : 0 \leq k < N\} \geq 0$. Note that c_k is on a slope defined by f^b: $\exists t \in B^{N-1}$ s.t. $c_k \subset S^*_\infty f^b(t)$. It follows that $c_k \leq_C in(f)$ $(0 \leq k < N)$. Otherwise $\exists a \in f^b$ s.t. $a \not\leq_C in(f)$, which is a contradiction. Therefore

$$in(f) \geq_C \left(\sum_{0 \leq j < N} (e/x_j) \right)^{K'} c,$$

which implies that $\vee_C f_b \geq_C c$. Since $c \geq_C a$, we obtain $\vee_C f_b \geq_C a$. □

Now we give the proof of proposition 9. Let $f \in PL$. We define subsets of L by

$$in_\infty(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] > h^\infty_C[\wedge_C f^b, b] \right\},$$
$$out_\infty(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] < h^\infty_C[\wedge_C f^b, b] \right\},$$
$$bd_\infty(f) := \left\{ be^{-h_T[f,b]/N} : b \in \pi_{\mathbf{R}}(L) \text{ s.t. } h_T[f,b] = h^\infty_C[\wedge_C f^b, b] \right\}.$$

Proof (of proposition 9).

(1) Let $t \in In(f) + Bd(f)$. Since $\wedge_C f^b \geq^\infty_C f^b$, Lemma 10 implies that $h^\infty_C[\wedge_C f^b, \pi_{\mathbf{R}}(a_i)] \leq h_T[f, \pi_{\mathbf{R}}(a_i)]$ $(0 \leq i < N)$. Thus $In_\infty(f) \supset In(f) + Bd(f)$.

Next we show that $Out(f) \subset Out_\infty(f) + Bd_\infty(f)$. If $In(f) + Bd(f) = \emptyset$, then $Out(f) = Out_\infty(f) + Bd_\infty(f) = B^{N-1}$. Suppose that $In(f) + Bd(f) \neq \emptyset$. Let $t \in Out(f)$ and $Sf(t) = [a,g]_T$. By lemma 11, $\wedge_C f^b \geq^\infty_C c \Leftrightarrow f^b \geq^\infty_C c$ for $\forall c \in out(f)$. Therefore

$$h^\infty_C[\wedge_C f^b, b] = h^\infty_C[f^b, b] \text{ for } \forall b \in \pi_{\mathbf{R}}(out(f)).$$

It follows that $S^*_\infty \wedge_C f^b(t) = S^*_\infty f^b(t)$ and denote the value by $[a', g']_C$. Since $\pi_{\mathbf{R}}(f^b) \subset \pi(In(f) + Bd(f))$, lemma 10 implies that

$$\exists c \in in(f) + bd(f) \text{ s.t. } S^*_\infty c(t) = [a', g']_C.$$

Suppose that $|Sf|(t) \geq |S^*_\infty \wedge_C f^b|(t)$. Then $a \geq_T a'$ and $h_T[a, \pi_{\mathbf{R}}(c)] > h_T[f, \pi_{\mathbf{R}}(c)]$. It is a contradiction since $a \in f$. Thus $|Sf|(t) < |S^*_\infty \wedge_C f^b|(t)$. By proposition 11, $t \in Out_\infty(f) + Bd_\infty(f)$.

(2) First we show that $Bd(f) \neq \emptyset \Rightarrow Bd_\infty(f) \neq \emptyset$.
Let $t \in Bd(f)$ and $Sf(t) = [a,g]_T$. Set $a_0 = a$ and $a_i = a x_{g(0)} x_{g(1)} \cdots x_{g(i-1)}$ $(0 < i < N)$. By proposition 10, $a_0 \in out(f)$ and $a_{N-1} \in in(f)$. Let $S^* \vee_C f_b(t) = [a'_0, a'_1, a'_2, \ldots, a'_{N-1}]$, where $\pi_{\mathbf{R}}(a'_i) = \pi_{\mathbf{R}}(a_i)$ $(0 \leq i < N)$. Then

$$a'_{N-1} \geq_C a'_{N-2} \geq_C \ldots \geq_C a'_0.$$

By lemma 9 and lemma 10, $\exists k$ s.t. $a'_k = a_k \in L^*$ and

$$S^*_\infty f^b(t) = [a'_k, a'_{k-1}, \dots, a'_0, a'_{N-1}e^{N-1}, \dots, a'_{k+1}e^{N-1}].$$

Note that $a_k \in bd(f)$. Thus $0 < k < N-1$. By lemma 8,

$$|a'_0|_L = |a_0|_L + k + (N-1)k < |a_0|_L + N(N-1)$$

We claim that $\exists x_l$ s.t. $a_0(e/x_l) \in out(f)$. Otherwise $a_0(e/x_j) \in in(f)$ $(0 \le j < N)$ since $a_0(e/x_j) \notin L^*$ $(0 \le j < N)$. It follows that $a_0 \in in(f)$, which is a contradiction. Then lemma 11 implies that

$$\begin{aligned} h^\infty_C[\wedge_C f^b, \pi_{\mathbf{R}}(a_0(e/x_l))] &= h^\infty_C[f^b, \pi_{\mathbf{R}}(a_0(e/x_l))] \\ &= -|a'_0|_L + (N-1)(N-1). \end{aligned}$$

On the other hand,

$$\begin{aligned} h_T[f, \pi_{\mathbf{R}}(a_0(e/x_l))] &= h_T[a_0, \pi_{\mathbf{R}}(a_0(e/x_l))] \\ &= -|a_0|_L - (N-1). \end{aligned}$$

By the above inequality, we obtain

$$\begin{aligned} &h^\infty_C[\wedge_C f^b, \pi_{\mathbf{R}}(a_0(e/x_l)] - h_T[f, \pi_{\mathbf{R}}(a_0(e/x_l)] \\ &= |a_0|_L - |a'_0|_L + N(N-1) > 0. \end{aligned}$$

That is, $a_0(e/x_l) \in out_\infty(f)$. Since $|a_0|_L < |a'_0|_L$, $a_0 \in in_\infty(f)$. Thus

$$\{\pi([a_0, g]_T) : g \in S_N \text{ s.t. } g(l) = N-1\} \subset Bd_\infty(f).$$

In particular, $Bd_\infty(f) \ne \emptyset$.

Next we show that $Bd(f) \ne \emptyset \Leftarrow Bd_\infty(f) \ne \emptyset$.
Let $t \in Bd_\infty(f)$ and $Sf(t) = [a, g]_T$. Set $a_0 = a$ and $a_i = a x_{g(0)} x_{g(1)} \cdots x_{g(i-1)}$ $(0 < i < N)$. By proposition 11, $a_0 \in in_\infty(f)$ and $a_{N-1} \in out_\infty(f)$. In particular, $a_0 \in L^*$. Let $S^*_\infty \wedge_C f^b(t) = [a'_0, a'_1, a'_2, \dots, a'_{N-1}]$, where $\pi_{\mathbf{R}}(a'_i) = \pi_{\mathbf{R}}(a_i)$ $(0 \le i < N)$. Then

$$a'_{N-1} \ge_C a'_{N-2} \ge_C \dots \ge_C a'_0.$$

Since $Bd_\infty(f) \subset Out(f)$ and $a_0 \notin L^*$, $a_0 \in out(f)$. Note that $\pi_{\mathbf{R}}(f^b) \subset \pi_{\mathbf{R}}(in(f) + bd(f))$, from which $a'_0 \notin f^b$ follows. Moreover $a_{N-1} \in out_\infty(f)$ implies that $a'_0 \notin \wedge_C f^b$. Thus

$$\exists c \in f^b, d_j \ge 0 \ (0 \le j < N) \text{ s.t. } \begin{cases} a'_0 = c \prod_{0 \le j < N} (x_j/e)^{d_j} \\ \min_{0 \le j < N}\{d_j\} = 0 \end{cases}.$$

Let

$$U = \left\{ a'_0 \prod_{0 \le j < N} (e/x_j)^{m_j} : 0 \le m_j \le d_j \right\} \subset L.$$

Since $a'_0 \notin L^*$, $\pi_{\mathbf{R}}(U) \subset \pi_{\mathbf{R}}(in(f)+out(f))$. And $c \in f^b$ implies that $\pi_{\mathbf{R}}(c) \in \pi_{\mathbf{R}}(in(f))$. Therefore $\exists c' \in U$ and x_l s.t.

$$\pi_{\mathbf{R}}(c') \in \pi_{\mathbf{R}}(out(f)) \text{ and } \pi_{\mathbf{R}}(c'(e/x_l)) \in \pi_{\mathbf{R}}(in(f)).$$

Thus

$$\{\pi([c', g]_T) : g \in S_N \text{ s.t. } g(l) = N-1\} \subset Bd(f).$$

In particular, $Bd(f) \neq \emptyset$. □

Corollary 1. *Let $f \in PL$ and suppose that $Bd(f) = \emptyset$. Then*

(1) $Out(f) = \left\{t \in B^{N-1} : |S^* \vee_C f_b|(t) \leq |Sf|(t)\right\}$,
$In(f) = \left\{t \in B^{N-1} : |S^* \vee_C f_b|(t) > |Sf|(t)\right\}$.

(2) $Out(f) = \left\{t \in B^{N-1} : |S^*_\infty \wedge_C f^b|(t) > |Sf|(t)\right\}$,
$In(f) = \left\{t \in B^{N-1} : |S^*_\infty \wedge_C f^b|(t) \leq |Sf|(t)\right\}$.

Proof. It follows from proposition 10 and 11 immediately. □

3.3.3. Algebra of prehetero numbers Now we consider factorization of prehetero numbers in $\mathbf{PHN}^{N-1}$. We define "prime" prehetero numbers and show that any prehetero number is factored into "primes" in $\mathbf{PHN}^{N-1}$.

Definition 25 (PL factor and PL prime). *Let $f, g \in \mathbf{PHN}^{N-1}$. g is a* PL factor *of f if $g \subset f$ and $In(g) \subset In(f)$. f is* PL prime *if f has no PL factors but itself.*

Theorem 3 (Factorization in $\mathbf{PHN}^{N-1}$). *Let $f \in \mathbf{PHN}^{N-1}$. Then there exists a unique set of PL primes $p_i \in \mathbf{PHN}^{N-1}$ $(0 \leq i < m)$ s.t.*

$$f = p_0 \oplus p_1 \oplus \cdots \oplus p_{m-1},$$
$$In(f) = In(p_0) + In(p_1) + \cdots + In(p_{m-1}) \quad (\textit{disjoint union}).$$

Proof. Let $g \in \mathbf{PHN}^{N-1}$ be a PL factor of f. We show that there exists a unique PL factor g' of f s.t.

$$f = g \oplus g',$$
$$In(f) = In(g) + In(g') \quad \text{(disjoint union)}.$$

Applying the claim successively, we obtain the theorem.

Set

$$g' := \max_C \left\{ \sum_{t \in In(f) \cap Out(g)} a\left(t, \min\left\{|S^* \vee_C f_b|(t), |S^*_\infty \wedge_C g^b|(t)\right\}\right) \right\},$$

$$g'' := \max_T \left\{ \sum_{t \in In(f) \cap Out(g)} a(t, |Sf|(t)) \right\}.$$

First we show that $g' = \vee_C g''_b$. Since $\vee_C f_b \geq_C g'$,

$$\exists K > 0 \text{ s.t. } f_b \geq_C \left(\sum_{0 \leq i < N} (e/x_i) \right)^K g'.$$

Note that $g'' = \max_T \{(in(f) \cup bd(f)) \backslash in(g)\}$. If $a \in in(g)$, then $g' \not\geq_C a$. Thus $g'' \geq_C \left(\sum_{0 \leq i < N}(e/x_i)\right)^K g'$. Since $g' \subset L^*$,

$$g''_b \geq_C \left(\sum_{0 \leq i < N} (e/x_i) \right)^K g'.$$

That is, $\vee_C g''_b \geq_C g'$.

Suppose that $g' \neq \vee_C g''_b$. By the definition of g'', $\exists a \in in(g)$ s.t. $\vee_C g''_b \geq_C a$. On the other hand, $g'' \not\geq_C a$ for $\forall a \in in(g)$. It implies that $\vee_C g''_b \not\geq_C a$ for $\forall a \in in(g)$, which is a contradiction. Thus $g' = \vee_C g''_b$.

By corollary 1,

$$\begin{aligned} Out(f) &= \left\{ t \in B^{N-1} : |S^* \vee_C f_b|(t) \leq |Sf|(t) \right\}, \\ In(f) &= \left\{ t \in B^{N-1} : |S^* \vee_C f_b|(t) > |Sf|(t) \right\}, \\ Out(g) &= \left\{ t \in B^{N-1} : |S^*_\infty \wedge_C g^b|(t) > |Sg|(t) \right\}, \\ In(g) &= \left\{ t \in B^{N-1} : |S^*_\infty \wedge_C g^b|(t) \leq |Sg|(t) \right\}. \end{aligned}$$

It follows that

$$\begin{aligned} Out(g'') &\supset Out(f) \cup In(g), \\ In(g'') &\supset In(f) \cap Out(g). \end{aligned}$$

Thus $Bd(g'') = \emptyset$. The result follows immediately. $\square$

For an element of $\mathbf{PHN}^{N-1}$, we also define action of PL.

Definition 26 (Action on $\mathbf{PHN}^{N-1}$). *Let $f_0 \in \mathbf{PHN}^{N-1}$ and $f \in PL$.*

$$f_0 * f := \begin{cases} \max_T\{f_0 + f\} & \textit{if } \max_T\{f_0 + f\} \in \mathbf{PHN}^{N-1}, \\ \textit{undefined} & \textit{else}. \end{cases}$$

$$f_0 * (-f) := \begin{cases} \max_T\{f_0 - f\} & \textit{if } \max_T\{f_0 - f\} \in \mathbf{PHN}^{N-1}, \\ \textit{undefined} & \textit{else}. \end{cases}$$

For any element $f \in \mathbf{AHN}^{N-1}$, there exists another element $f' \in \mathbf{AHN}^{N-1}$ s.t $f \oplus f' \in \mathbf{PHN}^{N-1}$. Thus one can define action of PL on f as action on $f + f'$ with the help of "catalysis" f'.

Definition 27 (Stabilizer). *Let $f \in \mathbf{PHN}^{N-1}$. Set $G(f) = \{a : a \in G_+(f)$ or $-a \in G_-(f)\}$, where*

$$G_+(f) := \{a \in L : a \not\leq_T f, In(f * a) = In(f)\} \subset L,$$
$$G_-(f) := \{a \in L : a \in f, In(f * (-a)) = In(f)\} \subset L.$$

We call $G(f)$ the stabilizer *of f.*

Using the action defined above, we can define "arithmetic equations" over $\mathbf{PHN}^{N-1}$. In particular, we consider the following type of equations.

Definition 28 (Reaction equation over $\mathbf{PHN}^{N-1}$). *Let $f, g \in \mathbf{PHN}^{N-1}$ and $\alpha, h \in PL$. For indeterminates X and Y, we consider an equation*

$$RE[g, \alpha, f, h] : Y \oplus g = ((\alpha \otimes X) \oplus f) * h,$$

where $h \subset G((\alpha \otimes X) \oplus f)$.

Note that $In(Y \oplus g) = In((\alpha \otimes X) \oplus f)$.

Geometrically the equation $RE[g, \alpha, f, h]$ is interpreted as follows:

(1) Translate X by α and attach it to f.
(2) Add h, which causes a reaction between $\alpha \otimes X$ and f. The "shape" of $(\alpha \otimes X) \oplus f$ is preserved under the reaction.
(3) As a consequence of the reaction, Y and g are produced.

In this way we obtain Y and g from X and f via action h.

Example 6. Let $N = 3$ and set $x = x_0,\ y = x_1,\ z = x_2$. Let f, f_0, f_1, f_2 and f_3 be the polynomials given in section 1.2. Then

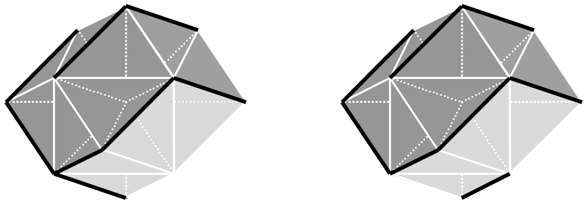
$$G_+(f) = \{x^2 y^{-4} z^{-1}, x^2 y^{-4} z^{-2}, y^{-3} z^{-1}\}, G_-(f) = \emptyset.$$

Since $f = f_0 \oplus f_1 \oplus f_2$ and $f * x^2 y^{-4} z^{-1} = f_3$, $(X, Y) = (f_0, f_3)$ gives a solution of $RE[0, 1, f_1 \oplus f_2, x^2 y^{-4} z^{-1}]$.

Example 7 (Decomposition of a dodecahedron). Let $N = 4$ and set $x = x_0$, $y = x_1$, $z = x_2$ and $w = x_3$. Let

$$f_0 = xyz + xyw + xzw + yzw \in \mathbf{PHN}^3 \text{ (Fig.3(b))},$$
$$f_1 = xyw + yzw + xz \in \mathbf{PHN}^3 \text{ (Fig.4(a))},$$
$$f_2 = xyw + xz + yz \in \mathbf{PHN}^3 \text{ (Fig.4(b))},$$
$$f_3 = xyw + xz + yz + wz \in \mathbf{PHN}^3 \text{ (Fig.3(a))}.$$

Then f_0, f_1 and f_2 are PL primes but not primes: $f_0 = g_0 \oplus g_1 \oplus g_2 \oplus g_3$, $f_1 = g_{10} \oplus g_{11} \oplus g_{12}$ and $f_2 = g_{20} \oplus g_{21}$, where g_0, g_1, g_2 and g_3 are given in example 5 and $g_{10} = xyw + yzw + xz$, $g_{11} = xyz + xyw + yzw$, $g_{12} = xyw + xzw + yzw$, $g_{20} = xyw + yzw + xz$ and $g_{21} = xyz + xzw + yw$. On the other hand, f_3 is not only PL prime but also prime.



(a) (b)

Fig. 4. Factoring of a dodecahedron (2)

Since $f_1 = f_0 * xz$, $f_2 = f_0 * (xz + yz)$ and $f_3 = f_0 * (xz + yz + wz)$,

$$(g_0 \oplus g_1 \oplus g_2 \oplus g_3) * xz = g_{10} \oplus g_{11} \oplus g_{12},$$
$$(g_0 \oplus g_1 \oplus g_2 \oplus g_3) * (xz + yz) = g_{20} \oplus g_{21},$$
$$(g_0 \oplus g_1 \oplus g_2 \oplus g_3) * (xz + yz + wz) = f_3.$$

4. D^2-moduli and proteins

4.1. D^2-moduli space

Definition 29 (D^2-code of a prime). *Let $f \in \mathbf{AHN}^{N-1}$ be a prime and $obt(f) = \{t[i]\}$. We call $cd(f) := \{D^2 f(t[i])\}$ the D^2-code of f. Since the orbit $obt(f)$ is closed, $cd(f)$ is not determined uniquely. Occasionally we denote $cd(f) = \{c[i] : i \in [0, k]\}$ by $c[0] - c[1] - \cdots - c[k]$.*

Remark 4. We identify $obt(f)$ with a finite orbit whose length is equal to the period of $obt(f)$.

Let $\mathcal{M}'$ be the set of all finite sequences of $\{+1, -1\}$. To eliminate the ambiguity mentioned above, we introduce an equivalence relation $\sim_c$ in $\mathcal{M}'$. Let $\{c'[i]\}, \{c[i]\} \in \mathcal{M}'$ have the same length L. Then $\{c'[i]\} \sim_c \{c[i]\}$ if and only if they coincide with each other by rotational shift, i.e., $\exists k \in \mathbf{Z}$ s.t. $c'[i + k \bmod L] = c[i]$, by inversion, i.e., $-c'[i] = c[i]$, by reversion, i.e., $c'[i] = c[-i]$ or by any combination of them.

Definition 30 (D^2-moduli space). *We call the quotient set $\mathcal{M} := \mathcal{M}'/\sim_c$ the D^2-moduli space. An element of $\mathcal{M}$ is called D^2-code. For $c \in \mathcal{M}'$, $[c] \in \mathcal{M}$ denotes the equivalence class determined by c.*

For any prime $f \in \mathbf{AHN}^{N-1}$, $[cd(f)] \in \mathcal{M}$ is determined uniquely.

Definition 31 (m-dimensional implementation of a D^2-code). *Let $c = \{c[i] : 0 \leq i < L\} \in \mathcal{M}'$ and $\{t[i] : 0 \leq i < L\} \subset B^m$ $(m \geq 2)$ be an orbit. We call $\{t[i]\}$ an m-dimensional implementation of $[c]$ if $D^2(t[i]) = c[i]$ $(0 \leq i < L)$ and denote it by $V_m(c, t[0])$:*

$$V_m(c, t) := \{t[i]\} \subset B^m \text{ s.t. } t[0] = t \text{ and } D^2 t[i] = c[i] \ (0 \leq i < L).$$

An affine hetero number is defined by a single polynomial, i.e., “affine”. On the other hand, an implementation of a D^2-code is obtained by patching affine orbits together as in the case of “manifold”. That is, implementations are “locally affine” and defined by a set of polynomials. Note that any local orbit is “affine” and an orbit is obtained by patching local orbits together.

Definition 32 (Scheme of an implementation). *Let $c = \{c[i] : 0 \le i < L\} \in \mathcal{M}'$ and $V_m(c,t) = \{t[i]\} \subset B^m$ $(m > 2)$ for $t \in B^m$. Then there exists a covering $\{[l_{0j}, l_{1j}] : 0 \le j < k\}$ of $[0, L-1] \subset \mathbf{Z}$ s.t.*

$\{t[i] : i \in [l_{0j}, l_{1j}]\}$ *is an orbit of the vector field $X(f_j)$ induced by f_j,*

where $f_j := \max_T\{a[i] : i \in [l_{0j}, l_{1j}]\}$. We set

$$shm^m(c,t) := \{(f_j, [l_{0j}, l_{1j}]) : 0 \le j < k\}$$

and call it the scheme *of the implementation $V_m(c,t)$.*

If $c \in \mathcal{M}'$ is a D^2-code of some prime $f \in \mathbf{AHN}^m$ and $V_m(c,t) = obt(f)$ for some $t \in B^m$, then $shm^m(c,t) = \{(f, [0, k-1])\}$, where k is the period of $obt(f)$.

In the following examples we denote $+1$ by U and -1 by D.

Example 8 (Atom). Let $c = D - U - D - U - D - U \in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \ge 2)$. Then $V_m(c,t) \in ALP^m$ and

$$shm^m(c,t) = \{(1 + x_0/x_m + x_0/x_{m-1}, [0,5])\}.$$

Example 9. Let $c =$

$$\begin{aligned} &D - D - D - D - D - D - D - U - D \\ &\qquad - U - U - U - U - U - U - U - D - U \end{aligned}$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \ge 2)$.

(1) 2-dim. implementation. $V_2(c,t) \in ALP^2$ and

$$shm^2(c,t) = \{(1 + x_0/x_1 + {x_0}^4{x_1}^3/x_2, [0,17])\}.$$

(2) 3-dim. implementation. $V_3(c,t) \in ALP^3$ and

$$shm^3(c,t) = \{(1 + x_0/x_2 + {x_0}^3{x_1}^2{x_2}^2/x_3, [0,17])\}.$$

(3) 4-dim. implementation. $V_4(c,t) \notin ALP^4$ and

$$shm^4(c,t) = \{(1 + x_0/x_3 + {x_0}^2{x_1}^2{x_2}^2x_3/x_4, [0,17])\}.$$

(4) 5-dim. implementation. $V_5(c,t) \notin ALP^5$ and

$$shm^5(c,t) = \{(1 + x_1/x_4 + {x_0}^2{x_1}^2x_2x_3x_4/x_5, [0,17])\}.$$

(5) 6-dim. implementation. $V_6(c,t) \in ALP^6$ and

$$shm^6(c,t) = \{(1 + x_0/x_5 + {x_0}^2 x_1 x_2 x_3 x_4 x_5/x_6, [0,17])\}.$$

(6) 7-dim. implementation. $V_7(c,t) \notin ALP^7$ and

$$shm^7(c,t) = \{(1 + x_0/x_5 + x_0 x_1 x_2 x_3 x_4 x_5 x_6/x_7, [0,17])\}.$$

Example 10. Let $c =$

$$\begin{aligned} &D - D - D - D - D - U - D - U - U \\ &\qquad - U - U - U - D - D - U - D - U - U \end{aligned}$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c,t) \notin ALP^2$ and

$$\begin{aligned} shm^2(c,t) = \{&(1 + x_0/x_1 + {x_0}^3{x_1}^2/x_2, [0,12]), \\ &(x_2 + x_0/x_1 + {x_0}^3{x_1}^2/x_2, [1,17])\}. \end{aligned}$$

(2) 3-dim. implementation. $V_3(c,t) \in ALP^3$ and

$$shm^3(c,t) = \{(1 + x_0 x_1/x_2 + {x_0}^2{x_1}^2 x_2/x_3, [0,17])\}.$$

(3) 4-dim. implementation. $V_4(c,t) \notin ALP^4$ and

$$\begin{aligned} shm^4(c,t) = \{&(1 + x_0/x_3 + {x_0}^2 x_1 x_2 x_3/x_4, [0,12]), \\ &(x_1 + x_0/x_3 + {x_0}^2 x_1 x_2 x_3/x_4, [1,17])\}. \end{aligned}$$

(4) 5-dim. implementation. $V_5(c,t) \notin ALP^5$ and

$$shm^5(c,t) = \{(1 + x_0 x_1/x_3 + x_0 x_1 x_2 x_3 x_4/x_5, [0,17])\}.$$

(5) 6-dim. implementation. $V_6(c,t) \notin ALP^6$ and

$$shm^6(c,t) = \{(1 + x_0 x_1/x_5 + x_0 x_1 x_2 x_3 x_4/x_6, [0,17])\}.$$

(6) 7-dim. implementation. $V_7(c,t) \notin ALP^7$ and

$$shm^7(c,t) = \{(1 + x_0 x_1/x_6 + x_0 x_1 x_2 x_3 x_4/x_7, [0,17])\}.$$

Example 11. Let $c =$

$$\begin{aligned} &D - D - D - D - U - D - U - U - U \\ &\qquad - D - D - D - U - D - U - U - U - U \end{aligned}$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c,t) \notin ALP^2$ and

$$shm^2(c,t) = \{(1 + {x_0}^2{x_1}^2/x_2, [0,8]),\\ (x_0 + x_1 + {x_0}^2{x_1}^2/x_2, [1,9]),\\ ({x_0}^2x_2 + x_1 + {x_0}^2{x_1}^2/x_2, [2,15]),\\ ({x_0}^2x_2 + x_1 + x_0x_1/x_2, [8,16]),\\ ({x_0}^2x_2 + x_1/x_2, [9,17])\}.$$

(2) 3-dim. implementation. $V_3(c,t) \in ALP^3$ and

$$shm^3(c,t) = \{(1 + {x_0}^2x_1x_2/x_3 + {x_0}^2x_1x_3/x_2, [0,17])\}.$$

(3) 4-dim. implementation. $V_4(c,t) \notin ALP^4$ and

$$shm^4(c,t) = \{(1 + x_0x_1x_2x_3/x_4, [0,8]),\\ (x_0 + x_1 + x_1x_2x_3 + x_0x_1x_2x_3/x_4, [1,16]),\\ (x_0/x_4 + x_1x_2x_3, [9,17])\}.$$

(4) 5-dim. implementation. $V_5(c,t) \notin ALP^5$ and

$$shm^5(c,t) = \{(1 + x_0x_1x_2x_3/x_5, [0,8]),\\ (x_0 + x_1 + x_1x_2x_3 + x_0x_1x_2x_3/x_5, [1,16]),\\ (x_0/x_5 + x_1x_2x_3, [9,17])\}.$$

(5) 6-dim. implementation. $V_6(c,t) \notin ALP^6$ and

$$shm^6(c,t) = \{(1 + x_0x_1x_2x_3/x_6, [0,8]),\\ (x_0 + x_1 + x_1x_2x_3 + x_0x_1x_2x_3/x_6, [1,16]),\\ (x_0/x_6 + x_1x_2x_3, [9,17])\}.$$

(6) 7-dim. implementation. $V_7(c,t) \notin ALP^7$ and

$$shm^7(c,t) = \{(1 + x_0x_1x_2x_3/x_7, [0,8]),\\ (x_0 + x_1 + x_1x_2x_3 + x_0x_1x_2x_3/x_7, [1,16]),\\ (x_0/x_7 + x_1x_2x_3, [9,17])\}.$$

Example 12. Let $c =$

$$D - D - D - U - U - U - D - D - D\\ - U - U - U - D - D - D - U - U - U$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c,t) \in ALP^2$ and

$$shm^2(c,t) = \{(1 + {x_0}^2/{x_1}^2 + {x_0}^2/{x_2}^2, [0,17])\}.$$

(2) 3-dim. implementation. $V_3(c,t) \notin ALP^3$ and

$$shm^3(c,t) = \{(1 + x_2/x_3 + {x_2}^2/{x_3}^2 + {x_2}^3/{x_3}^3, [0,17])\}.$$

(3) 4-dim. implementation. $V_4(c,t) \in ALP^4$ and

$$shm^4(c,t) = \{(1 + x_2/x_3 + x_2/x_4, [0,17])\}.$$

(4) 5-dim. implementation. $V_5(c,t) \in ALP^5$ and

$$shm^5(c,t) = \{(1 + x_2/x_4 + x_2/x_5, [0,17])\}.$$

(5) 6-dim. implementation. $V_6(c,t) \in ALP^6$ and

$$shm^6(c,t) = \{(1 + x_2/x_5 + x_2/x_6, [0,17])\}.$$

(6) 7-dim. implementation. $V_7(c,t) \in ALP^7$ and

$$shm^7(c,t) = \{(1 + x_2/x_6 + x_2/x_7, [0,17])\}.$$

Example 13. Let $c =$

$$\begin{aligned} &D - D - U - D - U - U - D - D - U \\ &\qquad - U - D - D - U - D - U - U - D - U \end{aligned}$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c,t) \notin ALP^2$ and

$$\begin{aligned} shm^2(c,t) = \{(1 + x_2/x_0 &+ x_0x_1/x_2 + x_1{x_2}^2/{x_0}^2 \\ &+ {x_2}^2/(x_0x_1), [0,17])\}. \end{aligned}$$

(2) 3-dim. implementation. $V_3(c,t) \in ALP^3$ and

$$shm^3(c,t) = \{(1 + x_0/x_2 + x_0x_1/x_3, [0,17])\}.$$

(3) 4-dim. implementation. $V_4(c,t) \in ALP^4$ and

$$shm^4(c,t) = \{(1 + x_0/x_3 + x_0x_1/x_4, [0,17])\}.$$

(4) 5-dim. implementation. $V_5(c,t) \in ALP^5$ and

$$shm^5(c,t) = \{(1 + x_0/x_4 + x_0x_1/x_5, [0,17])\}.$$

(5) 6-dim. implementation. $V_6(c,t) \in ALP^6$ and

$$shm^6(c,t) = \{(1 + x_0/x_5 + x_0x_1/x_6, [0,17])\}.$$

(6) 7-dim. implementation. $V_7(c,t) \in ALP^7$ and

$$shm^7(c,t) = \{(1 + x_0/x_6 + x_0x_1/x_7, [0,17])\}.$$

Table 1. Distribution of primes (affine loops)

length	$\cup_{m\leq 7} ALP^m$	ALP^2	ALP^3	ALP^4	ALP^5	ALP^6	ALP^7
6	1	1	1	1	1	1	1
10	1	1	0	0	0	0	0
12	3	0	3	2	2	2	2
14	2	2	0	1	0	0	0
16	2	0	0	1	2	1	1
18	10	5	6	3	3	4	3
20	6	0	0	4	3	3	4
22	14	11	0	3	3	1	1
24	19	0	17	4	6	9	7
26	36	27	0	5	6	3	5
28	20	0	0	13	10	10	10
30	122	78	42	10	5	10	6
32	29	0	0	10	17	11	14
34	256	234	0	12	6	7	9
36	173	0	118	41	28	30	25
38	821	778	0	29	25	15	15
40	77	0	0	24	30	30	35
42	3298	2831	391	49	48	33	21
44	187	0	0	122	42	59	54
46	11288	11122	0	91	70	35	42
48	1526	0	1301	109	100	100	77
total	17891	15090	1879	534	407	364	332

4.2. Statistical survey of primes

Now we consider the relation among primes of various dimensions. Let $f \in \mathbf{AHN}^m$ $(m \geq 2)$ be a prime and $obt(f) = \{t[i]\}$. Then $V_m(cd(f), t[0]) = obt(f)$ gives an m-dimensional implementation of $[cd(f)]$. On the other hand, $V_{m'}(cd(f), t')$ may not be any loop if $m' \neq m$.

Here we show the distribution of m-dimensional primes $(2 \leq m < 8)$. Primes are considered as implementations of D^2-codes and we only deal with D^2-codes of length less than 49.

Definition 33 (Type of D^2-code). *Let $c \in \mathcal{M}'$, $i_0, i_1, \cdots, i_k \in \mathbf{Z}$ and set $j_l = i_l$ or $\bar{i_l}$ $(0 \leq l \leq k)$. We call $[c] \in \mathcal{M}$ is* $\mathrm{type}\ (j_0, j_1, \ldots, j_k)$ *if its i_l-dimensional implementation is affine hetero number when $j_l = i_l$ and its i_l-dimensional implementation is not affine hetero number when $j_l = \bar{i_l}$:*

$$p(V_{i_l}(c, t_l)) \in \mathbf{AHN}^{i_l} \quad \text{if } j_l = i_l,$$
$$p(V_{i_l}(c, t_l)) \notin \mathbf{AHN}^{i_l} \quad \text{if } j_l = \bar{i_l},$$

where t_l is any tile of B^{i_l} $(0 \leq l < k)$. Note that the type of $[c]$ is not depend on the choice of c.

Table 2. Distribution of types. (D^2-codes of length less than 49 are grouped.)

type	D^2-codes	type	D^2-codes
$(2,3,4,5,6,7)$	1	$(\bar{2},3,4,5,6,7)$	16
$(2,3,4,5,6,\bar{7})$	1	$(\bar{2},3,4,5,6,\bar{7})$	0
$(2,3,4,5,\bar{6},7)$	0	$(\bar{2},3,4,5,\bar{6},7)$	0
$(2,3,4,5,\bar{6},\bar{7})$	0	$(\bar{2},3,4,5,\bar{6},\bar{7})$	0
$(2,3,4,\bar{5},6,7)$	1	$(\bar{2},3,4,\bar{5},6,7)$	2
$(2,3,4,\bar{5},6,\bar{7})$	1	$(\bar{2},3,4,\bar{5},6,\bar{7})$	8
$(2,3,4,\bar{5},\bar{6},7)$	0	$(\bar{2},3,4,\bar{5},\bar{6},7)$	0
$(2,3,4,\bar{5},\bar{6},\bar{7})$	2	$(\bar{2},3,4,\bar{5},\bar{6},\bar{7})$	0
$(2,3,\bar{4},5,6,7)$	0	$(\bar{2},3,\bar{4},5,6,7)$	12
$(2,3,\bar{4},5,6,\bar{7})$	0	$(\bar{2},3,\bar{4},5,6,\bar{7})$	4
$(2,3,\bar{4},5,\bar{6},7)$	0	$(\bar{2},3,\bar{4},5,\bar{6},7)$	0
$(2,3,\bar{4},5,\bar{6},\bar{7})$	0	$(\bar{2},3,\bar{4},5,\bar{6},\bar{7})$	10
$(2,3,\bar{4},\bar{5},6,7)$	0	$(\bar{2},3,\bar{4},\bar{5},6,7)$	17
$(2,3,\bar{4},\bar{5},6,\bar{7})$	7	$(\bar{2},3,\bar{4},\bar{5},6,\bar{7})$	56
$(2,3,\bar{4},\bar{5},\bar{6},7)$	0	$(\bar{2},3,\bar{4},\bar{5},\bar{6},7)$	1
$(2,3,\bar{4},\bar{5},\bar{6},\bar{7})$	5	$(\bar{2},3,\bar{4},\bar{5},\bar{6},\bar{7})$	1735
$(2,\bar{3},4,5,6,7)$	2	$(\bar{2},\bar{3},4,5,6,7)$	46
$(2,\bar{3},4,5,6,\bar{7})$	0	$(\bar{2},\bar{3},4,5,6,\bar{7})$	1
$(2,\bar{3},4,5,\bar{6},7)$	0	$(\bar{2},\bar{3},4,5,\bar{6},7)$	3
$(2,\bar{3},4,5,\bar{6},\bar{7})$	1	$(\bar{2},\bar{3},4,5,\bar{6},\bar{7})$	9
$(2,\bar{3},4,\bar{5},6,7)$	0	$(\bar{2},\bar{3},4,\bar{5},6,7)$	14
$(2,\bar{3},4,\bar{5},6,\bar{7})$	1	$(\bar{2},\bar{3},4,\bar{5},6,\bar{7})$	7
$(2,\bar{3},4,\bar{5},\bar{6},7)$	0	$(\bar{2},\bar{3},4,\bar{5},\bar{6},7)$	14
$(2,\bar{3},4,\bar{5},\bar{6},\bar{7})$	52	$(\bar{2},\bar{3},4,\bar{5},\bar{6},\bar{7})$	352
$(2,\bar{3},\bar{4},5,6,7)$	0	$(\bar{2},\bar{3},\bar{4},5,6,7)$	60
$(2,\bar{3},\bar{4},5,6,\bar{7})$	0	$(\bar{2},\bar{3},\bar{4},5,6,\bar{7})$	4
$(2,\bar{3},\bar{4},5,\bar{6},7)$	0	$(\bar{2},\bar{3},\bar{4},5,\bar{6},7)$	12
$(2,\bar{3},\bar{4},5,\bar{6},\bar{7})$	6	$(\bar{2},\bar{3},\bar{4},5,\bar{6},\bar{7})$	219
$(2,\bar{3},\bar{4},\bar{5},6,7)$	0	$(\bar{2},\bar{3},\bar{4},\bar{5},6,7)$	30
$(2,\bar{3},\bar{4},\bar{5},6,\bar{7})$	2	$(\bar{2},\bar{3},\bar{4},\bar{5},6,\bar{7})$	71
$(2,\bar{3},\bar{4},\bar{5},\bar{6},7)$	3	$(\bar{2},\bar{3},\bar{4},\bar{5},\bar{6},7)$	98
$(2,\bar{3},\bar{4},\bar{5},\bar{6},\bar{7})$	15005	$total$	17891

Table1 shows the number of D^2-codes whose m-dimensional implementation is an affine loop for $2 \leq m < 8$. In table2 D^2-codes of length less than 49 are grouped by type.

4.3. D^2-approximation of DNA and proteins

Finally we give the “second approximation” of DNA and proteins. A D^2-code gives the second approximation of any shape. We call it the D^2-*approximation* of the shape. Using the D^2-approximation, it is possible to classify proteins objectively. (See ‘Protein taxonomy’ in section 1.2.)

The correspondence between DNA and tile is given by

$$\text{one nucleotide} \Longleftrightarrow \text{one tile.}$$

Table 3. The D^2-code of amino-acids.

Code	$D^2t[i-1]$	$D^2t[i]$	$D^2t[i+1]$
0	D	D	D
1	U	D	D
2	D	U	D
3	U	U	D
4	D	D	U
5	U	D	U
6	D	U	U
7	U	U	U

In nature there are 4 kinds of nucleotides: adenine (A), cytosine (C), guanine (G) and thymine (T). They are coded by gradient of tiles, that is, up (U) and down (D).

The correspondence between protein and tile is given by

$$\text{one amino-acid} \Longleftrightarrow \text{three tiles.}$$

(Exactly speaking, a bond between amino-acids corresponds to three tiles.) In nature there are 20 kinds of amino-acids. They are coded by triplets of U/D. As a code table of amino-acids, we use Table 3.

Note that the coding is consistent with the actual genetic code, where a single amino-acid is coded by three nucleotides (codon). That is,

$$\text{one amino-acid} \Longleftrightarrow \text{three nucleotides.}$$

First we give an example of D^2-approximation of DNA.

Example 14 (DNA (helix)). DNA consists of double-stranded, nucleotide-paired helical molecules and forms a right-handed helical staircase. Roughly speaking, it is a cylinder with diameter about 20 Å.

When DNA is under dehydrated nonphysiological condition (A-DNA), there are an average of 10.9 nucleotide (base) pairs per turn of the helix, which corresponds to an average helical-twist angle of 33.1^{o} ($10.9 \times 33.1 = 360^{\mathrm{o}}$) from one nucleotide pair to the next. The spacing along the helix axis from one pair to the next is 2.9 Å.

When DNA is fully hydrated (B-DNA), there are an average of 10 nucleotide (base) pairs and 33.1^{o} and 3.4 Å, respectively ([2]). B-DNA is the most common form in cells.

On the other hand, the D^2-approximation given below has 12 tiles per turn of the helix.

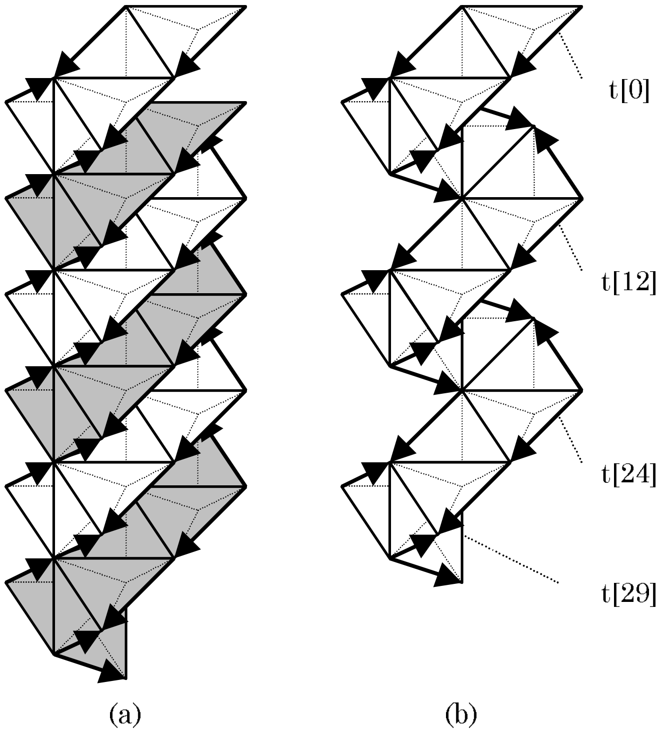


Fig. 5. DNA (helix)

Let $c =$

$$D - D - D - D - U - U - D - D - D - D$$
$$- U - U - D - D - D - D - U - U - D - D$$
$$- D - D - U - U - D - D - D - D - U - U$$

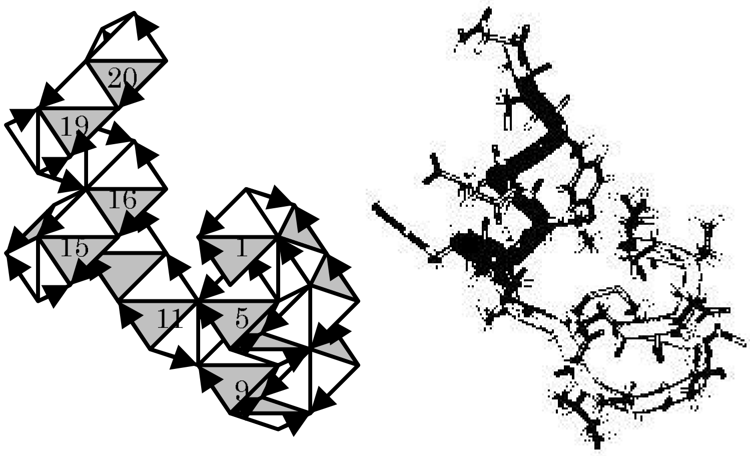
$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c, t) \notin ALP^2$ and

$$shm^2(c, t) = \{(1 + x_0{x_1}^2/x_2 + {x_0}^2{x_1}^4/{x_2}^2 + {x_0}^3{x_1}^6/{x_2}^3 + {x_0}^4{x_1}^8/{x_2}^4 + {x_0}^5{x_1}^10/{x_2}^5, [0, 29])\}.$$

(2) 3-dim. implementation. $V_3(c, t) \notin ALP^3$ and

$$\begin{aligned} shm^3(c, t) = \{&(1 + {x_0}^2x_1/x_3, [0, 7]), \\ &(x_0x_1 + {x_0}^2{x_2}^2 + {x_0}^2x_1/x_3, [2, 13]), \\ &({x_0}^2x_1x_2 + {x_0}^2{x_2}^2 + {x_0}^4x_1{x_2}^2/x_3, [8, 19]), \\ &({x_0}^3x_1{x_2}^2 + {x_0}^4{x_2}^4 + {x_0}^4x_1{x_2}^2/x_3, [14, 25]), \\ &({x_0}^4x_1{x_2}^3 + {x_0}^4{x_2}^4 + {x_0}^6x_1{x_2}^4/x_3, [20, 29])\}. \end{aligned}$$



(a) (b)

Fig. 6. 2HIU (chain A). The figure (b) is prepared using **WebLab Viewer** (Molecular Simulations Inc.).

(3) 4-dim. implementation. $V_4(c,t) \notin ALP^4$ and

$$\begin{aligned} & shm^4(c,t) \\ & = \{(1 + x_0x_1x_3/x_4, [0,7]), \\ & \quad (x_0x_1 + x_0x_2{x_3}^2 + x_0x_1x_3/x_4, [2,13]), \\ & \quad (x_0x_1x_2x_3 + x_0x_2{x_3}^2 + {x_0}^2x_1x_2{x_3}^3/x_4, [8,19]), \\ & \quad ({x_0}^2x_1x_2{x_3}^2 + {x_0}^2{x_2}^2{x_3}^4 + {x_0}^2x_1x_2{x_3}^3/x_4, [14,25]), \\ & \quad ({x_0}^2x_1{x_2}^2{x_3}^3 + {x_0}^2{x_2}^2{x_3}^4 + {x_0}^3x_1{x_2}^2{x_3}^5/x_4, [20,29])\}. \end{aligned}$$

The shape of the 3-dim. implementation of the D^2-code is shown in Fig.5(b). Two copies of the helix form a DNA-like double-helix as shown in Fig.5(a). In the figures arrows indicate the direction of “down (D)”.

Next we give two examples of D^2-approximation of proteins. We remark that the examples are rough approximation of proteins. Therefore they may not be the most suitable ones. In particular, functional aspect of the proteins is not considered at all. Note also that proteins are flexible objects, which may complicate the situation.

In the following ‘2HIU’ and ‘1EPI’ are the PDB IDs which are used to retrieve data from the PDB (Protein Data Bank).

Example 15 (2HIU (chain A)). Hormone (Insulin, human). Let $c =$

$$\begin{aligned} & D - U - D - U - U - U - U - D - D - U \\ & \quad - U - D - D - U - D - U - U - U - U - D \\ & \quad - D - U - U - D - D - D - D - U - U - U \end{aligned}$$

Table 4. The D^2-code of 2HIU.

No.	0	1	2	3	4	5	6	7
Amino acid	GLY	ILE	VAL	GLU	GLN	CYS	CYS	THR
D^2-code	2	7	4	6	2	7	4	6
No.	8	9	10	11	12	13	14	15
Amino acid	SER	ILE	CYS	SER	LEU	TYR	GLN	LEU
D^2-code	0	7	0	0	6	2	7	4
No.	16	17	18	19	20	21	22	23
Amino acid	GLU	ASN	TYR	CYS	ASN	–	–	–
D^2-code	7	4	7	4	–	–	–	–

$$
\begin{aligned}
&-D-D-D-D-D-D-U-U-D-D\\
&-U-D-U-U-U-U-D-D-U-U\\
&-U-U-D-D-U-U-U-U-D-D
\end{aligned}
$$

$\in \mathcal{M}'$ and $t = [1, id]_B \in B^m$ $(m \geq 2)$.

(1) 2-dim. implementation. $V_2(c,t) \notin ALP^2$ and

$$
\begin{aligned}
&shm^2(c,t)\\
&= \{(x_0/x_2 + 1/(x_0{x_1}^2) + {x_2}^2/{x_1}^4 + x_2/({x_0}^2{x_1}^3), [0,17]),\\
&\quad (x_2/{x_0}^5 + 1/({x_0}^3{x_1}^2) + {x_2}^2/{x_1}^4 + 1/({x_0}^2{x_1}^3), [11,32]),\\
&\quad (x_2/{x_0}^5 + x_2/({x_0}^3{x_1}^2) + {x_1}^4/({x_0}^2x_2) + {x_1}^3/{x_0}^4, [24,44]),\\
&\quad (x_1x_2/{x_0}^6 + x_2/{x_0}^8 + {x_1}^4/({x_0}^2x_2) + {x_1}^2/{x_0}^4, [38,59])\}.
\end{aligned}
$$

(2) 3-dim. implementation. $V_3(c,t) \notin ALP^3$ and

$$
\begin{aligned}
&shm^3(c,t)\\
&= \{(x_0/x_3 + 1/(x_1{x_2}^2) + 1/(x_0{x_2}^2), [0,12]),\\
&\quad (1/(x_2x_3) + 1/(x_1{x_2}^2) + 1/(x_0{x_2}^2), [5,16]),\\
&\quad (1/(x_2x_3) + 1/(x_0x_1{x_2}^3) + 1/({x_0}^2{x_2}^3)\\
&\qquad\qquad + x_1/(x_0x_2x_3), [11,27]),\\
&\quad (1/({x_0}^2x_2x_3) + 1/(x_0x_1{x_2}^3) + 1/({x_0}^2{x_2}^3)\\
&\qquad\qquad + x_1x_2/{x_3}^2, [14,40]),\\
&\quad ({x_1}^2x_2/(x_0x_3) + x_1/x_3 + x_1x_2/{x_3}^2, [34,43]),\\
&\quad ({x_1}^2x_2/(x_0x_3) + 1/{x_3}^2 + x_2/{x_3}^3, [38,49]),
\end{aligned}
$$

$$(x_2/(x_0{x_3}^4) + 1/{x_3}^2 + x_2/(x_1{x_3}^4), [44, 55]),$$
$$(x_2/(x_0{x_3}^4) + 1/({x_1}^2{x_3}^4), [50, 59])\}.$$

(3) 4-dim. implementation. $V_4(c, t) \notin ALP^4$ and

$$\begin{aligned}
&shm^4(c, t)\\
&= \{(x_0/x_4 + 1/(x_1x_2x_3) + 1/(x_1{x_3}^2), [0, 12]),\\
&\quad (1/(x_1x_2x_3) + 1/(x_3x_4) + 1/(x_1{x_3}^2), [5, 16]),\\
&\quad (1/(x_0x_1x_2{x_3}^2) + 1/(x_3x_4) + 1/(x_0{x_1}^2{x_3}^2)\\
&\qquad + x_2/({x_1}^2{x_3}^2x_4), [11, 32]),\\
&\quad ({x_2}^2/(x_0x_1) + 1/({x_1}^2{x_3}^2) + x_2/({x_1}^2{x_3}^2x_4), [24, 40]),\\
&\quad ({x_2}^2/(x_0x_1) + {x_2}^2/(x_1x_3) + x_2, [34, 43]),\\
&\quad ({x_2}^2/(x_0{x_1}^2x_4) + {x_2}^2/(x_0x_1x_3x_4) + x_2, [38, 49]),\\
&\quad ({x_2}^2/({x_0}^2{x_1}^3x_4) + {x_2}^2/(x_0x_1x_3x_4)\\
&\qquad + x_2/({x_0}^2{x_1}^2x_4), [44, 55]),\\
&\quad ({x_2}^2/({x_0}^2{x_1}^3x_3{x_4}^2) + x_2/({x_0}^2{x_1}^2x_4), [50, 59])\}.
\end{aligned}$$

The shape of the 3-dim. implementation of the D^2-code is shown in Fig.6(a). The experimentally obtained shape of 2HIU (chain A) is shown in Fig.6(b). Using Table 3, the D^2-code of amino-acids is given by Table 4.

Example 16 (1EPI). Epidermal growth factor (EGF, mouse). Let $c =$

$$\begin{aligned}
&D-U-U-D-U-U-D-U-U-D\\
&\quad -D-D-U-D-D-D-U-U-U-D\\
&\quad -D-U-D-U-U-D-D-U-U-U\\
&\quad -D-D-D-U-D-D-U-D-D-U\\
&\quad -U-U-D-U-U-U-U-D-U-U\\
&\quad -D-D-D-U-U-D-D-D-U-U\\
&\quad -U-D-D-U-D-U-U-D-U-U\\
&\quad -U-U-D-D-D-D-U-D-D-U\\
&\quad -U-D-U-U-D-U-U-U-D-D\\
&\quad -D-D-U-U-D-D-D-U-U-U\\
&\quad -D-D-D-D-D-D-U-D-D-D\\
&\quad -D-U-D-D-D-D-D-U-U-D\\
&\quad -D-D-D-D-D-U-U-U-U-D\\
&\quad -D-U-U-U-D-U-U-D-D-D\\
&\quad -D-D-U-D-U-U-D-U-D-D
\end{aligned}$$

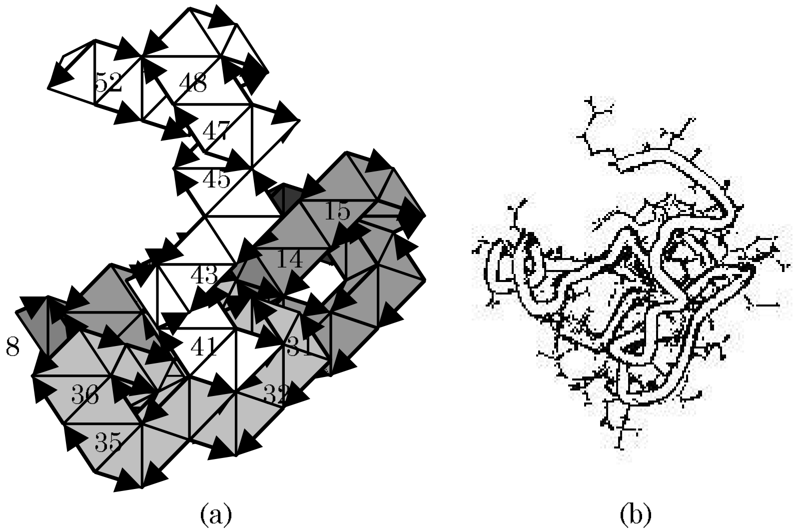


Fig. 7. 1EPI. The figure (b) is prepared using **WebLab Viewer** (Molecular Simulations Inc.).

$$-U-U-U-D-U-U$$

$\in \mathcal{M}'$ and $t=[1,id]_B \in B^3$. We show a 3-dim. implementation only. $V_3(c,t) \notin ALP^3$ and

$$\begin{aligned}
&shm^3(c,t)\\
&\quad=\{(1+x_0/(x_2x_3)+x_0/(x_1x_3),[0,4]),\\
&\qquad(x_0/(x_2x_3)+1/(x_1x_3)+x_0/(x_1{x_3}^2),[2,7]),\\
&\qquad(1/(x_1x_3)+x_0/(x_1x_2{x_3}^2)+{x_0}^2/({x_1}^2x_3),[5,13]),\\
&\qquad(x_0/(x_1x_3)+{x_0}^2/({x_1}^2x_3)+{x_0}^3x_2/(x_1{x_3}^3)\\
&\qquad\qquad\qquad+{x_0}^3/({x_1}^2{x_3}^2),[11,26]),\\
&\qquad({x_0}^2x_2/(x_1{x_3}^2)+{x_0}^3x_2/(x_1{x_3}^3)+{x_0}^3/({x_1}^2{x_3}^2),[17,27]),\\
&\qquad({x_0}^2/(x_1{x_3}^2)+{x_0}^3x_2/(x_1{x_3}^3)+{x_0}^3/({x_1}^2{x_3}^2),[18,28]),\\
&\qquad({x_0}^2/(x_1{x_3}^3)+x_0x_2/{x_3}^2+{x_0}^3/({x_1}^2{x_3}^2),[22,34]),\\
&\qquad({x_0}^2x_2/(x_1{x_3}^2)+x_0x_2/{x_3}^2+{x_0}^2/x_3,[32,37]),\\
&\qquad({x_0}^2x_2/{x_3}^2+x_1/x_3+x_0x_1/{x_3}^2+{x_0}^2/x_3,[35,43]),\\
&\qquad(x_0x_1/{x_3}^2+1/(x_0x_2x_3)+1/(x_1x_2x_3),[41,48]),\\
&\qquad(x_0/(x_1{x_2}^2x_3)+1/(x_0x_2x_3)+1/(x_1x_2{x_3}^2),[46,56]),
\end{aligned}$$

Table 5. The D^2-code of 1EPI.

No.	0	1	2	3	4	5	6	7	8
Amino acid	ASN	SER	TYR	PRO	GLY	CYS	PRO	SER	SER
D^2-code	3	3	3	0	4	3	4	5	4
No.	9	10	11	12	13	14	15	16	17
Amino acid	TYR	ASP	GLY	TYR	CYS	LEU	ASN	GLY	GLY
D^2-code	7	0	4	4	7	3	6	6	1
No.	18	19	20	21	22	23	24	25	26
Amino acid	VAL	CYS	MET	HIS	ILE	GLU	SER	LEU	ASP
D^2-code	4	3	4	5	5	7	0	2	3
No.	27	28	29	30	31	32	33	34	35
Amino acid	SER	TYR	THR	CYS	ASN	CYS	VAL	ILE	GLY
D^2-code	3	3	4	1	4	3	4	0	2
No.	36	37	38	39	40	41	42	43	44
Amino acid	TYR	SER	GLY	ASP	ARG	CYS	GLN	THR	ARG
D^2-code	0	4	0	6	0	1	7	1	6
No.	45	46	47	48	49	50	51	52	53
Amino acid	ASP	LEU	ARG	TRP	TRP	GLU	LEU	ARG	–
D^2-code	6	0	2	6	4	7	3	–	–

$$
\begin{aligned}
&({x_0}^2/(x_1{x_2}^2{x_3}^2) + {x_0}^2/(x_2{x_3}^3) + x_0/(x_1{x_2}^2x_3) \\
&\qquad\qquad + 1/(x_1x_2x_3), [51, 67]), \\
&({x_0}^2/(x_2{x_3}^3) + {x_0}^2/(x_1{x_2}^2{x_3}^2) + x_0/(x_1x_2{x_3}^2), [59, 68]), \\
&({x_0}^2/(x_1{x_3}^4) + {x_0}^2/(x_1{x_2}^2{x_3}^2) + 1/({x_1}^2x_2{x_3}^3), [66, 77]), \\
&(x_0/(x_1{x_3}^3) + {x_0}^2/(x_1{x_3}^4) + {x_0}^2/(x_2{x_3}^4) \\
&\qquad\qquad + x_0/(x_2{x_3}^3), [75, 82]), \\
&({x_0}^2/(x_2{x_3}^4) + x_0/(x_1x_2{x_3}^3) + x_0/({x_2}^2{x_3}^3), [80, 85]), \\
&(1/({x_2}^2{x_3}^4) + x_0/(x_1x_2{x_3}^3) + {x_0}^2/({x_2}^3{x_3}^3), [83, 95]), \\
&(x_0/({x_2}^2{x_3}^3) + {x_0}^3/({x_2}^3{x_3}^4) + {x_0}^2/({x_2}^3{x_3}^3) \\
&\qquad\qquad + {x_0}^5{x_1}^2/(x_2{x_3}^5), [90, 107]),
\end{aligned}
$$

$$
\begin{aligned}
&({x_0}^4{x_1}^2/(x_2{x_3}^4) + {x_0}^5{x_1}^2/(x_2{x_3}^5) \\
&\qquad + {x_0}^6{x_1}^3/({x_2}^2{x_3}^3), [105, 112]), \\
&({x_0}^5{x_1}^5/(x_2{x_3}^2) + {x_0}^6{x_1}^3/({x_2}^2{x_3}^3) \\
&\qquad + {x_0}^6{x_1}^3/(x_2{x_3}^4), [110, 120]), \\
&({x_0}^6{x_1}^4/(x_2{x_3}^2) + {x_0}^5{x_1}^5/(x_2{x_3}^2) + {x_0}^6{x_1}^7/{x_3}^4 \\
&\qquad + {x_0}^5{x_1}^7x_2/{x_3}^4, [115, 132]), \\
&({x_0}^6{x_1}^7/{x_3}^4 + {x_0}^5{x_1}^6x_2/{x_3}^4 + {x_0}^4{x_1}^7x_2/{x_3}^4, [127, 135]), \\
&({x_0}^6{x_1}^8{x_2}^2/{x_3}^5 + {x_0}^5{x_1}^6x_2/{x_3}^4 + {x_0}^4{x_1}^7/{x_3}^4, [133, 145]), \\
&({x_0}^5{x_1}^8x_2/{x_3}^4 + {x_0}^6{x_1}^7x_2/{x_3}^4 + {x_0}^6{x_1}^7{x_2}^2/{x_3}^5, [140, 151]), \\
&({x_0}^5{x_1}^8x_2/{x_3}^4 + {x_0}^6{x_1}^7x_2/{x_3}^4 + {x_0}^5{x_1}^7{x_2}^2/{x_3}^5 \\
&\qquad + {x_0}^6{x_1}^7{x_2}^2/{x_3}^6, [143, 154]), \\
&({x_0}^5{x_1}^7{x_2}^2/{x_3}^5 + {x_0}^6{x_1}^7x_2/{x_3}^6, [152, 155])\}.
\end{aligned}
$$

The shape of the 3-dim. implementation of the D^2-code is shown in Fig.7(a). The experimentally obtained shape of 1EPI is shown in Fig.7(b). Using Table 3, the D^2-code of amino-acids is given by Table 5.

References

1. D.Baker, *A surprising simplicity to protein folding*, Nature 405, 39-42(2000)
2. C.Branden and J.Tooze, Introduction to Protein Structure. Garland Publishing Inc., New York. 1998.
3. A.P.Coppa, *Three-dimensional folded chain structures*, US Patent 4,142,321 Mar.6, 1979.
4. C.M.Dobson and M. Karplus, *The foundation of protein folding: bringing together theory and experiment*, Curr. Opin. Struct. Biol. 9, 92-101(1999).
5. K.Keirshenbaum, R.N.Zuckermann and K.A.Dill, *Designing polymers that mimic biomolecules*, Curr. Opin. Struct. Biol. 9, 530-535(1999).
6. A.Maritan, C.Micheletti, A.Trovato and J.R.Banavar, *Optimal shapes of compact strings*, Nature 406, 287-290(2000).
7. W.R.Taylor, *A deeply knotted protein structure and how it might fold*, Nature 406, 916-919(2000).
8. W.R.Taylor, *A 'periodic table' for protein structure*, Nature 416, 657-660(2002).
9. C.A.Orengo, et al. Nucleic Acids REs. 27, 275-279(1999).
10. B.Robson, *Beyond proteins*, TIBTECH Aug.8 1999(vol. 17), 311-315.
11. D.Vitkup, E.Melamud, J.Moult and C.Sander, *Completeness in structural genomics*, Nature Struct. Biol. 8, 559-565(2001).
12. D.M.Webster(ed.), Methods in Molecular Biology 143, PROTEIN STRUCTURE PREDICTION *Methods and protocols.* Humana Press Inc. Totowa, New Jersey. 2000.
13. CATH, Univ. College, London, UK, http://www.biochem.ucl.ac.uk/bsm/cath/
14. SCOP, Univ. of Cambridge UK, http://scop.mrc-lmb.cam.ac.uk/scop/
15. Structural Genomics. Nature Struct. Biol. 7 supplement(2000).
16. The Protein Data bank, http://www.rcsb.org/pdb/

Symbol index

Subject index