New algorithms for sampling closed and/or confined equilateral polygons

Jason Cantarella and Clayton Shonkwiler

University of Georgia

BIRS Workshop
Entanglement in Biology
November, 2013

Definition

A random (open) polygon in $\mathbb{R}^3$ is a set of edge vectors $\vec{e}_1, \ldots, \vec{e}_n$ sampled independently from the unit sphere. We call this sample space

$$\mathrm{Arm}(n) := \underbrace{S^2 \times \cdots \times S^2}_{n \text{ times}}$$

Definition

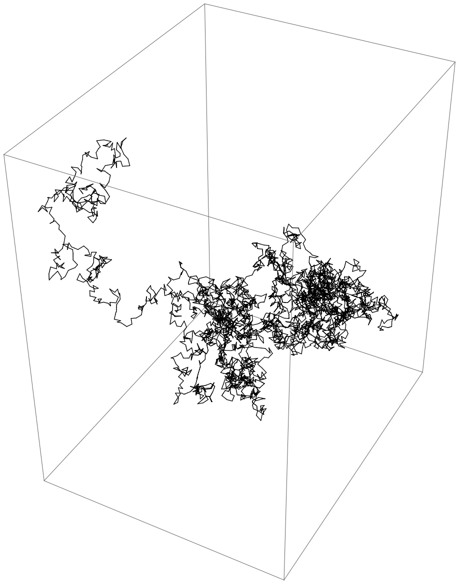
A random (open) polygon in $\mathbb{R}^3$ is a set of edge vectors $\vec{e}_1, \ldots, \vec{e}_n$ sampled independently from the unit sphere. We call this sample space

$$\mathrm{Arm}(n) := \underbrace{S^2 \times \cdots \times S^2}_{n \text{ times}}$$

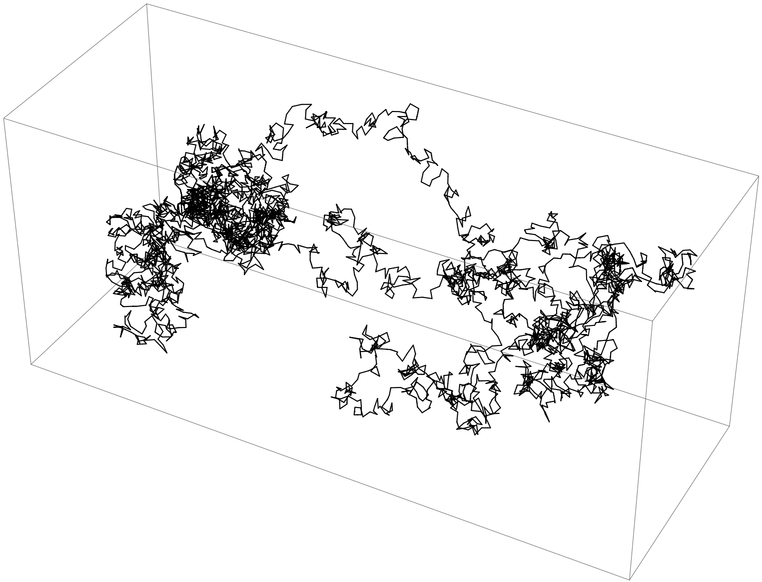
Definition

A random closed polygon conditions these samples on the hypothesis that $\sum \vec{e}_i = \vec{0}$, or samples from the submanifold of $\mathrm{Arm}(n)$ where $\sum \vec{e}_i = 0$, which we denote $\mathrm{Pol}(n)$.

Open Equilateral Random Polygon with 3,500 edges



Closed Equilateral Random Polygon with 3,500 edges



- What is the joint pdf of edge vectors in a closed walk?

Classical Problems

- What is the joint pdf of edge vectors in a closed walk?
- What can we prove about closed random walks?
  - What is the marginal distribution of a single chord length?
  - What is the joint distribution of several chord lengths?
  - What is the expectation of radius of gyration?
  - What is the expectation of total curvature?

- What is the joint pdf of edge vectors in a closed walk?
- What can we prove about closed random walks?
    - What is the marginal distribution of a single chord length?
    - What is the joint distribution of several chord lengths?
    - What is the expectation of radius of gyration?
    - What is the expectation of total curvature?
- How do we sample closed equilateral random walks?
    - What if the walk is confined to a sphere? (Confined DNA)
    - What if the edge lengths vary? (Loop closures)
    - Can we get error bars?

- What is the joint pdf of edge vectors in a closed walk?
- What can we prove about closed random walks?
  - What is the marginal distribution of a single chord length?
  - What is the joint distribution of several chord lengths?
  - What is the expectation of radius of gyration?
  - What is the expectation of total curvature?
- How do we sample closed equilateral random walks?
  - What if the walk is confined to a sphere? (Confined DNA)
  - What if the edge lengths vary? (Loop closures)
  - Can we get error bars?

Point of Talk

New sampling algorithms backed by deep and robust mathematical framework. Guaranteed to converge. Hard math, relatively easy code.

(Incomplete?) History of Sampling Algorithms

- 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)

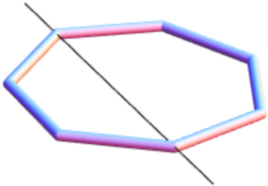
(Incomplete?) History of Sampling Algorithms

- Markov Chain Algorithms
  - crankshaft (Vologoskii et al. 1979, Klenin et al. 1988)
    - convergence to correct pdf unproved
  - polygonal fold (Millett 1994)
    - convergence to correct pdf unproved
- Direct Sampling Algorithms
  - triangle method (Moore et al. 2004)
    - samples a subset of closed polygons
  - generalized hedgehog method (Varela et al. 2009)
    - unproved whether this is correct pdf
  - sinc integral method (Moore et al. 2005, Diao et al. 2011)
    - requires sampling from complicated 1-d polynomial PDFs

Definition

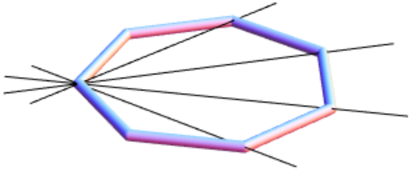
A *fold move* or *bending flow* rotates an arc of the polygon around the axis its endpoints.

The polygonal fold Markov chain selects arcs and angles at random and folds repeatedly.



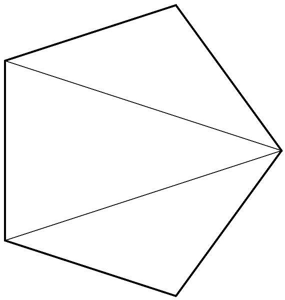
Definition

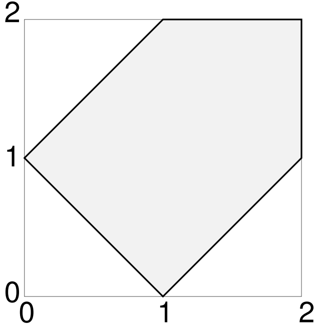
Given an (abstract) triangulation of the n-gon, the folds on any two chords commute. A *dihedral angle* move rotates around all of these chords by independently selected angles.



Definition

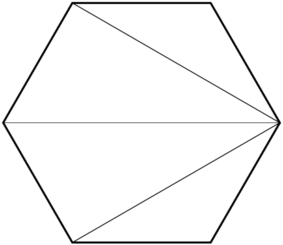
A abstract triangulation T of the n-gon picks out $n-3$ nonintersecting chords. The lengths of these chords obey triangle inequalities, so they lie in a convex polytope in $\mathbb{R}^{n-3}$ called the *triangulation polytope* $\mathcal{P}$.

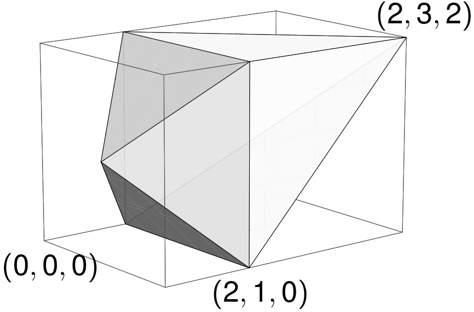




Definition

A abstract triangulation T of the n-gon picks out $n - 3$ nonintersecting chords. The lengths of these chords obey triangle inequalities, so they lie in a convex polytope in $\mathbb{R}^{n-3}$ called the *triangulation polytope* $\mathcal{P}$.

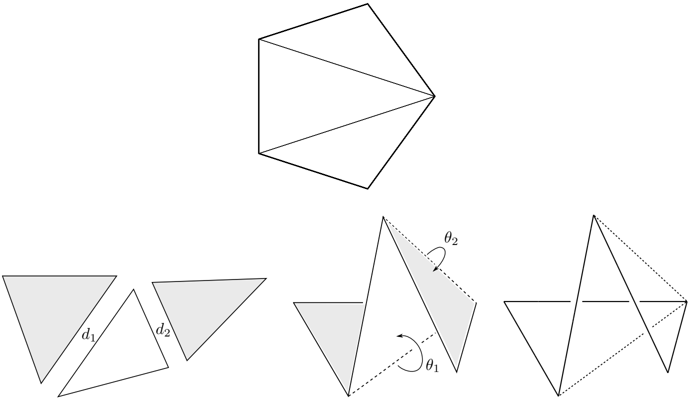




Definition

If $\mathcal{P}$ is the triangulation polytope and T^{n-3} is the torus of $n-3$ dihedral angles, then there are *action-angle coordinates*:

$$\alpha\colon \mathcal{P} \times T^{n-3} \to \mathrm{Pol}(n)/\,\mathrm{SO}(3)$$



Theorem (with Shonkwiler)

*α pushes the **standard probability measure** on $\mathcal{P} \times T^{n-3}$ forward to the **correct probability measure** on $\operatorname{Pol}(n)/\operatorname{SO}(3)$.*

Theorem (with Shonkwiler)

α *pushes the* ***standard probability measure*** *on* $\mathcal{P} \times T^{n-3}$ *forward to the* ***correct probability measure*** *on* $\operatorname{Pol}(n)/\operatorname{SO}(3)$.

Proof.

Millson-Kapovich toric symplectic structure on polygon space + Duistermaat-Heckmann theorem + Hitchin's theorem on compatibility of Riemannian and symplectic volume on symplectic reductions of Kähler manifolds + Howard-Manon-Millson analysis of polygon space. □

Theorem (with Shonkwiler)

*α pushes the **standard probability measure** on $\mathcal{P} \times T^{n-3}$ forward to the **correct probability measure** on* $\operatorname{Pol}(n)/\operatorname{SO}(3)$.

Proof.

Millson-Kapovich toric symplectic structure on polygon space + Duistermaat-Heckmann theorem + Hitchin's theorem on compatibility of Riemannian and symplectic volume on symplectic reductions of Kähler manifolds + Howard-Manon-Millson analysis of polygon space. □

Corollary

Any sampling algorithm for convex polytopes is a sampling algorithm for closed equilateral polygons.

Proposition (with Shonkwiler)

The joint pdf of the $n-3$ *chord lengths in an abstract triangulation of the n-gon in a closed random equilateral polygon is Lesbegue measure on the triangulation polytope.*

Proposition (with Shonkwiler)

The joint pdf of the $n-3$ chord lengths in an abstract triangulation of the n-gon in a closed random equilateral polygon is Lesbegue measure on the triangulation polytope.

The marginal pdf of a single chordlength is a piecewise-polynomial function given by the volume of a slice of the triangulation polytope in a coordinate direction.

These marginals derived by Moore/Grosberg 2004 and Diao/Ernst/Montemayor/Ziegler 2011.

Proposition (with Shonkwiler)

The joint pdf of the $n - 3$ chord lengths in an abstract triangulation of the n-gon in a closed random equilateral polygon is Lesbegue measure on the triangulation polytope.

The marginal pdf of a single chordlength is a piecewise-polynomial function given by the volume of a slice of the triangulation polytope in a coordinate direction.

These marginals derived by Moore/Grosberg 2004 and Diao/Ernst/Montemayor/Ziegler 2011.

Corollary (with Shonkwiler)

The expectation of any function of a collection of non-intersecting chordlengths can be computed by integrating over the triangulation polytope.

Proposition (with Shonkwiler)

The expected length of a chord skipping k edges in an n-gon is the $k-1$st coordinate of the center of mass of the fan triangulation polytope.

We can check these centers of mass against the first moments of the MG-DEMZ chordlength marginals:

Proposition (with Shonkwiler)

The expected length of a chord skipping k edges in an n-gon is the $k-1$st coordinate of the center of mass of the fan triangulation polytope.

We can check these centers of mass against the first moments of the MG-DEMZ chordlength marginals:

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	$\frac{112{,}121}{91{,}035}$	$\frac{127{,}059}{91{,}035}$	$\frac{133{,}337}{91{,}035}$	$\frac{133{,}337}{91{,}035}$	$\frac{127{,}059}{91{,}035}$	$\frac{112{,}121}{91{,}035}$	

Theorem (with Shonkwiler)

At least $1/2$ of equilateral hexagons are unknotted.

Theorem (with Shonkwiler)

At least 1/2 of equilateral hexagons are unknotted.

Proof.

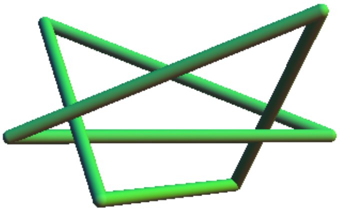
Consider the triangulation of the hexagon given by joining vertices 1, 3, and 5 by diagonals and the corresponding action-angle coordinates.

Using a result of Calvo, in either this triangulation or the $2-4-6$ triangulation, dihedral angles $\theta_1, \theta_2, \theta_3$ of a hexagonal trefoil must all be either between 0 and π or between π and 2π. Therefore, the fraction of knots is no bigger than

$$2\frac{\mathrm{Vol}([0,\pi]^3)+\mathrm{Vol}([\pi,2\pi]^3)}{\mathrm{Vol}(T^3)}=\frac{2\pi^3}{8\pi^3}=\frac{1}{2}$$

Theorem (with Shonkwiler)

At least $1/2$ of equilateral hexagons are unknotted.



To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

Definition (Hit-and-run Sampling Markov Chain)

Given $\vec{p}_k \in \mathcal{P} \subset \mathbb{R}^n$,

To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

Definition (Hit-and-run Sampling Markov Chain)

Given $\vec{p}_k \in \mathcal{P} \subset \mathbb{R}^n$,

1. Choose a random direction $\vec{v}$ uniformly on S^{n-1}.

To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

Definition (Hit-and-run Sampling Markov Chain)

Given $\vec{p}_k \in \mathcal{P} \subset \mathbb{R}^n$,

1. Choose a random direction $\vec{v}$ uniformly on S^{n-1}.
2. Let ℓ be the line through $\vec{p}_k$ in direction $\vec{v}$.

To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

Definition (Hit-and-run Sampling Markov Chain)

Given $\vec{p}_k \in \mathcal{P} \subset \mathbb{R}^n$,

1. Choose a random direction $\vec{v}$ uniformly on S^{n-1}.
2. Let ℓ be the line through $\vec{p}_k$ in direction $\vec{v}$.
3. Choose $\vec{p}_{k+1}$ uniformly on $\ell \cap \mathcal{P}$.

To repeat myself:

Action-angle coordinates reduce sampling equilateral polygon space to the (solved) problem of sampling a convex polytope.

Definition (Hit-and-run Sampling Markov Chain)

Given $\vec{p}_k \in \mathcal{P} \subset \mathbb{R}^n$,

1. Choose a random direction $\vec{v}$ uniformly on S^{n-1}.
2. Let ℓ be the line through $\vec{p}_k$ in direction $\vec{v}$.
3. Choose $\vec{p}_{k+1}$ uniformly on $\ell \cap \mathcal{P}$.

Theorem (Smith, 1984)

The m-step transition probability of hit-and-run starting at any point $\vec{p}$ in the interior of $\mathcal{P}$ converges geometrically to Lesbegue measure on $\mathcal{P}$ as $m \to \infty$.

Definition (TSMCMC(β))

Given a triangulation T of the n-gon and associated polytope $\mathcal{P}$. If $x_k = (\vec{p}_k, \vec{\theta}_k) \in \mathcal{P} \times T^{n-3}$, define x_{k+1} by

- Update $\vec{p}_k$ by a hit-and-run step on $\mathcal{P}$ with probability β.
- Replace $\vec{\theta}_k$ with a new uniformly sampled point in T^{n-3} with probability $1 - \beta$.

At each step, construct the corresponding polygon $\alpha(x_k)$ using action-angle coordinates.

Definition (TSMCMC(β))

Given a triangulation T of the n-gon and associated polytope $\mathcal{P}$. If $x_k = (\vec{p}_k, \vec{\theta}_k) \in \mathcal{P} \times T^{n-3}$, define x_{k+1} by

- Update $\vec{p}_k$ by a hit-and-run step on $\mathcal{P}$ with probability β.
- Replace $\vec{\theta}_k$ with a new uniformly sampled point in T^{n-3} with probability $1 - \beta$.

At each step, construct the corresponding polygon $\alpha(x_k)$ using action-angle coordinates.

Proposition (with Shonkwiler)

Starting at any polygon, the m-step transition probability of TSMCMC(β) converges geometrically to the standard probability measure on $\mathrm{Pol}(n)/SO(3)$.

Error Analysis for Integration with TSMCMC(β)

Suppose f is a function on polygons. If a run R of TSMCMC(β) produces $x_1, \ldots, x_m$, let

$$\text{SampleMean}(f; R, m) := \frac{1}{m} \sum_{k=1}^{m} f(\alpha(x_k))$$

be the sample average of the values of f over the run.

Because TSMCMC(β) converges geometrically, we have

[1] w denotes weak convergence, $E(f)$ is the expectation of f

Error Analysis for Integration with TSMCMC(β)

Suppose f is a function on polygons. If a run R of TSMCMC(β) produces $x_1, \ldots, x_m$, let

$$\text{SampleMean}(f; R, m) := \frac{1}{m} \sum_{k=1}^{m} f(\alpha(x_k))$$

be the sample average of the values of f over the run.

Because TSMCMC(β) converges geometrically, we have

Theorem (Markov Chain Central Limit Theorem)

If f is square-integrable, there exists a real number $\sigma(f)$ so that[1]

$$\sqrt{m}(\text{SampleMean}(f; R, m) - E(f)) \xrightarrow{w} \mathcal{N}(0, \sigma(f)^2),$$

the Gaussian with mean 0 and standard deviation $\sigma(f)^2$.

[1] w denotes weak convergence, $E(f)$ is the expectation of f

Given a length-m run R of TSMCMC and a square integrable function $f\colon M \to \mathbb{R}$, we can compute SampleMean($f; R, m$), there is a statistically consistent estimator called the **Geyer IPS Estimator** $\bar{\sigma}_m(f)$ for $\sigma(f)$.

According to the estimator, a 95% confidence interval for the expectation of f is given by

$$E(f) \in \text{SampleMean}(f; R, m) \pm 1.96\bar{\sigma}_m(f)/\sqrt{m}.$$

TSMCMC(β) Error Bars

Given a length-m run R of TSMCMC and a square integrable function $f\colon M \to \mathbb{R}$, we can compute SampleMean$(f; R, m)$, there is a statistically consistent estimator called the **Geyer IPS Estimator** $\bar{\sigma}_m(f)$ for $\sigma(f)$.

According to the estimator, a 95% confidence interval for the expectation of f is given by

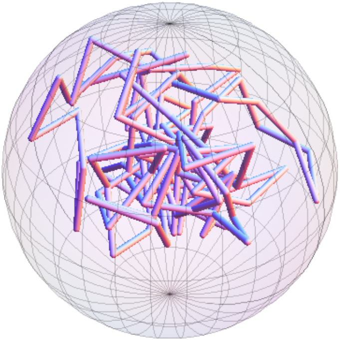
$$E(f) \in \text{SampleMean}(f; R, m) \pm 1.96\bar{\sigma}_m(f)/\sqrt{m}.$$

Experimental Observation

With 95% confidence, we can say that the fraction of knotted equilateral hexagons is between 1.1 and 1.5 in 10,000.

Definition

A polygon $p \in \text{Pol}(n; \vec{r})$ is in *rooted spherical confinement* of radius r if each diagonal length $d_i \leq r$. Such a polygon is contained in a sphere of radius r centered at the first vertex.



Proposition (with Shonkwiler)

Polygons in rooted spherical confinement of radius r have action-angle coordinates given by the polytope

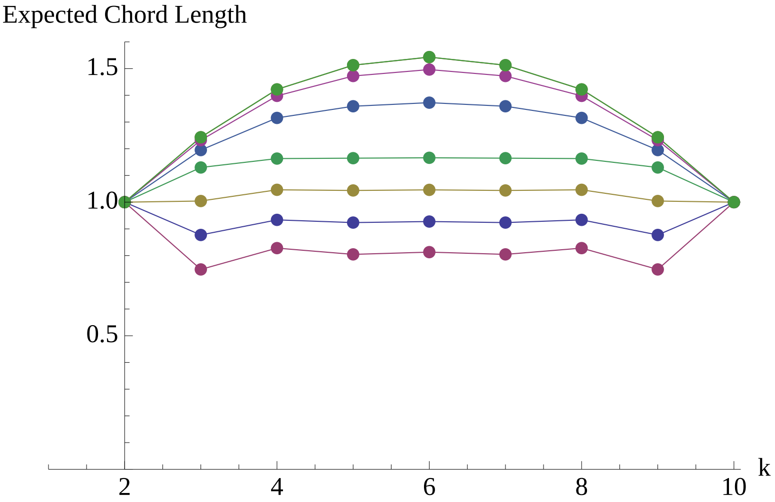
$$0 \leq d_1 \leq 2 \qquad \begin{matrix} 1 \leq d_i + d_{i+1} \\ |d_i - d_{i+1}| \leq 1 \end{matrix} \qquad 0 \leq d_{n-3} \leq 2$$

with the additional linear inequalities

$$d_i \leq r.$$

These polytopes are simply subpolytopes of the fan triangulation polytopes. Many other confinement models are possible!

Expected Chordlength Theorem for Confined 10-gons



Confinement radii are 1.25, 1.5, 1.75, 2, 3, 4, and 5.

Thank you!

Thank you for listening!

References

- *Probability Theory of Random Polygons from the Quaternionic Viewpoint*
  Jason Cantarella, Tetsuo Deguchi, and Clayton Shonkwiler
  arXiv:1206.3161
  Communications on Pure and Applied Mathematics
  (2013), doi:10.1002/cpa.21480.

- *The Expected Total Curvature of Random Polygons*
  Jason Cantarella, Alexander Y Grosberg, Robert Kusner, and Clayton Shonkwiler
  arXiv:1210.6537.

- *The symplectic geometry of closed equilateral random walks in 3-space*
  Jason Cantarella and Clayton Shonkwiler
  arXiv:1310.5924.