

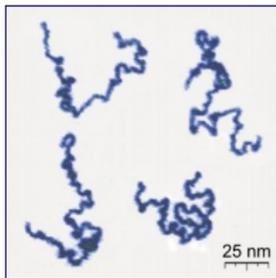
The differential geometry of spaces of polygons and polymers

Jason Cantarella

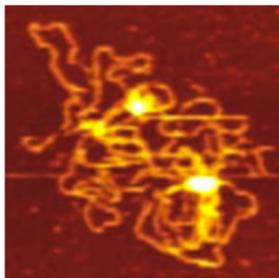
University of Georgia

Geometry and Shape Analysis in Biological Sciences
IMS Singapore, June 2017

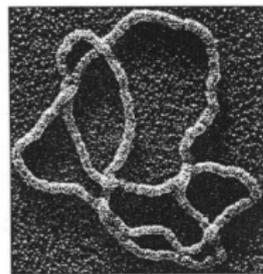
Space curves model useful biological shapes



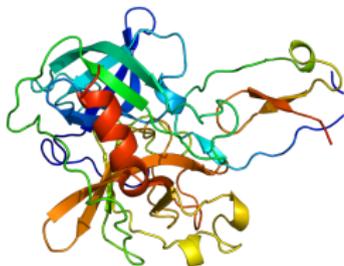
Protonated P2VP



Plasmid DNA



Knotted DNA



Elafin



Plant tendril



Cowpea root

- Replace curves with polygons.
- Use the (differential, symplectic, algebraic) geometry of polygon space.
- Add constraints (closed, fixed edglength, confined, different topology) as needed.
- Geometric structure \rightarrow efficient algorithms.

Definition

The quaternions \mathbb{H} are the skew-algebra over \mathbb{R} defined by adding \mathbf{i} , \mathbf{j} , and \mathbf{k} so that

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ijk} = -1$$

Proposition

Unit quaternions (S^3) double-cover $SO(3)$ (orthonormal frames for 3-space) via the Hopf map.

$$\text{Hopf}(q) = (\bar{q}\mathbf{i}q, \bar{q}\mathbf{j}q, \bar{q}\mathbf{k}q),$$

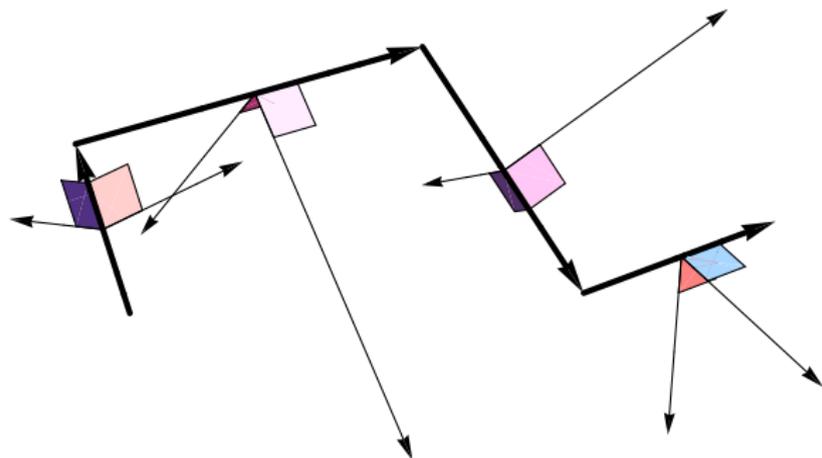
where the entries turn out to be purely imaginary quaternions, and hence vectors in \mathbb{R}^3 .

Constructing polygons from quaternions: Arms

Given a vector $\vec{q} \in \mathbb{H}^n$, we can construct a polygon:

$$\underbrace{(q_1, \dots, q_n)}_{\text{vector of quaternions}} \rightarrow \underbrace{(\text{Hopf}(q_1), \dots, \text{Hopf}(q_n))}_{\text{vector of frames}} \rightarrow \underbrace{(\bar{q}_1 \mathbf{i} q_1, \dots, \bar{q}_n \mathbf{i} q_n)}_{\text{vector of edges}}$$

We call the polygon $\text{Hopf}(\vec{q})$.



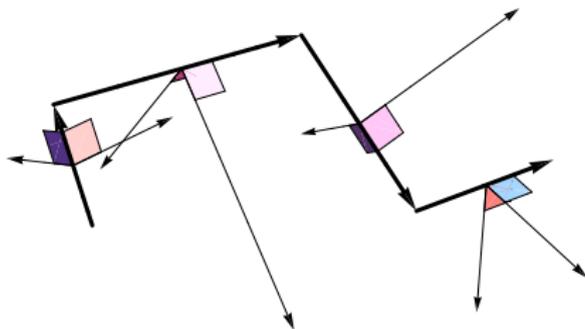
Proposition

framed n -gons of total length 1 $\iff S^{4n-1} \subset \mathbb{H}^n$

Proof.

The length of i -th edge is $|\bar{q}_i i q_i| = |q_i|^2$.

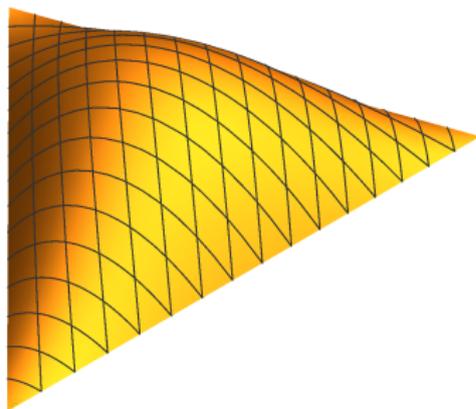
□



Proposition (with Shonkwiler)

The distribution of edges in the quaternionic model is:

- *directions are sampled independently, uniformly on $(S^2)^n$.*
- *lengths are sampled by the Dirichlet $(2, \dots, 2)$ distribution on the simplex $\{\vec{x} | x_i \geq 0, \sum x_i = 1\}$.*



\iff pdf is $\sim x_1 x_2 \cdots x_n$

Framed space polygon shape space

Alignment is an irritating problem in shape comparison: this geometric structure makes it easy.

Proposition

Multiplying \vec{q} by w rotates polygon by matrix $\text{Hopf}(w) \in \text{SO}(3)$.

Conclusion

Framed, length 1, space polygons (up to trans/rot) \iff

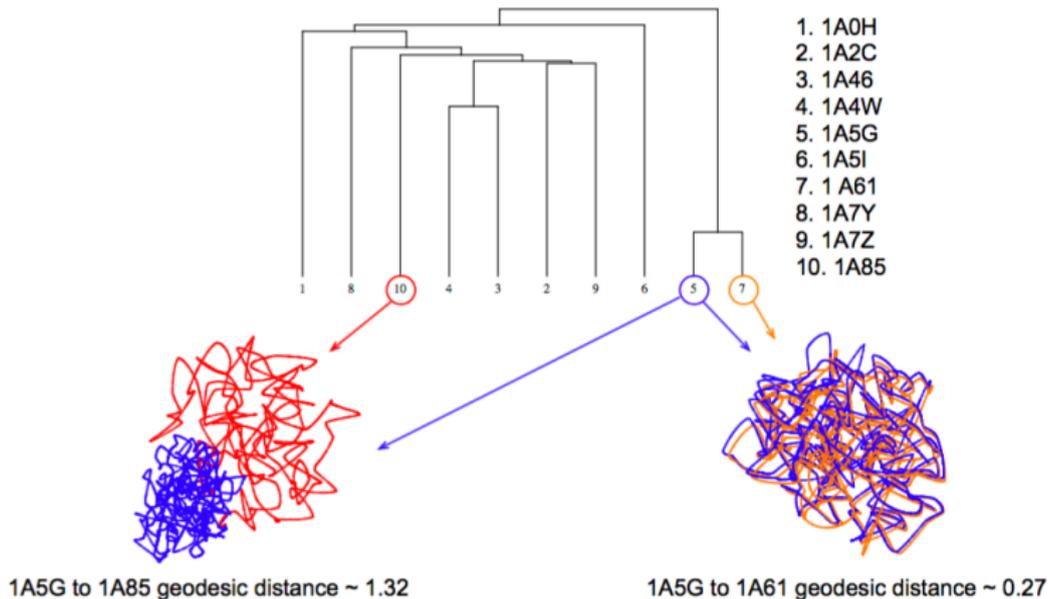
$$\mathbb{H}P^n = S^{4n-1} / (\vec{q} \simeq w\vec{q}, w \in \mathbb{H})$$

The metric on $\mathbb{H}P^n$ then gives a translation and rotation invariant distance function for space polygons:

$$\text{dist}(P, Q) = \text{acos} \sqrt{\frac{\langle P, Q \rangle \langle Q, P \rangle}{\langle P, P \rangle \langle Q, Q \rangle}}$$

Shape comparison of proteins from PDB

We can use the $\mathbb{H}IP^n$ distance to cluster protein shapes. Here is a proof-of-concept experiment with 10 proteins done by Tom Needham (Ohio State).



1A5G and 1A61 in the PDB

RCSB PDB MyPDB

Display Files - Download Files -

1A5G

HUMAN THROMBIN COMPLEXED WITH NOVEL SYNTHETIC PEPTIDE MIMETIC INHIBITOR AND HIRUGEN

DOI: [10.2210/pdb1a5g/pdb](https://doi.org/10.2210/pdb1a5g/pdb)

Classification: [HYDROLASE / HYDROLASE INHIBITOR](#)

Deposited: 1998-02-16 Released: 1998-05-27

Deposition author(s): [St Charles, R.](#), [Tulinsky, A.](#), [Kahn, M.](#)

Organism: [Homo sapiens](#) | [Hirudo medicinalis](#)

Structural Biology Knowledgebase: 1A5G (>24 annotations) [SBKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.06 Å

R-Value Work: 0.158

wwPDB Validation [3D Report](#) [Full Report](#)

Metric	Percentile Ranks	Value
Clashscore	~10	27
Ramachandran outliers	~10	2.2%
Sidechain outliers	~10	8.4%

Worse | Better
■ Percentile relative to all X-ray structures
□ Percentile relative to X-ray structures of similar resolution

Contact Us

RCSB PDB MyPDB

Display Files - Download Files -

1A61

THROMBIN COMPLEXED WITH A BETA-MIMETIC THIAZOLE-CONTAINING INHIBITOR

DOI: [10.2210/pdb1a61/pdb](https://doi.org/10.2210/pdb1a61/pdb)

Classification: [HYDROLASE / HYDROLASE INHIBITOR](#)

Deposited: 1998-03-05 Released: 1998-06-17

Deposition author(s): [St Charles, R.](#), [Matthews, J.H.](#), [Zhang, E.](#), [Tulinsky, A.](#), [Kahn, M.](#)

Organism: [Homo sapiens](#) | [Hirudo medicinalis](#)

Structural Biology Knowledgebase: 1A61 (>24 annotations) [SBKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.2 Å

R-Value Work: 0.148

wwPDB Validation [3D Report](#) [Full Report](#)

Metric	Percentile Ranks	Value
Clashscore	~10	38
Ramachandran outliers	~10	0.7%
Sidechain outliers	~10	12.7%
RSRZ outliers	~10	0.3%

Worse | Better
■ Percentile relative to all X-ray structures
□ Percentile relative to X-ray structures of similar resolution

Contact Us

A natural question is: how statistically significant is a distance of 0.27? In this setting, the answer is classical:

Proposition (based on Skriganov)

The probability that two random open polygons of n edges are within distance d of one another is given by

$$\sin^{4n} d$$

So we expect most completely unrelated polygons to be at distances which are very close to $\pi/2$, which is easy to check with experiments on random polygons.

Closed framed space polygons

Every quaternion $q = a + bj$, where $a, b \in \mathbb{C}$. This means that we can take *complex* vectors (\vec{a}, \vec{b}) corresponding to a quaternionic vector \vec{q} .

Proposition (Hausmann/Knutson)

P is closed, length 2 \iff the vectors (\vec{a}, \vec{b}) are Hermitian orthonormal.

Proof.

$$\text{Hopf}(a + bj) = (\overline{a + bj})i(a + bj) = i(|a|^2 - |b|^2 + 2\bar{a}bj)$$

so we have

$$\sum \text{Hopf}(a + bj) = 0 \iff \sum |a|^2 = \sum |b|^2, \sum \bar{a}b = 0.$$



Closed, rel. framed space poly shapes

Conclusion (Hausmann/Knutson)

Closed, framed space polygons \iff *Stiefel manifold* $V_2(\mathbb{C}^n)$.

Proposition (Hausmann/Knutson)

The action of the matrix group $U(2)$ *on* $V_2(\mathbb{C}^n)$

- *rotates the polygon in space* ($SU(2)$ *action*) **and**
- *spins all vectors of the frame* ($U(1)$ *action*).

Conclusion (Hausmann/Knutson)

Closed, rel. framed space polygons of length 2 \iff
Grassmannian of 2-planes in complex n-space $G_2(\mathbb{C}^n)$.

Metric on the Complex Grassmannian

An invariant distance between closed polygons is easy to compute in linear time:

Proposition

Given two closed, (relatively) framed polygons as $n \times 2$ complex matrices $Y_1 = (\vec{a}_1, \vec{b}_1)$ and $Y_2 = (\vec{a}_2, \vec{b}_2)$, let $\cos \theta_1, \cos \theta_2$ be the singular values of $Y_1^T Y_2$.

The singular values are invariant under translation and rotation and allow us to construct several metrics on polygon space:

$$d_{geo}(Y_1, Y_2) = \sqrt{\theta_1^2 + \theta_2^2}, \quad d_{chord}(Y_1, Y_2) = \sqrt{\sin^2 \theta_1 + \sin^2 \theta_2}$$

This provides a space curve version of well-known construction for plane curves:

Theorem (Younes–Michor–Shah–Mumford)

Roughly speaking,

$$\left\{ \begin{array}{l} \text{Contours in} \\ \mathbb{R}^2 \text{ modulo} \\ \text{similarities} \end{array} \right\} \longleftrightarrow \text{Gr}(2, C^\infty(S^1, \mathbb{R})),$$

where $\text{Gr}(k, V)$ is the Grassmannian of k -dimensional linear subspaces of the vector space V .



Total Curvature of Space Polygons

Proposition (with Grosberg, Kusner, Shonkwiler)

The expected value of total turning angle for an n -turn

- *open polygon is*

$$\frac{\pi}{2}n$$

- *closed polygon is*

$$\frac{\pi}{2}n + \frac{\pi}{4} \frac{2n}{2n-3}.$$

Proposition (with Grosberg, Kusner, Shonkwiler)

At least 1/3 of rel. framed hexagons and 1/11 of rel. framed heptagons are unknots.

To describe the space of equilateral polygons, we first recall that $G_2(\mathbb{C}^n)$ is a symplectic space; in fact it is the symplectic reduction of \mathbb{C}^{2n} by $U(2)$.

$$G_2(\mathbb{C}^n) = \underbrace{\mathbb{C}^{2n} // U(2)}_{U(2) \text{ acts on vectors } \vec{a}, \vec{b}}$$

Proposition (Knutson-Hausmann, Millson-Kapovich)

(Unframed) equilateral polygon space is a symplectic space; it is the symplectic reduction of $G_2(\mathbb{C}^n)$ by $U(1)^{n-1}$:

$$\text{ePol}_n = \underbrace{G_2(\mathbb{C}^n) // U(1)^{n-1}}_{U(1)^{n-1} \text{ rotates each frame around edge}}$$

Symplectic spaces are the right setting for classical mechanics:

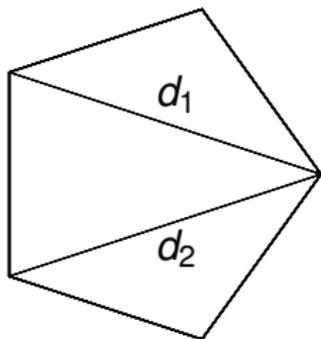
Theorem (Duistermaat-Heckmann, stated informally)

On a $2m$ -dimensional (symplectic) space,

*d continuous, commuting (Hamiltonian) symmetries \rightarrow
 d conserved quantities (momenta)*

joint distribution is cts, piecewise-polynomial, degree $\leq m - d$.

Polygons (up to rotation) are $2n - 6 = 2(n - 3)$ dimensional. Rotations around $n - 3$ chords d_i by $n - 3$ angles θ_i commute.



Theorem (with Shonkwiler)

The joint distribution of d_1, \dots, d_{n-3} and $\theta_1, \dots, \theta_{n-3}$ are all uniform (on their domains).

Proof.

Check D-H theorem applies (hard part).

Then count: $m = n - 3$ and we have $n - 3$ symmetries, so the pdf of the momenta d_i is piecewise polynomial of degree \leq

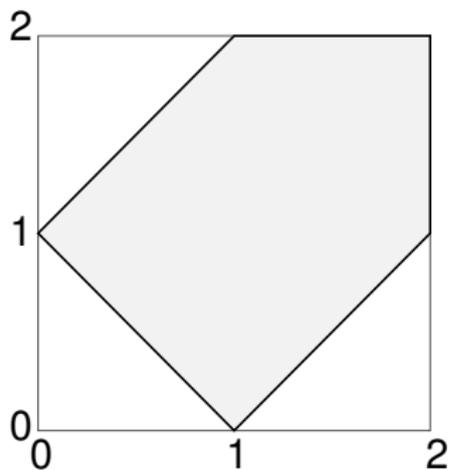
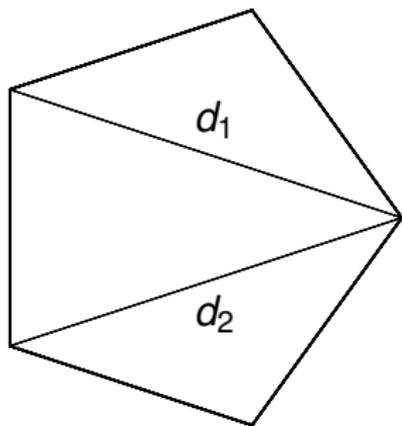
$$m - (n-3) = (n-3) - (n-3) = 0.$$

The pdf is continuous, so this means it's constant. □

What is the domain of the d_i ?

Definition

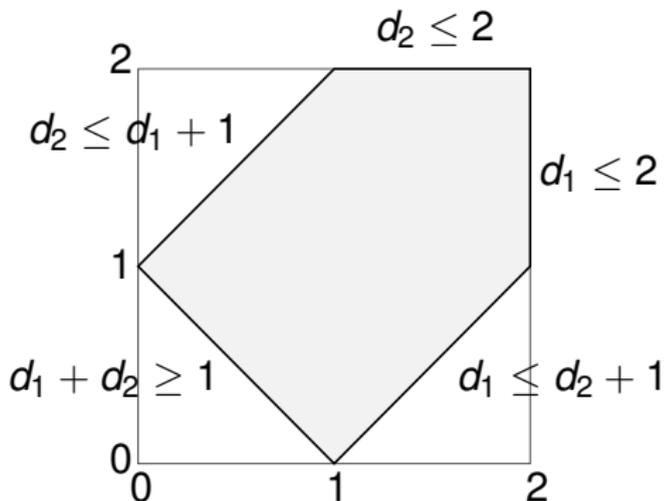
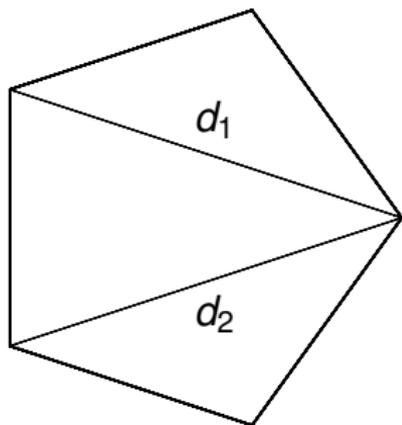
The momenta d_1, \dots, d_{n-3} obey triangle inequalities which determine an $n - 3$ dimensional polytope $\mathcal{P}_n \subset \mathbb{R}^{n-3}$. This is called the **moment polytope**.



What is the domain of the d_i ?

Definition

The momenta d_1, \dots, d_{n-3} obey triangle inequalities which determine an $n - 3$ dimensional polytope $\mathcal{P}_n \subset \mathbb{R}^{n-3}$. This is called the **moment polytope**.

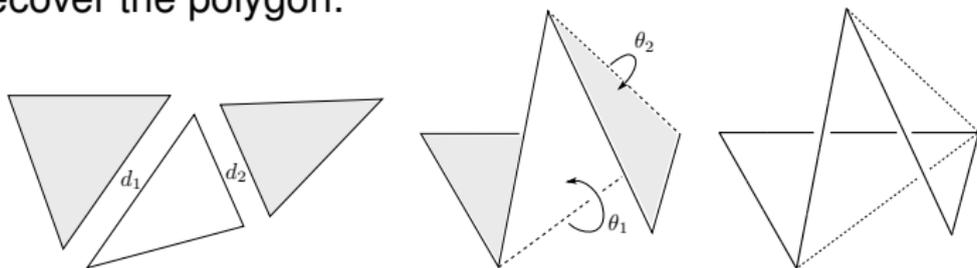


Definition

The d_i and θ_i are *action-angle coordinates* on polygon space. In these coordinates, the volume form is simple:

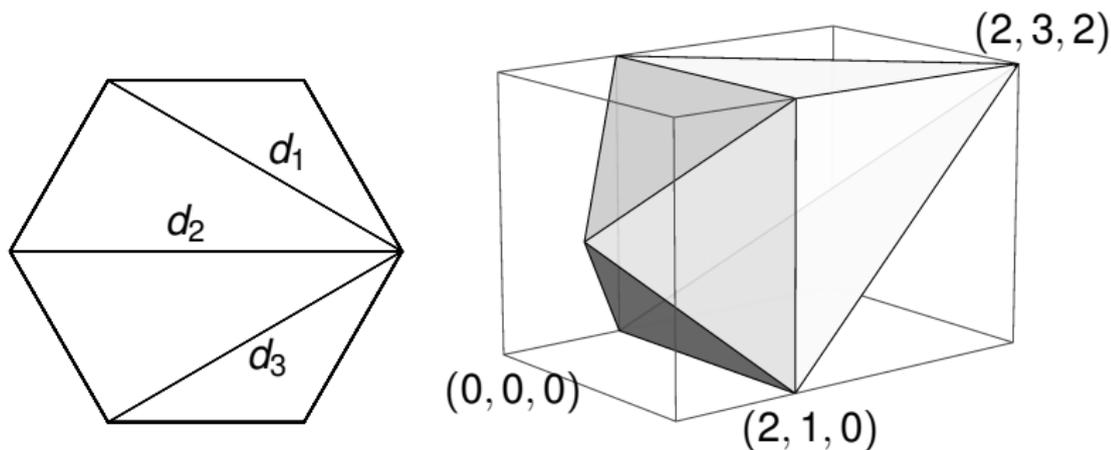
$$d\text{Vol} = dd_1 \wedge \dots \wedge dd_{n-3} \wedge d\theta_1 \wedge \dots \wedge d\theta_{n-3}.$$

To recover the polygon:



- Build the triangles from the edgelengths.
- Put the first one in a standard position.
- Place the rest using the dihedral angles.

Structure of the Moment Polytope



The polytope \mathcal{P}_n is defined by the triangle inequalities:

$$0 \leq d_1 \leq 2 \quad 1 \leq d_i + d_{i+1} \quad 0 \leq d_{n-3} \leq 2 \\ |d_i - d_{i+1}| \leq 1$$

Theorem (with Duplantier, Shonkwiler, Uehara)

A direct sampling algorithm for equilateral closed polygons with expected performance $O(n^{5/2})$ per sample.

If we let

$$s_i = d_i - d_{i-1}, \text{ for } 1 \leq i \leq n - 2$$

and $s_i \in [-1, 1]$, then d_i automatically have $|d_i - d_{i-1}| \leq 1$.

Proposition (with Duplantier, Shonkwiler, Uehara)

If we build d_i from s_i sampled uniformly in $[-1, 1]^n$, the d_i obey all triangle inequalities with probability $\sim 6\sqrt{6/\pi} n^{-3/2}$.

So rejection sample to build d_i , sample θ_i directly, and reassemble the polygon as above.

Diagonal sampling in 3 lines of code

```
RandomDiagonals[n_] :=  
  Accumulate[  
    Join[{1}, RandomVariate[UniformDistribution[{-1, 1}],  
      n]]];  
  
InMomentPolytopeQ[d_] :=  
  And[Last[d] ≥ 0, Last[d] ≤ 2,  
    And @@ (Total[#] ≥ 1 & /@ Partition[d, 2, 1])];  
  
DiagonalSample[n_] := Module[{d},  
  For[d = RandomDiagonals[n], ! InMomentPolytopeQ[d], ,  
    d = RandomDiagonals[n]];  
  d[[2 ;;]]  
];
```

Not first direct sampling algorithm (Grosberg-Moore, Diao-Ernst-Montemayor-Ziegler), but numerically stable, simple and fast.

Expected Value of Chord Lengths

Proposition (with Shonkwiler)

The expected length of a chord skipping k edges in an n -edge equilateral polygon is the $(k - 1)$ st coordinate of the center of mass of the moment polytope.

n	$k = 2$	3	4	5	6	7	8
4	1						
5	$\frac{17}{15}$	$\frac{17}{15}$					
6	$\frac{14}{12}$	$\frac{15}{12}$	$\frac{14}{12}$				
7	$\frac{461}{385}$	$\frac{506}{385}$	$\frac{506}{385}$	$\frac{461}{385}$			
8	$\frac{1,168}{960}$	$\frac{1,307}{960}$	$\frac{1,344}{960}$	$\frac{1,307}{960}$	$\frac{1,168}{960}$		
9	$\frac{112,121}{91,035}$	$\frac{127,059}{91,035}$	$\frac{133,337}{91,035}$	$\frac{133,337}{91,035}$	$\frac{127,059}{91,035}$	$\frac{112,121}{91,035}$	
10	$\frac{97,456}{78,400}$	$\frac{111,499}{78,400}$	$\frac{118,608}{78,400}$	$\frac{120,985}{78,400}$	$\frac{118,608}{78,400}$	$\frac{111,499}{78,400}$	$\frac{97,456}{78,400}$

Expected Value of Chord Lengths

Proposition (with Shonkwiler)

The expected length of a chord skipping k edges in an n -edge equilateral polygon is the $(k - 1)$ st coordinate of the center of mass of the moment polytope.

$$E(\text{chord}(37, 112)) =$$

2586147629602481872372707134354784581828166239735638
002149884020577366687369964908185973277294293751533
821217655703978549111529802222311915321645998238455
195807966750595587484029858333822248095439325965569
561018977292296096419815679068203766009993261268626
707418082275677495669153244706677550690707937136027
424519117786555575048213829170264569628637315477158
307368641045097103310496820323457318243992395055104
 ≈ 4.60973

Current: Equilateral polygons in other dimensions

Proposition

The space of closed equilateral polygons in \mathbb{R}^k is the quotient of (almost all) of the space $(S^{k-1})^n$ by the (diagonal) action of the Möbius group on S^{k-1} .

Proof.

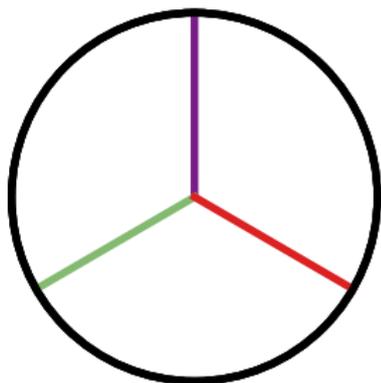
For each n point cloud on S^{k-1} (where fewer than $n/2$ points coincide) there is a unique (rotation-free) Möbius transformation which takes the center of mass to the origin. This associates a unique closed equilateral polygon with almost every polygonal arm. □

Current Project

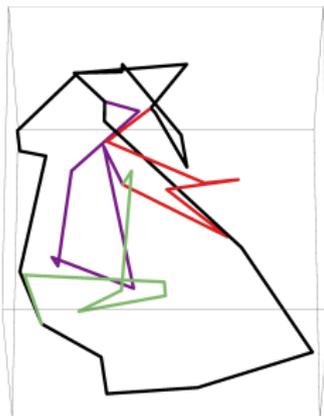
Use this structure to provide robust coordinates for equilateral polygons in all dimensions.

Current: Topologically Constrained Random Walks

A **topologically constrained random walk** (TCRW) is a collection of random walks in \mathbb{R}^3 whose components are required to realize the edges of some fixed multigraph.

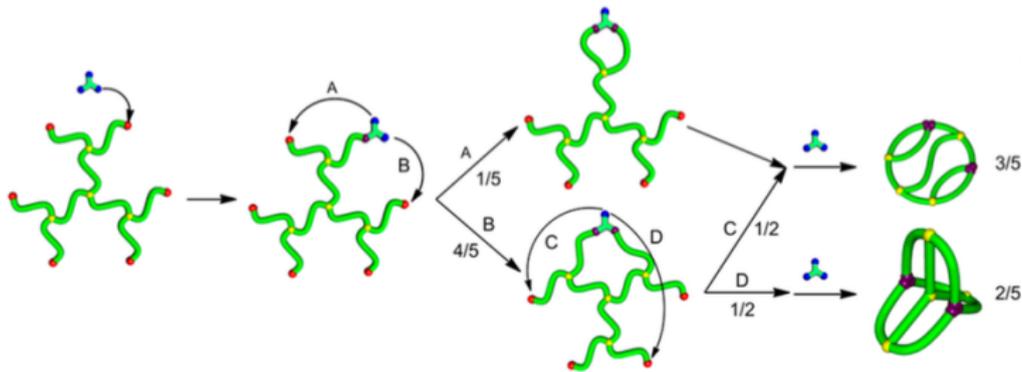


Abstract graph



TCRW

Current: Topologically Constrained Random Walks



Tezuka Lab, Tokyo Institute of Technology

A synthetic $K_{3,3}$!

Funding from Simons Foundation, NSF, Issac Newton Institute,

- *Probability Theory of Random Polygons from the Quaternionic Viewpoint* with T. Deguchi, and C. Shonkwiler
Comm. on Pure & Applied Mathematics **67** (2014)
- *The Expected Total Curvature of Random Polygons* with A. Y. Grosberg, R. Kusner, and C. Shonkwiler
American Journal of Mathematics **137** (2015)
- *The Symplectic Geometry of Closed Equilateral Random Walks in 3-space* with C. Shonkwiler
Annals of Applied Probability **26** (2016).
- *A Fast Direct Sampling Algorithm for Equilateral Closed Polygons* with B. Duplantier, C. Shonkwiler, and E. Uehara.
Journal of Physics A **49** (2016).