Applications of Geometry to Constrained Random Walks and Polymer Models

Clayton Shonkwiler

Colorado State University
shonkwiler.org

Geometry for Signal Processing and Machine Learning
October 14, 2016

Theorem (Younes–Michor–Shah–Mumford)

Roughly speaking,

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

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

Consequences: shape comparison, explicit geodesics, etc.

Theorem (Younes–Michor–Shah–Mumford)

Roughly speaking,

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

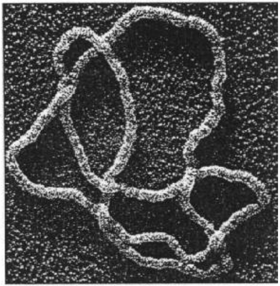
where $Gr(k, V)$ is the Grassmannian of k-dimensional linear subspaces of the vector space V.

Consequences: shape comparison, explicit geodesics, etc.

What is the analogous structure for closed curves in $\mathbb{R}^3$?

Statistical Physics Point of View

A ring polymer in solution takes on an ensemble of random shapes, with topology (knot type!) as the unique conserved quantity.



Knotted DNA
Wassermann et al.
Science **229**, 171–174

Statistical Physics Point of View

A ring polymer in solution takes on an ensemble of random shapes, with topology (knot type!) as the unique conserved quantity.

Physics Setup

> *Modern polymer physics is based on the analogy between a polymer chain and a random walk.*
>
> *—Alexander Grosberg*

Statistical Physics Point of View

A ring polymer in solution takes on an ensemble of random shapes, with topology (knot type!) as the unique conserved quantity.

Physics Setup

> *Modern polymer physics is based on the analogy between a polymer chain and a random walk.*
> —*Alexander Grosberg*

In this talk I will discretize and talk about *polygons* in $\mathbb{R}^3$, or loop random walks; most theorems generalize to smooth curves.

Question 0

What is the space of all possible conformations of n-gons in $\mathbb{R}^3$*?*

Question 0

What is the space of all possible conformations of n-gons in $\mathbb{R}^3$*?*

...and can you sample it uniformly?

Question 0

What is the space of all possible conformations of n-gons in $\mathbb{R}^3$*?*

...and can you sample it uniformly?
...and compute distances/geodesics/etc.?

Question 0

What is the space of all possible conformations of n-gons in $\mathbb{R}^3$*?*

...and can you sample it uniformly?
...and compute distances/geodesics/etc.?

Sampling Algorithms for Polygons:

- Markov Chain Algorithms
  - crankshaft (Vologoskii 1979, Klenin 1988)
  - polygonal fold (Millett 1994)
- Direct Sampling Algorithms
  - triangle method (Moore 2004)
  - generalized hedgehog method (Varela 2009)
  - sinc integral method (Moore 2005, Diao 2011)

Which (If Any) is Right?

Table 17 The number of distinct HOMFLY polynomials produced by each algorithm. Since the number of distinct knots sharing the same polynomial is small, this invariant is an suitable surrogate for knot type

Method	Sample size	Distinct HOMFLY
PFM	100 000 000	2 219
CRM	10 000 000	6 110
Hedgehog	10 000 000	1 111
Triangle	10 000 000	3 505

Alvarado, Calvo, Millett, *J. Stat. Phys.* **143** (2011), 102–138

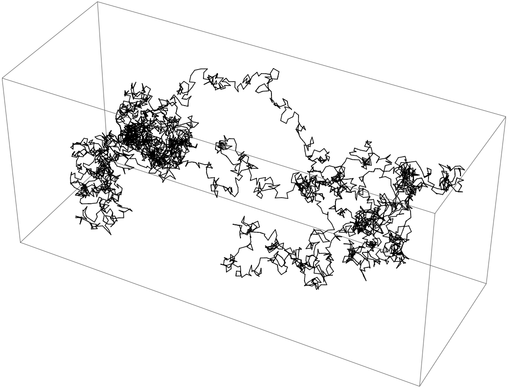
Ansatz

random closed curve $\iff$ *random point in some (nice!) conformation space*

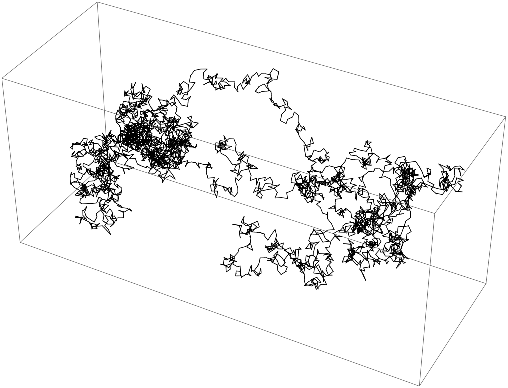
Scientific Idea

Knowledge of the (differential, symplectic, algebraic) geometry of these conformation spaces leads to both theorems and fast numerical algorithms for studying and comparing closed curves in $\mathbb{R}^3$*.*

Think of an n-step random walk up to translation as a collection of edge vectors $\vec{e}_1, \ldots, \vec{e}_n$.



Think of an n-step random walk up to translation as a collection of edge vectors $\vec{e}_1, \ldots, \vec{e}_n$.



The condition that it forms a closed loop is then

$$\vec{e}_1 + \ldots + \vec{e}_n = \vec{0}.$$

Clever Idea (in frames language)

Think of each edge vector $\vec{e}_i$ *as the* ***design vector*** *of* $\begin{bmatrix} a_i \\ b_i \end{bmatrix} \in \mathbb{C}^2$*:*

$$\vec{e}_i = \begin{bmatrix} |a_i|^2 - |b_i|^2 \\ -\operatorname{Im}(2\bar{a}_i b_i) \\ \operatorname{Re}(2\bar{a}_i b_i) \end{bmatrix}$$

Polygons are (Covered by) Parseval Frames

Clever Idea (in frames language)

Think of each edge vector $\vec{e}_i$ as the design vector of $\begin{bmatrix} a_i \\ b_i \end{bmatrix} \in \mathbb{C}^2$*:*

$$\vec{e}_i = \begin{bmatrix} |a_i|^2 - |b_i|^2 \\ -\operatorname{Im}(2\bar{a}_i b_i) \\ \operatorname{Re}(2\bar{a}_i b_i) \end{bmatrix}$$

Each random walk $(\vec{e}_1, \ldots, \vec{e}_n)$ corresponds to a frame

$$F = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{bmatrix} \text{ on } \mathbb{C}^2.$$

The squared norms of the columns are the edgelengths.

Polygons are (Covered by) Parseval Frames

Clever Idea (in frames language)

Think of each edge vector $\vec{e}_i$ *as the design vector of* $\begin{bmatrix} a_i \\ b_i \end{bmatrix} \in \mathbb{C}^2$*:*

$$\vec{e}_i = \begin{bmatrix} |a_i|^2 - |b_i|^2 \\ -\operatorname{Im}(2\bar{a}_i b_i) \\ \operatorname{Re}(2\bar{a}_i b_i) \end{bmatrix}$$

Each random walk $(\vec{e}_1, \ldots, \vec{e}_n)$ corresponds to a frame

$$F = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{bmatrix} \text{ on } \mathbb{C}^2.$$

The squared norms of the columns are the edgelengths.

Theorem (Hausmann–Knutson '97, Copenhaver et al. '14)

The random walk forms a closed polygon if and only if the frame is a λ*-tight frame (i.e.* $FF^T = \lambda I_2$*).*

The space of length-n Parseval frames ($\lambda = 1$) on $\mathbb{C}^2$ is the Stiefel manifold $St(2, \mathbb{C}^n)$, which is a very nice space...

The space of length-n Parseval frames ($\lambda = 1$) on $\mathbb{C}^2$ is the Stiefel manifold $St(2, \mathbb{C}^n)$, which is a very nice space...

Theorem (with Cantarella and Deguchi)

The space of n-gons admits a transitive $U(n)$ action, and so has a natural Haar measure. This can be sampled in linear time.

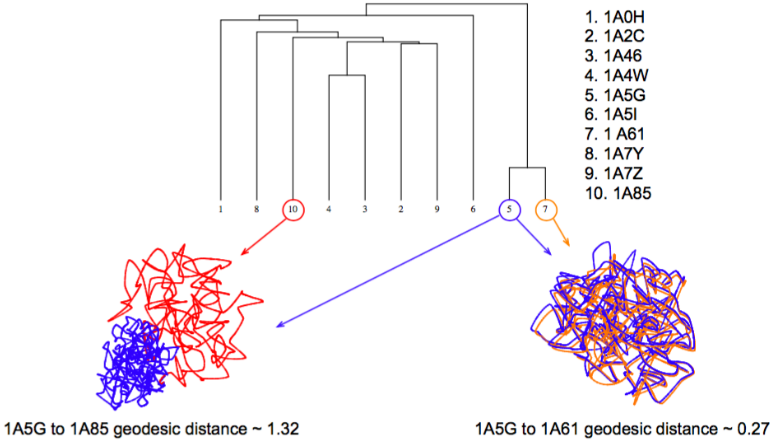
Theorem (with Cantarella, Grosberg, and Kusner)

The average turning angle at a vertex in a random n-gon is exactly $\frac{\pi}{2} + \frac{\pi}{4}\frac{2}{2n-3}$. This implies knot probability bounds.

Theorem (with Cantarella, Needham, and Stewart)

The probability that a random triangle is obtuse is exactly $\frac{3}{2} - \frac{3\ln 2}{\pi} \approx 0.838$.

Protein Registration and Clustering



Passing to the Grassmannian $Gr(2, \mathbb{C}^n)$ gives an orientation-independent representation of curves; cluster by geodesic distance (Needham '16).

Equilateral Space Polygons

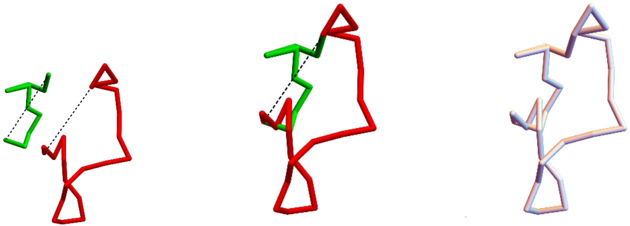
Polymer physicists are most interested in equilateral random walks, where all edges are the same length. Think of these as discretized arclength-parametrized curves.

In frames language, an equilateral polygon corresponds to a *unit norm* tight frame.

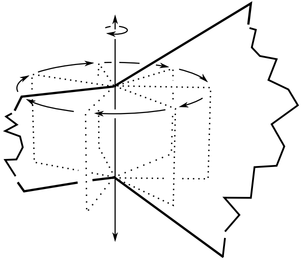
Polymer physicists are most interested in equilateral random walks, where all edges are the same length. Think of these as discretized arclength-parametrized curves.
In frames language, an equilateral polygon corresponds to a *unit norm* tight frame.

Idea

A random closed n-edge polygon is a k-edge random walk and an $n-k$-edge random walk, conditioned on having the same end-to-end distance.

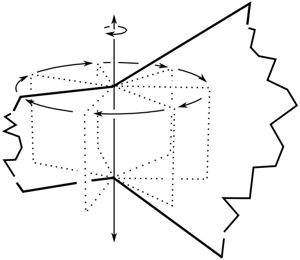


We can rotate the k-edge walk and the $(n - k)$-edge walk with respect to each other



This is a continuous symmetry of the polygon space.

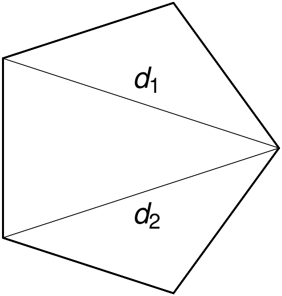
We can rotate the k-edge walk and the $(n-k)$-edge walk with respect to each other



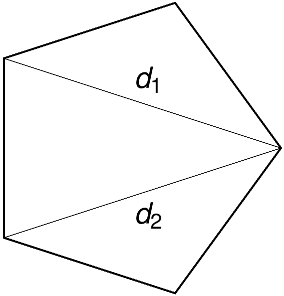
This is a continuous symmetry of the polygon space.

Should have associated conserved quantity or *moment map*.

Rotations around $n-3$ chords d_i by $n-3$ angles θ_i commute.

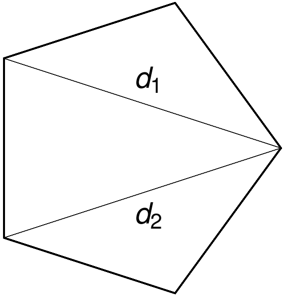


Rotations around $n-3$ chords d_i by $n-3$ angles θ_i commute.



But the space of random equilateral polygons up to translation and rotation is $(2n-6)$-dimensional!

Rotations around $n-3$ chords d_i by $n-3$ angles θ_i commute.



But the space of random equilateral polygons up to translation and rotation is $(2n-6)$-dimensional!

Theorem (Kapovich–Millson '96, Hausmann–Knutson '97)

The space of equilateral n-gons in $\mathbb{R}^3$ up to translation and rotation is toric symplectic and arises as the symplectic reduction of $Gr(2,\mathbb{C}^n)$ by the torus of diagonal matrices in $U(n)$.

A Drastic (Measure-Theoretic) Simplification

Theorem (with Cantarella)

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

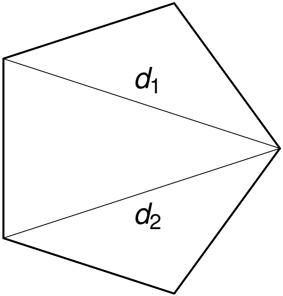
A Drastic (Measure-Theoretic) Simplification

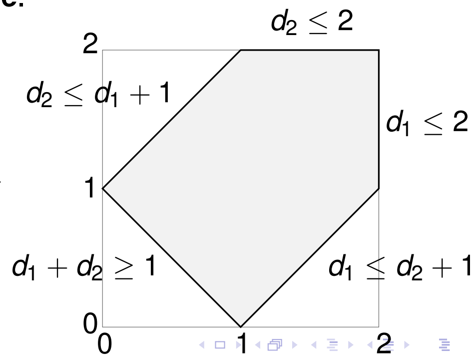
Theorem (with Cantarella)

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

Definition

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





You can compute the expected value of any sum of functions of single chordlengths this way (but we don't understand the structure of the answers yet):

$$E(\text{HydrodynamicRadius}(15)) = \frac{4}{3883074281625}(427763619147 + 11873777090560 \log\left(\frac{7}{6}\right) - 11591065307360 \log\left(\frac{6}{5}\right) + 2915195692640 \log\left(\frac{5}{4}\right) - 173574718240 \log\left(\frac{4}{3}\right) + 1072203440 \log\left(\frac{3}{2}\right) - 10010 \log(2)) \approx 0.768279$$

Theorem (with Cantarella, Duplantier, and Uehara)

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

Theorem (with Cantarella, Duplantier, and 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 Cantarella, Duplantier, and 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.

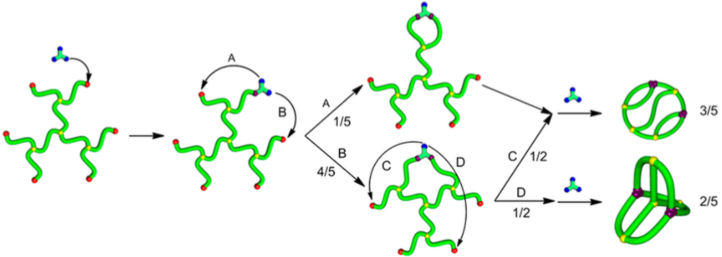
Diagonal sampling in 3 lines of code

```
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 ;;]]
  ];
```

A Question



Tezuka Lab, Tokyo Institute of Technology

Question

What special geometric structures exist on the conformation spaces of multicurves with more complicated topologies?

Thank you for listening!

References

- J. Cantarella, T. Deguchi, and C. Shonkwiler
  Communications on Pure and Applied Mathematics **67** (2014), no. 10, 1658–1699.

- J. Cantarella, A. Y. Grosberg, R. Kusner, and C. Shonkwiler
  American Journal of Mathematics **137** (2015), no. 2, 411–438.

- J. Cantarella and C. Shonkwiler
  Annals of Applied Probability **26** (2016), 549–596.

- J. Cantarella, B. Duplantier, C. Shonkwiler, and E. Uehara
  Journal of Physics A **49** (2016), 275202.

Funding: Simons Foundation, Isaac Newton Institute

`http://shonkwiler.org`