In this supplementary document, we provide details of our proposed Wormhole Hamiltonian Monte Carlo (WHMC) algorithm and prove its convergence to the stationary distribution. For simplicity, we assume that $G(\theta) \equiv I$. Our results can be extended to more general Riemannian metrics.

In what follows, we first prove the convergence of our method when an external vector field is added to the dynamics. Next, we prove the convergence for our final algorithm, where besides an external vector field, we include an auxiliary dimension along which the network of wormholes are constructed. Finally, we provide our algorithm to identify regeneration times.

A Adjustment of Metropolis acceptance probability in WHMC with vector field

As mentioned in the paper, in high dimensional problems with isolated modes, the effect of wormhole metric could diminish fast as the sampler leaves one mode towards another mode. To avoid this issue, we have extended our method by including a vector field, $f(\theta, v)$, which depends on a vicinity function illustrated in Figure 1. The resulting dynamics facilitates movements between modes,

$$\dot{\theta} = v + f(\theta, v) \quad (A.1)$$
$$\dot{v} = -\nabla_\theta U(\theta) \quad (A.2)$$

We solve (A.1) using the generalized leapfrog integrator [Leimkuhler and Reich, 2004; Girolami and Calderhead, 2011]:

$$v(\ell+1/2) = v(\ell) - \frac{\varepsilon}{2} \nabla_\theta U(\theta(\ell)) \quad (A.3)$$
$$\theta(\ell+1) = \theta(\ell) + \varepsilon[v(\ell+1/2) + f(\theta(\ell), v(\ell+1/2)) + f(\theta(\ell+1), v(\ell+1/2))]/2 \quad (A.4)$$

where $\ell$ is the index for leapfrog steps, and $v(\ell+1/2)$ denotes the current value of $v$ after half a step of leapfrog. The implicit equation (A.3) can be solved by the fixed point iteration.

The integrator (A.2)-(A.4) is time reversible and numerically stable; however, it is not volume preserving. To fix this issue, we can adjust the Metropolis acceptance probability with the Jacobian determinant of the mapping $\hat{T}$ given by (A.2)-(A.4) in order to satisfy the detailed balance condition. Denote $z = (\theta, v)$. Given the corresponding Hamiltonian function, $H(z)$, we define $P(dz) = \exp(-H(z))dz$ and prove the following proposition [See Green, 1995].
**Proposition A.1** (Detailed Balance Condition with determinant adjustment). Let \( z' = \hat{T}_L(z) \) be the proposal according to some time reversible integrator \( \hat{T}_L \) for dynamics \( \theta^{(L+1)} \). Then the detailed balance condition holds given the following adjusted acceptance probability:

\[
\tilde{\alpha}(z, z') = \min \left\{ 1, \frac{\exp(-H(z'))}{\exp(-H(z))} | \det d\hat{T}_L | \right\}
\]  

(A.5)

**Proof.** To prove detailed balance, we need to show

\[
\tilde{\alpha}(z, z') \mathbb{P}(dz) = \tilde{\alpha}(z', z) \mathbb{P}(dz')
\]  

(A.6)

The following steps show that this condition holds:

\[
\tilde{\alpha}(z, z') \mathbb{P}(dz) = \min \left\{ 1, \frac{\exp(-H(z'))}{\exp(-H(z))} \left| \frac{dz'}{dz} \right| \right\} \exp(-H(z))dz = \min \left\{ 1, \frac{\exp(-H(z'))}{\exp(-H(z'))} \left| \frac{dz}{dz'} \right| \right\} \exp(-H(z))dz = \tilde{\alpha}(z', z) \mathbb{P}(dz')
\]

We implement (A.2)-(A.4) for \( L \) steps to generate a proposal \( z^{(L+1)} \) and accept it with the following adjusted probability:

\[
\alpha_{VF}(z^{(1)}, z^{(L+1)}) = \min\{1, \exp(-H(z^{(L+1)}) + H(z^{(1)})\} \det J_{VF} \}
\]

where the Jacobian determinant, \( \det J_{VF} = \prod_{n=1}^{L} \left| \frac{dz^{(n)}}{dz^{(n-1)}} \right| = \prod_{n=1}^{L} \left| \frac{\partial(\theta^{(n+1)}, v^{(n)})}{\partial(\theta^{(n)}, v^{(n)})} \right| \), can be calculated through the following wedge product:

\[
d\theta^{(n+1)} \wedge dv^{(n)} = [I - \frac{\xi}{2} \nabla_{\theta^{(n)}} f(\theta^{(n)}), v^{(n+1/2)}]^{-1} [I + \frac{\xi}{2} \nabla_{\theta^{(n)}} f(\theta^{(n)}), v^{(n+1/2)}] d\theta^{(n)} \wedge dv^{(n)}
\]

with \( \nabla_{\theta^{(n)}} f(\theta, v) = v^T \nabla m(\theta)^T \) [See Lan et al., 2012, for more details.]
## B WHMC in the augmented $D + 1$ dimensional space

Suppose that the current position, $\tilde{\theta}$, of the sampler is near a mode denoted as $\tilde{\theta}_0$. A network of wormholes connects this mode to all the modes in the opposite world $\tilde{\theta}_k^*, k = 1, \cdots, K$. Wormholes in the augmented space starting from this mode may still interfere each other since they intersect. To resolve this issue, instead of deterministically weighing wormholes by the vicinity function (6), we use the following random vector field $\tilde{f}(\tilde{\theta}, \tilde{v})$:

$$
\tilde{f}(\tilde{\theta}, \tilde{v}) \sim \left\{ \begin{array}{ll}
(1 - \sum_k m_k(\tilde{\theta})) & \delta_{\tilde{v}}(\cdot) + \sum_k m_k(\tilde{\theta}) \delta_{2(\tilde{\theta}_k^* - \tilde{\theta})/e}(\cdot), & \text{if } \sum_k m_k(\tilde{\theta}) < 1 \\
\sum_k m_k(\tilde{\theta}) \delta_{2(\tilde{\theta}_k^* - \tilde{\theta})/e}(\cdot) & , & \text{if } \sum_k m_k(\tilde{\theta}) \geq 1
\end{array} \right.
$$

where $e$ is the stepsize, $\delta$ is the Kronecker delta function, and $m_k(\tilde{\theta}) = \exp\{-V_k(\tilde{\theta})/(DF)\}$. Here, the vicinity function $V_k(\tilde{\theta})$ along the $k$-th wormhole is defined similarly to Equation (6),

$$
V_k(\tilde{\theta}) = \langle \tilde{\theta} - \tilde{\theta}_0, \tilde{\theta} - \tilde{\theta}_k^* \rangle + |\langle \tilde{\theta} - \tilde{\theta}_0, \tilde{v}_{W_k} \rangle| |\langle \tilde{\theta} - \tilde{\theta}_k^*, \tilde{v}_{W_k} \rangle|
$$

where $\tilde{v}_{W_k} = (\tilde{\theta}_k^* - \tilde{\theta}_0)/\|\tilde{\theta}_k^* - \tilde{\theta}_0\|$.

For each update, $\tilde{f}(\tilde{\theta}, \tilde{v})$ is either set to $\tilde{v}$ or $2(\tilde{\theta}_k^* - \tilde{\theta})/e$ according to the position dependent probabilities defined in terms of $m_k(\tilde{\theta})$. Therefore, we write the Hamiltonian dynamics in the extended space as follows:

$$
\dot{\tilde{\theta}} = \tilde{f}(\tilde{\theta}, \tilde{v}) \\
\dot{\tilde{v}} = -\nabla_{\tilde{\theta}} U(\tilde{\theta}) \quad \text{(B.1)}
$$

We use (A.2)-(A.4) to numerically solve (B.1), but replace (A.3) with the following equation:

$$
\tilde{\theta}^{(\ell+1)} = \tilde{\theta}^{(\ell)} + e/2[\tilde{f}(\tilde{\theta}^{(\ell+1)}, \tilde{v}^{(\ell+1/2)}) + \tilde{f}(\tilde{\theta}^{(\ell)}, \tilde{v}^{(\ell+1/2)})] \quad \text{(B.2)}
$$

Note that this is an implicit equation, which can be solved using the fixed point iteration approach [Leimkuhler and Reich 2004, Girolami and Calderhead 2011].

According to the above dynamic, at each leapfrog step, $\ell$, the sampler either stays at the vicinity of $\tilde{\theta}_0$ or proposes a move towards a mode $\tilde{\theta}_k^*$ in the opposite world depending on the values of $\tilde{f}(\tilde{\theta}^{(\ell)}, \tilde{v}^{(\ell+1/2)})$ and $\tilde{f}(\tilde{\theta}^{(\ell+1)}, \tilde{v}^{(\ell+1/2)})$. For example, if $\tilde{f}(\tilde{\theta}^{(\ell)}, \tilde{v}^{(\ell+1/2)}) = 2(\tilde{\theta}_k^* - \tilde{\theta}^{(\ell)})/e$, and $\tilde{f}(\tilde{\theta}^{(\ell+1)}, \tilde{v}^{(\ell+1/2)}) = \tilde{v}^{(\ell+1/2)}$, then equation (B.2) becomes

$$
\tilde{\theta}^{(\ell+1)} = \tilde{\theta}_k^* + \frac{e}{2} \tilde{v}^{(\ell+1/2)}
$$

which indicates that a move to the $k$-th mode in the opposite world has in fact occurred. Note that the movement $\tilde{\theta}^{(\ell)} \to \tilde{\theta}^{(\ell+1)}$ in this case is discontinuous since

$$
\lim_{e \to 0} \|\tilde{\theta}^{(\ell+1)} - \tilde{\theta}^{(\ell)}\| \geq 2h > 0
$$

Therefore, in such cases, there will be an energy gap, $\Delta E := H(\tilde{\theta}^{(\ell+1)}, \tilde{v}^{(\ell+1)}) - H(\tilde{\theta}^{(\ell)}, \tilde{v}^{(\ell)})$, between the two states. We need to adjust the Metropolis acceptance probability to account for the resulting
Further, we limit the maximum number of jumps within each iteration of MCMC (i.e., over \( L \) leapfrog steps) to 1 so the sampler can explore the vicinity of the new mode before making another jump. Algorithm \( \square \) provides the details of this approach.

We note that according to the definition of \( \tilde{f}(\theta, \hat{v}) \) and equation \( \square \), the jump occurs randomly. We use \( \ell' \) to denote the step at which the sampler jumps. That is, \( \ell' \) randomly takes a value in \( \{0, 1, \ldots, L\} \) depending on \( \tilde{f} \). When there is no jump along the trajectory, we set \( \Delta E = 0, \ell' = 0, \) and the algorithm reduces to standard HMC. In the following, we first prove the detailed balance condition when a jump happens, and then use it to prove the convergence of the algorithm to the stationary distribution.

When a mode jumping occurs at some fixed step \( \ell' \), we can divide the \( L \) leapfrog steps into three parts: \( \ell' - 1 \) steps continuous movement according to standard HMC, 1 step discontinuous jump, and \( L - \ell' \) steps according to standard HMC in the opposite world. Note that the Metropolis acceptance probability \( \alpha \) depends on \( \tilde{E} \) depending on \( \tilde{z} \).

That is, we only count the acceptance probability for the first \( \ell' - 1 \) and the last \( L - \ell' \) steps of continuous movement in standard HMC; at \( \ell' + 1 \) step, we “reset” the energy level by accounting for the energy gap \( \Delta E \). Therefore, the following proposition is true.

**Proposition B.1 (Detailed Balance Condition with energy adjustment).** When a discontinuous jump happens in WHMC, we have the following detailed balance condition with the adjusted acceptance probability \( \square \).

\[
\alpha_{RF}(\tilde{z}(1), \tilde{z}(L+1)) = \min \{1, \exp(-H(\tilde{z}(L+1)) + H(\tilde{z}(1))) \} = \min \left\{1, \exp \left( -\sum_{\ell=1}^{L} (H(\tilde{z}(\ell+1)) - H(\tilde{z}(\ell))) \right) \right\}
\]

Each summand \( H(\tilde{z}(\ell+1)) - H(\tilde{z}(\ell)) \) is small \( (\mathcal{O}(\varepsilon^3)) \) except for the \( \ell' \)-th one, where there is an energy gap, \( \Delta E \). Given that the jump happens at the \( \ell' \)-th step, we should remove the \( \ell' \)-th summand from the acceptance probability:

\[
\alpha_{RF}(\tilde{z}(1), \tilde{z}(L+1)) = \min \left\{1, \exp \left( -\sum_{\ell \neq \ell'} (H(\tilde{z}(\ell+1)) - H(\tilde{z}(\ell))) \right) \right\}
\]

That is, we only count the acceptance probability for the first \( \ell' - 1 \) and the last \( L - \ell' \) steps of continuous movement in standard HMC; at \( \ell' + 1 \) step, we “reset” the energy level by accounting for the energy gap \( \Delta E \). Therefore, the following proposition is true.

**Theorem 1 (Stationarity of WHMC).** The samples given by WHMC (algorithm \( \square \)) have the target distribution \( \pi(\cdot) \) as its stationary distribution.

**Proof.** Let \( \tilde{z}^* = \tilde{T}_{1:L}(\tilde{z}) \). Suppose \( \tilde{\theta}^* \sim f(\tilde{\theta}^*) \). We want to prove that \( f(\cdot) = \pi(\cdot) \) through \( E_{\tilde{f}}[h(\tilde{\theta}^*)] = E_{\pi}[h(\tilde{\theta}^*)] \) for any square integrable function \( h(\tilde{\theta}^*) \).

---

\( ^1 \)Proposition \( \square \) does not apply because Jacobian determinant is not well defined for discontinuous movement.
Note that \( \hat{\theta} \) is either an accepted proposal or the current state after a rejection. Therefore,

\[
E_f[h(\hat{\theta}^*)] = \iint h(\hat{\theta}^*) \alpha(\hat{T}_{1:L}^{-1}(\hat{z}^*), \hat{z}^*) \mathbb{P}(d\hat{T}_{1:L}^{-1}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1+\ell:L}^{-1}(\hat{z}^*)) + (1 - \alpha(\hat{z}^*, \hat{T}_{1:L}(\hat{z}^*))) \mathbb{P}(d\hat{z}^*) \mathbb{P}(d\hat{T}_{1,\ell'}(\hat{z}^*)) \] 

= \int h(\hat{\theta}^*) \mathbb{P}(d\hat{z}^*) + 
\iint h(\hat{\theta}^*) \alpha(\hat{T}_{1:L}^{-1}(\hat{z}^*), \hat{z}^*) \mathbb{P}(d\hat{T}_{1:L}^{-1}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1+\ell:L}^{-1}(\hat{z}^*)) - \alpha(\hat{z}^*, \hat{T}_{1:L}(\hat{z}^*)) \mathbb{P}(d\hat{z}^*) \mathbb{P}(d\hat{T}_{1,\ell'}(\hat{z}^*)) \] 

Therefore, it suffices to prove that

\[
\iiint h(\hat{\theta}^*) \mathbb{P}(d\hat{T}_{1:L}^{-1}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1+\ell:L}^{-1}(\hat{z}^*)) = \iiint h(\hat{\theta}^*) \alpha(\hat{z}^*, \hat{T}_{1:L}(\hat{z}^*)) \mathbb{P}(d\hat{z}^*) \mathbb{P}(d\hat{T}_{1,\ell'}(\hat{z}^*)) \]  
\tag{B.5} 

Based on its construction, \( \hat{T}_{\ell} \) is time reversible for all \( \ell \). Denote the involution \( \nu : (\hat{\theta}, \hat{v}) \mapsto (\hat{\theta}, -\hat{v}) \). We have \( \hat{T}_{\ell}^{-1}(\hat{z}^*) = \nu \hat{T}_{\ell} \nu(\hat{z}^*) \). Further, because \( E \) is quadratic in \( \hat{v} \), we have \( H(\nu(\hat{z})) = H(\hat{z}) \). Therefore \( \alpha(\nu(\hat{z}), \hat{z}') = \alpha(\hat{z}, \nu(\hat{z}')) \). Then the left hand side of (B.5) becomes

\[
\text{LHS} = \iiint h(\hat{\theta}^*) \alpha(\nu \hat{T}_{1:L} \nu(\hat{z}^*), \nu(\hat{z}^*)) \mathbb{P}(d\nu \hat{T}_{1:L} \nu(\hat{z}^*)) \mathbb{P}(d\nu \hat{T}_{1+\ell:L} \nu(\hat{z}^*)) d\ell' 
\]
\[
= \iiint h(\hat{\theta}^*) \alpha(\hat{T}_{1:L}(\hat{z}^*), \nu(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1:L}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1+\ell:L}(\hat{z}^*)) d\ell' 
\]
\[
\nu(\hat{z}^*) \mapsto \hat{z}^* \iiint h(\hat{\theta}^*) \alpha(\hat{T}_{1:L}(\hat{z}^*), \hat{z}^*) \mathbb{P}(d\hat{T}_{1:L}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1+\ell:L}(\hat{z}^*)) d\ell' 
\]

On the other hand, by the detailed balance condition (B.4), the left hand side of (B.5) becomes

\[
\text{RHS} = \iiint h(\hat{\theta}^*) \alpha(\hat{T}_{1:L}(\hat{z}^*), \hat{z}^*) \mathbb{P}(d\hat{T}_{1:L}(\hat{z}^*)) \mathbb{P}(d\hat{T}_{1,\ell'-1}(\hat{z}^*)) d\ell' 
\]

Note that \( \mathbb{P}(d\hat{T}_{1+\ell:L}(\hat{z}^*)) = \mathbb{P}(d\hat{T}_{1,\ell'-1}(\hat{z}^*)) \) since both \( \hat{T}_{1+\ell:L} \) and \( \hat{T}_{1,\ell'-1} \) are leapfrog steps of standard HMC (no jump). The difference in the numbers of leapfrog steps (i.e., \( L - \ell' \) and \( \ell' - 1 \)) does not affect the stationarity since the number of leapfrog steps can be randomized in HMC [Neal, 2010]. Therefore, we have LHS = RHS, which proves (B.5). \( \square \)
Algorithm 1 Wormhole Hamilton Monte Carlo (WHMC)

Prepare the modes $\hat{\theta}^*_k$, $k = 1, \cdots K$
Set $\hat{\theta}^{(1)} =$ current $\hat{\theta}$
Sample velocity $\hat{\mathbf{v}}^{(1)} \sim \mathcal{N}(0, \mathbf{I}_{D+1})$
Calculate $E(\hat{\mathbf{v}}^{(1)}, \hat{\theta}^{(1)}) = U(\hat{\theta}^{(1)}) + K(\hat{\mathbf{v}}^{(1)})$
Set $\Delta \log \det = 0$, $\Delta E = 0$, Jumped = false.

for $n = 1$ to $L$ do
  $\hat{\mathbf{v}}^{(\ell + \frac{1}{2})} = \hat{\mathbf{v}}^{(\ell)} - \frac{\epsilon}{2} \nabla_{\hat{\theta}} U(\hat{\theta}^{(\ell)})$
  if Jumped then
    $\hat{\theta}^{(\ell + 1)} = \hat{\theta}^{(\ell)} + e^{\hat{\mathbf{v}}^{(\ell + \frac{1}{2})}}$
  else
    Find the closest mode $\hat{\theta}^*_0$ and build a network connecting it to all modes $\hat{\theta}^*_k$, $k = 1, \cdots K$ in the opposite world
    for $m = 1$ to $M$ do
      Calculate $m_k(\hat{\theta}^{(m)})$, $k = 1, \cdots K$
      Sample $u \sim \text{Unif}(0, 1)$
      if $u < 1 - \sum_k m_k(\hat{\theta}^{(m)})$ then
        Set $f(\hat{\theta}^{(m)}, \hat{\mathbf{v}}^{(\ell + \frac{1}{2})}) = \hat{\mathbf{v}}^{(\ell + \frac{1}{2})}$
      else
        Choose one of the $k$ wormholes according to probability $\{m_k/\sum_{k'} m_{k'}\}$ and set
        $f(\hat{\theta}^{(m)}, \hat{\mathbf{v}}^{(\ell + \frac{1}{2})}) = 2(\hat{\theta}^*_k - \hat{\theta}^{(m)})/e$
      end if
    end for
    $\hat{\theta}^{(m + 1)} = \hat{\theta}^{(\ell)} + \frac{e}{2} [f(\hat{\theta}^{(m)}, \hat{\mathbf{v}}^{(\ell + \frac{1}{2})}) + f(\hat{\theta}^{(\ell)}, \hat{\mathbf{v}}^{(\ell + \frac{1}{2})})]$
  end if
end for

$\hat{\theta}^{(L + 1)} = \hat{\theta}^{(M + 1)}$
$\hat{\mathbf{v}}^{(L + 1)} = \hat{\mathbf{v}}^{(\ell + \frac{1}{2})} - \frac{\epsilon}{2} \nabla_{\hat{\theta}} U(\hat{\theta}^{(L + 1)})$
If a jump has occurred, set Jumped = true and calculate energy gap $\Delta E$.

Calculate $E(\hat{\theta}^{(L + 1)}, \hat{\mathbf{v}}^{(L + 1)}) = U(\hat{\theta}^{(L + 1)}) + K(\hat{\mathbf{v}}^{(L + 1)})$
$p = \exp\{-E(\hat{\theta}^{(L + 1)}, \hat{\mathbf{v}}^{(L + 1)}) + E(\hat{\theta}^{(1)}, \hat{\mathbf{v}}^{(1)}) + \Delta E\}$
Accept or reject the proposal $(\hat{\theta}^{(L + 1)}, \hat{\mathbf{v}}^{(L + 1)})$ according to $p$
Algorithm 2 Regeneration in Wormhole Hamiltonian Monte Carlo

Initially search modes $\theta_1, \cdots, \theta_k$

for $n = 1$ to $L$

Sample $\tilde{\theta} = (\theta, \theta_{D+1})$ as the current state according to WHMC (algorithm 1).

Fit a mixture of Gaussians $q(\theta)$ with known modes, Hessians and relative weights. Propose $\theta^* \sim q(\cdot)$ and accept it with probability $\alpha = \min \left\{ 1, \frac{\pi(\theta^*)/q(\theta^*)}{\pi(\theta)/q(\theta)} \right\}$.

if $\theta^*$ accepted then

Determine if $\theta^*$ is a regeneration using (9)-(12) with $\theta_t = \theta$ and $\theta_{t+1} = \theta^*$.

if Regeneration occurs then

Search new modes by minimizing $U_r(\theta, T)$; if new modes are discovered, update the mode library, wormhole network, and $q(\theta)$.

Discard $\theta^*$, sample $\theta^{(\ell+1)} \sim Q(\cdot)$ as in (12) using rejection sampling.

else

Set $\theta^{(\ell+1)} = \theta^*$.

end if

else

Set $\theta^{(\ell+1)} = \tilde{\theta}$.

end if

end for

References


