6. Appendix

6.1. Properties of the phylogenetic posterior distribution

Assumption 2.3 for the phylogenetic posterior distribution. Recall that $L(\tau, q)$ denotes the likelihood function of the tree $T = (\tau, q)$, we have

$$U(\tau, q) = -\log L(\tau, q) - \log \pi_0(\tau, q)$$

Since $-\log \pi_0(\tau, q)$ is assumed to satisfy the Assumption 2.3, we just need to prove that the phylogenetic likelihood function is smooth while each orthant and is continuous on the whole space.

Without loss of generality, we consider the case when a single branch length of some edge e is contracted to zero. To investigate the changes in the likelihood function and its derivatives, we first fix all other branches, partition the set of all extensions of ψ according to their labels at the end points of e, and split $E(T)$ into two sets of edges E_{left} and E_{right} corresponding to the location of the edges with respect to e. The likelihood function of the tree $T = (\tau, q)$ can be rewritten as

$$\begin{aligned} L(T) &= \prod_{s=1}^{S} \sum_{ij} \sum_{a \in \mathcal{A}_{ij}} \left(\prod_{(u,v) \in E_{\text{left}}} P^{uv}_{a_u a_v}(q_{uv}) \right) \\ &\quad \times \eta(i) P^e_{ij}(t) \times \left(\prod_{(u,v) \in E_{\text{right}}} P^{uv}_{a_u a_v}(q_{uv}) \right) \end{aligned}$$

where t is the branch length of e, η is the stationary distribution, $\mathcal{A}_{ij}$ denotes the set of all extensions of ψ for which the labels at the left end point and the right end point of e are i and j, respectively. By grouping the products over E_{left} and E_{right}, the stationary frequency $\eta(\cdot)$, and the sum over a in a single term b^s_{ij}, we can define the one-dimensional log-likelihood function as a univariate function of the branch length of e

$$L_T(t) = \prod_{s=1}^{S} \left(\sum_{ij} b^s_{ij} P^e_{ij}(t) \right).$$

Consider the tree T' obtained by collapsing edge e of the tree T to zero. The likelihood of T' can be written as

$$L(T') = \prod_{s=1}^{S} \left(\sum_{i=j} b^s_{ij} P^e_{ij}(0) \right) = \prod_{s=1}^{S} \left(\sum_{i} b^s_{ii} \right)$$

since $P_{ij}(0) = 1$ if $i = j$ and 0 otherwise. Thus

$$\lim_{t \to 0} L_T(t) = L(T').$$

Since this is true for all (τ, q) and $t \in E(\tau, q)$, we deduce that the likelihood function is continuous up to the boundary of each orthant, and thus, is continuous on the whole tree space. Moreover, using the same arguments, we can prove that likelihood function is smooth up to the boundary of each orthant.

Now fixing all but two branch lengths t_e, t_f, the likelihood can be rewritten as

$$L_T(t_e, t_f) = \prod_{s=1}^{S} \left(\sum_{ij} b^s_{ij}(t_e) P^f_{ij}(t_f) \right)$$

and the derivative of the log likelihood is

$$\frac{1}{L_T(t_e, t_f)} \frac{\partial L_T}{\partial t_f}(t_e, t_f) = \sum_{s=1}^{S} \frac{\sum_{ij} b^s_{ij}(t_e) (P^f_{ij})'(t_f)}{\sum_{ij} b^s_{ij}(t_e) P^f_{ij}(t_f)}.$$

By using the same argument as above, we have that $b^s_{ij}(t_e)$ is continuous in t_e up to zero and so

$$\lim_{t_e \to 0} \frac{1}{L_T(t_e, t_f)} \frac{\partial L_T}{\partial t_f}(t_e, t_f) = \frac{1}{L(T')} \frac{\partial L}{\partial t_f}(T').$$

Thus, when a Hamiltonian particle crosses a boundary between orthants, partial derivatives of the energy function with respect to positive branch lengths are continuous. □

6.2. Theoretical properties of the leap-prog integrator

Proof of Lemma 3.5. Note that for PPHMC, in a single leap-prog step γ of finite size ϵ, the algorithm only re-evaluates the gradient of the energy function at the end of the step when the final position q' has been fixed, and changes in topology on the path have no effect on the changes of position and momentum. Thus, the projection $\tilde{\gamma}$ of γ to the (q, p) space is just a deterministic reflected Hamiltonian path. As a result, for any $s^{(1)} = (\tau^{(1)}, q^{(1)}, p^{(1)}), s^{(2)} = (\tau^{(2)}, q^{(2)}, p^{(2)}) \in R(s)$, we have $(q^{(1)}, p^{(1)}) = (q^{(2)}, p^{(2)})$. This, along with the fact that set of topologies is countable, implies that $R(s)$ is countable.

Now denote by $\{t^{(1)} < t^{(2)} < \ldots < t^{(n)} < \ldots \leq \epsilon\}$ the set of time points at which $\tilde{\gamma}$ hits the boundary. Since this set is strictly increasing, it is countable. Moreover, the τ-component of γ is only updated with finite choices at $\{t^{(i)}\}$. This implies that $K(s)$ is countable.

Finally, consider any leap-prog step γ that connects s and s' through infinite number of topological changes. We note that at each $t^{(i)}$, the next topology is chosen among $x^{(i)} \geq 2$ neighboring topologies. Denote by $P_\gamma(s, s')$ the probability of moving from s to s' via path γ, we have

$$P_\gamma(s, s') \leq \prod_{i=1}^{\infty} \frac{1}{x^{(i)}} = 0.$$

Since $K(s)$ is countable, we deduce that $P_\infty(s, s') = 0$. □

Proof of Lemma 2.1. Consider any possible path γ that connects s and s'. By definition, one can find a sequence of augmented states $(s = s^{(0)}, s^{(1)}, s^{(2)}, \dots, s^{(k)} = s')$ such that γ can be decomposed into segments on which the topology does not change. From standard result about Hamiltonian dynamics, the Hamiltonian is constant on each segment.

For PPHMC, since the potential energy is continuous across the boundary and the magnitude of the momentum does not change when moving from one orthant to another one, we deduce that the Hamiltonian is constant along that path.

Similarly, for PPHMC with surrogates, the algorithm is designed in such a way that any changes in potential energy is balanced by a change in momentum, which conserves the total energy from one segment to another. We also deduce that the Hamiltonian is constant along the whole path. □

Proof of Lemma 3.1. Define $\sigma(\tau, q, p) := (\tau, q, -p)$. Consider any possible leap-prog step γ that connects s and s'; say the sequences of augmented states $(s = s^{(0)}, s^{(1)}, s^{(2)}, \dots, s^{(k)} = s')$, topologies $(\tau = \tau^{(0)}, \tau^{(1)}, \tau^{(2)}, \dots, \tau^{(k)} = \tau')$ and times $(t = t^{(0)}, t^{(1)}, t^{(2)}, \dots, t^{(k)} = t')$ decompose γ into segments on which the topology is unchanged. Denote by $P_\gamma(s, s')$ the probability of moving from s to s' via path γ, we have

$$P_\gamma(s, s') = \prod_i \mathbb{P}(s^{(i+1)} \mid s^{(i)}, t^{(i+1)} - t^{(i)}) \times \prod_j \mathbb{P}(\tau^{(j+1)} \mid \tau^{(j)}),$$

where each sub-step of the algorithm is a leapfrog update ($\phi^{(i)}$) with some momentum reversing ($\sigma^{(i)}$), that is $s^{(i+1)} = \sigma^{(i)}(\phi^{(i)}(s^{(i)}))$ and $\sigma^{(i)}$ is a map that changes the sign of some momentum coordinates.

If we start the dynamics at $\sigma(s^{(i+1)})$, then since the particle is crossing the boundary, the momenta corresponding to the crossing coordinates are immediately negated and the system is instantly moved to the augmented state

$$\sigma^{(i)}\sigma(s^{(i+1)}) = \sigma\sigma^{(i)}(s^{(i+1)}) = \sigma(\phi^{(i)}(s^{(i)})).$$

A standard result about reversibility of Reflective Hamiltonian dynamics implies that the system starting at $\sigma(\phi^{(i)}(s^{(i)}))$ will end at $\sigma(s^{(i)})$ after the same period of time $t^{(i+1)} - t^{(i)}$. We deduce that

$$\begin{aligned} &\mathbb{P}(s^{(i+1)} \mid s^{(i)}, t^{(i+1)} - t^{(i)})) \\ =\ &\mathbb{P}(\sigma(s^{(i)}) \mid \sigma(s^{(i+1)}), t^{(i+1)} - t^{(i)})). \end{aligned}$$

On the other hand, at time $t^{(j)}$, $(\tau^{(j)}, q^{(j)})$ and $(\tau^{(j+1)}, q^{(j)})$ are neighboring topologies, hence

$$\mathbb{P}(\tau^{(j+1)} \mid \tau^{(j)}) = \frac{1}{|\mathcal{N}(\tau^{(j)}, q^{(j)})|} = \frac{1}{|\mathcal{N}(\tau^{(j+1)}, q^{(j)})|} = \mathbb{P}(\tau^{(j)} \mid \tau^{(j+1)}).$$

Therefore

$$P_\gamma(s, s') = P_\gamma(\sigma(s'), \sigma(s))$$

for any path γ. This completes the proof. □

Proof of Lemma 3.2. We denote by C the set of pairs $(s, s') \in A \times B$ such that $P(s, s') > 0$. Let us consider any possible leap-prog step γ that connects $s \in A$ and $s' \in B$ crossing a finite number of boundaries and the sequences of augmented states $(s = s^{(0)}, s^{(1)}, s^{(2)}, \dots, s^{(k)} = s')$, topologies $(\tau = \tau^{(0)}, \tau^{(1)}, \tau^{(2)}, \dots, \tau^{(k)} = \tau')$, times $(t = t^{(0)}, t^{(1)}, t^{(2)}, \dots, t^{(k)} = t')$ and indices $\alpha = (\alpha^{(0)}, \alpha^{(1)}, \dots, \alpha^{(k)})$ (each $\alpha^{(i)}$ is a vector of ± 1 entries characterizing the coordinates crossing zero in each sub-step) that decompose γ into segments on which the topology is unchanged. By grouping the members of C by the value of α and ω, we have:

$$C = \bigcup_{(\alpha,\omega)} C_{\alpha,\omega}.$$

Because there will typically be many paths between s and s', the $C_{\alpha,\omega}$ need not be disjoint. However, we can modify the (countable number of) sets by picking one set for each (s, s') and dropping it from the rest, making a collection of disjoint sets $\{C_j\}$ such that each C_j is a subset of some $C_{\alpha,\omega}$ and

$$C = \bigcup_{j \in J} C_j.$$

We will write $s \in A_j(s')$ and $s' \in B_j(s)$ if $(s, s') \in C_j$ and denote

$$A_j = \bigcup_{s' \in B} A_j(s') \quad \text{and} \quad B_j = \bigcup_{s \in A} B_j(s).$$

We note that although the leap-prog algorithm is stochastic, if (α, ω) has been pre-specified, the whole path depends deterministically on the initial momentum. Thus, by denoting the projection of $C_{\alpha,\omega}$ to A by $A_{\alpha,\omega}$, we have that the transformation $\phi_{\alpha,\omega}$ that maps s to s' is well-defined on $A_{\alpha,\omega}$. Since the projection of the particles (in a single leap-prog step) to the (q, p) space is exactly Reflective Hamiltonian Monte Carlo on $\mathbb{R}^n_{\geq 0}$. Using Lemma 1, Lemma 2 and Theorem 1 in (Afshar and Domke, 2015), we deduce that the determinant of the Jacobian of $\phi_{\alpha,\omega}$ is 1.

Now consider any $j \in J$ such that $C_j \subset C_{\alpha,\omega}$. Because $P(s, s') = P(s', s)$ for all $s, s' \in \mathbb{T}$ and the determinant of

the Jacobian of $\phi_{\alpha,\omega}$ is 1, we have

$$\begin{aligned}\int_{B_j} \sum_{s\in A_j(s')} P(s',s)\,ds' &= \int_{B_j} P(s',\phi^{-1}_{\alpha,\omega}(s'))\,ds' \\ &= \int_{A_j} P(\phi_{\alpha,\omega}(s),s)\,ds \\ &= \int_{A_j} P(s,\phi_{\alpha,\omega}(s))\,ds \\ &= \int_{A_j} \sum_{s'\in B_j(s)} P(s,s')\,ds. \end{aligned} \tag{6.1}$$

Denote

$$A^* = \bigcup_j A_j \qquad \text{and} \qquad B^* = \bigcup_j B_j.$$

Summing (6.1) over all possible values of j gives

$$\int_{B^*} \sum_{s\in A(s')} P(s',s)\,ds' = \int_{A^*} \sum_{s'\in B(s)} P(s,s')\,ds.$$

Moreover, we note that for $s \notin A^*$, $B(s) = \emptyset$. Similarly, if $s' \notin B^*$, $A(s') = \emptyset$. We deduce that

$$\int_{B} \sum_{s\in A(s')} P(s',s)\,ds' = \int_{A} \sum_{s'\in B(s)} P(s,s')\,ds.$$

□

Proof of Lemma 3.3. By definition of k, for any state $(\tau',q') \in B$, we can find a sequence of topologies $(\tau = \tau^{(0)},\tau^{(1)},\tau^{(2)},\dots,\tau^{(k)} = \tau')$ for some $l \le k$ such that $\tau^{(i)}$ and $\tau^{(i+1)}$ are adjacent topologies. From the construction of the state space, let $(\tau^{(i)},q^{(i)})$ denote a state on the boundary between the orthants for the two topologies $\tau^{(i)}$ and $\tau^{(i+1)}$. Moreover, since $(\tau^{(i)},q^{(i)})$ and $(\tau^{(i+1)},q^{(i+1)})$ lie in the same orthant, we can find momentum values $p^{(i)}$ and $(p^{(i)})'$ such that

$$P((\tau^{(i)},q^{(i)},p^{(i)}) \to (\tau^{(i+1)},q^{(i+1)},(p^{(i+1)})')) > 0$$

for all i. That is, we can get from $(\tau^{(i)},q^{(i)},p^{(i)})$ to $(\tau^{(i+1)},q^{(i+1)},(p^{(i+1)})')$ by a sequence of leapfrog steps $\Sigma^{(i)}$ with length T. By joining the $\Sigma^{(i)}$'s, we obtain a path Σ of k PPHMC steps that connects $(\tau^{(0)},q^{(0)})$ and (τ',q'). □

Proof of Lemma 3.4. For a path Σ of k PPHMC steps connecting $(\tau^{(0)},q^{(0)})$ and (τ',q'), we define $F_\Sigma = \{(\tau^{(0)},q^{(0)}),(\tau^{(1)},q^{(1)}),\dots,(\tau^{(n_\omega)},q^{(n_\omega)})\}$, where $(\tau^{(i)},q^{(i)})$ denotes the state on Σ that joins the topologies $\tau^{(i)}$ and $\tau^{(i+1)}$. We first note that although our leap-prog algorithm is stochastic, if the sequence of topologies crossed by a path Σ has been pre-specified, the whole path depends deterministically on the sequence of momenta $p = (p^{(0)},\dots,p^{(m)})$ along Σ. Thus, the functions

$$\phi_{i,\omega}(p) := q^{(i)} \qquad \forall p \in I_{B,\omega},$$

are well-defined.

We will prove that $\phi_{n_\omega,\omega}$ is Lipschitz by induction on n_ω. For the base case $n_\omega = 0$, the sequence ω is of length 1, which implies no topological changes along the path. The leap-prog algorithm reduces to the baseline leapfrog algorithm and from standard results about HMC on Euclidean spaces (see, e.g., Cances et al., 2007), we deduce that $\phi_{1,\omega}$ is Lipschitz.

Now assume that the results holds true for $n_\omega = n$. Consider a sequence ω of length $n+1$. For all $(\tau',q') \in B_\omega$, let $\Sigma(\tau',q')$ be a (k,T)-path connecting $(\tau^{(0)},q^{(0)})$ and (τ',q'). We recall that

$$\begin{aligned}&F_{\Sigma(\tau',q')} \\ &= \{(\tau^{(0)},q^{(0)}),(\tau^{(1)},\phi_{1,\omega}(p)),\dots,(\tau^{(n_\omega)},\phi_{n_\omega,\omega}(p))\},\end{aligned}$$

where $\phi_{n_\omega,\omega}(p)) = (\tau',q')$, is the set of states that join the topologies on the path $\Sigma(\tau',q')$.

Define $\omega' = \{\tau^{(0)},\tau^{(1)},\dots,\tau^{(n_\omega-1)}\}$ and $B' = \phi_{n_\omega-1}(I_{B,\omega})$, the induction hypothesis implies that the function $\phi_{n_{\omega'},\omega'} = \phi_{n_\omega-1,\omega}$ is Lipschitz on $I_{B',\omega'} = I_{B,\omega}$.

On the other hand, since $(\tau^{(n)},q^{(n)})$ and $(\tau^{(n+1)},q^{(n+1)})$ belong to the same topology, the base case implies that $q^{(n+1)}$ is a Lipschitz function in p and $q^{(n)} = \phi_{n_\omega-1,\omega}(p)$. Since compositions of Lipschitz functions are also Lipschitz, we deduce that $\phi_{n_\omega,\omega}$ is Lipschitz.

Since Lipschitz functions map zero measure sets to zero measure sets (see, e.g., Section 2.2, Theorem 2 and Section 2.4, Theorem 1 of Evans and Gariepy, 2015), this implies $\mu(B_\omega) = 0$ which completes the proof. □

6.3. Ergodicity of PPHMC

Proof of Lemma 3.6 . We denote

$$\begin{aligned}\nu(\tau,q,p) &= \frac{1}{Z}\exp(-U(\tau,q))\exp\left(-K(p)\right) \\ &= \frac{1}{Z}\exp(-H(\tau,q,p))\end{aligned}$$

and refer to it as the canonical distribution.

It is straightforward to check that for all $s,s' \in \mathbb{T}$, we have $\nu(s)r(s,s') = \nu(s')r(s',s)$. Lemma 3.2 implies that

$$P(s,ds')ds = P(s',ds)ds'$$

in term of measures. This gives the detailed balance condition

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

for all $A, B \subset \mathbb{T}$.

We deduce that every update step of the second step of PPHMC satisfies detailed balance with respect to ν and hence, leaves ν invariant. On the other hand, since ν is a function of $|p|$, the negation of the momentum p at the end of the second step also fixes ν. Similarly, in the fist step, p is drawn from its correct conditional distribution given q and thus leaves ν invariant.

Since the target distribution π is the marginal distribution of ν on the position variables, PPHMC also leaves π invariant. □

6.4. Approximation error of reflective leapfrog algorithm

In this section, we investigate the local approximation error of the reflective leapfrog algorithm (Afshar and Domke, 2015) without using surrogates. Recall that V^+ and V^- are the restrictions of the potential function V on the sets $\{x_1 \geq 0\}$ and $\{x_1 \leq 0\}$, and we assume that V^+ and V^- are smooth up to the boundary of their domains.

Consider a reflective leapfrog step with potential energy function V starting at $(q^{(0)}, p^{(0)})$ (with $q_1^{(0)} > 0$), ending at $(q^{(1)}, p^{(1)})$ (with $q_1^{(1)} < 0$) and hitting the boundary at x (with $x_1 = 0$, i.e., a refraction event happens on the hyperplane of the first component).

Proof of Proposition 3.1. Let p and p' denote the half-step momentum of a leapfrog step before and after the refraction events, respectively. Recall that in a leapfrog approximation with refraction at $x_1 = 0$, we have

$$p_i^{(0)} = p_i + \frac{\epsilon}{2} \frac{\partial V}{\partial q_i}(q^{(0)}), \qquad p_i^{(1)} = p_i' - \frac{\epsilon}{2} \frac{\partial V}{\partial q_i}(q^{(1)}),$$

where

$$p_1' = \sqrt{p_1^2 - 2dV(x)},$$

$p_i' = p_i$ for $i > 1$, and $dV(x) = V^-(x) - V^+(x)$ denotes the change in potential energy across the hyper-plane.

The change in kinetic energy after this leapfrog step is

$$\Delta K = -dV(x) - \frac{\epsilon}{2} \sum_i \left(p_i \frac{\partial V}{\partial q_i}(q^{(0)}) + p_i' \frac{\partial V}{\partial q_i}(q^{(1)}) \right) + \frac{\epsilon^2}{8} \sum_i \left(\left(\frac{\partial V}{\partial q_i}(q^{(1)}) \right)^2 - \left(\frac{\partial V}{\partial q_i}(q^{(0)}) \right)^2 \right).$$

We can bound the second-order term by

$$\left(\frac{\partial V}{\partial q_i}(q^{(1)}) \right)^2 - \left(\frac{\partial V}{\partial q_i}(q^{(0)}) \right)^2 = 2 \int_0^\epsilon \frac{\partial V}{\partial q_i}(q^{(0)} + tp) \frac{\partial^2 V}{\partial q_i^2}(q^{(0)} + tp) p_i \, dt = \mathcal{O}(\epsilon) \cdot \sup_{z, W = V^+, V^-} \frac{\partial W}{\partial q_i}(z) \frac{\partial^2 W}{\partial q_i^2}(z).$$

On the other hand for the potential energy,

$$\begin{aligned} \Delta V &= V(q^{(1)}) - V(q^{(0)}) \\ &= V(q^{(1)}) - V^-(x) + dV(x) + V^+(x) - V(q^{(0)}) \\ &= dV(x) + \int_{\epsilon_1}^{\epsilon} \nabla V(q^{(0)} + tp) \cdot p \, dt + \int_0^{\epsilon_1} \nabla V(q^{(0)} + tp) \cdot p \, dt \end{aligned}$$

where ϵ_1 and $\epsilon_2 := \epsilon - \epsilon_1$ denote the integration times before and after refraction. By the trapezoid rule for integration,

$$\begin{aligned} \Delta V = dV(x) &+ \sum_{i>1} \frac{\epsilon}{2} \left(p_i \frac{\partial V}{\partial q_i}(q^{(0)}) + p_i' \frac{\partial V}{\partial q_i}(q^{(1)}) \right) \\ &+ \frac{p_1' \epsilon_2}{2} \frac{\partial V}{\partial q_1}(q^{(1)}) + \frac{p_1' \epsilon_2}{2} \frac{\partial V^-}{\partial q_1}(x) \\ &+ \frac{p_1 \epsilon_1}{2} \frac{\partial V}{\partial q_1}(q^{(0)}) + \frac{p_1 \epsilon_1}{2} \frac{\partial V^+}{\partial q_1}(x) \\ &+ \mathcal{O}(\epsilon^3) \cdot \sup_z \sum_{i, W = V^+, V^-} \left(\frac{\partial^3 W}{\partial q_i^3}(z) \right). \end{aligned}$$

We recall that the error of the trapezoid rule on $[a, b]$ with resolution h is a constant multiple of $h^2(b-a)$, which is of order ϵ^3 in our case. We deduce that

$$\begin{aligned} \Delta H = \Delta V + \Delta K = &- \frac{p_1' \epsilon_1}{2} \frac{\partial V}{\partial q_1}(q^{(1)}) + \frac{p_1' \epsilon_2}{2} \frac{\partial V^-}{\partial q_1}(x) \\ &- \frac{p_1 \epsilon_2}{2} \frac{\partial V}{\partial q_1}(q^{(0)}) + \frac{p_1 \epsilon_1}{2} \frac{\partial V^+}{\partial q_1}(x) \\ &+ \mathcal{O}(\epsilon^3). \end{aligned}$$

Using Taylor expansion, we have

$$\frac{\partial V}{\partial q_1}(q^{(1)}) = \frac{\partial V^-}{\partial q_1}(x) + \mathcal{O}(\epsilon),$$

and

$$\frac{\partial V}{\partial q_1}(q^{(0)}) = \frac{\partial V^-}{\partial q_1}(x) + \mathcal{O}(\epsilon).$$

This implies

$$\Delta H = (\epsilon_2 - \epsilon_1) \left(p_1' \frac{\partial V^-}{\partial q_1}(x) - p_1 \frac{\partial V^+}{\partial q_1}(x) \right) + \mathcal{O}(\epsilon^2).$$

In general, there is no dependency between ϵ_1 and ϵ_2, and the only cases where ΔH is not of order $\mathcal{O}(\epsilon)$ are when

$$\sqrt{p_1^2 - 2dV(x)}\,\frac{\partial V^-}{\partial q_1}(x) - p_1\frac{\partial V^+}{\partial q_1}(x) = 0.$$

In order for this to be true for all p, we need to have either

$$dV(x) = 0 \qquad \text{and} \qquad \frac{\partial V^-}{\partial q_1}(x) = \frac{\partial V^+}{\partial q_1}(x),$$

or

$$\frac{\partial V^-}{\partial q_1}(x) = \frac{\partial V^+}{\partial q_1}(x) = 0.$$

In both cases, the first derivative of V with respect to the first component must be continuous. This completes the proof. □

6.5. Estimated posterior tree distributions for the simulated data

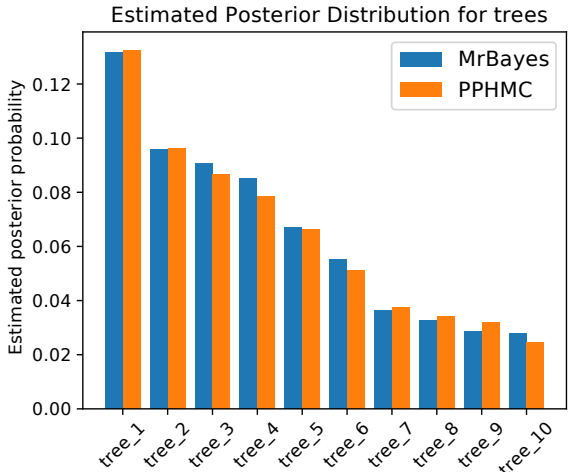

Figure 3. Estimated posterior distributions for the top 10 trees from the ground truth for MrBayes (blue) and PPHMC (orange), respectively.

6.6. Coordinate systems for branch lengths on trees

In this section we verify Assumption 2.1 for phylogenetic trees. Further explanation of the framework used here can be found in (Semple and Steel, 2003; Bryant, 2004).

Assume we are considering phylogenetic trees on N leaves, and that those leaves have labels $[N] := \{1, \ldots, N\}$. Every possible edge in such a phylogenetic tree can be described by its corresponding *split*, which is a partition of $[N]$ into two non-empty sets, by removing that edge of the tree and observing the resulting partitioning of the leaf labels. If a split can be obtained by deleting such an edge of a given phylogenetic tree, we say that the tree *displays* that split. We use a vertical bar (|) to denote the division between the two sets of the bipartition. For example, if we take the unrooted tree with four leaves such that 1 and 2 are sister to one another, the tree displays splits $1|234$, $12|34$, $134|2$, $124|3$, and $123|4$. Two splits $A|B$ and $C|D$ on the same leaf set are called *compatible* if one of $A \cap C$, $B \cap C$, $A \cap D$, or $B \cap D$ is empty. A set of splits that are pairwise compatible can be displayed on a phylogenetic tree (Buneman, 1971), and in fact the set of pairwise compatible sets of splits is in one-to-one correspondence with the set of (potentially multifurcating) unrooted phylogenetic trees.

When a single branch length goes to zero, $\mathcal{N}(\tau, q)$ will have three elements: τ itself and its two NNI neighbors. When multiple branch lengths go to zero, one can re-expand branch lengths for any set of splits that are compatible with each other and with the splits that did not originally go to zero. This generalizes the NNI condition. However, the correspondence between the branches that went to zero and the newly expanded branches is no longer obvious.

One can define such a correspondence using a global splits-based coordinate system. Namely, such a coordinate system can be achieved by indexing branch length vectors by splits, with the proviso that for any two incompatible splits r and s, one of q_r or q_s is zero. We could have used such a coordinate system for this paper, such that branch length vectors q would live in $\mathbb{R}^{2^{N-1}}$.

However, for simplicity of notation, we have indexed the branch lengths (e.g. in Algorithm 1) with integers $[n]$ corresponding to the actual branches of a phylogenetic tree. Thus our branch length vectors q live in $2N - 3$ dimensions. One can use a total order on the splits to unambiguously define which branches map to which others when the HMC crosses a boundary. We will describe how this works when two branch lengths, q_i and q_j, go to zero. The extension to more branch lengths is clear.

Our branch indices $i, j \in [2N - 3]$ are always associated with a phylogenetic tree τ with numbered edges. For any branch index i on τ, one can unambiguously take the split s_i. Assume without loss of generality that $s_i < s_j$ in the total order on splits. Now, when q_i and q_j go to zero, one can transition to a new tree τ' which may differ from τ by up to two splits. We assume without loss of generality that these are actually new splits (if not, we are in a previously defined setting) which we call s'_1 and s'_2 such that $s'_1 < s'_2$. We carry all of the branch indices for branches that aren't shrinking to zero across to τ'. Then map branch i in τ to the branch in τ' corresponding to the split s'_1, and branch j to the branch in τ' corresponding to the split s'_2. Thus, for example, the momentum q_i in the τ orthant is carried over to this corresponding q_i in the τ' orthant.