

Article

# Dihedral Reductions of Cyclic DNA Sequences

Marlos A.G. Viana

Symmetry Studies Laboratory, University of Illinois at Chicago Eye Center, Chicago, IL 60612, USA;  
E-Mail: viana@uic.edu

Academic Editor: Slavik Jablan

Received: 26 January 2014 / Accepted: 18 December 2014 / Published: 16 January 2015

---

**Abstract:** The data-analytic methodology of dihedral reductions for cyclic orbits of distinct-base codons is described both in terms of Fourier analysis over the dihedral groups and in (algebraically equivalent) terms of canonical projections. Numerical evaluations are presented for discrete and continuous scalar data indexed by cyclic orbits.

**Keywords:** Fourier analysis; group representations; dihedral groups; cyclic orbits; canonical projections; nucleotides; group orbits; invariance

---

## 1. Introduction

The role of group-theoretic arguments in biology has a long history, in which R.A. Fisher's classification of segregation genotypes in the theory of polysomic inheritance is a classical example [1]. In the theory of experimental designs, it was also Fisher who demonstrated the explicit usefulness of cyclic groups in the theory of confounding in factorial experiments [2,3], now widely used in biology and genetics studies. In recent decades, the theory and applications of algebraic methods in statistics and probability became a well-established area of interest, e.g., [4,5].

In structural biology as well, applications of symmetry arguments have been used to formulate working hypotheses and to suggest explanation and prediction, e.g., [6–10]. An explicit connection between symmetry arguments and data analytic reasoning in structural biology can be exemplified by the study [11] of the evolutionary importance of purine and pyrimidine content in the human immunodeficiency virus type 1, based on the statistical assessment of the frequency diversity of cyclic sets, defined as the ratio

$$x_{\mathcal{O}} = \frac{\max_{\mathcal{O}} f}{\min_{\mathcal{O}} f} \quad (1)$$

of extreme (max/min) frequency counts ( $f$ ) in the cyclic set  $\mathcal{O}$  evaluated over a given region of the genome. In the context of symmetry studies, the frequency diversity is just but one of the many possible data summaries *indexed* by cyclic sets or orbits. In the present communication, the frequency diversity, as well as cyclic summaries such as the raw sum

$$x_{\mathcal{O}} = \sum_{\mathcal{O}} f \quad (2)$$

of frequency counts along the orbit, will be shown to share similar algebraic and data-analytic structures.

More specifically, the present communication is aimed at showing that there is a broader group-theoretic and data-analytic framework within which the methodology described in [11] can be identified and further utilized, thus leading to eventually richer biological interpretations and explanatory narratives.

The framework of interest (Symmetry Studies) was described originally in [12] and is briefly reviewed in [13]. We will also refer to [14] for notions of Fourier analysis over the finite groups relevant to the present applications. See also [15–18] for related discussions, and [19,20] for applications in the field of linear optics.

This paper is divided as follows. The basic definitions, assumptions and notations are introduced in the next section. The cyclic reductions are discussed in Section 3. Numerical evaluations are presented in Section 4. Additional background material is presented in Appendices A, B, and C.

## 2. Definitions, Assumptions and Notation

Any DNA sequence in length of  $\ell$  base pairs (bp) can be represented as a point in the set  $\mathcal{A}^L$  of all mappings

$$S : L = \{1, \dots, \ell\} \mapsto \mathcal{A} = \{\text{A,G,C,T}\}$$

where, typically, permutations  $\tau$  in subgroups  $G$  of the full symmetric group  $S_{\ell}$  act on the left according to

$$S\tau^{-1}. \quad (3)$$

The inverse of the group element appearing in the group action  $\tau \cdot s = s\tau^{-1}$  is necessary so its defining property  $\eta \cdot (\tau \cdot s) = (\eta\tau) \cdot s$  can be verified. Permutations  $\sigma$  in subgroups  $H$  of  $S_4$  act on the right according to

$$\sigma S, \quad (4)$$

and subgroups  $G \times H$  of the direct product group  $S_{\ell} \times S_4$  act bilaterally according to

$$\sigma S\tau^{-1}. \quad (5)$$

In what follows, we will indicate by

$$s = \{S\gamma^{-1}; \gamma \in C_{\ell}\} \quad (6)$$

the orbit generated by the left action of the cyclic group  $C_{\ell}$  on a given sequence  $S \in \mathcal{A}^L$ . We will refer, generically, to these sets as cyclic orbits or cyclic sets.

Throughout this communication, sequences written in lower case will always indicate the cyclic orbit generated by the corresponding sequence, to be written in upper case. For example,

$$act = \{ACT, CTA, TAC\}, \quad agct = \{AGCT, GCTA, CTAG, TAGC\}.$$

It follows directly from (4) that two sequences  $S$  and  $F$  are *complementary* if

$$F = \sigma S,$$

with

$$\sigma = (AT)(GC)$$

the permutation in  $S_4$ , written in cycle notation, representing the standard DNA complementarity of base pairs. For example, ACT and TGA are complementary sequences, and in that case we also say that their corresponding orbits  $act$  and  $tga$  are also complementary. Specifically,

$$F = \sigma S \Rightarrow f = \{F\gamma^{-1}; \gamma \in C_\ell\} = \{\sigma S\gamma^{-1}; \gamma \in C_\ell\} = \sigma\{S\gamma^{-1}; \gamma \in C_\ell\} = \sigma f.$$

### 2.1. Injective Sequences

The dihedral reductions to be considered in this communication are obtained for DNA sequences in length of three (or codons) composed of distinct bases  $\{A, G, C, T\}$ . That is, for the *injective* mappings into  $\mathcal{A}$  with domain  $L = \{1, 2, 3\}$ . These sequences account for the 24 distinct injective codons factored into 8 distinct cyclic orbits of length three.

Although the group actions (3)–(5) are defined for all mappings in  $\mathcal{A}^L$ , the resulting data-analytic applications may need to be adapted when non-injective sequences are included, due to the fact that the resulting actions may no longer be transitive. In that case, the data analysis is carried piecewise within the transitive parts [12]. In addition, because the actions on the injective sequences are faithful, any experimental results indexed by the points in the orbit are in one-to-one correspondence with the group elements, and can consequently be indexed by the group elements themselves. It is in the resulting group algebra structure that Fourier transforms can naturally be defined.

### 2.2. Scalar Measurements

Throughout this paper it will be opportune to distinguish the following types of experimental data:

- Data indexed by sequences,  $x : \mathcal{A}^L \mapsto \mathbb{R}$ , indicated by  $x_S$ ;
- Data indexed by cyclic orbits,  $x : \mathcal{O} \mapsto \mathbb{R}$ , indicated by  $x_s$ ;
- Data indexed by group elements,  $x : G \mapsto \mathbb{R}$ , indicated by  $x_\tau, x_\sigma, \dots$

For example, the frequency diversity Equation (1) for  $act$  in terms of frequency counts  $x_S$  over a given region of the genome is given by

$$x_{act} = \frac{\max\{x_{ACT}, x_{CTA}, x_{TAC}\}}{\min\{x_{ACT}, x_{CTA}, x_{TAC}\}}, \quad (7)$$

whereas the raw sum Equation (2) for the same orbit is

$$x_{act} = x_{ACT} + x_{CTA} + x_{TAC}. \quad (8)$$

### 2.3. Orbit Invariance

Every symmetry orbit has an intrinsic arbitrariness in the choice of its generating point, so that the resulting orbit is the same regardless of its generator. For example, recalling Equation (6),

$$act = cta = tac.$$

Therefore, one would want the corresponding data summaries

$$x_{act}, \quad x_{cta}, \quad x_{tac}$$

to be *stable*, or *invariant*, under different choices of orbit generators. Obviously Equations (7) and (8) are both orbit invariants. This is a universal requirement that applies to all summaries obtained from data indexed by symmetry orbits.

A class of data summaries with this (orbit) invariance property, as shown in [14], is given precisely by the Fourier transforms

$$\langle x, \xi \rangle = \sum_{\tau \in G} x_{\tau} \xi_{\tau},$$

evaluated at the (irreducible) representations  $\xi$  of  $G$ . The invariance property says that, regardless of the different orbit relabelings  $\tau x$ , their Fourier transforms  $\langle \tau x, \xi \rangle$  stay bound to certain well-defined (irreducible representation) subspaces of the original data module or vector space [14]. That is,

$$\langle \tau x, \xi \rangle = \xi_{\tau} \langle x, \xi \rangle, \quad (9)$$

so that the transforms *reduce* as the corresponding irreducible characters.

A class of (faithful) group actions on the cyclic orbits that allows us to identify  $x_{\tau}$  with  $x_s$  and evaluate the Fourier transforms (or orbit invariants) will be introduced in the next section.

### 2.4. Dihedral Orbits

The dihedral groups  $D_n$ , for  $n = 3, 4, \dots$ , can be realized as the group

$$C_n = \{1, r, r^2, \dots, r^{n-1}\}$$

of rotations of a regular  $n$ -side polygon, adjoined with the corresponding reversals

$$C_n h = \{h, rh, r^2 h, \dots, r^{n-1} h\},$$

or  $h$ -mirrored rotations, giving  $D_n$  a (non-commutative) group structure of order  $2n$ . In addition, when  $n = 2$ ,

$$D_2 = \{1, r, h, rh\} \simeq C_2 \times C_2,$$

also known as the Klein four-group, is often considered as one of the dihedral groups. This commutative group, when realized as the symmetries of a regular rectangle, describes *vertical* and *horizontal* line reflections and their two-fold  $180^\circ$  rotation composition. For completion, we also include  $D_1 = \{1, h\}$ , where  $h^2 = 1$ , and  $D_0 = \{1\}$  into the dihedral class as well.

The action (3) under  $G = D_3$  gives four distinct dihedral orbits

$$act \cup tca, \quad tga \cup agt, \quad gct \cup tcg, \quad cga \cup agc,$$

where

$$\begin{aligned} act &= \{ACT, CTA, TAC\}, \quad tca = \{TCA, CAT, ATC\}, \\ tga &= \{TGA, GAT, ATG\}, \quad agt = \{AGT, GTA, TAG\}, \\ gct &= \{GCT, CTG, TGC\}, \quad tcg = \{TCG, CGT, GTC\}, \end{aligned}$$

and

$$cga = \{CGA, GAC, ACG\}, \quad agc = \{AGC, GCA, CAG\}$$

are the cyclic orbit components, accounting for the 24 distinct injective codons.

Similarly, the action (3) under  $G = D_4$  gives three distinct dihedral orbits

$$actg \cup gtca, \quad agct \cup tcga, \quad atgc \cup cgta,$$

where

$$\begin{aligned} actg &= \{ACTG, CTGA, TGAC, GACT\}, \quad gtca = \{GTCA, TCAG, CAGT, AGTC\}, \\ agct &= \{AGCT, GCTA, CTAG, TAGC\}, \quad tcga = \{TCGA, CGAT, GATC, ATCG\}, \end{aligned}$$

and

$$atgc = \{ATGC, TGCA, GCAT, CATG\}, \quad cgta = \{CGTA, GTAC, TACG, ACGT\}$$

are the cyclic orbit components, accounting for the 24 distinct injective sequences in length of four.

### 3. Invariant Reductions

In Diagrams (10) and (11),  $D_3$  rotations and reversals are shown sideways along the rows of the diagrams and complementary orbits are shown along columns, so that each box is labeled by a cyclic orbit. We shall refer to the orbits in each of the diagrams simply as *conjugated* orbits.

In addition, each orbit is labeled by the **polarity** ( $\oplus, \ominus$ ) of the sequence's strand and by the encoding sense ( $\rightarrow$ ) or anti-sense ( $\leftarrow$ ) **direction** with which a gene or protein product reads off the sequence. More specifically, following (10), if any point in the *act* orbit is labeled with a **positive** polarity and with a reading **sense** direction *then*:

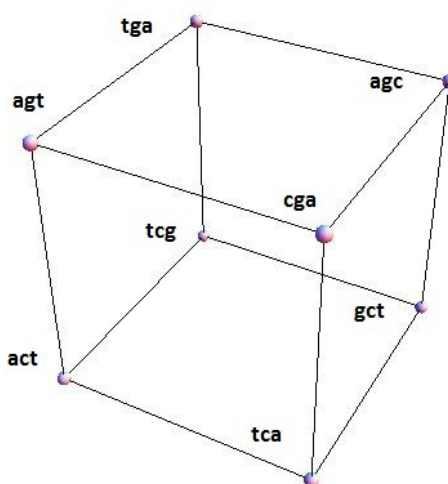
- The corresponding point in the *tca* orbit has **positive** polarity and the reading is in the **anti-sense** direction;
- The corresponding point in the *tga* orbit has **negative** polarity and the reading is in the **sense** direction, and;
- The corresponding point in the *agt* orbit has **negative** polarity and the reading is in the **anti-sense** direction.

$$\begin{array}{ccc}
 \boxed{act} & \oplus, \rightarrow & \boxed{tca} \\
 \parallel & & \parallel \\
 \boxed{tga} & \ominus, \rightarrow & \boxed{agt} \\
 & \oplus, \leftarrow & \oplus, \leftarrow
 \end{array} \quad (10)$$

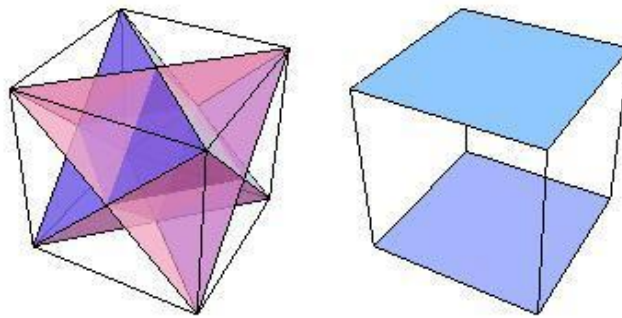
Diagram (11) shows the complementary orbits *gct* and *agc*, with the same polarity and direction interpretation as in Diagram (10).

$$\begin{array}{ccc}
 \boxed{gct} & \oplus, \rightarrow & \boxed{tcg} \\
 \parallel & & \parallel \\
 \boxed{cga} & \ominus, \rightarrow & \boxed{agc} \\
 & \oplus, \leftarrow & \oplus, \leftarrow
 \end{array} \quad (11)$$

Figure 1 shows a configuration space for the conjugated cyclic orbits of Diagrams (10) and (11), relative to which Figure 2 shows, respectively on the left and right images, the common direction and common polarity configuration subspaces. In this configuration space (obviously not unique), same-direction subspaces span two intercepting tetrahedrons, whereas same-polarity subspaces span two parallel faces of the configuration space.



**Figure 1.** A configuration space for the conjugated cyclic orbits.



**Figure 2.** Common-direction (**left**) and common-polarity (**right**) configuration subspaces for the conjugated cyclic orbits of Diagrams (10) and (11).

### 3.1. $D_2$ -Invariant Reductions

There is a transitive faithful action of  $C_2 \times C_2 \simeq D_2$  on the cyclic orbits  $s$  of (10), given by

$$\sigma S \tau^{-1}, \quad S \in s,$$

with

$$\sigma \in \{1, (\text{AT})(\text{GC})\} \simeq C_2, \quad \tau \in \{1, (13)\} \simeq C_2.$$

Specifically,

$\tau$	$\sigma$	$act$	$tca$	$tga$	$agt$
1	1	$act$	$tca$	$tga$	$agt$
1	$(\text{AT})(\text{GC})$	$tga$	$agt$	$act$	$tca$
(13)	1	$tca$	$act$	$agt$	$tga$
(13)	$(\text{AT})(\text{GC})$	$agt$	$tga$	$tca$	$act$

As a consequence, any experimental data  $x_s$  indexed by the orbits ( $s$ ) in the diagram can be reduced by the tools of dihedral Fourier analysis over  $D_2$ . We emphasize that the transitivity and faithfulness of the  $D_2$  action on the set of orbits is necessary to identify the orbits with the group elements and then proceed to the determination of the ( $D_2$ ) orbit invariants using the Fourier transforms. Following [14], these four one-dimensional transforms are simply

$$x_{act} + x_{tca} + x_{tga} + x_{agt},$$

$$x_{act} + x_{tca} - x_{tga} - x_{agt},$$

$$x_{act} - x_{tca} + x_{tga} - x_{agt},$$

$$x_{act} - x_{tca} - x_{tga} + x_{agt},$$

and constitute a set of one-dimensional orbit invariants for the data [14]. Similar reductions can then be obtained for (11).

### 3.2. Entropy Invariants

The orbit invariants determined above are functions of *any* scalars  $x_s$  obtained over the orbit  $s$ , such as its diversity (7), its raw sum (8), or its total molecular weight. When  $x_s$  are positive integers, such as the sum of frequency counts over the orbit  $s$ , then the entropy (Ent) of the observed distributions ( $\mathcal{L}$ ) of frequency counts given by,

$$\text{Ent } \mathcal{L}_1 = \text{Ent} (x_{act}, x_{tca}, x_{tga}, x_{agt}), \quad (12)$$

$$\text{Ent } \mathcal{L}_{act+tca} = \text{Ent} (x_{act} + x_{tca}, x_{tga} + x_{agt}), \quad (13)$$

$$\text{Ent } \mathcal{L}_{act+tga} = \text{Ent} (x_{act} + x_{tga}, x_{tca} + x_{agt}), \quad (14)$$

and

$$\text{Ent } \mathcal{L}_{act+agt} = \text{Ent} (x_{act} + x_{agt}, x_{tca} + x_{tga}) \quad (15)$$

are also orbit invariants [17,21]. In Section 4 these orbit invariants are evaluated for Diagrams (10) and (11) to describe a specific DNA sequence and query it for potential structural variations along the genome.

### 3.3. $D_4$ -Invariant Reductions

There are three non-equivalent transitive right actions  $\sigma s$  of  $D_4$  on the set

$$\mathcal{O}_3 = \{act, tca, tga, agt, gct, tcg, cga, agc\}$$

of all injective cyclic orbits in length of three, jointly reducing Diagrams (10) and (11), generated by:

- $D_4 \simeq \langle (ATGC), (AG) \rangle$ ,
- $D_4 \simeq \langle (AGCT), (AC) \rangle$ ,
- $D_4 \simeq \langle (ACTG), (AT) \rangle$ .

The action of  $D_4 \simeq \langle (ATGC), (AG) \rangle$  is given by:

		<i>act</i>	<i>tca</i>	<i>tga</i>	<i>agt</i>	<i>gct</i>	<i>tcg</i>	<i>cga</i>	<i>agc</i>
	$\sigma$	1	2	3	4	5	6	7	8
1	1	1	2	3	4	5	6	7	8
$r$	(ATGC)	4	3	5	6	8	7	1	2
$r^2$	(AG)(TC)	6	5	8	7	2	1	4	3
$r^3$	(ACGT)	7	8	2	1	3	4	6	5
$h$	(AG)	5	6	4	3	1	2	8	7
$rh$	(AT)(GC)	3	4	1	2	7	8	5	6
$r^2h$	(TC)	2	1	7	8	6	5	3	4
$r^3h$	(AC)(TG)	8	7	6	5	4	3	2	1

(16)

thus showing that any experimental data  $x_s$  indexed by  $\mathcal{O}_3$  can be reduced by dihedral Fourier transforms over  $D_4$  or the corresponding canonical projections decompositions. These two views, both leading to the identification of the orbit invariants, are outlined next.



### 3.4. Canonical Projections

The linear representation of  $D_4$  in  $\mathbb{R}^8$  defined by (16) is given by the permutation matrices associated with the rotations

$$\{1, (1467)(2358), (16)(47)(25)(38), (1764)(2853)\},$$

and with the reversals

$$\{(15)(26)(34)(78), (12)(37)(48)(56), (13)(24)(57)(68), (18)(27)(36)(45)\},$$

and hence generated by

$$< (1467)(2358), (15)(26)(34)(78) > \simeq D_4. \quad (17)$$

Similarly, the action generated by  $D_4 \simeq < (\text{AGCT}), (\text{AC}) >$  yields a linear representation of

$$< (1485)(2376), (12)(35)(46)(78) > \simeq D_4, \quad (18)$$

whereas the action generated by  $D_4 \simeq < (\text{ACTG}), (\text{AT}) >$  yields a linear representations of

$$< (1537)(2648), (12)(34)(58)(67) > \simeq D_4. \quad (19)$$

The resulting canonical projections  $\mathcal{P}_\xi$ , indexed by the irreducible representations

$$\xi \in \{1, \alpha, \gamma_+, \gamma_-, \beta\}$$

of  $D_4$ , evaluated for the representation of (17), are shown in Appendix C, along with the components  $x' \mathcal{P}_\xi x$  of the resulting decomposition

$$||x||^2 = \sum_{\xi} x' \mathcal{P}_\xi x \quad (20)$$

of the sum of squares. More generally,  $||x||^2 = \sum_{\xi} x' \mathcal{P}_\xi \Phi x$ , where  $x' \Phi x$  is an Euclidean fundamental form [22]. These decompositions are often used for the statistical analysis of continuous data (analysis of variance).

#### 3.4.1. Interpretation of the Components $x' \mathcal{P}_\xi x$

The particular representation (17) leads to the following interpretation of each of the non-trivial (orbit invariant) components  $x' \mathcal{P}_\xi x$  of  $||x||^2$ , in terms of the combinations of polarity ( $\oplus, \ominus$ ) and direction ( $\rightarrow, \leftarrow$ ),

$$x' = \underbrace{\left( \frac{\oplus}{\rightarrow}, \frac{\ominus}{\leftarrow}, \frac{\oplus}{\leftarrow}, \frac{\ominus}{\rightarrow} \right)}_{\text{rotations}}, \underbrace{\left( \frac{\oplus}{\rightarrow}, \frac{\ominus}{\rightarrow}, \frac{\oplus}{\leftarrow}, \frac{\ominus}{\leftarrow} \right)}_{\text{reversals}}$$

corresponding to the components of

$$x' = \underbrace{(act, agt, tcg, cga)}_{\text{rotations}}, \underbrace{(gct, tga, tca, agc)}_{\text{reversals}}$$

where, here for simplicity of notation, we let  $s$  (the labels) indicate  $x_s$  (the data indexed by that labels).

The projection  $\mathcal{P}_\alpha$  identifies a one-dimensional invariant comparing the overall mean effects

$$\left(\frac{\oplus}{\rightarrow} + \frac{\ominus}{\leftarrow} + \frac{\oplus}{\leftarrow} + \frac{\ominus}{\rightarrow}\right) \text{ vs. } \left(\frac{\oplus}{\rightarrow} + \frac{\ominus}{\rightarrow} + \frac{\oplus}{\leftarrow} + \frac{\ominus}{\leftarrow}\right)$$

between rotations and reversals. The projection  $\mathcal{P}_{\gamma_+}$  identifies a one-dimensional invariant *combining* the overall within-rotation sensitivity to *polarity given overall direction* variation as assessed by

$$\left(\frac{\oplus}{\rightarrow} - \frac{\ominus}{\leftarrow} + \frac{\oplus}{\leftarrow} - \frac{\ominus}{\rightarrow}\right)$$

with the corresponding within-reversal variation assessed by

$$\left(\frac{\oplus}{\rightarrow} - \frac{\ominus}{\rightarrow} + \frac{\oplus}{\leftarrow} - \frac{\ominus}{\leftarrow}\right).$$

The projection  $\mathcal{P}_{\gamma_-}$  identifies a one-dimensional invariant *contrasting* the same variation described above. Lastly, the projection  $x'\mathcal{P}_\beta x$  identifies a two-dimensional invariant assessing *direction given polarity* effects in terms of

$$\left(\frac{\oplus}{\rightarrow} - \frac{\oplus}{\leftarrow}\right), \text{ and } \left(\frac{\ominus}{\rightarrow} - \frac{\ominus}{\leftarrow}\right).$$

### 3.5. Dihedral Fourier Analysis

Reading from the column under the *act* orbit in (16), the points in the group algebra  $\mathbb{C}D_4$  are given by

$$x = act \, 1 + agt \, r + tcg \, r^2 + cga \, r^3 + gct \, h + tga \, rh + tca \, r^2h + agc \, r^3h, \quad (21)$$

from which we obtain the corresponding Fourier transforms

$$\begin{aligned} \langle x, 1 \rangle &= act + agt + tcg + cga + gct + tga + tca + agc, \\ \langle x, \alpha \rangle &= act + agt + tcg + cga - (gct + tga + tca + agc), \\ \langle x, \gamma_+ \rangle &= (act - agt + tcg - cga) + (gct - tga + tca - agc), \\ \langle x, \gamma_- \rangle &= (act - agt + tcg - cga) - (gct - tga + tca - agc), \\ \langle x, \beta \rangle &= \begin{pmatrix} act + gct - tca - tcg & -agc - agt + cga + tga \\ -agc + agt - cga + tga & act - gct + tca - tcg \end{pmatrix}. \end{aligned}$$

Parseval's equality

$$\|x\|^2 = \sum_{\xi \in \hat{D}_n} \frac{n_\xi}{2n} \|\langle x, \xi \rangle\|^2$$

establishes the correspondence with the decomposition Equation (20) obtained in terms of the canonical projections.

It is opportune to remark here that in the definition of Equation (21) we arbitrarily assigned the identity in  $D_4$  to  $x_{act}$ . Any of the other potential assignments would be precisely a relabeling of the orbit's starting point. The Fourier transforms, however, would remain orbit invariant, in the sense of Equation (9).

## 4. Numerical Evaluations

In this section we apply the cyclic reductions described in Section 3 to specific complete genomes of the human immunodeficiency virus type 1 and the hepatitis C virus.

#### 4.1. Relative Entropy Study of the HIV1 BRUCG Isolate

Following Section 3.2, the data indexed by the cyclic orbits are simply the sums  $x_s$  of the frequency counts  $x_S$  with which the sequence  $S$  occurs in a given region of the genome, that is,

$$x_s = \sum_{S \in s} x_S.$$

The frequency counts were evaluated by scanning the genome one base at a time in the 5′–3′ direction. The sequence in FASTA format was downloaded from the NCBI website (<http://www.ncbi.nlm.nih.gov>). Computations were evaluated using the Symmetry Computing Toolbox (Symmetry Computing Toolbox, ©M.Viana). This particular HIV1 isolate, used here for numerical illustration only, also appears in the study of the HIV1’s evolutionary properties [11].

The frequency counts were obtained for the complete genome of human immunodeficiency virus type 1, isolate BRU (LAV-1), sequence ID gi:326417, accession number K02013.1, HIVBRUCG. See [23]. The full 9229 bp-long sequence was partitioned into six equal-length adjacent regions numbered 1–6, where the cyclic summaries  $x_s$  were evaluated. The frequency counts for the conjugated cyclic orbits of *act* corresponding to Diagram (10) are shown in (22) and (23), whereas in (24) and (25) show the frequency counts for the conjugated cyclic orbits of *gct* corresponding to Diagram (11).

$s$	$S$	1	2	3	4	5	6
<i>act</i>	ACT	17	26	16	23	20	22
	CTA	24	16	18	27	14	20
	TAC	12	25	23	16	23	22
<i>tca</i>	TCA	23	25	25	24	22	18
	CAT	25	23	25	35	20	16
	ATC	19	20	21	20	18	17

(22)

$s$	$S$	1	2	3	4	5	6
<i>tga</i>	TGA	19	22	17	20	20	23
	GAT	25	30	21	27	18	26
	ATG	27	27	28	39	29	18
<i>agt</i>	AGT	27	30	45	30	32	25
	GTA	20	29	33	26	37	19
	TAG	33	34	42	48	30	31

(23)

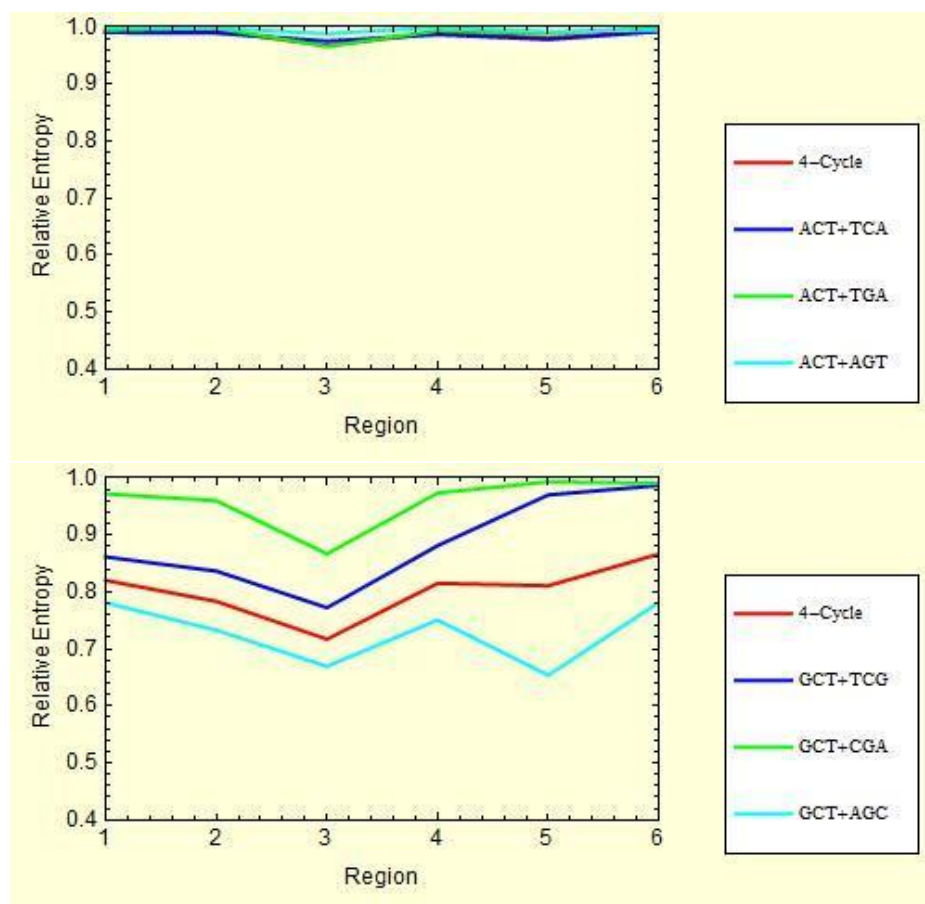
$s$	$S$	1	2	3	4	5	6
<i>gct</i>	GCT	21	10	8	15	20	38
	CTG	17	21	15	17	28	36
	TGC	18	9	10	17	23	22
<i>tcg</i>	TCG	5	1	2	1	0	6
	CGT	2	1	1	1	3	3
	GTC	7	6	8	9	9	5

(24)

$s$	$S$	1	2	3	4	5	6
$cga$	CGA	10	2	1	4	2	12
	GAC	29	26	19	26	15	26
	ACG	4	1	3	2	6	7
$agc$	AGC	44	24	34	35	28	40
	GCA	38	21	40	34	35	23
	CAG	51	58	53	39	39	36

(25)

Figure 3 shows the resulting relative entropy invariants, as defined in Section 3.2, both for the *act* and for the *gct* conjugated cyclic orbits.



**Figure 3.** The top profiles show (respectively to the legends shown top to bottom) the relative entropy for the joint distributions of  $(act, tca, tga, agt)$ ;  $(act + tca, tga + agt)$ ;  $(act + tga, tca + agt)$ , and  $(act + agt, tca + tga)$ . The bottom profiles display the corresponding results for the conjugated orbits of *gct*.

For example, reading from Region 3, in Equations (24) and (25), we have

$$x_{gct} = 8 + 15 + 10 = 33, \quad x_{agc} = 34 + 40 + 53 = 127,$$

$$x_{tcg} = 2 + 1 + 8 = 11, \quad x_{cga} = 1 + 19 + 3 = 23,$$

so that the resulting relative entropy invariants, from Equations (12)–(15), are given by

$$\begin{aligned}\text{Ent}(33, 127, 11, 23) / \max \text{Ent} &= 0.717, \\ \text{Ent}(33 + 11, 127 + 23) / \max \text{Ent} &= 0.772, \\ \text{Ent}(33 + 127, 11 + 23) / \max \text{Ent} &= 0.669, \\ \text{Ent}(33 + 23, 127 + 11) / \max \text{Ent} &= 0.866,\end{aligned}$$

as displayed in the profiles shown in the bottom part of Figure 3. The top part of Figure 3 shows the corresponding profiles for the GCT conjugate cyclic orbits.

### Interpretations:

Reading again from Diagrams (10) and (11), it follows that the invariant

$$\text{Ent } \mathcal{L}_{act+tca} = \text{Ent}(act + tca, tga + agt)$$

introduced in Equation (13) is a measure of **polarity** uncertainty; the invariant

$$\text{Ent } \mathcal{L}_{act+tga} = \text{Ent}(act + tga, tca + agt)$$

in Equation (14) is a measure of **direction** uncertainty, whereas the invariant

$$\text{Ent } \mathcal{L}_{act+agt} = \text{Ent}(act + agt, tca + tga)$$

introduced in Equation (15) is a measure of the interaction between direction and polarity. In summary:

- Polarity uncertainty:  $\text{Ent } \mathcal{L}_{act+tca} = \text{Ent}(act + tca, tga + agt)$ ;
- Direction uncertainty:  $\text{Ent } \mathcal{L}_{act+tga} = \text{Ent}(act + tga, tca + agt)$ ;
- Interaction:  $\text{Ent } \mathcal{L}_{act+agt} = \text{Ent}(act + agt, tca + tga)$ .

### 4.2. Statistical Assessment

The statistical assessment of the entropy can be obtained by numerically evaluating its sampling distribution based on 10,000 randomly generated observations from the posterior (Beta) distribution conjugated to binomial likelihood for the data, relative to the uniform prior probability distribution. Based on the resulting sampling distribution, a numerical evaluation of a posterior 95% credibility interval (CI) for the relative entropy can be obtained.

For example, reading from Region 5, in (24) and (25), we have,

$$x_{gct} = 71, \quad x_{tcg} = 12, \quad x_{cga} = 23, \quad x_{agc} = 102,$$

so that, from Equations (12)–(15), the polarity relative entropy is

$$\text{Ent}(83, 125) / \max \text{Ent} = 0.97, \quad 95\% \text{ CI} = (0.921518, 0.996971)$$

the direction relative entropy is,

$$\text{Ent}(94, 114) / \max \text{Ent} = 0.99, \quad 95\% \text{ CI} = (0.962473, 0.999979),$$

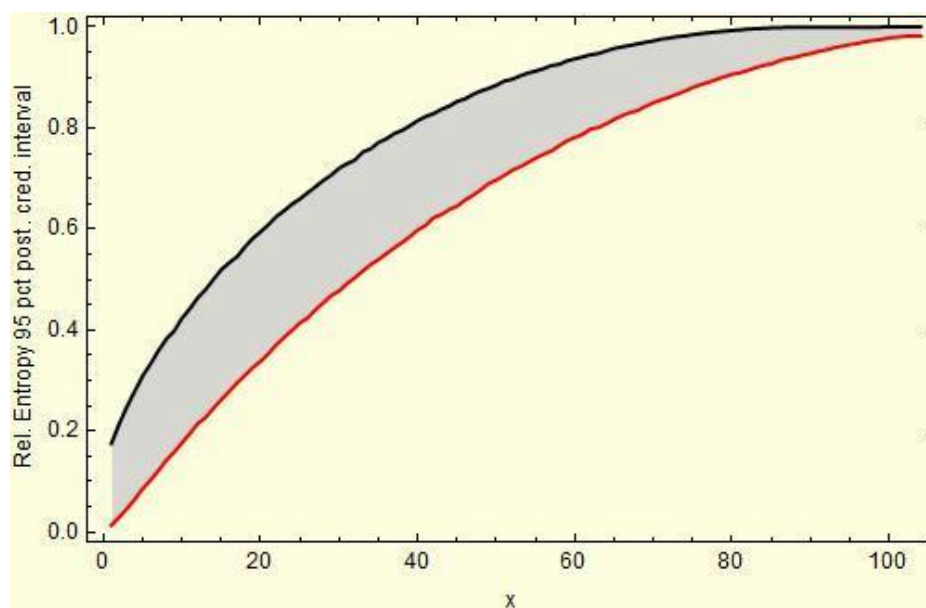
and

$$\text{Ent}(173, 35) / \max \text{Ent} = 0.65, \quad 95\% \text{ CI} = (0.540566, 0.769399)$$

is the relative entropy for the polarity-direction interaction or residual term. The credibility intervals thus suggest that the drop in polarity uncertainty in Region 5 is statistically distinct from the other two uncertainties and from the relative entropy of a uniform distribution of the same total binomial size, namely (104, 104), which is (0.983009, 0.999997). Figure 4 shows the posterior 95% credibility bands for the relative entropy of binomial distributions of total sample size

$$n = x_{gct} + x_{tcg} + x_{cga} + x_{agc} = 208,$$

from which the above credibility intervals can be identified. The range of the monogram is half the total binomial sample because of the entropy (orbit) invariance property.



**Figure 4.** Posterior 95% credibility intervals for the relative entropy of binomial distributions of total sample size  $n = 208$ .

#### 4.3. Orbit Diversity Decomposition for the HIV1 Samples

In this section we apply the canonical decomposition introduced in Section 3.4 and evaluated in Appendix C to reduce the diversity data shown in Equation (1) indexed by the joint set of conjugated orbits (the conjugated orbits of *act* adjoined to the conjugated orbits of *gct*), using the  $D_4$  action defined in Section 3.3. The orbit diversity for the joint set of conjugated orbits is shown in (26) for each sequence in the sample of 10 Brazilian sequences referenced in Appendix B.1.

s	act	tca	tga	agt	gct	tcg	cga	agc
1	1.1261	1.4286	1.5167	1.4295	1.0408	8.0000	9.9231	1.4536
2	1.0721	1.4000	1.4783	1.4714	1.3295	6.8571	6.6190	1.4719
3	1.1417	1.3009	1.4286	1.4114	1.2824	6.7143	8.4444	1.3915
4	1.0500	1.3796	1.5841	1.2727	1.1212	6.1250	9.2000	1.5480
5	1.1250	1.3398	1.4016	1.2911	1.4235	3.7500	6.3810	1.4787
6	1.0769	1.3495	1.4914	1.2500	1.2442	8.8000	6.0870	1.4595
7	1.0420	1.3431	1.5000	1.2638	1.2292	4.6364	6.3333	1.4530
8	1.0087	1.3960	1.4696	1.2485	1.3210	6.0000	7.5789	1.5435
9	1.0168	1.3267	1.5877	1.2911	1.2651	4.5000	7.2105	1.5359
10	1.0609	1.1557	1.4237	1.3270	1.3103	7.8000	5.7917	1.3590

(26)

The inclusion of the error (due to sampling variability) term in the canonical decomposition for the sample is obtained by tensoring the decomposition induced by the representation of interest, shown in Appendix C, with the *standard* canonical decomposition [12] (Chapter 4)

$$I_n = \mathcal{A} + \mathcal{Q},$$

where  $\mathcal{A}$  is the  $n \times n$  projection matrix with all entries equal to  $1/n$ ,  $I_n$  is the  $n \times n$  identity matrix and  $\mathcal{Q} = I - \mathcal{A}$ . The canonical decomposition for the sample is then

$$I_{gn} = I_g \otimes I_n = (\mathcal{P}_1 + \mathcal{P}_\alpha + \mathcal{P}_{\gamma_+} + \mathcal{P}_{\gamma_-} + \mathcal{P}_\beta) \otimes (\mathcal{A} + \mathcal{Q})$$

from which we obtain the (multivariate) analysis of variance shown in 27, where  $x'(\mathcal{P}_\xi \otimes \mathcal{A})x$  are the sample *mean effects*,  $x'(\mathcal{P}_\xi \otimes \mathcal{Q})x$  the sampling error terms, and  $x'(I_g \otimes \mathcal{Q})x$  the total sampling error and  $g$  is the group order. More specifically, it is assumed that

$$x = (x_\tau)_{\tau \in G} \in \mathbb{R}^{gn}, \quad x_\tau \sim N(\mu_\tau \otimes e, I_n), \quad x \sim N(\mu \otimes e, \Sigma \otimes I_n),$$

where  $e' = (1, \dots, 1)$  with  $n$  components,  $\mu$  is the vector of dihedral means, and  $\Sigma$  is the dihedral covariance structure. It then follows, for all  $\mathcal{P}_\xi$ , that

$$(\mathcal{P} \otimes \mathcal{A})x \sim N((\mathcal{P} \otimes \mathcal{A})(\mu \otimes e), (\mathcal{P} \otimes \mathcal{A})(\Sigma \otimes I_n)(\mathcal{P}' \otimes \mathcal{A}')).$$

Because  $\mathcal{A}' = \mathcal{A}$ ,  $\mathcal{A}^2 = \mathcal{A}$  and  $\mathcal{A}e = e$ , we have

$$(\mathcal{P} \otimes \mathcal{A})x \sim N((\mathcal{P}\mu) \otimes e, (\mathcal{P}\Sigma\mathcal{P}') \otimes \mathcal{A}).$$

Similarly,

$$(\mathcal{P} \otimes \mathcal{Q})x \sim N((\mathcal{P}\mu) \otimes e, (\mathcal{P}\Sigma\mathcal{P}') \otimes \mathcal{Q}).$$

The degrees of freedom in each case are obtained by the traces of the corresponding projections, which are also equal to the dimension of the projecting (invariant) subspaces. Under suitable parametric assumptions the magnitude of the ratios

$$\frac{(x'(\mathcal{P}_\xi \otimes \mathcal{A})x)/\text{tr}((\mathcal{P}_\xi \otimes \mathcal{A}))}{(x'(I_g \otimes \mathcal{Q})x)/\text{tr}(I_g \otimes \mathcal{Q})}$$

can be assessed by (typically non-central) F-distributions with  $n_{\xi}^2$  and  $g(n-1)$  degrees of freedom. The corresponding underlying parametric hypotheses  $\mu' \mathcal{P}_{\xi} \mu = 0$  are those introduced earlier in Section 3.4.

$\mathcal{M}$	$x' \mathcal{M} x$	$\text{tr } \mathcal{M}$	$x' \mathcal{M} x / \text{tr } \mathcal{M}$
$\mathcal{P}_1 \otimes \mathcal{A}$	584.8	1	584.8
$\mathcal{P}_{\alpha} \otimes \mathcal{A}$	138.25	1	138.25
$\mathcal{P}_{\gamma_+} \otimes \mathcal{A}$	3.41	1	3.41
$\mathcal{P}_{\gamma_-} \otimes \mathcal{A}$	1.09	1	1.09
$\mathcal{P}_{\beta} \otimes \mathcal{A}$	319.5	4	79.88
$\mathcal{P}_1 \otimes \mathcal{Q}$	6.0	9	0.67
$\mathcal{P}_{\alpha} \otimes \mathcal{Q}$	6.6	9	0.74
$\mathcal{P}_{\gamma_+} \otimes \mathcal{Q}$	4.8	9	0.53
$\mathcal{P}_{\gamma_-} \otimes \mathcal{Q}$	4.4	9	0.49
$\mathcal{P}_{\beta} \otimes \mathcal{Q}$	20.8	36	0.579
$I_8 \otimes \mathcal{Q}$	42.7	72	0.592
$\ x\ ^2$	1097.7	80	—

(27)

Under large-sample parametric assumptions and independent dihedral covariance structure it follows that, with the exception of the contrast associated with  $\gamma_-$ , all F-ratios are significantly high (statistically distinct from zero).

#### 4.4. Orbit Diversity Decomposition for the HCV Samples

This section replicates the methods described in Section 4.3 for a sample of 10 Brazilian hepatitis C sequences. The orbit diversity for the joint set of conjugated orbits for each sequence in the sample is shown in (28). Their accession numbers are referenced in Appendix B.1.

s	act	tca	tga	agt	gct	tgc	cga	agc
1	1.3519	1.3770	1.2581	1.3175	1.1888	1.3739	1.5962	1.1007
2	1.4787	1.3629	1.6196	1.3175	1.0699	1.2456	1.5133	1.2992
3	1.3889	1.3359	1.6170	1.2535	1.0960	1.2521	1.5043	1.2889
4	1.3796	1.3206	1.6170	1.2899	1.0909	1.2521	1.5043	1.2889
5	1.3925	1.3130	1.6000	1.2857	1.0909	1.2500	1.5043	1.2985
6	1.4510	1.3178	1.5914	1.2571	1.0821	1.2521	1.5259	1.2985
7	1.3211	1.4052	1.3409	1.2787	1.2240	1.6082	1.6667	1.0922
8	1.3333	1.4123	1.2809	1.3492	1.2204	1.5631	1.6979	1.1064
9	1.3578	1.4425	1.2697	1.3710	1.2363	1.5534	1.6875	1.1268
10	1.2895	1.4298	1.2967	1.4098	1.1959	1.5094	1.6939	1.1241

(28)



The corresponding analysis of variance decomposition is shown in (29).

$\mathcal{M}$	$x' \mathcal{M} x$	$\text{tr } \mathcal{M}$	$x' \mathcal{M} x / \text{tr } \mathcal{M}$
$\mathcal{P}_1 \otimes \mathcal{A}$	146.76	1	146.76
$\mathcal{P}_\alpha \otimes \mathcal{A}$	0.30	1	0.30
$\mathcal{P}_{\gamma_+} \otimes \mathcal{A}$	0.09	1	0.09
$\mathcal{P}_{\gamma_-} \otimes \mathcal{A}$	0.0001	1	0.0001
$\mathcal{P}_\beta \otimes \mathcal{A}$	0.93	4	0.233
$\mathcal{P}_1 \otimes \mathcal{Q}$	0.02	9	0.003
$\mathcal{P}_\alpha \otimes \mathcal{Q}$	0.16	9	0.018
$\mathcal{P}_{\gamma_+} \otimes \mathcal{Q}$	0.15	9	0.017
$\mathcal{P}_{\gamma_-} \otimes \mathcal{Q}$	0.18	9	0.020
$\mathcal{P}_\beta \otimes \mathcal{Q}$	0.23	36	0.0065
$I_8 \otimes \mathcal{Q}$	0.7	72	0.0104
$  x  ^2$	148.84	80	—

(29)

It should be evident, by comparing the magnitude of the F-ratios,

virus	$\mathcal{P}_\alpha$	$\mathcal{P}_{\gamma_+}$	$\mathcal{P}_{\gamma_-}$	$\mathcal{P}_\beta$
HIV1	238.342	5.889	1.882	137.962
HCV	28.846	8.653	0.009	22.115

identified by the decompositions in (27) and (29), that the two viruses have significantly distinct joint cyclic diversity profiles. Additional numerical studies are referenced in Appendix B.

## 5. Summary

In this communication we constructed dihedral  $D_2$  reduction of conjugate injective cyclic orbits in length of three, a dihedral  $D_4$  reduction of their combined set, and a dihedral  $D_3$  reduction of the set of conjugate injective cyclic orbits in length of four. In each case, the experimental scalar data can be any summary obtained over the cyclic orbits, such as the sum or an extreme value of the frequency counts over the cyclic orbit, the entropy of a frequency distribution over the orbit, its amino acid content, or, as in [11], the orbit's frequency diversity. In the case of matrix data, the data-analytic methods of group rings, instead of group algebras would then be the appropriate methodology [14].

## Acknowledgements

The author is thankful to the referees' comments and clarifying suggestions.

## Appendix

### A. HIV1 and HCV Sequences

The following are the accession numbers for the HIV1 and HCV sequences considered in the present study:

HIV1 seq id	i	HCV seq id	i
<a href="#">EF637046</a>	1	<a href="#">EF032890</a>	1
<a href="#">EF637047</a>	2	<a href="#">EF032891</a>	2
<a href="#">EF637048</a>	3	<a href="#">EF032892</a>	3
<a href="#">EF637049</a>	4	<a href="#">EF032893</a>	4
<a href="#">EF637050</a>	5	<a href="#">EF032894</a>	5
<a href="#">EF637051</a>	6	<a href="#">EF032895</a>	6
<a href="#">EF637053</a>	7	<a href="#">EF032896</a>	7
<a href="#">EF637054</a>	8	<a href="#">EF032898</a>	8
<a href="#">EF637056</a>	9	<a href="#">EF032899</a>	9
<a href="#">EF637057</a>	10	<a href="#">EF032900</a>	10

### B. Additional Studies

#### B.1. Relative Entropy Study of 10 Brazilian HIV1 Sequences

The relative entropy evaluations illustrated above in Section 4.1 were replicated for a sample of 10 Brazilian HIV1 sequences, referenced in Appendix A. The raw frequency counts and the corresponding relative entropy profiles for each of 10 sequences are linked in [24].

#### B.2. Relative Entropy Study of 10 Brazilian HCV Sequences

Similarly to the study for the HIV1, a sample of 10 Brazilian hepatitis C sequences was evaluated for their relative entropy. The sequences are referenced in Appendix A. The raw frequency counts and the corresponding relative entropy profiles along each genome are linked in [25]. The relative entropy invariant profiles clearly highlight the structural differences between the two types of viruses.

#### B.3. Relative Entropy Study of Random Reference Sequences

It is statistically useful to compare the cyclic reductions obtained for HIV1's isolate described above with those from random DNA sequences of comparable lengths. The results, based on 20 random sequences, shown in [26], clearly indicate that the observed variations in relative entropy (invariants) for the conjugated *gct* orbits, both for HIV1 and HCV sequences, are well below what one would expect to observe for random sequences of comparable lengths.

### C. Canonical Projections

The following are the canonical projections

$$\mathcal{P}_\xi = \frac{n_\xi}{g} \sum_{\tau} \chi_{\tau^{-1}}^\xi \rho_\tau,$$

evaluated for the permutation representation  $\rho$  described in Section 3.4, along with the corresponding components  $x' \mathcal{P}_\xi x$  of  $\|x\|^2$ , with the entries of  $x$  indexed in correspondence with  $(act, tca, tga, agt, gct, tcg, cga, agc)$ . Here:

$$\mathcal{P}_1 = \frac{1}{8} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

gives an overall mean,

$$\mu' \mathcal{P}_1 \mu = \frac{1}{8} (\mu_{act} + \mu_{agc} + \mu_{agt} + \mu_{cga} + \mu_{gct} + \mu_{tca} + \mu_{tcg} + \mu_{tga})^2,$$

$$\mathcal{P}_\alpha = \frac{1}{8} \begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

gives the (*handedness*) contrast

$$\mu' \mathcal{P}_\alpha \mu = \frac{1}{8} (\mu_{act} - \mu_{agc} + \mu_{agt} + \mu_{cga} - \mu_{gct} - \mu_{tca} + \mu_{tcg} - \mu_{tga})^2$$

between the cyclic summaries

$$\{act, agt, tcg, cga\},$$

indexed by the rotations and the summaries

$$\{gct, tga, tca, agc\}$$

indexed by reversals as a sum of within-rotation and within-reversal variability;

$$\mathcal{P}_{\gamma+} = \frac{1}{8} \begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \end{pmatrix}$$

gives the sum

$$\mu' \mathcal{P}_{\gamma+} \mu = \frac{1}{8} (\mu_{act} - \mu_{agc} - \mu_{agt} - \mu_{cga} + \mu_{gct} + \mu_{tca} + \mu_{tcg} - \mu_{tga})^2$$

of within-rotation and within-reversal variability, whereas

$$\mathcal{P}_{\gamma-} = \frac{1}{8} \begin{pmatrix} 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \end{pmatrix}$$

gives the linear contrast

$$\mu' \mathcal{P}_{\gamma-} \mu = \frac{1}{8} (\mu_{act} + \mu_{agc} - \mu_{agt} - \mu_{cga} - \mu_{gct} - \mu_{tca} + \mu_{tcg} + \mu_{tga})^2$$

between within-rotation and within-reversal variability. Lastly,

$$\mathcal{P}_{\beta} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

gives the sum

$$\mu' \mathcal{P}_{\beta} \mu = \frac{1}{2} (\mu_{act} - \mu_{tcg})^2 + (\mu_{agc} - \mu_{tga})^2 + (\mu_{agt} - \mu_{cga})^2 + (\mu_{gct} - \mu_{tca})^2$$

of two contrasts in dimension of 2.

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Fisher, R.A. The theory of linkage in polysomic inheritance. *Philos. Trans. Roy. Soc. London. Ser. B* **1947**, *233*, 55–87.
2. Fisher, R.A. The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen.* **1942**, *11*, 341–353.
3. Fisher, R.A. A system of confounding for factors with more than two alternatives, giving completely orthogonal cubes and higher powers. *Ann. Eugen.* **1945**, *12*, 283–290.
4. Viana, M. *Algebraic Methods in Statistics and Probability*; Contemporary Mathematics; Richards, D., Ed.; American Mathematical Society: Providence, RI, USA, 2001; Volume 287.
5. Viana, M. *Algebraic Methods in Statistics and Probability II*; Contemporary Mathematics; Wynn, H., Ed.; American Mathematical Society: Providence, RI, USA, 2010; Volume 516.
6. Findley, G.L.; Findley, A.M.; McGlynn, S.P. Symmetry characteristics of the genetic code. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 7061–7065.
7. Sergienko, I.V.; Gupal, A.M.; Vagis, A.A. Symmetry in encoding genetic information in DNA. *Cybern. Syst. Anal.* **2011**, *47*, 408–414.
8. Hornos, J.E.; Braggion, L.; Magini, M.; Forger, M. Symmetry preservation in the evolution of the genetic code. *IUBMB Life* **2004**, *56*, 125–130.
9. Zandi, R.; Reguera, D.; Bruinsma, R.F.; Gelbart, W.M.; Rudnick, J.; Reiss, H. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15556–15560.
10. Finkel, D.L. HIV-1 ancestry primordial expansions of RRE and RRE- related sequences. *J. Theor. Biol.* **1992**, *3*, 285–302.
11. Doi, H. Importance of purine and pyrimidine content of local nucleotide sequences (six bases long) for evolution of human immunodeficiency virus type 1. *Evolution* **1991**, *88*, 9282–9286.
12. Viana, M. *Symmetry Studies, an Introduction to the Analysis of Structured Data in Applications*; Cambridge University Press: New York, NY, USA, 2008.
13. Souza, D.J.; Chaves, L.M.; Viana, M.A.G. Symmetries in symbolic sequences. *Rev. Bras. Biom.* **2010**, *1*, 73–86.
14. Viana, M.; Lakshminarayanan, V. *Dihedral Fourier Analysis, Data-Analytic Aspects and Applications*; Lecture Notes in Statistics; Springer: New York, NY, USA, 2013; Volume 206.
15. Viana, M. Canonical invariants for three-candidate preference rankings. *Can. Appl. Math. Q.* **2007**, *15*, 203–222.
16. Viana, M. *Canonical Decompositions and Invariants for Data Analysis*; Elsevier: Amsterdam, The Netherlands, 2009.
17. Viana, M. Symmetry studies and decompositions of entropy. *Entropy* **2006**, *8*, 88–109.
18. Viana, M. Symmetry orbits and their data-analytic properties. *Rev. Mat.* **2013**, *20*, 155–166.

19. Viana, M. Dihedral Polynomials. In *Mathematical Optics: Classical, Quantum, and Computational Methods*; Lakshminarayanan, V., Calvo, M.L., Alieva, T., Eds.; CRC Press: Boca Raton, FL, USA, 2012.
20. Viana, M.; Lakshminarayanan, V. Symmetry studies of refraction data. *J. Modern Opt.* **2014**, *61*, 138–146.
21. Viana, M. Symmetry-Related Decompositions of Uncertainty. In *Proceedings of the XI Brazilian Meeting on Bayesian Statistics*, Amparo-SP, Brazil, 18–22 March 2012; Stern, J., Lauretto, M., Polpo, A., Diniz, M., Eds.; American Institute of Physics: Melville, NY, USA, 2012; doi:10.1063/1.4759582
22. Cartan, E. *The Theory of Spinors*; MIT Press: Cambridge, MA, USA, 1966.
23. Human immunodeficiency virus type 1, isolate BRU, complete genome (LAV-1). Available online: <http://www.ncbi.nlm.nih.gov/nuccore/K02013.1> (accessed on 23 December 2014).
24. Relative entropy study of 10 Brazilian HIV1 sequences. Available online: <https://app.box.com/s/ui9nt6uu6pitc6mxl6cg> (accessed on 23 December 2014).
25. Relative entropy study of 10 Brazilian HCV sequences. Available online: <https://app.box.com/s/zgpv0fkpd1pz9c30ql1q> (accessed on 23 December 2014).
26. Relative entropy study of random reference sequences. Available online: <https://www.box.com/s/g7rguk0dy3ben5w7e93x> (accessed on 23 December 2014).

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).