Convergence and curvature of phylogenetic Markov chains

Alex Gavryushkin

21st October 2015

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Ricci-Ollivier curvature

Definition (Ollivier [2009])

Let $(\mathcal{T}, d)$ be a metric (tree) space with a random walk

$$m = (m_T)_{T \in \mathcal{T}}.$$

Let $T, R \in \mathcal{T}$ be two distinct points (trees). The Ricci-Ollivier curvature of $(\mathcal{T}, d, m)$ along $\overrightarrow{TR}$ is

$$\kappa_m(T, R) = 1 - \frac{W(m_T, m_R)}{d(T, R)},$$

where $W(\cdot\,,\cdot)$ is the earth mover's distance.

Curvature of Markov chains on graphs

Theorem (Ollivier [2009])

If $(\mathcal{T}, d)$ is a geodesic space then curvature is a local property.

Definition

Let $(\mathcal{T}, d)$ be a graph with a Markov chain m. Then the *curvature of the Markov chain m* on the graph $\mathcal{T}$ is the greatest number χ_m such that

$$\chi_m \leq \kappa_m(T, R) \text{ for adjacent } T \text{ and } R.$$

Trivial observation

Under a distance-one random walk, the following is true for any finite metric d and any pair of points x, y:

$$\frac{-2}{d(x,y)} \leq \kappa(x,y) \leq \frac{2}{d(x,y)}.$$

Random walks

For now, we consider three simplest random walks on various phylogenetic tree spaces.

- Metropolis-Hastings random walk: Choose a tree from the one neighbourhood and accept it with probability $\min(1, \frac{|N_1(T_{old})|}{|N_1(T_{new})|})$.
- Uniform random walk.
- Uniform p-lazy random walk, where p is the laziness probability.

Theorem (Whidden and Matsen [2015])

(1) *The curvature of the uniform random walk on the SPR graph on rooted trees with n leaves is bounded from below by*

$$\frac{-n^2+2n}{3.5n^2-15n+16} \geq -2/5$$

(2) *Subtract* $1/6$ *to get a lower bound on the curvature of the Metropolis-Hastings random walk.*

Theorem (G, Whidden, and Matsen [2015])

(1) *The curvature of the p-lazy uniform random walk on the SPR graph, the NNI graph, the* τ*-graph, and the discrete* τ*-space on rooted trees with n leaves is bounded from below by*

$$-p\,\frac{n-3}{n-2}\,.$$

(2) *Subtract* $2/3$ *to get a lower bound on the curvature of the Metropolis-Hastings random walk.*

Theorem (Whidden and Matsen [2015])

The maximum curvature of the uniform random walk on the SPR graph between two adjacent trees with n leaves is

$$\frac{6n - 17}{3n^2 - 13n + 14}.$$

Theorem (G, Whidden, and Matsen [2015])

The curvature of a uniform random walk on the discrete τ-space satisfies

$$\kappa_{d_\tau}(T, R) \leq \frac{1}{2(n-2)}.$$

Theorem (G, Whidden, and Matsen [2015])

Let $\{T_n \mid n \in \mathbb{N}\}$ and $\{S_n \mid n \in \mathbb{N}\}$ be two sequences of phylogenetic trees such that $d(T_n, R_n) = 1$ for all n. Then

$$\lim_{n \to \infty} \kappa_n(T_n, S_n) = 0$$

for the uniform random walk on the SPR graph, the NNI graph, the τ-graph, and the discrete τ-space.*

---

*For the SPR graph, we have to bound the size of the subtree which is getting moved.

Thank you for your attention!

Yann Ollivier.
Ricci curvature of Markov chains on metric spaces.
J. Functional Analysis, 256, 3, 810–864, 2009.

Alex Gavryushkin and Alexei Drummond.
The space of ultrametric phylogenetic trees.
arXiv preprint arXiv:1410.3544, 2014.

Chris Whidden and Frederick A. Matsen IV.
Quantifying MCMC exploration of phylogenetic tree space.
Systematic Biology, doi:10.1093/sysbio/syv006, 2015.

Chris Whidden and Frederick A. Matsen IV.
Ricci-Ollivier curvature of two random walks on rooted phylogenetic subtree-prune-regraft graph.
To appear in the proceedings of the *Thirteenth Workshop on Analytic Algorithmics and Combinatorics*, 2015.

Alex Gavryushkin, Chris Whidden, and Frederick A. Matsen IV.
Random walks over discrete time-trees.
To appear on the *arXiv*, 2015.