Genome rearrangements in the real moduli spaces $\overline{M}_{0,n}$.

Gefry Barad
Department of Mathematics, Johns Hopkins University
Baltimore MD 21218, gefry@math.jhu.edu, http://www.math.jhu.edu/˜gefry

December 4, 2002

Abstract

We aknowledge the presence of the the Chromosome Inversion Problem in some spaces connected with Algebraic Geometry. There are at least two benefits of this approach: there are not many articles dedicated to circular permutations. A reason could be that the related Algebraic Geometry is non-trivial. Another benefit is the following: maybe the geometric approach will lead to a geometric proof of the diameter of the circular permutations by reversals. We give a practical applications of reversals in the study of these spaces. The opposite road, from Geometry to Computational Biology, will be analyzed in a forthcoming paper

A.M.S. Classification: 68Q25, 68Q05, 14P25
Keywords: Pevzner-Hannenhalli Theory, sorting by reversals, moduli spaces, homology

1 Introduction.

This paper is part of a project aimed at establishing a stronger connection between the Computational Biology community , and the pure field of Algebraic Geometry, on the other side.

We are stimulated by the following approaches:

1. Mulmuley and Sohoni ' 2001 approach of the P vs. NP Conjecture via Geometric Invariant Theory. [14]. Mulmuley used some results from Algebraic Geometry to prove upper bounds connected with CS.
2. The study of Thurston, Sleator and Tarjan on the rotation graph, using hyperbolic geometry. We still do not have a pure combinatorics proof of their result. Their work was motivated by Dynamic Optimality Conjecture. It is among the very few articles which show that it is possible to have geometric proofs for combinatorics problems.

Our spaces came from Geometric Invariant Theory [13], which is the study of actions of algebraic groups on algebraic varieties. They contain the graph of Thurston , Sleator and Tarjan. The Chromosome inversion problem and Pevzner Theory was fully observable at the early stage of our pure math investigation. It took a couple of months and luck to understand that the non-trivial combinatorics of these spaces is not explicitly studied in Algebraic Geometry and it is much better understood since 1996.[Pevzner, Bafna, Caprara, Tarjan, Shamir]

2 The Chromosome Inversion Graph G(n). An old problem

The Chromosome Inversion Problem was first stated in 1982 by Watterson, Ewens, Hall and Morgan, in a desire to construct a phylogenetic tree of evolution, and to define a genetic distance [18]. In their paper , a first attempt of an algorithm and a conjecture of a diameter of a graph G(n) appeared. A fundamental approach came in 1996, in the seminal article of Bafna and Pevzner [2]. See also [15].

2.1 Mathematical formulations

In the paper [18] , the objects of study are circular chromosomes, without a fixed 12 o'clock position, without top and bottom. Mirror images are the same. Mathematically:

Let G_n be the following graph. The vertices of G_n are the right coset classes $S_n/D(n)$, of the symmetric group, modulo the dihedral group . $D(n)$ is the symmetry group of the regular n-gon.

T is a set of $n(n-3)/2$ involutions from S_n. For every pair (k, l), $0 < k < l < n$, where (k,l) is different from (1, n-1), we define d:=d(k, l)(x)=l+k-x , if x is between k and l. Otherwise, x is a fixed point.

Two vertices are joined by an edge if there are two permutations a and b in their equivalence classes, and if there is t in T, such that a=bt.

Remark. For any r in D(n) and t in T, rtr^{-1} is in T.

Remark. Any graph has a metric , induced by the smallest number of edges between any 2 vertices. Applying the first remark, we can say that the distance between 2 vertices which contain the permutations σ_1 and σ_2 is k, if there are $d_1, d_2, ..., d_k$ several reversals of T,and R in $D(n)$ such that $\sigma_1 = \sigma_2 \circ R \circ d_1 \circ d_2 \circ ... \circ d_k$. ($\circ$ means composition of functions), and k is minimal with this property .

In the case of Genome Rearrangements studied by Bafna and Pevzner, the vertices of a graph are given by all permutations of S(n). Instead of $n(n-3)/2$ reversals, the full set of $n(n-1)/2$ reversals are considered.

Remark. In spite of their similarity, we did not find the mathematical equivalence between the problems above. Instead, we can perform the steps of the solution of the Gollan Conjecture .

Lemma 2.1.1. *The diameter of the graph G_n is $\leq$ n-2. (The "ratchet" algorithm).*

Proof ([18]): Let a and b be two vertices of the graph. They can be represented as labelled points of a circle , the labelling being given by two permutations from a and b. Starting from a point A on the second circle, we travel clockwise until we first come to a point not in the position from its first circle. We make a reversal, to create a neighborhood point of A as on the first circle. We consider the next labelled point which is not in the right position, and so on . We can begin with any A, and we need not to correct the position of the last point. So, the diameter of the graph G_n is $\leq$ n-2.

By a circular permutation, we mean a right coset class in the S_n, modulo the dihedral group. (i.e we label n points on a circle in the three dimensional space using permutations).

Breakpoint graph of a circular permutation σ. For every vertex of G_n, we will associate a bi-colored graph. Let i $\sim$ j if $|i-j| = 1$ or n-1. $\sigma = \sigma(1)\sigma(2)..\sigma(k)..\sigma(n)$. A pair of consecutive elements $\sigma(i)$ and $\sigma(i+1)$ is called a breakpoint if the absolute value of their difference is not 0 or n-1. Define an edge colored graph G(σ) with n vertices on the circle, labelled in the order given by the permutation. We join i and j by a black edge if (i,j) is a breakpoint of σ. We join i and j by a grey edge if i $\sim$ j , but i and j are not consecutive in σ.

Theorem 2.1.2. *(circular Gollan conjecture) For every n, the distance d(γ_n) between the vertices represented by the identity and by the Gollan permutation is at least n-3.*

Let us recall the Gollan permutation:

$$\gamma_n = \begin{cases} (3,1,5,2,7,4,...n-3,n-5,n-1,n-4,n,n-2) & n \;\; even \\ (3,1,5,2,7,4,...n-6,n-2,n-5,n,n-3,n-1) & n \;\; odd. \end{cases}$$

The theorem above is a consequence of the 2 fundamental lemmas from Pevzner's work [2] [15]. It is possible to apply these lemmas in the case of the circular breakpoint graph.

We prove:

1.There are at most 3 disjoint alternating cycles in G(γ_n).

2.d(a) $\geq$ b(a) - c(a), where b is the number of black edges of G(a), and c is the maximal number of disjoint alternating cycles of G(a).

Remark: Very few articles studied the circular permutations.

3 The Rotation Graph and the Associahedron.

There is a 1-1 correspondence between binary trees and triangulations of the n-gon, using n-3 non-intersection diagonals.

A diagonal flip is an operation that transforms one triangulation of a polygon into another: a diagonal inside the polygon is removed , creating a face with four sides. The opposite diagonal of this quadrilateral is inserted , restoring the diagram of a new triangulation.

Let $TG(n)$ be a graph with one node for each triangulation of an n-gon and an edge between 2 nodes if they are related by a diagonal flip.

Theorem 3.0.3 (Thurston, Sleator and Tarjan). *The diameter of $TG(n)$ is 2n-10, for n big enough.*

K_{n-1} is the n-3 dimensional convex polytope whose partial order set of its faces is isomorphic with the partial order set of an n-gon with a couple of non-intersecting diagonals. The partial orders are given by inclusions. There is such a convex polytope [Ziegler-Lectures on Polytope p.310].The codimension k faces of the associahedron are indexed by n-gons with k non-intersecting diagonals.

4 Gluing geometric objects

In this section we will combine the rotation graph (or the associahedron) and the graph of the chromosome inversion.

We will build a new space $\overline{M}_{0,n}(R)$, by gluing associahedra according to the rules below. (see Devadoss).

Remark. . The graph of the chromosomial inversion shows how the associahedra are glued together, using codimension one faces. The distance in the graph $G(n)$ shows the minimal number of associahedra placed between 2 given associahedra.

T is the following set of $n(n-3)/2$ involutions from S_n. For every pair (k, l), $0 < k < l < n$, where (k,l) is different from (1, n-1), we define d:=d(k, l)(x)=l+k-x , if x is between k and l. Otherwise, x is a fixed point. Let P be a fixed n-gon, with edges labeled 1,2,3...n. For every diagonal of P we can associate an element of T in the following way: any diagonal determines a partition of 1,2...n. Take the one which doesn't contain n: it's between 2 numbers, k and l. Then the associated d will be d(k,l), and we say that d(k,l) is supported by the diagonal of the n-gon P. **Throughout the paper, the word "diagonal" means a diagonal of the n-gon, or the reversal carried by the diagonal.**

Take $n!$ copies of K_{n-1}. For every permutation of S_n, label the edges of the n-gon with $\sigma(1), \sigma(2)...\sigma(k), ...\sigma(n)$. So the codimension k faces are labeled by decorated n-gons with k non-intersecting diagonals. Two codimension k faces of different K_{n-1}'s are identified (glued) if the permutations σ_1 and σ_2 which color the edges of the n-gons satisfy the following condition "flip" or gluing condition: there are $d_1, d_2, ..., d_i$ couple of elements of T, supported by the diagonals of the second face, such that $\sigma_1 = \sigma_2 \circ d_1 \circ d_2 \circ ... \circ d_i$. ($\circ$ means composition of functions).

The top dimensional faces (without diagonals) are identified by the action of D_n, the dihedral group. So we can begin with $(n-1)!/2$ copies of K_{n-1}, indexed over S_n/D_n. Two codimension k faces are identified if their classes modulo the dihedral group D_n contain 2 permutations which satisfy the flip condition from the previous paragraph.

It is a classical procedure in topology, to build spaces using gluing maps. Everything is possible if we are gluing in a space of big dimension (a generalization of the 3-dimensional procedure).

The new space above is called the moduli space of pointed real curves of genus 0, or real points of the Deligne-Mumford-Knudsen compactification of stable n-pointed curves of genus 0.

It is an aspherical space or an Eilenberg-MacLane space $K(\pi, 1)$, [3].: every spheroid can be deformed to a point. i.e any function from a sphere to this space is shrinkable in the ambient space, like the surface of a 3-dimensional ball which can be deformed into a point, using the ambient space.

5 A Topological Theorem which uses Reversals

Theorem 5.0.4. *The $(n-3)^{th}$ homology group of the real moduli space is zero. More exactly: $\overline{M}_{0,n}(R)$ are non-orientable compact smooth manifolds of dimension n-3.*

The proof is based on the following theorem (Spanier, page 206): if K is a finite n-dimensional orientable pseudomanifold, then $H_{n-1}(K)$ has no torsion.

We will assume that the moduli space is orientable. We will prove that the first homology group has Z_2-torsion. By Poincare duality, $H_{n-1}(K)$ has torsion. We will find an orientable manifold which does not satisfy Spanier's lemma.

So, the main point is to prove that the first homology group has Z_2-torsion. We will use Davis' description of the fundamental group:

The fundamental group π is described by Davis et al. as a kernel of an epimorphism f. Let A be the following group, given by generators and relations:

-involutory generators α_T, for each proper sub-interval T of [1, n-2]. The relations are: if the distance between the intervals X and Y is at least 2, then α_X and α_Y commute. If X $\subset$ Y, then $\alpha_Y \alpha_X \alpha_Y = \alpha_Z$. Z denotes the image of X under the order-reversing involution of Y.

-T is a set of $n(n-3)/2$ involutions from S_n, called reversals. For every pair (k, l), $0 < k < l < n$, where (k,l) is different from (1, n-1), we define d:=d(k, l)(x)=l+k-x , if x is between k and l. Otherwise, x is a fixed point.

The epimorphism f sends the generators of A into the corresponding reversals, which satisfy the same relations. The kernel of f is isomorphic with the fundamental group of $\overline{M}_{0,n}(R)$.

We are doing a simple diagram chasing:

Note: in Devadoss' research, he described reversals as transpositions. The consequences of this error are the following: -the imposibility to prove , using the diagram above, the fact that the first homology contains torsion (any transposition will go to an odd permutation, it is not in the kernel) -This discovery, of the reversals inside the real moduli space, lead us to Computational Biology where these permutations were first studied in the context of Gollan Conjecture. A wild guess of the diameter of the circular permutations by reversals appeard in an arxive preprint, where the working combinatorics was given by non-crossing partitions. In fact the combinatorics is given by reversals.

6 Future Work

6.1 a question of Davis

Recently, Krammer proved a long standing problem in group theory: the Braid groups are linear. A fundamental tool was a new description of the Braid group , having generators given by the set of non-crossing partitions of n elements. We can say that the generators are the vertices of the associahedron. The challenge is to answer a similar question from Davis et al. article, using Pevzner Theory of sorting by reversals, and Caprara's theorems on the relation between permutations and breaking point graphs. Davis' group is a group A, given by :

-involutory generators α_T, for each proper subinterval T of [1, n-2]. The relations are: if the distance between the intervals X and Y is at least 2, then α_X and α_Y commute. If X $\subset$ Y, then $\alpha_Y \alpha_X \alpha_Y = \alpha_Z$. Z denotes the image of X under the order-reversing involution of Y. The generators are the reversals from T. The relations are given by the relations among the reversals.

- Let G the Cayley graph of the symmetric group, with respect to the set of $n(n-1)/2$ transpositions (i,j). Take λ a cycle of length n, which is at the distance n-1 from the identity. Let Q be the set of all permutations which are on the geodesics from the identity to λ. The set Q, together with the relations given by permutation composition, is a presentation of the braid group B_n.

- The graph $G(n)$ from our second section, the Gollan permutation, and similar construction could play a role in the study of Davis's A- group.

6.2 Thurston , Tarjan and Sleator graph

A combinatorial proof of their results, or a proof which uses the spaces above, is our work in progress.

6.3 Signed permutations

. If we take the reversal graph of signed permutation, there is a space whose building blocks are glued in a manner prescribed by this graph: the edges show the incidence relations among the big cells. We are convinced that there is a nice interplay among the already developed theory from Computational Biology and its math counterpart, studied by Bott, Taubes, Axelrod and Singer. The geometric theory of these spaces is not so developed as the theory for $\overline{M}_{0,n}(R)$.

6.4 Computational Biology

The forces which drive Computational Biology, and Algebraic Geometry are so different, that it is too much to predict the future interactions among them. To speculate: we already know the power of Physics and Computer Sciences predictions in pure mathematics. Maybe the future will show predictions given by Genetics.

References

[1] Arbarello; Cornalba *Combinatorial and algebro-geometric cohomology classes on the moduli space of curves.* J.Alg.Geom. 5 (1996), no.4

[2] V. Bafna and P. Pevzner *Genome Rearrangements and sorting by reversals.* SIAM Journal of Computing Vol. 25, No 2 pp 272-289, April 1996.

[3] M.Davis, R.Scott, T.Januszkiewicz *Fundamental groups of blow-ups.* Advances in Mathematics, page 38, Remark 6.2.4 2002.

[4] S. Devadoss *Tessellations of moduli spaces and the mosaic operad..* Homotopy invariant algebraic structures 1998, pp.91-114 Contemp.Math. 239.

[5] M.Gromov *Hyperbolic Groups.* Essays in Group Theory (Gersten, ed.), Springer, Berlin 1988, 75-264

[6] M.Kapranov *Chow quotients of Grassmannians, I..* Adv.in Soviet Math 16(1993), 29-111.

[7] M.Kapranov, I.M. Gelfand, A.Zelevinsky *Discriminants, resultants, and multidimensional determinants.* Birkhuser Boston, Inc., Boston, MA, 1994.

[8] S.Keel *Intersection theory of moduli space of stable n-pointed curves of genus zero.* Trans. Amer. Math. Soc. **330** (1992), no. 2, 545-574.

[9] F. Knudsen *Projectivity of the moduli space of stable curves II.* Math. Scand. **52** (1983), 1225-1265.

[10] M.Kontsevich, Yu. Manin Quantum cohomology of a product. In *Inventiones mathematicae*, 124, 313-339 (1996)

[11] D. Krammer *Braid Groups are linear.* Annals of Mathematics 155 (2002), no. 1, 131-156.

[12] Yu.Manin, A.B. Goncharov *Multiple ζ-motives and moduli spaces* $\overline{\mathcal{M}}_{0,n}$. math.AG/0204102

[13] Mumford, D.; Fogarty, J.; Kirwan, F. *Geometric invariant theory. Third edition.* Ergebnisse der Mathematik und ihrer Grenzgebiete (2) 34. Springer-Verlag, Berlin, 1994.

[14] Munkres *Elements of Algebraic Topology* Addison-Wesley'1984

[15] P.A. Pevzner *Computational Molecular Biology.* A Bradford Book. MIT Press, Cambridge, MA, 2000. xviii+314 pp

[16] V. Poenaru *Discrete Groups.* Lecture Notes in Mathematics, Vol. 421, Springer-Verlag, Berlin-New York, 1974

[17] E.Spanier *Algebraic topology.* Springer-Verlag, New York-Berlin, 1981. xvi+528 pp

[18] G. Watterson, W. Ewens, Hall, A. Morgan *The Chromosome Inversion Problem.* Journal of Theoretical Biology , 1982 , 99. p.1-7.

[19] M. Yoshida *The democratic compactification of configuration spaces of point sets on the real projective line.* Kyushu J. of Math. 50 (1996) 493-512

[20] M.Yoshida, K.Mimachi *Intersection numbers of twisted cycles and the correlation functions of the conformal field theory II* . math.AG/0208097

[21] G.Ziegler *Lectures on Polytopes*, page 310. Springer'1995