GENERALIZED TOPOLOGICAL GROUPS AND GENETIC RECOMBINATION

S.A. AHMADI
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SISTAN AND BALUCHESTAN,
ZAHEDAN, IRAN
E-MAIL: SA.AHMDI@GMAIL.COM

(Received: 11 January 2013, Accepted: 27 May 2013)

Abstract. In this paper we introduce topological and algebraic structures on the space of DNA sequences. The idea of these structures comes from genetic recombination.

AMS Classification: 54H11, 92D20, 37B10.
Keywords: DNA sequence, Generalized group, Generalized isomorphism, Smale space, Topological group.

1. Introduction

DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people. The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences [7]. In order to apply mathematical techniques to DNA analysis, the set of DNA

JOURNAL OF DYNAMICAL SYSTEMS & GEOMETRIC THEORIES
VOL. 11, NUMBER 1-2 (2013) 51-58.
©TARU PUBLICATIONS

sequences must be transformed to a mathematical structure. This structure can be

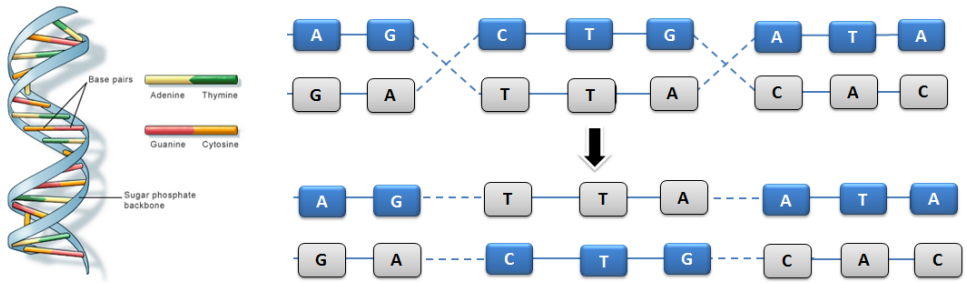


FIGURE 1. DNA recombination

an algebraic structure or a topological structure. Herein, we present an structure which contains both of topological and algebraic structures. The idea of these structures comes from the DNA recombination [1]. DNA (or genetic) recombination plays important roles in heredity and evolution. Recombinant DNA molecules are DNA sequences that result from the use of laboratory methods (molecular cloning) to bring together genetic material from multiple sources, creating sequences that would not otherwise be found in biological organisms. Recombinant DNA is possible because DNA molecules from all organisms share the same chemical structure; they differ only in the sequence of nucleotides within that identical overall structure. First problem in constructing an algebraic structure from recombination is that there is no DNA sequence which acts as the identity (that is, after recombination with all DNA sequences does not change them). This leads us to use the notion of generalized group.

2. PRELIMINARIES

2.1. **Generalized topological groups.** The theory of generalized groups was first introduced by M.R. Molaei [4, 5]. In a generalized group the identity element for an arbitrary element is a function of that element. More precisely let G be a non-empty set admitting an operation

$$* : G \times G \to G$$

$$(g, h) \mapsto g * h$$

called multiplication which satisfies the following conditions:

(1) $(g * h) * k = g * (h * k)$ for all $g, h, k \in G$;

(2) For each $g \in G$ there exists a unique $e(g) \in G$ such that

$$e(g) * g = g * e(g) = g;$$

(3) For each $g \in G$ there exists $g^{-1} \in G$ such that $g * g^{-1} = g^{-1} * g = e(g)$.

Then $(G, *)$ is called a *generalized group*.

Example 2.1. *Let*

$$G = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \ : \ a, b, c, d \in \mathbb{R} \right\}.$$

Then G with the operation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \mapsto \begin{bmatrix} a & b' \\ c' & d \end{bmatrix}$$

is a generalized group.

Theorem 2.2. [3] *For each $a \in G$ there exists a unique $a^{-1} \in G$.*

Theorem 2.3. [3] *Let $(G, *)$ be a generalized group and $g * h = h * g$ for all $g, h \in G$. Then $(G, *)$ is a group.*

Definition 2.4. [4] *$(G, *)$ is called a generalized topological group provided that:*

(1) $(G, *)$ is a generalized group;

(2) G is a Hausdorff topological space;

(3) The functions

$$\mu : G \times G \to G$$

$$(g, h) \mapsto g * h$$

and

$$\nu : G \to G$$

$$g \mapsto g^{-1}$$

are continuous maps.

2.2. **Shift spaces.** Information is often represented as a sequence of discrete symbols drawn from a fixed finite set. This finite set of *symbols* which we will call it *alphabet* is usually denoted by $\mathcal{A}$. In this case the *full $\mathcal{A}$-shift* is the collection of all bi-infinite sequences of symbols in $\mathcal{A}$ and is denoted by

$$\mathcal{A}^{\mathbb{Z}} = \{x = (x_i) : \ x_i \in \mathcal{A} \quad \forall i \in \mathcal{Z}\}.$$

Blocks of consecutive symbols will play a central role. A *block* over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. For example a typical block over $\mathcal{A} = \{a, b\}$ looks like *abaaabbabbba*. If x is a point in $\mathcal{A}^{\mathbb{Z}}$ and $i \leq j$, then we will denote the block of coordinates in x from position i to position j by $x_{[i,j]} = x_i x_{i+1} \cdots x_j$. If x is a point in $\mathcal{A}^{\mathbb{Z}}$ and w is a block over $\mathcal{A}$, we will say that w occurs in x if there are indices i and j so that $w = x_{[i,j]}$. We denote the empty block with ε. The *shift map* $\sigma : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ maps a point x to the point $y = \sigma(x)$ whose ith coordinate is $y_i = x_{i+1}$. Recall that the metric

$$d(w, w') = 2^{-l} \qquad \text{where} \quad l = \min\{|i| : \ w_i \neq w'_i\}$$

generates the product topology on $\mathcal{A}^{\mathbb{Z}}$.
For each $k \in \mathbb{N}$ and symbols $j_{-k}, j_{-k+1}, ..., j_k \in \mathcal{A}$ the open ball

$$C^k_{j_{-k},\cdots,j_k} = \{w : \ w_{[-k,k]} = j_{-k} j_{-k+1} \cdots j_k\}$$

is called a *k-central cylinder*.
Let $\mathcal{F}$ be collection of blocks over an alphabet $\mathcal{A}$, which we will think of as being the *forbidden blocks*. For any such $\mathcal{F}$, define $X_{\mathcal{F}}$ to be the subset of sequences in $\mathcal{A}^{\mathbb{Z}}$ which does not contain any block in $\mathcal{F}$. A *shift space* is a subset $X \subset \mathcal{A}^{\mathbb{Z}}$ such that $X = X_{\mathcal{F}}$ for some collection $\mathcal{F}$ of forbidden blocks over $\mathcal{A}$ [2].

2.3. **Smale spaces.** Ruelle developed in [6] an abstract theory of the so-called Smale spaces.

Definition 2.5. *Let f be a homeomorphism on a campact metric space (X, d). Then (X, f) is called a Smale space If the following conditions hold:*

(SS1) *There exist $\eta > 0$ and a continuous map*

$$[,] : \{(x, y) \in X \times X \ : \ d(x, y) < \eta\} \to X$$

such that $[x,x]=x$ *and*

$$[[x,y],z]=[x,z], \quad [x,[y,z]]=[x,z],$$

$$f([x,y])=[f(x),f(y)].$$

(SS2) *There exist* $\delta>0$ *and* $\nu\in(0,1)$ *such that*

$$d(f^n(y),f^n(z))\leq \nu^n d(x,y) \quad ; \quad y,z\in V_\delta^s(x), n>0,$$

$$d(f^{-n}(y),f^{-n}(z))\leq \nu^n d(x,y) \quad ; \quad y,z\in V_\delta^u(x), n>0.$$

where

$$V_\delta^s(x)=\{u \ : \ u=[u,x],\ d(x,u)<\delta\},$$

$$V_\delta^u(x)=\{v \ : \ v=[x,v],\ d(x,v)<\delta\}.$$

3. Main Results

Mathematically, we have at our disposal a 4 letter alphabet $\mathcal{D}=\{A,G,C,T\}$ or equivalently $\mathcal{D}=\{1,2,3,4\}$ to encode information in genetic. The space of DNA sequences is defined as $\Sigma=\{a=(a_i): \ a_i\in\mathcal{D}, i\in\mathbb{Z}\}$.

Definition 3.1. *Let* $\xi=\{n_i\}_{i\in\mathbb{Z}}$ *be a strictly increasing sequence of integer numbers so that* $n_{-1}<0\leq n_0$. *Then we define an operation*

$$*:\Sigma\times\Sigma\to\Sigma$$

$$(a,b)\mapsto c$$

where

$$c_{[n_i+1,n_{i+1}]}=\begin{cases} a_{[n_i+1,n_{i+1}]} & i \text{ is even} \\ b_{[n_i+1,n_{i+1}]} & i \text{ is odd} \end{cases}$$

Which is called a ξ*-action.*

Theorem 3.2. $(\Sigma,*)$ *is a topological generalized group.*

Proof. First we show that $(\Sigma,*)$ is a generalized group. For all $a,b,c\in\Sigma$ we have $(a*b)*c=a*c=a*(b*c)$. Suppose that $a\in\Sigma$ then by choosing $e(a)=a$ we deduce $a*e(a)=e(a)*a=a$ and it is easily follows that $e(a)$ is unique. Hence for each $a\in\Sigma$ we can put $a^{-1}=a$. It's remind to show that the maps $a\mapsto a^{-1}$ and $(a,b)\mapsto a*b$ are continuous. The first map is identity and therefore it is continuous.

For continuity of the second map given $\epsilon > 0$ there exists $N > 0$ such that $2^{-N} < \epsilon$. If $\delta = 2^{-N}$, $a, b, a', b' \in \Sigma$ and

$$D((a,b),(a',b')) = \max\{d(a,a'), d(b,b')\} < \delta,$$

then we have

$$a_{[-N,N]} = a'_{[-N,N]} \ , \ b_{[-N,N]} = b'_{[-N,N]}.$$

Therefore

$$(a * b)_{[-N,N]} = (a' * b')_{[-N,N]}.$$

So $d(a * b, a' * b') < \epsilon$. □

In especial case let $n_{i+1} - n_i = n_i - n_{i-1}$ for each $i \in \mathbb{Z}$. Then the *block shift map* $\phi : (\Sigma, *) \to (\Sigma, *)$, maps a point x to the point $y = \phi(x)$ where $y_{[n_{i-1}+1, n_i]} = x_{[n_i+1, n_{i+1}]}$. By this assumption we have the following theorem:

Theorem 3.3. *The map $\phi^2 : (\Sigma, *) \to (\Sigma, *)$ is a generalized isomorphism of topological groups.*

Proof. First we show that $\phi^2(a * b) = \phi^2(a) * \phi^2(b)$ for all $a, b \in \Sigma$. If $a, b \in \Sigma$, then

$$(\phi^2(a * b))_{[n_{i-1}+1, n_i]} = (a * b)_{[n_{i+1}+1, n_{i+2}]} = \begin{cases} a_{[n_{i+1}+1, n_{i+2}]} & i \text{ is odd} \\ b_{[n_{i+1}+1, n_{i+2}]} & i \text{ is even} \end{cases}$$

In other hand

$$\begin{aligned} (\phi^2(a) * \phi^2(b))_{[n_{i-1}+1, n_i]} &= \begin{cases} \phi^2(a)_{[n_{i-1}+1, n_i]} & i \text{ is odd} \\ \phi^2(b)_{[n_{i-1}+1, n_i]} & i \text{ is even} \end{cases} \\ &= \begin{cases} a_{[n_{i+1}+1, n_{i+2}]} & i \text{ is odd} \\ b_{[n_{i+1}+1, n_{i+2}]} & i \text{ is even} \end{cases} \end{aligned}$$

Hence ϕ^2 is a homomorphism. Obviously ϕ^2 is a bijection. So ϕ^2 is an isomorphism. It is remind to show that ϕ^2 is continuous. Given $\epsilon > 0$ there exist $N \in \mathbb{N}$ such that $2^{-N} < \epsilon$. Suppose that $K = n_i - n_{i-1}$, $i \in \mathbb{Z}$ and $\delta = 2^{-(N+2K)}$. Thus if $d(a,b) < \delta$ then $a_{[-(N+2K), N+2K]} = b_{[-(N+2K), N+2K]}$. Therefore $\phi^2(a)_{[-N,N]} = \phi^2(b)_{[-N,N]}$. So $d(\phi^2(a), \phi^2(b)) < \epsilon$. □

Proposition 3.4. *Let (X, f) be a Smale space. Then there exists $\delta > 0$ such that for each $x \in X$ the open ball $B_\delta(x)$ has a generalized topological group structure so that f is a generalized isomorphism of topological groups.*

Proof. Suppose that $\eta > 0$ be the constant of the condition SS2 in the definition 2.5. There exists $\delta \in (0, \eta)$ such that the inequality $d(x, y) < \delta$ implies that $d(f(x), f(y)) < \eta$, for each $x, y \in X$. If for each $a, b \in B_\delta(x)$ we define $a * b = [a, b]$, then for each $a, b, c \in B_\delta(x)$ we have $a * (b * c) = a * c = (a * b) * c$. Also for each $a \in B_\delta(x)$ we define $e(a) = a$. Obviously $a * e(a) = e(a) * a = a$. We must show that $e(a)$ is unique. For do this let $b \in B_\delta(x)$ satisfies the relation $a * b = b * a = a$. Then

$$a = b * a = b * (a * b) = b * b = b.$$

The inverse of each element is itself. Since for each $a, b \in B_\delta(x)$ we have $d(f(a), f(b)) < \eta$ then $f(a) * f(b)$ is well defined and $f(a * b) = f(a) * f(b)$. Therefore f is an isomorphism of generalized topological groups. □

Definition 3.5. *Let $k \in \mathbb{Z}$. We define a k-action as an action $\star : \Sigma \times \Sigma \to \Sigma$ so that*

$$(a \star b)_n = \begin{cases} b_n & n \leq k \\ a_n & n \geq k + 1 \end{cases}$$

Proposition 3.6. *(Σ, σ) is a Smale space.*

Proof. We define the bracket of $a, b \in \Sigma$ as $[a, b] = a \star b$. At this point, the bracket is defined for all pairs a and b. Also we have

$$[a, a] = a \star a = a, \qquad [a, [b, c]] = a \star (b \star c) = a \star c = [a, c], \qquad [[a, b], c] = [a, c].$$

Now we show that for each $a, b \in \Sigma$ with $d(a, b) < 2^{-k}$ we have $[\sigma(a), \sigma] = \sigma([a, b])$. If $d(a, b) < 2^{-k}$, then $a_{[-k,k]} = b_{[-k,k]}$. Therefore

$$[\sigma(a), \sigma(b)]_n = \begin{cases} \sigma(b)_n & n \leq k \\ \sigma(a)_n & n \geq k + 1 \end{cases} = \begin{cases} b_{n+1} & n \leq k \\ a_{n+1} & n \geq k + 1 \end{cases} = \sigma([a, b])_n$$

Finally we show that $\delta = 2^{-k}$ and $\nu = 2^{-1}$ satisfy the condition SS2 in definition 2.5. Suppose that $a, b, c \in \Sigma$, $a = [a, b]$, $a = [c, a]$ and $d(a, b), d(a, c) < \delta$. Then $a_{[-k,\infty)} = b_{[-k,\infty)} = c_{[-k,\infty)}$. Now let $d(b, c) = 2^N$ where N is the greatest integer (necessarily negative) so that $b_N \neq c_N$. It is easily follows that $d(\sigma(b), \sigma(c)) = 2^{N-1} = 2^{-1} d(b, c)$. □

Corollary 3.7. *Every $(k+1)$-central cylinder in Σ has a generalized topological group structure with the action $\star$ so that with respect to it the shift map σ is an isomorphism of generalized topological groups.*

References

[1] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, RNA-guided DNA assembly, Journal of Theoretical Biology, Vol. 248 Number 4, , 706-720 (2007).

[2] F. Blanchard, A. Maass, A. Nogueira, Topics in symbolic dynamics and applications, London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Vol. 279, 2000.

[3] M. Mehrabi , M.R. Molaei, A. Olomi, Generalized Subgroups and Homomorphisms, Arab Journal of Mathematical Sciences, Vol. 6 Number 2, 1-7 (2000).

[4] M.R. Molaei, Topological Generalized Groups, International Journal of Pure and Applied Mathematics, Vol. 2 Number 9, 1055-1060 (2000).

[5] M.R. Molaei, Generalized Structures Based on Completely Simple Semigroups, Hadronic Press, 2005.

[6] D. Ruelle, Thermodynamic Formalism, Encyclopedia of Math. and its Appl. 5. Addison-Wesley, Reading, MA, 1978.

[7] J. Schmutz, J. Martin, A. Terry, The DNA sequence and comparative analysis of human chromosome 5, Nature, Vol. 431 , 268-274 (2004).