Probabilistic Path Hamiltonian Monte Carlo

Vu Dinh,* Arman Bilge,* Cheng Zhang,* and Frederick A. Matsen IV

* Equal contribution Program in Computational Biology, Fred Hutchison Cancer Research Center Department of Statistics, University of Washington

Motivation

If the probabilistic model contains both continuous and structural discrete parameters, how can we sample from the posterior efficiently?

- Simple MCMCs are not efficient at sampling continuous parameters
- Advanced MCMCs, e.g. Hamiltonian Monte Carlo, can not handle discrete parameters

BL on Orthant Complexes

An *orthant complex* is a geometric object $\mathcal{X}$ obtained by gluing orthants of the same dimension

$$\mathcal{X} = \{(\tau, q) : \tau \in \Gamma,\ q \in \mathbb{R}^n_{\geq 0}\}$$

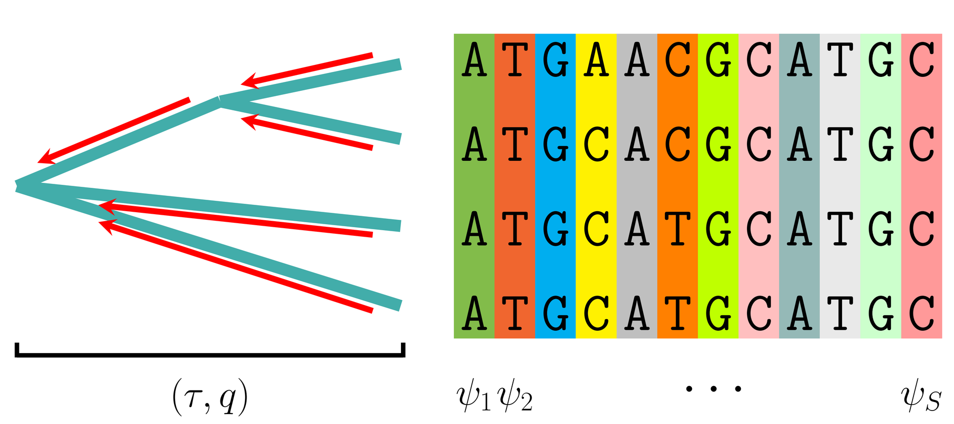
where Γ is a countable set. Given observations D and proper priors $\pi_0(\tau, q)$, the posterior

$$P(\tau, q|D) \propto L(D|\tau, q)\pi_0(\tau, q)$$

- $(\tau, q_\tau) = (\tau', q_{\tau'}) \Rightarrow q_\tau = q_{\tau'},\ \tau' \in \mathcal{N}(\tau, q_\tau)$
- The adjacency graph of $\mathcal{X}$ has finite diameter k.
- $U(\tau, q) = -\log P(\tau, q)$ is continuous and smooth up to the boundary.

A Challenging Example

Let (τ, q) be a phylogenetic tree and $\psi = \{\psi_i\}_{i=1}^S$ be the observed sequences over the leaves.



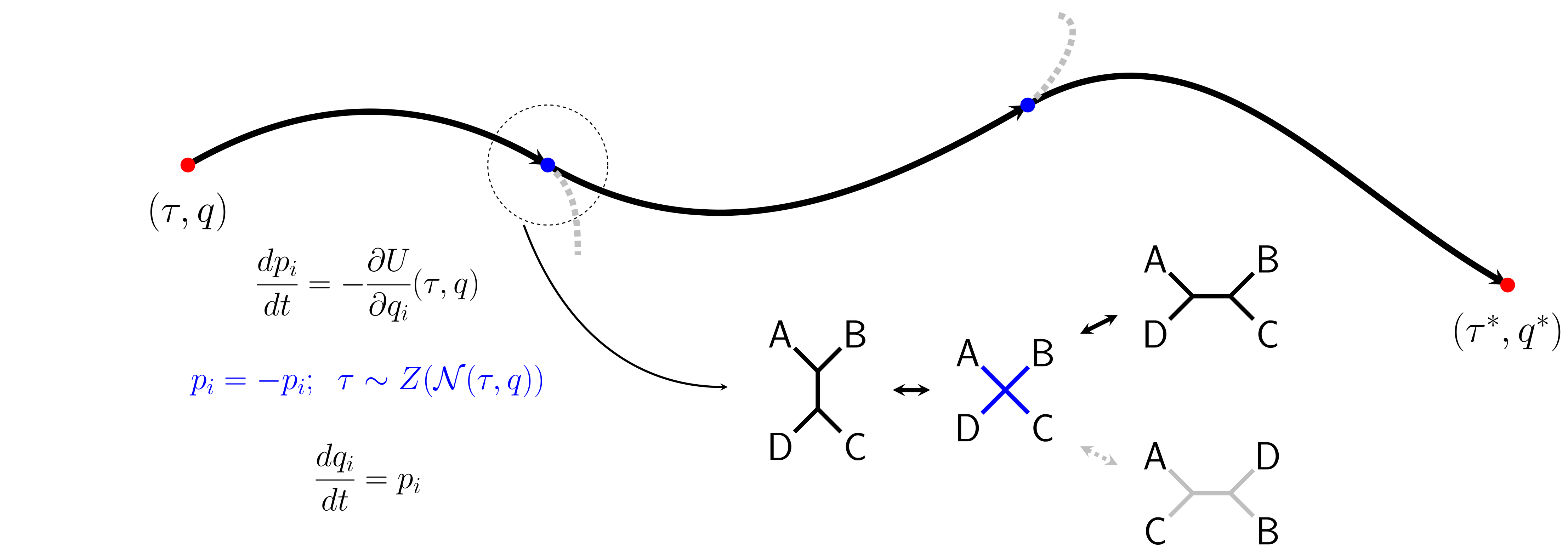
A continuous-time Markov chain is usually used to model the evolution history

$$L(\psi|\tau, q) = \prod_{s=1}^{S} \sum_{a^s} \eta(a^s_\rho) \prod_{(u,v)\in E(\tau,q)} P^{uv}_{a^s_u a^s_v}(q_{uv})$$

- Efficient computation via Felsenstein's pruning algorithm [1].
- Orthant Complexes: **NNI neighbors**.
- # possible topologies: $T(n)$, # leaves: n

$$T(n) = \frac{(2n-5)!}{(n-3)!\,2^{n-3}} = e^{\mathcal{O}(n \log n)}$$

Hamiltonian: $H(\tau, q, p) = U(\tau, q) + K(p),\quad U(\tau, q) = -\log P(\tau, q),\quad K(p) = \frac{1}{2}\|p\|^2$



Theoretical Properties

Augmented state: $s = (\tau, q, p)$. Let A, B be a pair of measurable sets in the augmented state space. We assume the following condition

$$P(\tau'|\tau, q) = P(\tau|\tau', q),\ \tau' \in \mathcal{N}(\tau, q)$$

which is true with *uniform* topology proposal Z.

1. **Probabilistic Reversibility.**

$$P((\tau, q, p), (\tau^*, q^*, p^*)) = P((\tau^*, q^*, -p^*), (\tau, q, -p))$$

2. **Stochastic Volume Presevation.**

$$\int_A \int_B P(s, s')ds'\,ds = \int_B \int_A P(s', s)ds\,ds'$$

3. **k-accessibility.** Any two states in $\mathcal{X}$ can be connected by k iterations of probabilistic path HMC.

Theorem. *Probabilistic Path HMC preserves the posterior and is ergodic.*

Surrogate Smoothing

∂U is usually non-differentiable on the boundary which lead to $\mathcal{O}(C\epsilon + T\epsilon^3)$ global numerical error, where C is the number of reflection/refraction events. We found a recipe by using a surrogate smoothing strategy and refraction in the leap-frog steps.

$$\tilde{U}(\tau, q) = U(\tau, G(q))$$

$$G(q) = (g_\delta(q_1), g_\delta(q_2), \ldots, g_\delta(q_{2n-3}))$$

and

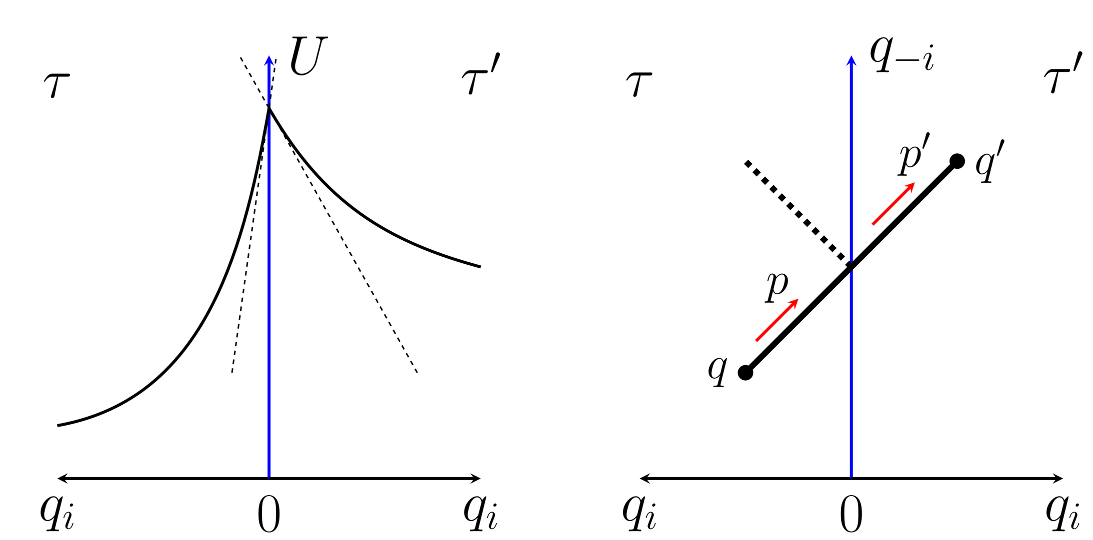
$$g_\delta(x) = \begin{cases} x, & x \geq \delta \\ \frac{1}{2\delta}(x^2 + \delta^2), & 0 \leq x < \delta \end{cases}$$

Reflection and Refraction

To maintain the desired properties of Hamiltonian dynamics, we adopted the *reflection* and *refraction* technique [2] when jumping between topologies.

- Reflection. $U(\tau, q)$ is continuous across boundary

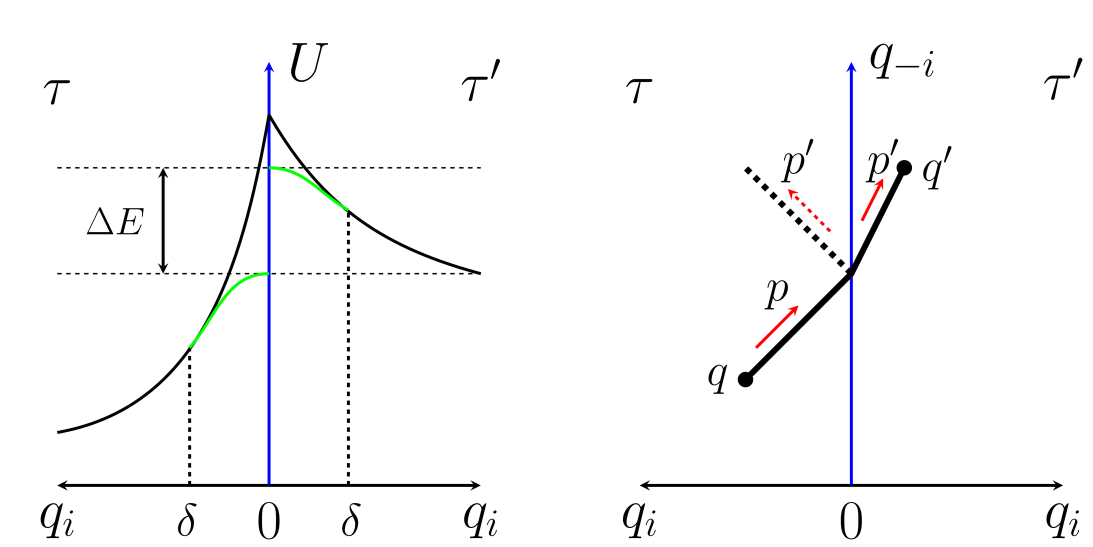
$$\Delta E = U(\tau', q) - U(\tau, q) = 0,\quad q_i = 0$$



Momentum & topology update:

$$p_i = -p_i, \quad \tau = \tau'$$

- Refraction. Surrogate evens the gradients while creating controllable energy gaps $\Delta E \neq 0$



Momentum & topology update:

$$(\tau, p_i) = \begin{cases} (\tau', \sqrt{\|p_i\|^2 - 2\Delta E}) & \|p_i\|^2 > 2\Delta E \\ (\tau, -p_i) & \text{otherwise} \end{cases}$$

Results

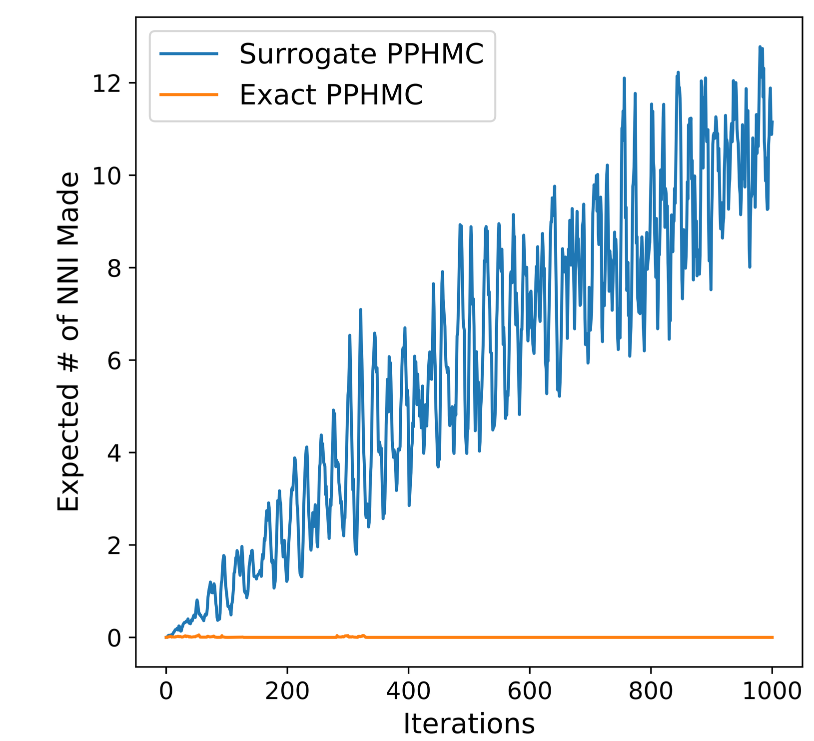


Figure 1: Expected number of NNI moves on a real data set.

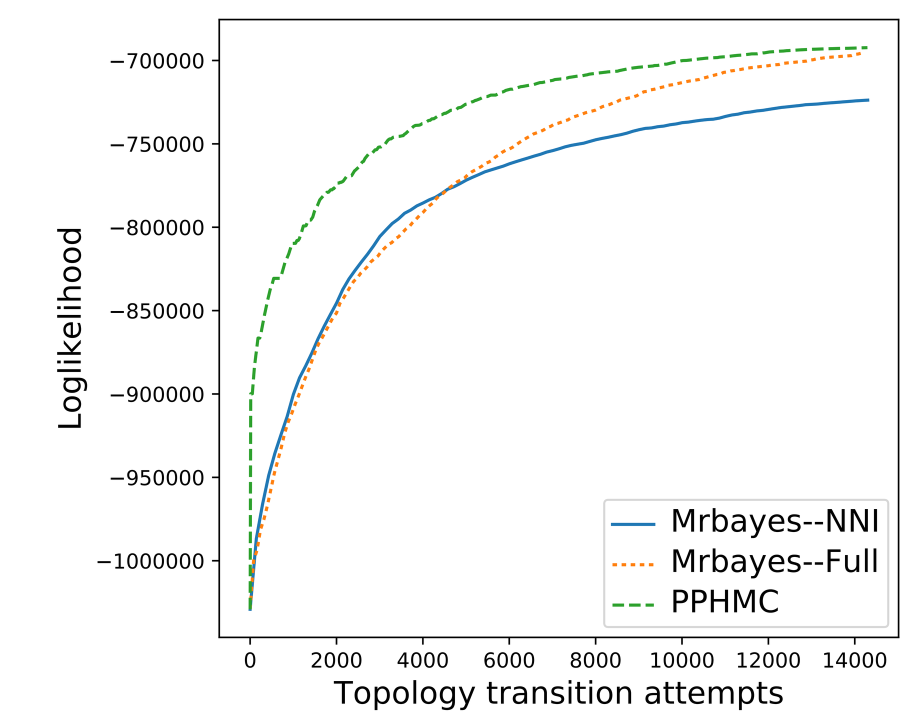


Figure 2: Loglikelihood vs topology transitions on a 1000 taxa simulated data set.

Contributions

- Extended HMC towards sampling both continuous and structural discrete parameters.
- Developed a smoothing surrogate function that enables long HMC paths with potential non-differentiable boundary transitions.

References

[1] J. Felsenstein, J. Mol. Evol., **17**(6) (1981), 368-376

[2] H. M. Afshar, J. Domke, NIPS (2015), 2989-2997

Acknowledgements