Simulating Constrained Random Walks for Applications to Polymer Models

Clayton Shonkwiler, Jason Cantarella, …

Colorado State University and University of Georgia
shonkwiler.org and jasoncantarella.com

SIAM LS16
July 14, 2016
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
Knot complexity in DNA from P4 tailless mutants

Arsuaga et al., *PNAS* 99 (2002), 5373–5377
Knot complexity in DNA from P4 tailless mutants

Arusuaga et al., *PNAS* 99 (2002), 5373–5377

Is this surprising?
Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology as the unique conserved quantity.

Protonated P2VP
Roiter/Minko
Clarkson University

Plasmid DNA
Alonso-Sarduy, Dietler Lab
EPF Lausanne
Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology as the unique conserved quantity.

Nano-structured hydrogel
Theoretical Framework

Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology 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
Simulating Random Walks is Easy...

A random walk with 3500 steps
Sampling Algorithms for (Equilateral) 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample size</th>
<th>Distinct HOMFLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFM</td>
<td>100 000 000</td>
<td>2 219</td>
</tr>
<tr>
<td>CRM</td>
<td>10 000 000</td>
<td>6 110</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>10 000 000</td>
<td>1 111</td>
</tr>
<tr>
<td>Triangle</td>
<td>10 000 000</td>
<td>3 505</td>
</tr>
</tbody>
</table>

**Main Ideas**

**Ansatz**

*random walk $\iff$ random point in some (nice!) moduli space*

**Scientific Idea**

*Use the (differential, symplectic, algebraic) geometry and topology of these moduli spaces to prove theorems and devise algorithms for studying 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.
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.

Tezuka Lab, Tokyo Institute of Technology

A synthetic $K_{3,3}$!
Random Walk Questions

- What is the joint distribution of steps in a TCRW?
- What can we prove about TCRWs?
  - What is the joint distribution of vertex–vertex distances?
  - What is the expectation of radius of gyration?
  - Spectrum of graph isotopy types?
- How do we sample TCRWs?
Closed Random Walks (a.k.a. Random Polygons)

The simplest multigraph with at least one edge is , which corresponds to a *classical* random walk, modeling a *linear* polymer.
Closed Random Walks (a.k.a. Random Polygons)

The simplest multigraph with at least one edge is \( \uparrow \), which corresponds to a *classical* random walk, modeling a *linear* polymer.

The next simplest multigraph is \( \bigcirc \), which yields a *closed* random walk (or *random polygon*), modeling a *ring* polymer.
Closed Random Walks (a.k.a. Random Polygons)

The simplest multigraph with at least one edge is \( \xrightarrow{\ } \), which corresponds to a classical random walk, modeling a linear polymer.

The next simplest multigraph is \( \bigcirc \), which yields a closed random walk (or random polygon), modeling a ring polymer.
Let Arm\((n)\) be the moduli space of random walks in \(\mathbb{R}^3\) consisting of \(n\) unit steps up to translation.

\[
\text{Arm}(n) \cong S^2 \times \ldots \times S^2.
\]
Let $\text{Arm}(n)$ be the moduli space of random walks in $\mathbb{R}^3$ consisting of $n$ unit steps up to translation.

$\text{Arm}(n) \cong S^2 \times \ldots \times S^2$.

Let $\text{Pol}(n) \subset \text{Arm}(n)$ be the submanifold of closed random walks (or *random polygons*); i.e., those walks which satisfy

$$\sum_{i=1}^{n} \mathbf{e}_i = \mathbf{0}.$$
Theorem (with Cantarella and Deguchi; Millett and Zirbel)

The expected value of the squared radius of gyration for an equilateral $n$-gon is

$$\mathbb{E}[\text{Gyradius}, \text{Pol}(n)] = \frac{n + 1}{12}.$$
Definition
An 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* $\mathcal{P}_n$. 

![Diagram of a triangulation polytope](image)
The Triangulation Polytope

Definition
An 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 \textit{triangulation polytope} $\mathcal{P}_n$. 

\begin{align*}
  d_1 + d_2 &\geq 1 \\
  d_2 &\leq d_1 + 1 \\
  d_1 &\leq d_2 + 1 \\
  d_2 &\leq 2 \\
  d_1 &\leq 2 \\
  d_1 + d_2 &\geq 1
\end{align*}
Definition
An 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* $\mathcal{P}_n$. 
Definition

If $P_n$ is the triangulation polytope and $T^{n-3} = (S^1)^{n-3}$ is the torus of $n - 3$ dihedral angles, then there are action-angle coordinates:

$$\alpha: P_n \times T^{n-3} \to \text{Pol}(n) / \text{SO}(3)$$
Theorem (with Cantarella)

\( \alpha \) pushes the **standard probability measure** on \( \mathcal{P}_n \times T^{n-3} \) forward to the **correct probability measure** on \( \text{Pol}(n) / \text{SO}(3) \).
Theorem (with Cantarella)

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

Corollary

Any sampling algorithm for $\mathcal{P}_n$ is a sampling algorithm for closed polygons.
Theorem (with Cantarella)

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

**Proof Ideas.**

Pol$(n)/\text{SO}(3)$ is (almost) a toric symplectic manifold; chordlengths and dihedral angles give action-angle coordinates, the triangulation polytope is the moment polytope, and the Duistermaat–Heckman theorem guarantees the pushforward measure on the moment polytope is uniform.

**Corollary**

*Any sampling algorithm for $\mathcal{P}_n$ is a sampling algorithm for closed polygons.*
The polytope $\mathcal{P}_n$ corresponding to the “fan triangulation” is defined by the triangle inequalities:

\begin{align*}
0 & \leq d_1 \leq 2 \\
1 & \leq d_i + d_{i+1} \\
|d_i - d_{i+1}| & \leq 1 \\
0 & \leq d_{n-3} \leq 2
\end{align*}
A Change of Coordinates

If we introduce fake chordlength \( d_0 = 1 = d_{n-2} \), and make the linear transformation

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

then \( \sum s_i = d_{n-2} - d_0 = 0 \), so \( s_{n-2} \) is determined by \( s_1, \ldots, s_{n-3} \).
A Change of Coordinates

If we introduce fake chordlength \( d_0 = 1 = d_{n-2} \), and make the linear transformation

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

then \( \sum s_i = d_{n-2} - d_0 = 0 \), so \( s_{n-2} \) is determined by \( s_1, \ldots, s_{n-3} \) and the inequalities

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

become

\[
-1 \leq s_i \leq 1, \quad -1 \leq \sum_{i=1}^{n-3} s_i \leq 1, \quad \sum_{j=1}^i s_j + \sum_{j=1}^{i+1} s_j \geq -1
\]

\[
|d_i - d_{i+1}| \leq 1 \quad d_i + d_{i+1} \geq 1
\]
Let $C_n \subset [-1, 1]^{n-3}$ be determined by the inequalities on the previous slide.
Let $C_n \subset [-1, 1]^{n-3}$ be determined by the inequalities on the previous slide.
The Key Result

Let $C_n \subset [-1, 1]^{n-3}$ be determined by the inequalities on the previous slide.
Let \( C_n \subset [-1, 1]^{n-3} \) be determined by the inequalities on the previous slide.

\[ C_6 \]

**Theorem (with Cantarella, Duplantier, Uehara)**

The probability that a point in the hypercube lies in \( C_n \) is asymptotic to

\[ \frac{6\sqrt{6}}{\sqrt{\pi}} \frac{1}{n^{3/2}}. \]
The Action-Angle Method

Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. Generate \((s_1, \ldots, s_{n-3})\) uniformly on \([-1, 1]^{n-3}\) \(O(n)\) time

2. Test whether \((s_1, \ldots, s_{n-3}) \in \mathcal{C}_n\) acceptance ratio \(\sim 1/n^{3/2}\)

3. Let \(s_{n-2} = -\sum s_i\) and change coordinates to get diagonal lengths

4. Generate dihedral angles from \(T^{n-3}\)

5. Build sample polygon in action-angle coordinates
The Action-Angle Method

Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. Generate \((s_1, \ldots, s_{n-3})\) uniformly on \([-1, 1]^{n-3}\) \(O(n)\) time

2. Test whether \((s_1, \ldots, s_{n-3}) \in C_n\) acceptance ratio \(\sim 1/n^{3/2}\)

3. Let \(s_{n-2} = -\sum s_i\) and change coordinates to get diagonal lengths

4. Generate dihedral angles from \(T^{n-3}\)

5. Build sample polygon in action-angle coordinates

Expected complexity \(\Theta(n^{5/2})\)
The Action-Angle Method

Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. Generate \((s_1, \ldots, s_{n-3})\) uniformly on \([-1, 1]^{n-3}\) \(O(n)\) time

2. Test whether \((s_1, \ldots, s_{n-3})\) \(\in C_n\) acceptance ratio \(\sim 1/n^{3/2}\)

3. Let \(s_{n-2} = -\sum s_i\) and change coordinates to get diagonal lengths

4. Generate dihedral angles from \(T^{n-3}\)

5. Build sample polygon in action-angle coordinates

Expected complexity \(\Theta(n^{5/2})\)

Implemented in \texttt{plcurve}, available at
http://www.jasoncantarella.com/wordpress/software/
### Knot Types of 10 Million 60-gons

#### Straight line: $e^{-e} n^{-7/4}$

![Graph showing probability distribution for knot types]

#### Algorithm Comparison

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>distinct HOMFLYs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFM$^1$</td>
<td>2219</td>
</tr>
<tr>
<td>CRM</td>
<td>6110</td>
</tr>
<tr>
<td>Hedgehog Method</td>
<td>1111</td>
</tr>
<tr>
<td>Triangle Method</td>
<td>3505</td>
</tr>
<tr>
<td>Action-Angle Method</td>
<td>$\geq 6371$</td>
</tr>
</tbody>
</table>

---

$^1$100 million samples, instead of 10 million.
Future Challenges
Question

*How to incorporate excluded volume into the model?*
Topologically Constrained Random Walks

Tezuka Lab, Tokyo Institute of Technology

**Question**

*What special geometric structures exist on the moduli space of topologically-constrained random walks patterned on a given graph?*
Thank you for listening!

Thanks to my collaborators for being smart and hard-working

Alexander Y Grosberg
Bertrand Duplantier
Erica Uehara
Jason Cantarella
Rob Kusner
Tetsuo Deguchi

Thanks to the Simons Foundation for funding
• J. Cantarella, T. Deguchi, and C. Shonkwiler

• J. Cantarella, A. Y. Grosberg, R. Kusner, and C. Shonkwiler

• J. Cantarella and C. Shonkwiler

• J. Cantarella, B. Duplantier, C. Shonkwiler, and E. Uehara

http://arxiv.org/a/shonkwiler_c_1
The Volume of $C_n$

Theorem (Khoi, Takakura, Mandini)

The volume of $C_n$ is

$$\frac{1}{2(n-3)!} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^{k+1} \binom{n}{k} (n-2k)^{n-3}$$
The Volume of $\mathcal{C}_n$

**Theorem (Khoi, Takakura, Mandini)**

The volume of $\mathcal{C}_n$ is

$$\frac{1}{2(n-3)!} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^{k+1} \binom{n}{k} (n-2k)^{n-3}$$

**Observation (Edwards, 1922)**

$$\text{Vol } \mathcal{C}_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \frac{\sin^n x}{x^{n-2}} \, dx$$
[T]here is in these days far too great a tendency on the part of teachers to push on their pupils so fast to the Higher Branches of Analysis or to Physical Mathematics that many have neither time nor opportunity for the cultivation of real personal proficiency, or for the acquirement of that individual manipulative skill which is essential to any real confidence of the student in his own power to conduct unaided investigation.

– Joseph Edwards, 1922
The Volume of $C_n$

Theorem (Khoi, Takakura, Mandini)

The volume of $C_n$ is

$$\frac{1}{2(n-3)!} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^{k+1} \binom{n}{k} (n-2k)^{n-3}$$

Observation (Edwards, 1922)

$$\text{Vol } C_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \frac{\sin^n x}{x^{n-2}} \, dx$$
The Volume of $C_n$

Theorem (Khoi, Takakura, Mandini)

*The volume of $C_n$ is*

$$\frac{1}{2(n-3)!} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^{k+1} \binom{n}{k} (n-2k)^{n-3}$$

Observation (Edwards, 1922)

$$\text{Vol } C_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \frac{\sin^n x}{x^{n-2}} \, dx$$

$$= \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \text{sinc}^n(x)x^2 \, dx$$
Asymptotic volume

Making the substitution \( x = y/\sqrt{n} \) gives

\[
\text{Vol} C_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \text{sinc}^n \left( \frac{y}{\sqrt{n}} \right) \frac{y^2 \, dy}{n^{3/2}}.
\]

Therefore,

\[
\text{Vol} C_n \bigg|_{-1, 1} \approx 3\sqrt{\frac{3}{\pi}} \left( \frac{1}{n^{3/2}} \right).
\]
Making the substitution $x = \frac{y}{\sqrt{n}}$ gives

$$\text{Vol}_{C_n} = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \text{sinc}^n \left( \frac{y}{\sqrt{n}} \right) \frac{y^2 \, dy}{n^{3/2}}.$$ 

But $\text{sinc} \left( \frac{y}{\sqrt{n}} \right) = 1 - \frac{y^2}{6n} + O \left( \frac{1}{n} \right)$ and

$$\lim_{n \to \infty} \left( 1 - \frac{a}{n} + O \left( \frac{1}{n} \right) \right)^n = e^{-a},$$

so we can approximate by

$$\frac{2^{n-1}}{2\pi} \frac{1}{n^{3/2}} \int_{-\infty}^{\infty} e^{-y^2/6} y^2 \, dy = 3 \sqrt{\frac{3}{\pi}} 2^{n-\frac{3}{2}} \frac{1}{n^{3/2}}.$$
Asymptotic volume

Making the substitution \( x = y / \sqrt{n} \) gives

\[
\text{Vol} C_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \text{sinc}^n \left( \frac{y}{\sqrt{n}} \right) \frac{y^2 \, dy}{n^{3/2}}.
\]

But \( \text{sinc} \left( \frac{y}{\sqrt{n}} \right) = 1 - y^2 / 6n + o \left( \frac{1}{n} \right) \) and

\[
\lim_{n \to \infty} \left( 1 - \frac{a}{n} + o \left( \frac{1}{n} \right) \right)^n = e^{-a},
\]

so we can approximate by

\[
\frac{2^{n-1}}{2\pi} \frac{1}{n^{3/2}} \int_{-\infty}^{\infty} e^{-y^2/6} y^2 \, dy = 3 \sqrt{\frac{3}{\pi}} 2^{-3/2} \frac{1}{n^{3/2}}.
\]

Therefore,

\[
\frac{\text{Vol} C_n}{\text{Vol} [-1, 1]^{n-3}} \sim 3 \sqrt{\frac{3}{\pi}} 2^{-3/2} \frac{1}{n^{3/2}} = 6 \sqrt{\frac{6}{\pi}} \frac{1}{n^{3/2}}.
\]