Appendices

A Simulated Tempering Metropolis–Hastings algorithm

Algorithm 1 Simulated Tempering Metropolis–Hastings Algorithm

Input: function f, inverse temperatures $\beta_1, \ldots, \beta_\ell$, partition function estimates $\widehat{Z}_1, \ldots, \widehat{Z}_\ell$, number of steps N, step size η, rate λ, initial covariance matrix Σ_0.
Sample $x_0 \sim \mathcal{N}(0, \Sigma_0)$
$i \leftarrow 1, x \leftarrow x_0, n \leftarrow 0$
while $n < N$ **do**
  Sample $u \sim$ Bernoulli(λ).
  if $u = 0$ **then**
    Propose $x' \sim \mathcal{N}(x, \eta I)$
    Sample $v \sim \mathrm{Uniform}(0, 1)$
    if $v < \min\left\{1, \frac{e^{-\beta_i f(x')}}{e^{-\beta_i f(x)}}\right\}$ **then**
      $x \leftarrow x'$
    end if
  else
    Propose $i' = i \pm 1$, each with probability $1/2$
    if $1 \leq i' \leq \ell$ **then**
      Sample $v \sim \mathrm{Uniform}(0, 1)$
      if $v < \min\left\{1, \frac{e^{-\beta_{i'} f(x)}/\widehat{Z}_{i'}}{e^{-\beta_i f(x)}/\widehat{Z}_i}\right\}$ **then**
        $i \leftarrow i'$
      end if
    end if
  end if
  $n \leftarrow n + 1$
end while
Output: Sample (x, i) collected at the N^{th} step.

Algorithm 2 Partition Function Estimation

Input: function f, inverse temperature sequence $\beta_1 < \cdots < \beta_L$ and number of samples s.
$\widehat{Z}_1 \leftarrow 1$
for $\ell = 1$ **to** L **do**
  Repeat Algorithm 1 until s samples $(x_j)_{j=1}^s$ are obtained at temperature level ℓ.
  $\widehat{Z_{\ell+1}} \leftarrow (\widehat{Z_\ell}/s) \sum_{j=1}^s e^{(-\beta_{\ell+1}+\beta_\ell) f(x_j)}$
end for

Remark 7. Algorithm 1 is always run for a fixed number of steps N and returns the sample obtained at the final step. In Algorithm 2, if this sample is not from the desired temperature level, Algorithm 1 is simply re-run for another N steps.

B Proofs for Section 2

B.1 Proof of Lemma 1

Proof. Clearly, $\sum_{i,j} \overline{P}(i,j) = 1$. Hence, it only remains to check the detailed balance condition

$$\overline{P}((i,j))\,\overline{M}((i,j),(i',j')) = \overline{P}((i',j'))\,\overline{M}((i',j'),(i,j))$$

for two types of moves. First, let i be fixed and consider $j \neq j'$. Then,

$$r_i w_{(i,j)} P_{(i,j)}(\mathcal{X}^0)\overline{M}((i,j),(i,j')) = r_i(1-\lambda)\int_{\mathcal{X}^0} w_{(i,j)} p_{(i,j)}(x) \frac{w_{(i,j')} p_{(i,j')}(x)}{p_i(x)} \mathrm{d}x,$$

which is clearly symmetric with respect to j and j'. Similarly, if j is fixed and $i' = i \pm 1$ (assuming both $i, i' \in [L]$), we have

$$\begin{aligned} r_i w_{(i,j)} P_{(i,j)}(\mathcal{X}^0)\overline{M}((i,j),(i',j)) &= \frac{\lambda}{2}\int_{\mathcal{X}^0} r_i w_{(i,j)} p_{(i,j)}(x) a((i,j,x),(i',j,x))\mathrm{d}x \\ &= \frac{\lambda}{2}\int_{\mathcal{X}^0} \min\left\{r_i w_{(i,j)} p_{(i,j)}(x), r_{i'} w_{(i',j)} p_{(i',j)}(x)\right\}\mathrm{d}x, \end{aligned}$$

which is symmetric with respect to i and i'. □

B.2 Proof of Theorem 1

We first prove an auxiliary lemma on the Dirichlet form of the simulated tempering Markov chain.

Lemma 3. *The $[L] \times \mathcal{X}^0$-restricted Dirichlet form $\mathcal{E}_{[L]\times\mathcal{X}^0}$ of the simulated tempering Markov chain M, defined in Definition 2, can be expressed by*

$$\mathcal{E}_{[L]\times\mathcal{X}^0}(g,g) = (1-\lambda)\sum_{i=1}^{L} r_i\, \mathcal{E}_{i,\mathcal{X}^0}(g_i,g_i) + \lambda\, \mathcal{E}^I_{\mathcal{X}^0}(g,g),$$

where $\mathcal{E}_{i,\mathcal{X}^0}$ is the $\mathcal{X}^0$-restricted Dirichlet form of the Markov chain M_i, $g \in \mathcal{L}^2([L]\times\mathcal{X}, P)$ with $g_i(x) = g(i,x)$ for each $i \in [L]$ and

$$\mathcal{E}^I_{\mathcal{X}^0}(g,g) = \frac{1}{4}\sum_{i,i'\in[L]:\, i'=i\pm1} \int_{\mathcal{X}^0} \big(g(i,x) - g(i',x)\big)^2 r_i\, p_i(x)\, a\big((i,x),(i',x)\big)\,\mathrm{d}x.$$

Proof. Since the stationary density of M is $p(i,x) = r_i p_i(x)$ and either x or i is fixed in each simulated tempering iteration, the restricted Dirichlet form $\mathcal{E}_{[L]\times\mathcal{X}^0}(g,g)$ can be expressed by

$$\begin{aligned} \mathcal{E}_{[L]\times\mathcal{X}^0}(g,g) = \frac{1}{2}\sum_{i=1}^{L}\int_{\mathcal{X}^0}\int_{\mathcal{X}^0} \big(g(i,x)-g(i,y)\big)^2 r_i\, p_i(x)\, M\big((i,x),(i,\mathrm{d}y)\big)\,\mathrm{d}x \\ + \frac{1}{2}\sum_{i,i'\in[L]:\, i'=i\pm1}\int_{\mathcal{X}^0} \big(g(i,x)-g(i',x)\big)^2 r_i\, p_i(x)\, M\big((i,x),(i',x)\big)\,\mathrm{d}x. \end{aligned} \tag{10}$$

Since $M\big((i,x),(i,\mathrm{d}y)\big) = (1-\lambda)M_i(x,\mathrm{d}y)$ and M_i has stationary density p_i,

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{L}\int_{\mathcal{X}^0}\int_{\mathcal{X}^0} \big(g(i,x)-g(i,y)\big)^2 r_i\, p_i(x)\, M\big((i,x),(i,\mathrm{d}y)\big)\,\mathrm{d}x \\ &= \frac{1-\lambda}{2}\sum_{i=1}^{L} r_i \int_{\mathcal{X}^0}\int_{\mathcal{X}^0} \big(g_i(x)-g_i(y)\big)^2 p_i(x)\, M_i\big(x,\mathrm{d}y\big)\mathrm{d}x = (1-\lambda)\sum_{i=1}^{L} r_i\, \mathcal{E}_{i,\mathcal{X}^0}\big(g_i,g_i\big). \end{aligned}$$

For the second term on the right-hand side of (10), note that

$$M\big((i,x),(i',x)\big) = \frac{\lambda}{2}a((i,x),(i',x)),$$

where the acceptance probability function a is given by (1). A straightforward calculation then concludes the proof of the lemma. □

Next, we prove two key lemmas about the Dirichlet form $\overline{\mathcal{E}}$ of the Markov chain $\overline{M}$ constructed in Definition 3. Let $\theta = P([L] \times \mathcal{X}^0)$. Recall that under Assumption 1, we can augment the stationary density to

$$p(i, j, x) = r_i w_{(i,j)} p_{(i,j)}(x).$$

We still denote the corresponding probability measure by P. Let P_0 denote the conditional probability measure given $X \in \mathcal{X}^0$, whose density is given by

$$p_0(i, j, x) = \frac{r_i w_{(i,j)} p_{(i,j)}(x)}{\theta}.$$

The Dirichlet form $\overline{\mathcal{E}}$ can be expressed by

$$\overline{\mathcal{E}}(\overline{g}, \overline{g}) = \overline{\mathcal{E}}^J(\overline{g}, \overline{g}) + \overline{\mathcal{E}}^I(\overline{g}, \overline{g}),$$

where

$$\overline{\mathcal{E}}^J(\overline{g}, \overline{g}) = \frac{1}{2\theta} \sum_{i=1}^{L} \sum_{j,j'=1}^{n} (\overline{g}(i,j) - \overline{g}(i,j'))^2\, r_i w_{(i,j)} P_{(i,j)}(\mathcal{X}^0)\, \overline{M}((i,j),(i,j')),$$

$$\overline{\mathcal{E}}^I(\overline{g}, \overline{g}) = \frac{1}{2\theta} \sum_{j=1}^{n} \sum_{i,i' \in [L]:\, i'=i\pm 1} (\overline{g}(i,j) - \overline{g}(i',j))^2\, r_i w_{(i,j)} P_{(i,j)}(\mathcal{X}^0)\, \overline{M}((i,j),(i',j)).$$

Lemma 4. *Suppose Assumption 1 holds. For any $g \in \mathcal{L}^2([L] \times \mathcal{X}, P)$, define $g_i \colon \mathcal{X} \to \mathbb{R}$ by $g_i(x) = g(i,x)$, and define $\overline{g} \colon [L] \times [n] \to \mathbb{R}$ by*

$$\overline{g}(i,j) = \int_{\mathcal{X}^0} g(i,x) \frac{p_{(i,j)}(x)}{P_{(i,j)}(\mathcal{X}^0)} \mathrm{d}x.$$

Then,

$$\overline{\mathcal{E}}^J(\overline{g}, \overline{g}) \le \frac{2(1-\lambda)}{\theta} \sum_{i=1}^{L} \sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{P_{(i,j)}(\mathcal{X}^0)} \operatorname{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i).$$

Proof. For every $x \in \mathcal{X}$, $i \in [L]$ and every pair $j, j' \in [n]$, the following inequality holds

$$\left(\overline{g}(i,j) - \overline{g}(i,j')\right)^2 \le 2\left[\left(\overline{g}(i,j) - g(i,x)\right)^2 + \left(\overline{g}(i,j') - g(i,x)\right)^2\right].$$

Hence, using the expression for $\overline{M}((i,j),(i,j'))$, we get

$$\begin{aligned}
\overline{\mathcal{E}}^J(\overline{g}, \overline{g}) &= \frac{1-\lambda}{2\theta} \sum_{i=1}^{L} \sum_{j,j'=1}^{n} (\overline{g}(i,j) - \overline{g}(i,j'))^2 r_i w_{(i,j)} \int_{\mathcal{X}^0} p_{(i,j)}(x) p_i(j'|x)\, \mathrm{d}x \\
&\le (1-\lambda) \sum_{i=1}^{L} \sum_{j,j'=1}^{n} r_i w_{(i,j)} \int_{\mathcal{X}^0} \left[\left(\overline{g}(i,j) - g(i,x)\right)^2 + \left(\overline{g}(i,j') - g(i,x)\right)^2\right] \frac{p_{(i,j)}(x)}{\theta} p_i(j'|x)\, \mathrm{d}x \\
&= (1-\lambda) \mathsf{E}_{\tilde{P}}\left[\left(\overline{g}(I,J) - g(I,X)\right)^2 + \left(\overline{g}(I,J') - g(I,X)\right)^2\right],
\end{aligned} \tag{11}$$

where $\tilde{P}$ denotes the joint probability measure of (I, J, J', X) with density

$$\tilde{p}(i,j,j',x) = \frac{r_i w_{(i,j)} p_{(i,j)}(x)}{\theta} p_i(j'|x) = \frac{r_i w_{(i,j)} w_{(i,j')} p_{(i,j)}(x) p_{(i,j')}(x)}{\theta\, p_i(x)},$$

for $i \in [L], j, j' \in [n], x \in \mathcal{X}^0$. Hence, under $\tilde{P}$, the joint distribution of (I, J, X) and that of (I, J', X) are both given by P_0, and thus

$$\mathsf{E}_{\tilde{P}}\left[\left(\overline{g}(I,J) - g(I,X)\right)^2 + \left(\overline{g}(I,J') - g(I,X)\right)^2\right] = 2\mathsf{E}_{P_0}\left[\left(\overline{g}(I,J) - g(I,X)\right)^2\right]. \tag{12}$$

Since $\overline{g}(i,j) = \mathsf{E}_{P_0}[g(I,X) \,|\, I=i, J=j]$, we find that

$$\begin{aligned}
\mathsf{E}_{P_0}\left[\left(\overline{g}(I,J) - g(I,X)\right)^2\right] &= \mathsf{E}_{P_0}\left[\mathrm{Var}_{P_0}(g(I,X) \,|\, I, J)\right] \\
&= \sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)} P_{(i,j)}(\mathcal{X}^0)}{\theta}\ \mathrm{Var}_{P_0}(g(I,X) \,|\, I=i, J=j) \\
&= \sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{\theta\, P_{(i,j)}(\mathcal{X}^0)}\ \mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i),
\end{aligned} \tag{13}$$

where in the last step we have used

$$\mathrm{Var}_{P_0}(g(I,X) \,|\, I=i, J=j) = \frac{1}{2}\int_{\mathcal{X}^0\times\mathcal{X}^0} [g(i,x) - g(i,y)]^2 \frac{p_{(i,j)}(x)}{P_{(i,j)}(\mathcal{X}^0)} \frac{p_{(i,j)}(y)}{P_{(i,j)}(\mathcal{X}^0)} \mathrm{d}x\mathrm{d}y.$$

The claim then follows from (11), (12) and (13). $\square$

Lemma 5. *Consider the setting of Lemma 4. We also have*

$$\overline{\mathcal{E}}^I(\overline{g},\overline{g}) \leq \frac{3\lambda}{\theta}\sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{P_{(i,j)}(\mathcal{X}^0)}\ \mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i) + \frac{3\lambda}{\theta}\mathcal{E}^I_{\mathcal{X}^0}(g,g).$$

Proof. For every $x \in \mathcal{X}$, every pair $i, i' \in [L]$, and each $j \in [n]$,

$$\left(\overline{g}(i,j) - \overline{g}(i',j)\right)^2 \leq 3\Big[\left(\overline{g}(i,j) - g(i,x)\right)^2 + \left(g(i,x) - g(i',x)\right)^2 + \left(\overline{g}(i',j) - g(i',x)\right)^2\Big].$$

Then, using the definition of $\overline{M}((i,j),(i',j))$ for $i' = i \pm 1$, we get

$$\begin{aligned}
\overline{\mathcal{E}}^I(\overline{g},\overline{g}) &= \frac{\lambda}{4\theta}\sum_{i,i'\in[L]:\ i'=i\pm1}\sum_{j=1}^{n} (\overline{g}(i,j) - \overline{g}(i',j))^2 r_i w_{(i,j)} \int_{\mathcal{X}^0} p_{(i,j)}(x) a((i,j,x),(i',j,x))\mathrm{d}x \\
&\leq \frac{3\lambda}{2}\mathsf{E}_{\tilde{P}}\Big[\left(\overline{g}(I,J) - g(I,X)\right)^2 + \left(g(I,X) - g(I',X)\right)^2 + \left(\overline{g}(I',J) - g(I',X)\right)^2\Big],
\end{aligned}$$

where $\tilde{P}$ is the probability measure of (I, I', J, X) with density

$$\tilde{p}(i,i',j,x) = \begin{cases} \dfrac{r_i w_{(i,j)} p_{(i,j)}(x)\, a\big((i,j,x),(i',j,x)\big)}{2\theta}, & \text{if } i' = i \pm 1, \\ 1 - \tilde{p}(i,i+1,j,x) - \tilde{p}(i,i-1,j,x), & \text{if } i' = i, \\ 0, & \text{otherwise.} \end{cases}$$

That is, we first draw $I, J, X \sim P_0$ and then update I' by proposing $I' = I \pm 1$ with equal probability and accept it with probability $a((i,j,x),(i',j,x))$. Since the update for I' given I, J, X is reversible with respect to P_0, we also have $(I', J, X) \sim P_0$. Hence,

$$\mathsf{E}_{\tilde{P}}\Big[\left(\overline{g}(I,J) - g(I,X)\right)^2\Big] = \mathsf{E}_{\tilde{P}}\Big[\left(\overline{g}(I',J) - g(I',X)\right)^2\Big] = \mathsf{E}_{P_0}\left[\left(\overline{g}(I,J) - g(I,X)\right)^2\right]$$

which has been characterized in (13). Finally,

$$\begin{aligned}
&\mathsf{E}_{\tilde{P}}\Big[\left(g(I,X) - g(I',X)\right)^2\Big] \\
&= \frac{1}{2\theta}\sum_{i,i'\in[L]:\ i'=i\pm1}\sum_{j=1}^{n}\int_{\mathcal{X}^0} (g(i,x) - g(i',x))^2 r_i w_{(i,j)} p_{(i,j)}(x) a((i,j,x),(i',j,x))\mathrm{d}x \\
&= \frac{1}{2\theta}\sum_{i,i'\in[L]:\ i'=i\pm1}\sum_{j=1}^{n}\int_{\mathcal{X}^0} (g(i,x) - g(i',x))^2 \min\left\{r_i w_{(i,j)} p_{(i,j)}(x),\ r_{i'} w_{(i',j)} p_{(i',j)}(x)\right\}\mathrm{d}x \\
&\leq \frac{1}{2\theta}\sum_{i,i'\in[L]:\ i'=i\pm1}\int_{\mathcal{X}^0} (g(i,x) - g(i',x))^2 \min\left\{r_i p_i(x),\ r_{i'} p_{i'}(x)\right\}\mathrm{d}x \\
&= \frac{2}{\theta}\mathcal{E}^I_{\mathcal{X}^0}(g,g),
\end{aligned}$$

where $\mathcal{E}^I_{\mathcal{X}^0}(g,g)$ is defined in Lemma 3. Note that in the inequality above, we have used that $\sum_j \min\{a_j, b_j\} \leq \min\{\sum_j a_j, \sum_j b_j\}$ for two non-negative sequences a_j, b_j. □

Proof of Theorem 1. Fix an arbitrary $g \in \mathcal{L}^2([L] \times \mathcal{X}, P)$. Define, for each i, $g_i \colon \mathcal{X} \to \mathbb{R}$ by $g_i(x) = g(i,x)$, and $\overline{g} \colon [L] \times [n] \to \mathbb{R}$ by

$$\overline{g}(i,j) = \int_{\mathcal{X}^0} g(i,x) \frac{p_{(i,j)}(x)}{P_{(i,j)(\mathcal{X}^0)}} \mathrm{d}x.$$

Note that $\overline{g}(i,j)$ is the conditional expectation of $g(I,X)$ given $I = i$ and $J = j$ under the joint probability measure P_0, and $\overline{P}(i,j)$ is the marginal probability of $I = i, J = j$ under P_0. Hence, by the law of total variance, Assumption 3 and Equation 13, we find that

$$\begin{aligned}\mathrm{Var}_{P_0}(g) &= \mathrm{Var}_{\overline{P}}(\overline{g}) + \sum_{i=1}^{L}\sum_{j=1}^{n} \overline{P}(i,j)\,\mathrm{Var}_{P_0}(g(I,X)\,|\,I=i, J=j) \\ &\leq C_3 \overline{\mathcal{E}}(\overline{g},\overline{g}) + \sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{\theta P_{(i,j)}(\mathcal{X}^0)}\,\mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i).\end{aligned}$$

Using $P_{(i,j)}(\mathcal{X}^0) \geq \phi$ and Assumption 2,

$$\sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{\theta P_{(i,j)}(\mathcal{X}^0)}\,\mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i) \leq \frac{C_2}{\theta\phi}\sum_{i=1}^{L} r_i \sum_{j=1}^{n} w_{(i,j)}\mathcal{E}_{(i,j),\mathcal{X}^0}(g_i,g_i) \leq \frac{C_1C_2}{\theta\phi}\sum_{i=1}^{L} r_i \mathcal{E}_{i,\mathcal{X}^0}(g_i,g_i).$$

Recall that $\overline{\mathcal{E}}(\overline{g},\overline{g}) = \overline{\mathcal{E}}^J(\overline{g},\overline{g}) + \overline{\mathcal{E}}^I(\overline{g},\overline{g})$. By Lemma 4,

$$\overline{\mathcal{E}}^J(\overline{g},\overline{g}) \leq \frac{2(1-\lambda)}{\theta}\sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{P_{(i,j)}(\mathcal{X}^0)}\,\mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i) \leq \frac{2(1-\lambda)C_1C_2}{\theta\phi}\sum_{i=1}^{L} r_i \mathcal{E}_{i,\mathcal{X}^0}(g_i,g_i).$$

By Lemma 5,

$$\begin{aligned}\overline{\mathcal{E}}^I(\overline{g},\overline{g}) &\leq \frac{3\lambda}{\theta}\sum_{i=1}^{L}\sum_{j=1}^{n} \frac{r_i w_{(i,j)}}{P_{(i,j)}(\mathcal{X}^0)}\,\mathrm{Var}_{P_{(i,j)},\mathcal{X}^0}(g_i) + \frac{3\lambda}{\theta}\mathcal{E}^I_{\mathcal{X}^0}(g,g) \\ &\leq \frac{3\lambda C_1C_2}{\theta\phi}\sum_{i=1}^{L} r_i \mathcal{E}_{i,\mathcal{X}^0}(g_i,g_i) + \frac{3\lambda}{\theta}\mathcal{E}^I_{\mathcal{X}^0}(g,g)\end{aligned}$$

Hence,

$$\frac{1}{\theta^2}\,\mathrm{Var}_{P,[L]\times\mathcal{X}^0}(g) = \mathrm{Var}_{P_0}(g) \leq \frac{3\lambda C_3}{\theta}\mathcal{E}^I_{\mathcal{X}^0}(g,g) + \frac{C_1C_2\left[(2+\lambda)C_3+1\right]}{\theta\phi}\sum_{i=1}^{L} r_i \mathcal{E}_{i,\mathcal{X}^0}(g_i,g_i).$$

Comparing with Lemma 3, we obtain the Poincaré inequality for M

$$\mathrm{Var}_{P,[L]\times\mathcal{X}^0}(g) \leq \max\left\{3\theta C_3,\ \frac{\theta C_1C_2}{\phi(1-\lambda)}\left((2+\lambda)C_3+1\right)\right\}\mathcal{E}_{[L]\times\mathcal{X}^0}(g,g)$$

which concludes the proof of the theorem. □

B.3 Proof for the Mixing Times

We first recall the mixing time bound given in Atchadé et al. [2011] using restricted spectral gaps.

Lemma 6 (Atchadé et al. [2011]). *Let K be a lazy, reversible Markov transition kernel on a state space Ω, with stationary distribution Π. Suppose the initial distribution Π_0 is absolutely continuous with respect to Π, and define*

$$f_0(\omega)\Pi(\mathrm{d}\omega) = \Pi_0(\mathrm{d}\omega).$$

Assume there exist constants $B > 1$ and $q > 2$ such that $\|f_0\|_{\mathcal{L}^q(\Pi)} \leq B$, where $\|\cdot\|_{\mathcal{L}^q(\Pi)}$ denotes the $\mathcal{L}^q$-norm with respect to Π. Let $\varepsilon \in (0,1)$. Further, suppose there exists a measurable subset $\Omega^0 \subseteq \Omega$ such that

$$\Pi(\Omega^0) \geq 1 - \left(\frac{\varepsilon}{20B^2}\right)^{q/(q-2)}.$$

Then, for

$$N \geq \frac{1}{\mathrm{SpecGap}_{\Omega^0}(K)} \log\left(\frac{2B^2}{\varepsilon^2}\right),$$

the total variation distance between the distribution of the Markov chain K after N steps and its stationary distribution Π is at most ε.

Proof of Lemma 2. The first part of Equation (6) follows directly from Lemma 6. To prove the second part of (6), we first note that the TV distance between P^N and P admits the following lower bound

$$\|P^N - P\|_{\mathrm{tv}} = \sum_{i=1}^{L} \int \left|P^N(i, \mathrm{d}x) - P(i, \mathrm{d}x)\right| \geq \int \left|p^N(i,x) - r_i p_i(x)\right| \mathrm{d}x, \quad \text{for all } i \in [L].$$

For each $i \in [L]$, let $r_{i,N} = P^N(i, \mathcal{X})$. The TV distance between P_i^N and P_i is bounded by

$$\begin{aligned}
\|P_i^N - P_i\|_{\mathrm{tv}} &= \left\|r_{i,N}^{-1} P^N(i,\cdot) \; - \; P_i\right\|_{\mathrm{tv}} = \int \left|r_{i,N}^{-1} p^N(i,x) \; - \; p_i(x)\right| \mathrm{d}x \\
&\leq \int \left|r_{i,N}^{-1} p^N(i,x) \; - \; r_i^{-1} p^N(i,x)\right| \mathrm{d}x + \int \left|r_i^{-1} p^N(i,x) \; - \; p_i(x)\right| \mathrm{d}x \\
&\leq \int \left|r_{i,N}^{-1} p^N(i,x) \; - \; r_i^{-1} p^N(i,x)\right| \mathrm{d}x + r_i^{-1} \|P^N - P\|_{\mathrm{tv}},
\end{aligned}$$

where the first inequality follows from the triangle inequality. For the first term in the last expression, using $r_{i,N} = P^N(i, \mathcal{X})$ we get

$$\int \left|r_{i,N}^{-1} p^N(i,x) \; - \; r_i^{-1} p^N(i,x)\right| \mathrm{d}x = r_i^{-1} \left|r_i - r_{i,N}\right| \leq \frac{r_i^{-1}}{2} \|P^N - P\|_{\mathrm{tv}},$$

where in the last step we use $r_i = P(A)$ and $r_{i,N} = P^N(A)$ with $A = \{i\} \times \mathcal{X}$.

Combining the above two displayed inequalities and using the first part of Equation (6), we get

$$\|P_i^N - P_i\|_{\mathrm{tv}} \;\; \leq \;\; \frac{3}{2r_i}\varepsilon \leq \frac{3}{2\min_{k \in [L]} r_k}\varepsilon.$$

This completes the proof. $\square$

C Appendix for Section 3

C.1 Comparison of STMH chain with approximate STMH chain

To compare the STMH chain defined in Definition 4 with the approximate STMH chain in Definition 5, it suffices to compare the stationary density p_i^* with $\widetilde{p}_i$ and transition kernel M_i^* with $\widetilde{M}_i$. For the former, we use Lemma 7.3 of Ge et al. [2018], which shows that varying the temperature is roughly the same as changing the variance of a Gaussian distribution. For the latter, we derive a bound in Lemma 8.

Lemma 7 (Lemma 7.3 of Ge et al. [2018])**.** *Let $0 < \beta \leq 1$, and suppose $w_1, \ldots, w_n > 0$ are weights such that $\sum_{i=1}^n w_i = 1$. Define the density functions*

$$\pi(x) \propto \left(\sum_{i=1}^n w_i \pi_i(x) \right)^\beta \quad \textit{and} \quad \widetilde{\pi}(x) \propto \sum_{i=1}^n w_i \pi_i^\beta(x),$$

where $\pi_1, \ldots, \pi_n$ are component densities. Then,

$$w_{\min} \cdot \widetilde{\pi}(x) \leq \pi(x) \leq \frac{1}{w_{\min}} \cdot \widetilde{\pi}(x),$$

where $w_{\min} := \min_{1 \leq i \leq n} w_i$.

Lemma 8. *For each $i \in [L]$, let M_i^* be the transition kernel defined in Definition 4, with transition density m_i^*. Also, let $\widetilde{M}_i$ be the transition kernel defined in Definition 5, with transition density $\widetilde{m}_i$. Assume that $w_{\min} := \min_{1 \leq j \leq m} w_j > 0$. Then, for all $x \neq y \in \mathbb{R}^d$, the following inequality holds*

$$\widetilde{m}_i(x, y) \ \leq \ \frac{1}{w_{\min}^2} m_i^*(x, y).$$

Proof. Let q denote the symmetric Gaussian proposal density used in the Metropolis–Hastings algorithms M_i^* and $\widetilde{M}_i$. Then, for $x \neq y$, the transition densities are given by

$$m_i^*(x, y) = q(x, y)\, \alpha_i^*(x, y), \quad \widetilde{m}_i(x, y) = q(x, y)\, \widetilde{\alpha}_i(x, y),$$

where

$$\alpha_i^*(x, y) = \min\left\{1, \ \frac{p_i^*(y)}{p_i^*(x)}\right\}, \quad \widetilde{\alpha}_i(x, y) = \min\left\{1, \ \frac{\widetilde{p}_i(y)}{\widetilde{p}_i(x)}\right\}.$$

Using Lemma 7, we have

$$\widetilde{p}_i(y) \leq \frac{1}{w_{\min}} p_i^*(y) \quad \text{and} \quad \widetilde{p}_i(x) \geq w_{\min}\, p_i^*(x),$$

which gives

$$\frac{\widetilde{p}_i(y)}{\widetilde{p}_i(x)} \leq \frac{1}{w_{\min}^2} \cdot \frac{p_i^*(y)}{p_i^*(x)}.$$

Hence,

$$\widetilde{m}_i(x, y) \leq \frac{1}{w_{\min}^2} q(x, y)\, \alpha_i^*(x, y) = \frac{1}{w_{\min}^2} m_i^*(x, y),$$

which completes the proof. □

From now on, we assume that $\mathcal{X}^0 \subseteq \mathbb{R}^d$ is a measurable subset. Our next result, Lemma 9, shows that it suffices to obtain a lower bound on the $[L] \times \mathcal{X}^0$-restricted spectral gap of the approximate STMH chain in order to derive a corresponding bound for the STMH chain.

Lemma 9. *Let M^* be the STMH chain defined in Definition 4, and let $\widetilde{M}$ be the approximate STMH chain defined in Definition 5. Assume that the mixture weights satisfy $w_{\min} := \min_{1 \leq j \leq n} w_j > 0$. Then, the $[L] \times \mathcal{X}^0$-restricted spectral gaps of M^* and $\widetilde{M}$ satisfy the inequality*

$$\mathrm{SpecGap}_{[L] \times \mathcal{X}^0}(\widetilde{M}) \ \leq \ \frac{1}{w_{\min}^5} \mathrm{SpecGap}_{[L] \times \mathcal{X}^0}(M^*).$$

Proof. Let p^* and $\widetilde{p}$ denote the stationary densities of the Markov chains M^* and $\widetilde{M}$, respectively. Then, for all $i \in [L]$ and $x \in \mathbb{R}^d$, we have

$$p^*(i,x) = r_i p_i^*(x) \quad \text{and} \quad \widetilde{p}(i,x) = r_i \widetilde{p}_i(x).$$

By Lemma 7, we have

$$w_{\min}\, \widetilde{p}_i(x) \leq p_i^*(x) \leq w_{\min}^{-1}\, \widetilde{p}_i(x) \tag{14}$$

for every (i,x), and the same inequality holds for $p^*(i,x)$ and $\widetilde{p}(i,x)$ since the weights $(r_i)_{i=1}^L$ are the same for P^* and $\widetilde{P}$. Let $\mathcal{E}^*_{[L]\times\mathcal{X}^0}, \widetilde{\mathcal{E}}_{[L]\times\mathcal{X}^0}$ denote the $[L]\times\mathcal{X}^0$-restricted Dirichlet forms associated with M^* and $\widetilde{M}$ respectively. Fix a function $g \in \mathcal{L}^2([L]\times\mathcal{X}, \widetilde{p})$ and define $g_i(x) := g(i,x)$ for each (i,x). By Definition 1, it suffices to show that

$$\mathrm{Var}_{P^*,[L]\times\mathcal{X}^0}(g) \leq \frac{1}{w_{\min}^2} \mathrm{Var}_{\widetilde{P},[L]\times\mathcal{X}^0}(g), \text{ and } \mathcal{E}^*_{[L]\times\mathcal{X}^0}(g,g) \leq \frac{1}{w_{\min}^3} \widetilde{\mathcal{E}}_{[L]\times\mathcal{X}^0}(g,g).$$

For the first inequality, it follows from (14) that

$$\begin{aligned}
\mathrm{Var}_{P^*,[L]\times\mathcal{X}^0}(g) &\leq \frac{1}{2w_{\min}^2} \sum_{i=1}^L \sum_{j=1}^L \int_{\mathcal{X}^0} \int_{\mathcal{X}^0} (g(i,x) - g(j,y))^2\, \widetilde{p}(i,x)\, \widetilde{p}(j,y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \frac{1}{w_{\min}^2} \mathrm{Var}_{\widetilde{P},[L]\times\mathcal{X}^0}(g).
\end{aligned}$$

By Lemma 3, we have

$$\widetilde{\mathcal{E}}_{[L]\times\mathcal{X}^0}(g,g) = (1-\lambda) \sum_{i=1}^L r_i\, \widetilde{\mathcal{E}}_{i,\mathcal{X}^0}(g_i,g_i) + \lambda\, \widetilde{\mathcal{E}}^I_{\mathcal{X}^0}(g,g), \tag{15}$$

and $\mathcal{E}^*_{[L]\times\mathcal{X}^0}$ can be decomposed analogously. We will bound the two terms on the right-hand side of Equation (15) separately. For the first term, we apply Lemmas 7 and 8 to get

$$\begin{aligned}
\widetilde{\mathcal{E}}_{i,\mathcal{X}^0}(g_i,g_i) &= \frac{1}{2} \int_{\mathcal{X}^0} \int_{\mathcal{X}^0} \big(g_i(x) - g_i(y)\big)^2\, \widetilde{p}_i(x)\, \widetilde{M}_i(x, \mathrm{d}y)\, \mathrm{d}x \\
&\leq \frac{1}{2w_{\min}^3} \int_{\mathcal{X}^0} \int_{\mathcal{X}^0} \big(g_i(x) - g_i(y)\big)^2\, p_i(x)\, M_i(x, \mathrm{d}y)\, \mathrm{d}x \\
&= \frac{1}{w_{\min}^3} \mathcal{E}^*_{i,\mathcal{X}^0}(g_i,g_i).
\end{aligned}$$

For the second term, we apply Lemma 7 to get

$$\begin{aligned}
\widetilde{\mathcal{E}}^I_{\mathcal{X}^0}(g,g) &= \frac{1}{4} \sum_{i,i' \in [L]\,:\, i'=i\pm 1} \int_{\mathcal{X}^0} \big(g_i(x) - g_{i'}(x)\big)^2 \min\left\{r_i \widetilde{p}_i(x),\, r_{i'} \widetilde{p}_{i'}(x)\right\}\, \mathrm{d}x \\
&\leq \frac{1}{4\, w_{\min}} \sum_{i,i' \in [L]\,:\, i'=i\pm 1} \int_{\mathcal{X}^0} \big(g_i(x) - g_{i'}(x)\big)^2 \min\left\{r_i p_i(x),\, r_{i'} p_{i'}(x)\right\}\, \mathrm{d}x \\
&= \frac{1}{w_{\min}}\, \mathcal{E}^{*,I}_{\mathcal{X}^0}(g,g).
\end{aligned}$$

Combining both bounds, we obtain that $\mathcal{E}^*_{[L]\times\mathcal{X}^0}(g,g) \leq w_{\min}^{-3} \widetilde{\mathcal{E}}_{[L]\times\mathcal{X}^0}(g,g)$, which concludes the proof. $\square$

C.2 Restricted Spectral Gap of the Approximate STMH Chain

We begin by introducing some notation. For each $i \in [L]$ and $j \in [n]$, define the j-th component of the density $\widetilde{p}_i$ as

$$\widetilde{p}_{(i,j)}(x) \propto \exp\left\{-\frac{\beta_i}{2}(x-\mu_j)^\top \Sigma^{-1}(x-\mu_j)\right\}, \tag{16}$$

so that $\widetilde{p}_i$ is a weighted mixture of the $\widetilde{p}_{(i,j)}$'s:

$$\widetilde{p}_i(x) \propto \sum_{j=1}^{n} w_j \widetilde{p}_{(i,j)}(x).$$

Let $\widetilde{M}_{(i,j)}$ denote the Metropolis–Hastings transition kernel targeting $\widetilde{p}_{(i,j)}$, with a symmetric Gaussian proposal density $q(x,y) = \mathcal{N}(y; x, \eta I)$, where $\eta > 0$ is the step size. We set

$$\mathcal{X}^0 := \left\{ x \in \mathbb{R}^d \ : \ \|x\| \leq R \right\},$$

where $R > 0$ is a fixed radius. To obtain a lower bound on the $[L] \times \mathcal{X}^0$-restricted spectral gap of the approximate STMH chain, we invoke Theorem 1. Assumption 1 holds by our construction of $\widetilde{P}$. The following lemmas verify the other two assumptions required for this theorem.

C.2.1 Validation of Condition (2) in Assumption 2

Lemma 10. *For each $i \in [L]$, let $\widetilde{p}_i$ be the density defined in Equation (9), and let $g_i \in \mathcal{L}^2(\mathcal{X}, \widetilde{p}_i)$. Then the following inequality holds*

$$\sum_{j=1}^{n} w_j \, \widetilde{\mathcal{E}}_{(i,j),\mathcal{X}^0}(g_i, g_i), \ \leq \ \widetilde{\mathcal{E}}_{i,\mathcal{X}^0}(g_i, g_i) \qquad \forall i \in [L],$$

where $\widetilde{\mathcal{E}}_{(i,j),\mathcal{X}^0}$ denotes the $\mathcal{X}^0$-restricted Dirichlet form of the kernel $\widetilde{M}_{(i,j)}$, and $\widetilde{\mathcal{E}}_{i,\mathcal{X}^0}$ denotes the $\mathcal{X}^0$-restricted Dirichlet form of the kernel $\widetilde{M}_i$, as defined in Definition 5.

In particular, for the approximate STMH chain $\widetilde{M}$ defined in Definition 5, condition (2) holds with constant $C_1 = 1$.

Proof. For any nonnegative real sequences $\{a_j\}$ and $\{b_j\}$, we have the inequality $\min\left\{\sum_j a_j, \ \sum_j b_j\right\} \ \geq \ \sum_j \min\{a_j, \ b_j\}$. Applying this to $\widetilde{p}_i = \sum_{j=1}^{n} w_j \widetilde{p}_{(i,j)}$, we obtain

$$\min\{\widetilde{p}_i(x), \widetilde{p}_i(z)\} \ \geq \ \sum_{j=1}^{n} w_j \min\left\{\widetilde{p}_{(i,j)}(x), \widetilde{p}_{(i,j)}(z)\right\}, \tag{17}$$

for all $x, z \in \mathbb{R}^d$ and $i \in [L]$. Let q denote the symmetric Gaussian proposal density associated with the Metropolis–Hastings algorithms $\widetilde{M}_i$ and $\widetilde{M}_{(i,j)}$. Then, applying Equation (17), we obtain

$$\begin{aligned}
\widetilde{\mathcal{E}}_{i,\mathcal{X}^0}(g_i, g_i) &= \frac{1}{2} \int_{\mathcal{X}^0} \int_{\mathcal{X}^0} (g_i(x) - g_i(z))^2 \, q(x,z) \min\{\widetilde{p}_i(x), \widetilde{p}_i(z)\} \, \mathrm{d}x \, \mathrm{d}z \\
&\geq \frac{1}{2} \int_{\mathcal{X}^0} \int_{\mathcal{X}^0} (g_i(x) - g_i(z))^2 \, q(x,z) \sum_{j=1}^{n} w_j \min\left\{\widetilde{p}_{(i,j)}(x), \widetilde{p}_{(i,j)}(z)\right\} \, \mathrm{d}x \, \mathrm{d}z \\
&= \sum_{j=1}^{n} w_j \widetilde{\mathcal{E}}_{(i,j),\mathcal{X}^0}(g_i, g_i).
\end{aligned}$$

This completes the proof of the Lemma. □

C.2.2 Validation of Condition (3) in Assumption 2

We lower bound the $\mathcal{X}^0$-restricted spectral gap of each Metropolis–Hastings chain $\widetilde{M}_{(i,j)}$ using the path method of Yuen [2000] in the following lemma.

Lemma 11. *Let $0 < \eta \leq R^2$. For each $i \in [L]$ and $j \in [n]$, the Markov chain $\widetilde{M}_{(i,j)}$ admits the following lower bound on its $\mathcal{X}^0$-restricted spectral gap*

$$\mathrm{SpecGap}_{\mathcal{X}^0}\big(\widetilde{M}_{(i,j)}\big) \ \geq \ \frac{\gamma_{\min}^{d/2} \eta^{3/2}}{13 R^{d+3}}.$$

In particular, for the approximate STMH chain $\widetilde{M}$ defined in Definition 5, condition (3) *holds with constant*

$$C_2 = \frac{13R^{d+3}}{\gamma_{\min}^{d/2}}.$$

Proof. We use the linear path method described in Section 2 of Yuen [2000]. This approach also extends to the restricted spectral gap setting; see, for example, Atchadé [2021] and Chang and Zhou [2024] where the canonical path method has been adapted to the restricted spectral gap in discrete spaces. For each pair $(x, y) \in \mathcal{X}^0 \times \mathcal{X}^0$, we construct a linear path connecting x to y, with all intermediate points lying in $\mathcal{X}^0$. Fix a step size $\delta > 0$, and define the number of steps along the path by

$$b_{xy} := \left\lceil \frac{\|x-y\|}{\delta} \right\rceil .$$

The path is then given by

$$\gamma_{xy} = \left(\gamma_{xy}^{(0)}, \dots, \gamma_{xy}^{(b_{xy})}\right),$$

where

$$\gamma_{xy}^{(i)} := \frac{(b_{xy}-i)x + iy}{b_{xy}}, \quad \text{for } 0 \le i \le b_{xy}.$$

Let $\Gamma := \{\gamma_{xy} : (x, y) \in \mathcal{X}^0 \times \mathcal{X}^0\}$ denote the collection of all such paths, and let E denote the set of all edges that appear in at least one path $\gamma_{xy} \in \Gamma$. The capacity of an edge $(u, v) \in E$ is given by

$$T_{(i,j)}(u, v) := \widetilde{p}_{(i,j)}(u)\, \widetilde{m}_{(i,j)}(u, v) = \widetilde{p}_{(i,j)}(u)\, q(u, v) \min\left\{1,\ \frac{\widetilde{p}_{(i,j)}(v)}{\widetilde{p}_{(i,j)}(u)}\right\},$$

where $\widetilde{m}_{(i,j)}$ denotes the transition density corresponding to the kernel $\widetilde{M}_{(i,j)}$, and $q(u, v)$ is the Gaussian proposal density associated with kernel $\widetilde{M}_{(i,j)}$. As shown in Section 2 of Yuen [2000], the set of paths Γ satisfies the regularity conditions and, for any $(u, v) \in \gamma_{xy}$, the associated Jacobian satisfies $J_{x,y}(u, v) = b_{xy}^d$ (see [Yuen, 2000, page 5] for details). Then, by Theorem 2.1 and Corollary 2.4 in Yuen [2000], we have

$$\mathrm{SpecGap}_{\mathcal{X}^0}\big(\widetilde{M}_{(i,j)}\big) \ge \frac{1}{A} \tag{18}$$

where

$$A = \operatorname*{ess\,sup}_{(u,v)\in E} \left\{ \frac{1}{T_{(i,j)}(u, v)} \sum_{\gamma_{xy} \ni (u,v)} |\gamma_{xy}|\, \widetilde{p}_{(i,j)}(x)\, \widetilde{p}_{(i,j)}(y)\, b_{xy}^d \right\},$$

and $|\gamma_{xy}|$ denotes the length of the path γ_{xy}. Since $\tilde{p}_{(i,j)}$ is log-concave, for any $(u, v) \in \gamma_{xy}$, we have

$$\min\left\{\widetilde{p}_{(i,j)}(x),\, \widetilde{p}_{(i,j)}(y)\right\} \le \min\left\{\widetilde{p}_{(i,j)}(u),\, \widetilde{p}_{(i,j)}(v)\right\}.$$

Hence, $T_{(i,j)}(u, v) \ge q(u, v) \min\{\widetilde{p}_{(i,j)}(x),\, \widetilde{p}_{(i,j)}(y)\}$, and we can upper bound A as

$$A \le b^{d+1} \cdot \operatorname*{ess\,sup}_{(u,v)\in E} \left\{ q(u, v)^{-1} \sum_{\gamma_{x,y} \ni (u,v)} \widetilde{p}_{(i,j)}(z_{x,y}) \right\}, \tag{19}$$

where $b := \max_{(x,y)\in\mathcal{X}^0\times\mathcal{X}^0} b_{x,y}$ and $z_{x,y}$ is defined as

$$z_{x,y} := \begin{cases} x, & \text{if } \max\{\widetilde{p}_{(i,j)}(x), \widetilde{p}_{(i,j)}(y)\} = \widetilde{p}_{(i,j)}(x), \\ y, & \text{otherwise.} \end{cases}$$

Note that

$$\widetilde{p}_{(i,j)}(z_{x,y}) \ \le\ \frac{\beta_i^{d/2}}{(2\pi\gamma_{\min})^{d/2}},$$

and the proposal density $q(u, v)$ is given by

$$q(u, v) \ =\ \frac{1}{(2\pi\eta)^{d/2}} \exp\left(-\frac{\|v-u\|^2}{2\eta}\right).$$

Substituting this into Equation (19), we obtain

$$A \;\le\; \frac{\beta_i^{d/2}\eta^{d/2}b^{d+3}}{\gamma_{\min}^{d/2}} \cdot \operatorname*{ess\,sup}_{(u,v)\in E}\left\{\exp\left(\frac{\|v-u\|^2}{2\eta}\right)\right\},$$

where we have also used that an edge (u,v) belongs to at most b^2 paths in Γ. Choose step size $\delta = \sqrt{5\eta}$, which yields

$$b \le \left\lceil \frac{2R}{\sqrt{5\eta}} \right\rceil \le \frac{R}{\sqrt{\eta}}.$$

Since $\beta_i \le 1$, we obtain that

$$A \;\le\; \frac{\beta_i^{d/2}R^{d+3}}{\gamma_{\min}^{d/2}\eta^{3/2}}e^{2.5} \le \frac{13R^{d+3}}{\gamma_{\min}^{d/2}\eta^{3/2}}.$$

From Equation (18), we get

$$\mathrm{SpecGap}_{\mathcal{X}^0}\big(\widetilde{M}_{(i,j)}\big) \;\ge\; \frac{1}{A} \;\ge\; \frac{\gamma_{\min}^{d/2}\eta^{3/2}}{13R^{d+3}},$$

which concludes the proof. □

C.2.3 Auxiliary Lemmas

To verify Assumption 3 and compute the constant C_3 in condition (4), we will need several lemmas. The proof of Lemma 12 is omitted.

Lemma 12 (Canonical Paths Bound)**.** *Let $\mathcal{S}$ be a finite state space, and let K be the transition kernel of a reversible Markov chain on $\mathcal{S}$ with stationary distribution Π and Dirichlet form $\mathcal{E}$. For each pair of distinct states $x, y \in \mathcal{S}$, let γ_{xy} denote a path from x to y consisting of valid transitions under K, i.e.,*

$$x = x_0 \to x_1 \to x_2 \to \cdots \to x_{n-1} \to x_n = y.$$

Let $\Gamma = \{\gamma_{xy} : x, y \in \mathcal{S},\ x \ne y\}$ be the collection of such paths for all distinct pairs (x,y). The edge congestion associated with Γ is defined as

$$\rho_e(\Gamma) = \max_{\substack{u,v\in\mathcal{S}\\ K(u,v)>0}} \frac{1}{\Pi(u)K(u,v)} \sum_{\substack{(u,v)\in\gamma_{xy}\\ \gamma_{xy}\in\Gamma}} \Pi(x)\Pi(y)|\gamma_{xy}|,$$

where $|\gamma_{xy}|$ denotes the length of the path γ_{xy}. Then, for any function $g : \mathcal{S} \to \mathbb{R}$, the following Poincaré inequality holds

$$\mathrm{Var}_\Pi(g) \le \rho_e(\Gamma)\,\mathcal{E}(g,g).$$

Lemma 13. *Let $\Pi, \tilde{\Pi}$ be two probability distributions (absolutely continuous with respect to each other) with density function $\pi, \tilde{\pi}$ respectively. Then,*

$$\int \min\{\pi(x), \tilde{\pi}(x)\}\,\mathrm{d}x = 1 - \frac{1}{2}\|\Pi - \tilde{\Pi}\|_{\mathrm{tv}} \ge 1 - \sqrt{\frac{1}{2}\mathrm{KL}(\Pi \,|\, \tilde{\Pi})}.$$

Lemma 14. *Let $|\Sigma|$ denote the determinant of a matrix Σ. The Kullback-Leiber divergence between two d-dimensional Gaussian distributions with equal means is given by*

$$\mathrm{KL}\left(N(\mu,\Sigma_1)\,|\,N(\mu,\Sigma_2)\right) = \frac{1}{2}\left\{\log\frac{|\Sigma_2|}{|\Sigma_1|} - d + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1)\right\}.$$

Lemma 15. *Let $D := \max\left\{\max_{k\in[n]}\|\mu_k\|,\ \sqrt{\gamma_{\min}}\right\}$. For each $i \in [L]$, $j \in [n]$, define*

$$\widetilde{p}_i(j\,|\,x) := \frac{w_j\,\widetilde{p}_{(i,j)}(x)}{\sum_{k=1}^n w_k\,\widetilde{p}_{(i,k)}(x)} \qquad \text{for all } x \in \mathbb{R}^d,$$

where $\widetilde{p}_{(i,j)}(x)$ denotes the density defined in Equation (16). Then, for all $i \in [L]$, $j \in [n]$, and $x \in \mathbb{R}^d$, the following inequalities hold

$$\widetilde{p}_{(i,j)}(x) \ge \left(\frac{\beta_i}{2\pi\gamma_{\max}}\right)^{d/2}\exp\left(-\frac{\beta_i}{2\gamma_{\min}}(\|x\|+D)^2\right), \tag{20}$$

$$\widetilde{p}_i(j\,|\,x) \ge w_j\exp\left(-\frac{\beta_i}{\gamma_{\min}}(\|x\|+D)^2\right). \tag{21}$$

Proof. To prove Equation (20), we write

$$\widetilde{p}_{(i,j)}(x) \geq \left(\frac{\beta_i}{2\pi\gamma_{\max}}\right)^{d/2} \exp\left(-\frac{\beta_i}{2}(x-\mu_j)^\top \Sigma^{-1}(x-\mu_j)\right),$$

where the inequality follows from the bound $|\Sigma| \leq \gamma_{\max}^d$. Next, we use the inequality

$$\|(x-\mu_j)^\top \Sigma^{-1}(x-\mu_j)\| \leq \frac{1}{\gamma_{\min}}\|x-\mu_j\|^2 \leq \frac{1}{\gamma_{\min}}(\|x\|+\|\mu_j\|)^2 \leq \frac{1}{\gamma_{\min}}(\|x\|+D)^2$$

to obtain

$$\widetilde{p}_{(i,j)}(x) \geq \left(\frac{\beta_i}{2\pi\gamma_{\max}}\right)^{d/2} \exp\left(-\frac{\beta_i}{2\gamma_{\min}}(\|x\|+D)^2\right).$$

This establishes Equation (20). To prove Equation (21), we define the function $\widetilde{J} : \mathbb{R}^d \to [n]$ by

$$\widetilde{J}(x) := \arg\max_{k\in[n]} \widetilde{p}_{(i,k)}(x).$$

It follows that for every $k \in [n]$,

$$\widetilde{p}_{(i,k)}(x) \leq \widetilde{p}_{(i,\widetilde{J}(x))}(x),$$

and therefore,

$$\sum_{k=1}^{n} w_k\, \widetilde{p}_{(i,k)}(x) \leq \sum_{k=1}^{n} w_k\, \widetilde{p}_{(i,\widetilde{J}(x))}(x) = \widetilde{p}_{(i,\widetilde{J}(x))}(x).$$

Substituting this upper bound into the definition of $\widetilde{p}_i(j \,|\, x)$, we get

$$\widetilde{p}_i(j \,|\, x) \geq \frac{w_j \widetilde{p}_{(i,j)}(x)}{\widetilde{p}_{(i,\widetilde{J}(x))}(x)}. \tag{22}$$

To simplify the ratio of Gaussian densities on the right-hand side, we expand the expression explicitly as

$$\frac{\widetilde{p}_{(i,j)}(x)}{\widetilde{p}_{(i,\widetilde{J}(x))}(x)} = \exp\left\{-\beta_i(\mu_{\widetilde{J}(x)}-\mu_j)^\top \Sigma^{-1}x - \frac{\beta_i}{2}\left(\mu_j^\top \Sigma^{-1}\mu_j - \mu_{\widetilde{J}(x)}^\top \Sigma^{-1}\mu_{\widetilde{J}(x)}\right)\right\}.$$

By definition of D, we have $\|\mu_{\widetilde{J}(x)} - \mu_j\| \leq 2D$. Using the Cauchy–Schwarz inequality, we obtain

$$\|(\mu_{\widetilde{J}(x)}-\mu_j)^\top \Sigma^{-1}x\| \leq \|\mu_{\widetilde{J}(x)}-\mu_j\| \cdot \|\Sigma^{-1}x\| \leq \frac{2D}{\gamma_{\min}}\|x\|,$$

and similarly,

$$\left\|\mu_j^\top \Sigma^{-1}\mu_j - \mu_{\widetilde{J}(x)}^\top \Sigma^{-1}\mu_{\widetilde{J}(x)}\right\| \leq \frac{1}{\gamma_{\min}}\left(\|\mu_j\|^2 + \|\mu_{\widetilde{J}(x)}\|^2\right) \leq \frac{2D^2}{\gamma_{\min}}.$$

Putting these together, we get

$$\frac{\widetilde{p}_{(i,j)}(x)}{\widetilde{p}_{(i,\widetilde{J}(x))}(x)} \geq \exp\left(-\frac{\beta_i}{\gamma_{\min}}(2D\|x\|+D^2)\right) \geq \exp\left(-\frac{\beta_i}{\gamma_{\min}}(\|x\|+D)^2\right). \tag{23}$$

Equations (22) and (23) together prove Equation (21). This completes the proof. □

C.2.4 Validation of Condition (4) in Assumption 3

Let $\widehat{M}$ denote the projected chain associated with the approximate STMH chain $\widetilde{M}$. We next establish a lower bound on the spectral gap of $\widehat{M}$ using the canonical paths method. Recall that we define $\mathcal{X}^0 = \{x \in \mathbb{R}^d : \|x\| \leq R\}$.

Lemma 16. *Let $\widetilde{M}$ denote the approximate STMH chain defined in Definition 5. Define the following parameters*

$$D := \max\left\{\max_{k\in[n]}\|\mu_k\|,\ \sqrt{\gamma_{\min}}\right\}, \quad r := \frac{\min_{i\in[L]} r_i}{\max_{i\in[L]} r_i}.$$

Suppose the following conditions hold

(i) Let $R \geq \sqrt{d}D$ be such that $P_{(i,j)}(\mathcal{X}^0) \geq 3/4$ for all $i \in [L]$ and $j \in [n]$,

(ii) $\beta_1 = \Theta\left(\gamma_{\min}/D^2\right)$ and $\beta_1 \leq 1$,

(iii) $\beta_{i+1}/\beta_i \leq 1 + 1/\sqrt{d}$ for all $i \in [L-1]$.

Let $\widehat{M}$ be the projected chain defined in Definition 3 associated with $\widetilde{M}$. Under these conditions, $\widehat{M}$ satisfies the spectral gap bound

$$\mathrm{SpecGap}(\widehat{M}) \ \geq \ \frac{3\min\{(1-\lambda),\lambda\}\, r^2}{64L^2\kappa^{d/2}\exp(cd)},$$

where $c > 0$ is a fixed constant. In particular, for the approximate STMH chain $\widetilde{M}$, condition (4) *holds with constant*

$$C_3 = \frac{64L^2\kappa^{d/2}\exp(cd)}{3\min\{(1-\lambda),\lambda\}\, r^2}.$$

Proof. We construct the canonical paths as follows. Fix two arbitrary states $x = (i,j)$, $y = (i',j') \in [L] \times [n]$ with $i \leq i'$.

(a) If $j = j'$, let γ_{xy} be $(i,j) \to (i+1,j) \to \ldots \to (i',j)$.

(b) If $j \neq j'$, let γ_{xy} be $(i,j) \to (i-1,j) \to \ldots \to (1,j) \to (1,j') \to (2,j') \to \ldots \to (i',j')$.

Define γ_{yx} as the reverse of γ_{xy}. Let Γ denote the collection of such paths over all distinct pairs (x,y). Let $i \in [L]$, $j \in [n]$, and $i' = i \pm 1 \in [L]$. From the definition of $\widehat{M}$, we have

$$\widehat{M}((i,j),(i',j)) = \frac{\lambda}{2}\int_{\mathcal{X}^0} \frac{\widetilde{p}_{(i,j)}(x)}{\widetilde{P}_{(i,j)}(\mathcal{X}^0)} \cdot \widetilde{a}\big((i,j,x),(i',j,x)\big)\,\mathrm{d}x,$$

where the acceptance probability is given by

$$\widetilde{a}\big((i,j,x),(i',j,x)\big) = \min\left\{\frac{r_{i'}\,\widetilde{p}_{(i',j)}(x)}{r_i\,\widetilde{p}_{(i,j)}(x)},\ 1\right\}.$$

Hence, the probability of transitioning from state (i,j) to $(i-1,j)$ under the projected chain $\widehat{M}$ is given by

$$\widehat{M}((i,j),(i-1,j)) = \frac{\lambda}{2\widetilde{P}_{(i,j)}(\mathcal{X}^0)}\int_{\mathcal{X}^0} \min\left\{\frac{r_{i-1}}{r_i}\,\widetilde{p}_{(i-1,j)}(x),\ \widetilde{p}_{(i,j)}(x)\right\}\,\mathrm{d}x,$$

for all $i \in \{2,\ldots,L\}$ and $j \in [n]$. Since $r_{i-1}/r_i \geq r$ by definition and $\widetilde{P}_{(i,j)}(\mathcal{X}^0) \leq 1$, we have

$$\widehat{M}((i,j),(i-1,j)) \geq \frac{\lambda r}{2}\int_{\mathcal{X}^0} \min\left\{\widetilde{p}_{(i-1,j)}(x),\ \widetilde{p}_{(i,j)}(x)\right\}\,\mathrm{d}x. \tag{24}$$

By Lemma 13 and Lemma 14,

$$\begin{aligned}\int_{\mathbb{R}^d} \min\left\{\widetilde{p}_{(i-1,j)}(x),\ \widetilde{p}_{(i,j)}(x)\right\}\,\mathrm{d}x &\geq 1 - \sqrt{\frac{1}{2}\mathrm{KL}(\widetilde{P}_{(i,j)} \,|\, \widetilde{P}_{(i-1,j)})}\\ &\geq 1 - \frac{\sqrt{d}}{2}\sqrt{f\left(\frac{\beta_i}{\beta_{i-1}}\right)},\ \text{where } f(x) = x - 1 - \log x.\end{aligned}$$

For $x \geq 1$, we have $f(x) \leq (x-1)^2/2$. Hence, if $\beta_i/\beta_{i-1} - 1 = 1/\sqrt{d}$, then

$$\int_{\mathbb{R}^d} \min\left\{\widetilde{p}_{(i-1,j)}(x),\ \widetilde{p}_{(i,j)}(x)\right\}\,\mathrm{d}x \geq \frac{1}{2}.$$

Condition (i) implies

$$\int_{\mathcal{X}^0} \min\left\{\widetilde{p}_{(i-1,j)}(x),\ \widetilde{p}_{(i,j)}(x)\right\}\,\mathrm{d}x \geq \frac{1}{4}.$$

Substituting the above bound in Equation (24), we obtain

$$\widehat{M}((i,j),(i-1,j)) \geq \frac{\lambda r}{8}.$$

Similarly, we can derive the bound

$$\widehat{M}((i,j),(i+1,j)) \geq \frac{\lambda r}{8}, \qquad \text{for all } i \in [L-1],\ j \in [n].$$

Next, we derive a lower bound on the transition probability from $(1,j)$ to $(1,j')$ in the projected chain $\widehat{M}$, which is given by

$$\widehat{M}((1,j),(1,j')) = (1-\lambda) \int_{\mathcal{X}^0} \frac{\widetilde{p}_{(1,j)}(x)}{\widetilde{P}_{(1,j)}(\mathcal{X}^0)} \cdot \widetilde{p}_1(j' \,|\, x)\, \mathrm{d}x, \qquad j,j' \in [n].$$

By applying Lemma 15 and noting that $\widetilde{P}_{(1,j)}(\mathcal{X}^0) \leq 1$, we obtain

$$\widehat{M}((1,j),(1,j')) \geq (1-\lambda) w_{j'} \left(\frac{\beta_1}{2\pi\gamma_{\max}}\right)^{d/2} \int_{\mathcal{X}^0} \exp\left(-\frac{2\beta_1}{\gamma_{\min}}(\|x\|+D)^2\right) \mathrm{d}x.$$

By condition (ii), there exist fixed constants $c_1, c_2 > 0$ such that

$$c_1 \frac{\gamma_{\min}}{D^2} < \beta_1 < c_2 \frac{\gamma_{\min}}{D^2}.$$

Let $\mathcal{X}^D := \{x = (x_1, \ldots, x_d) \in \mathbb{R}^d : |x_i| \leq D \text{ for all } i \in [d]\} \subseteq \mathcal{X}^0$. Then for any $x \in \mathcal{X}^D$, we have $\|x\| + D \leq 2\sqrt{d}D$, which implies

$$\exp\left(-\frac{2\beta_1}{\gamma_{\min}}(\|x\|+D)^2\right) \geq \exp\left(-\frac{8\beta_1 d D^2}{\gamma_{\min}}\right) \geq \exp(-8c_2 d).$$

Therefore, we obtain the following lower bound

$$\widehat{M}((1,j),(1,j')) \geq (1-\lambda) w_{j'} (c_1)^{d/2} \left(\frac{\gamma_{\min}}{2\pi\gamma_{\max} D^2}\right)^{d/2} \exp(-8c_2 d) \cdot \mathrm{Vol}(\mathcal{X}^D),$$

where $\mathrm{Vol}(\mathcal{X}^D)$ denotes the volume of the cube $\mathcal{X}^D$. Substituting $\mathrm{Vol}(\mathcal{X}^D) = (2D)^d$, we get

$$\begin{aligned}\widehat{M}((1,j),(1,j')) &\geq (1-\lambda) w_{j'} (c_1)^{d/2} \left(\frac{\gamma_{\min}}{2\pi\gamma_{\max} D^2}\right)^{d/2} \exp(-8c_2 d) \cdot (2D)^d \\ &\geq (1-\lambda) w_{j'} \cdot \kappa^{-d/2} \cdot \exp(-cd), \end{aligned} \tag{25}$$

where $c > 0$ is a fixed constant.

Let γ_{xy} be a path between any two vertices $x, y \in [L] \times [n]$. Then, $|\gamma_{xy}| \leq 2L$. We now derive an upper bound on the edge congestion $\rho_e(\Gamma)$ defined in Lemma 12. Let $z, w \in [L] \times [n]$.

(a) Let $z = (i,j)$ and $w = (i-1,j)$. Then the edge (z,w) is used only by paths between vertices x and y such that one lies in the set $S := \{i, \ldots, L\} \times \{j\}$, and the other in S^c. Therefore, its contribution to the edge congestion is bounded by

$$\frac{\sum_{(x,y)\in\Gamma:((i,j),(i-1,j))\in\gamma_{xy}} |\gamma_{xy}|\, \widehat{P}(x)\, \widehat{P}(y)}{\widehat{P}((i,j))\, \widehat{M}((i,j),(i-1,j))} \leq \frac{(2L)\, \widehat{P}(S)\, \widehat{P}(S^c)}{\widehat{P}((i,j))\, \widehat{M}((i,j),(i-1,j))},$$

where

$$\widehat{P}((\ell,k)) = r_\ell w_k \frac{\widetilde{P}_{(\ell,k)}(\mathcal{X}^0)}{\widetilde{P}([L]\times\mathcal{X}^0)}, \qquad \ell \in [L],\ k \in [n]$$

denotes the stationary distribution of the projected chain $\widehat{M}$. We have the following bounds

$$\frac{\widehat{P}(S)}{\widehat{P}((i,j))} = \frac{\widehat{P}(\{i,\ldots,L\}\times\{j\})}{\widehat{P}((i,j))} \leq \frac{4L}{3r}, \qquad \widehat{P}(S^c) \leq 1, \qquad \widehat{M}((i,j),(i-1,j)) \geq \frac{\lambda r}{8},$$

where the first bound follows from condition (i). Combining these, we conclude

$$\frac{\sum_{(x,y)\in\Gamma:((i,j),(i-1,j))\in\gamma_{xy}} |\gamma_{xy}|\, \widehat{P}(x)\, \widehat{P}(y)}{\widehat{P}((i,j))\, \widehat{M}((i,j),(i-1,j))} \leq \frac{64L^2}{3\lambda r^2}.$$

Similarly, we obtain the same bound for $z = (i,j)$ and $w = (i+1,j)$.

(b) Let $z = (1, j)$ and $w = (1, j')$. Then the edge (z, w) is used only by paths between vertices x and y such that one of them lies in the set $[L] \times \{j\}$ and the other in $[L] \times \{j'\}$. Therefore, its contribution to edge congestion is bounded by

$$\frac{\sum_{(x,y)\in\Gamma:((1,j),(1,j'))\in\gamma_{xy}} |\gamma_{x,y}|\, \widehat{P}(x)\, \widehat{P}(y)}{\widehat{P}((1,j))\, \widehat{M}((1,j),(1,j'))} \leq \frac{(2L)\, \widehat{P}([L]\times\{j\})\, \widehat{P}([L]\times\{j'\})}{\widehat{P}((1,j))\, \widehat{M}((1,j),(1,j'))}.$$

We now bound each term on the right-hand side. By condition (i), we have

$$\widehat{P}([L]\times\{j\}) \leq \frac{4L}{3r}\, \widehat{P}((1,j)), \qquad \widehat{P}([L]\times\{j'\}) \leq \frac{4w_{j'}}{3},$$

and from Equation (25), we have

$$\widehat{M}((1,j),(1,j')) \geq (1-\lambda) w_{j'} \cdot \kappa^{-d/2} \cdot \exp(-cd).$$

Combining these, we obtain

$$\frac{\sum_{(x,y):((1,j),(1,j'))\in\gamma_{xy}} |\gamma_{xy}|\, \widehat{P}(x)\, \widehat{P}(y)}{\widehat{P}((1,j))\, \widehat{M}((1,j),(1,j'))} \leq \frac{32L^2\kappa^{d/2}\exp(cd)}{9(1-\lambda)r}.$$

Let $\lambda \in (0,1)$ be a fixed constant. Thus, the edge congestion associated with Γ is bounded by

$$\rho_e(\Gamma) \leq \frac{64L^2\kappa^{d/2}\exp(cd)}{3\min\{(1-\lambda),\lambda\}\, r^2}.$$

From Lemma 12, projected chain $\widehat{M}$ satisfies the Poincaré inequality

$$\mathrm{Var}_{\widehat{P}}(\widehat{g}) \leq \frac{64L^2\kappa^{d/2}\exp(cd)}{3\min\{(1-\lambda),\lambda\}\, r^2}\, \widehat{\mathcal{E}}(\widehat{g},\widehat{g}), \quad \forall\, \widehat{g}\colon [L]\times[n] \to \mathbb{R},$$

where $\widehat{\mathcal{E}}$ denotes the Dirichlet form associated with the projected chain $\widehat{M}$. This completes the proof of the lemma. $\square$

C.2.5 Restricted Spectral Gap Bound

We now invoke Theorem 1 to bound the $[L] \times \mathcal{X}^0$-restricted spectral gap of the approximate STMH chain $\widetilde{M}$, as formalized in the next lemma.

Lemma 17. *Let $\widetilde{M}$ denote the approximate STMH chain, as defined in Definition 5, with λ being a fixed constant. Under the same conditions as in Lemma 16, the $[L] \times \mathcal{X}^0$-restricted spectral gap of $\widetilde{M}$ admits the following lower bound*

$$\mathrm{SpecGap}_{[L]\times\mathcal{X}^0}(\widetilde{M}) \;\geq\; \Omega\left(\frac{\gamma_{\min}^{d/2} r^2 \eta^{3/2}}{R^{d+3} L^2 \kappa^{d/2}\exp(cd)}\right),$$

where $c > 0$ is a fixed constant.

Proof. For the approximate STMH chain $\widetilde{M}$, Assumption 2 is satisfied with constants $C_1 = 1$ and

$$C_2 = \frac{13R^{d+3}}{\gamma_{\min}^{d/2}\eta^{3/2}}.$$

Additionally, Assumption 3 holds with

$$C_3 = \frac{64L^2\kappa^{d/2}\exp(cd)}{3\min\{(1-\lambda),\lambda\}\, r^2},$$

where $c > 0$ is a fixed constant. Since λ is treated as fixed, combining these with Theorem 1 completes the proof of the lemma. $\square$

C.3 Restricted Spectral Gap of the STMH chain

We now establish a lower bound on the $[L] \times \mathcal{X}^0$-restricted spectral gap of the STMH chain M^*, as formalized in next lemma.

Lemma 18. *Let M^* be the STMH chain, as defined in Definition 4, with λ being a fixed constant. Under the same conditions as in Lemma 16, the $[L] \times \mathcal{X}^0$-restricted spectral gap of M^* admits the lower bound*

$$\mathrm{SpecGap}_{[L]\times\mathcal{X}^0}(M^*) \;\geq\; \Omega\left(\frac{w_{\min}^5 \gamma_{\min}^{d/2} r^2 \eta^{3/2}}{R^{d+3} L^2 \kappa^{d/2} \exp(cd)}\right),$$

where $c > 0$ is a fixed constant.

Proof. This follows directly from Lemma 9 and Lemma 17. $\square$

C.4 Estimation of Partition Functions

C.4.1 Assumptions on the Parameters

We now describe how to choose the algorithm parameters—number of temperature levels L, inverse temperature sequence $(\beta_i)_{i=1}^L$, temperature-swap rate λ, proposal step size η, initial covariance matrix Σ_0, number of iterations N—so that the STMH algorithm achieves the asserted time complexity. Recall that $\kappa = \gamma_{\max}/\gamma_{\min}$.

$$L = \Theta\left[\kappa\left\{D^2 + \log w_{\min}^{-1} + d\,(1 + \log\kappa)\right\}\log\left(\frac{D^2}{\gamma_{\min}}\right) + 1\right], \tag{26}$$

$$\beta_1 = \Theta\left(\frac{\gamma_{\min}}{D^2}\right),\ \frac{\beta_{i+1}}{\beta_i} \leq \min\left\{1 + \frac{1}{\sqrt{d}},\ \frac{\gamma_{\min}}{D^2 + 2\gamma_{\max} d\nu}\right\} \text{ for } i \in [L-1], \tag{27}$$

$$\text{where } \nu = 1 + \log\kappa + \frac{2}{d}\log\left(\frac{2}{w_{\min}}\right),$$

$$R = D + \sqrt{d\kappa D^2} + \sqrt{2\kappa D^2 \log\left(\frac{20\,e^6 L^2 \kappa^d}{w_{\min}^2 \varepsilon}\right)}, \tag{28}$$

$$N \geq \frac{C' L^4 R^d \kappa^{d/2} \exp(c'd)}{\gamma_{\min}^{d/2} w_{\min}^5} \log\left(\frac{L^2\kappa^d}{\varepsilon^2 w_{\min}^2}\right),\ \text{for some fixed constants } c', C' > 0, \tag{29}$$

$$\Sigma_0 = \sigma_0^2 I,\ \text{where } \sigma_0^2 = \Theta\left(\frac{\gamma_{\min}}{\beta_1}\right), \tag{30}$$

$$\lambda \text{ is any fixed constant}, \tag{31}$$

$$\eta \geq R^2. \tag{32}$$

C.4.2 Auxiliary Lemmas

Lemma 19. *Let $L > 0$ be an integer. Assume the partition–function estimates $\widehat{Z}_1, \ldots, \widehat{Z}_L$ satisfy*

$$\frac{\widehat{Z}_i/Z_i}{\widehat{Z}_1/Z_1} \in \left[(1 - \tfrac{1}{L})^{i-1},\, (1 + \tfrac{1}{L})^{i-1}\right], \qquad \text{for all } i \in [L]. \tag{33}$$

Define

$$r_i \;:=\; \frac{Z_i/\widehat{Z}_i}{\sum_{k=1}^{L} Z_k/\widehat{Z}_k}, \qquad \text{for all } i \in [L].$$

Then,

$$\frac{e^{-2}}{L} \;\leq\; r_i \;\leq\; \frac{e^2}{L}, \qquad \text{for all } i \in [L]. \tag{34}$$

Moreover,

$$r := \frac{\min_{i\in[L]} r_i}{\max_{i\in[L]} r_i} \geq e^{-4}.$$

Proof. For each $i \in [L]$, define $b_i := Z_i/\widehat{Z}_i$, and denote their sum by S,

$$S := \sum_{k=1}^{L} b_k.$$

Then $r_i = b_i/S$. From Equation (33) we have, for every $i \in [L]$,

$$(1 + \tfrac{1}{L})^{-(i-1)} \;\leq\; \frac{b_i}{b_1} \;\leq\; (1 - \tfrac{1}{L})^{-(i-1)}, \tag{35}$$

which gives

$$L\, b_1\, (1 + \tfrac{1}{L})^{-(L-1)} \;\leq\; S \;\leq\; L\, b_1\, (1 - \tfrac{1}{L})^{-(L-1)}. \tag{36}$$

Combining Equations (35) with (36), we get

$$r_i = \frac{b_i}{S} \;\geq\; \frac{b_1 (L+1)^{-(L-1)}}{L\, b_1\, (L-1)^{-(L-1)}} = \frac{1}{L}\Big(\frac{L-1}{L+1}\Big)^{L-1},$$

and

$$r_i \;\leq\; \frac{b_1 (L-1)^{-(L-1)}}{L\, b_1\, (L+1)^{-(L-1)}} = \frac{1}{L}\Big(\frac{L+1}{L-1}\Big)^{L-1}.$$

Define

$$C_L \;:=\; \Big(\frac{L+1}{L-1}\Big)^{L-1}.$$

Taking logarithms and using $\log(1+x) \leq x$ for all $x > -1$, we obtain

$$\log C_L = (L-1)\, \log\!\Big(1 + \tfrac{2}{L-1}\Big) \;\leq\; (L-1)\Big(\tfrac{2}{L-1}\Big) \;\leq\; 2,$$

so $C_L \leq e^2$. Likewise $C_L^{-1} \geq e^{-2}$. Therefore

$$\frac{e^{-2}}{L} \;\leq\; r_i \;\leq\; \frac{e^2}{L}, \qquad \text{for all } i \in [L],$$

establishing (34). This further implies that $r \geq e^{-4}$, completing the proof of the lemma. □

Lemma 20. *Let $L > 0$ be an integer, and suppose β_1 and σ_0 satisfy Equation* (27) *and Equation* (30)*, respectively. Then the initial density is given by*

$$p^0(1,\cdot) = \mathcal{N}\left(0, \sigma_0^2 I\right), \qquad p^0(i,\cdot) = 0 \quad \text{for all } i \in [L] \setminus \{1\},$$

with the corresponding distribution denoted by P^0. The stationary density of the STMH chain M^*, as defined in Definition 4, is

$$p(i,x) = r_i\, p_i^*(x), \quad i \in [L],\, x \in \mathbb{R}^d,$$

where $(r_i)_{i=1}^L$ are defined in Lemma 19, and the component densities $(p_i^*)_{i\in[L]}$ are given in Equation (8)*. Let P denote the corresponding stationary distribution. Define $f_0 := dP^0/dP$. Then,*

$$\|f_0\|^2_{\mathcal{L}^2(P)} \leq \frac{c_1\, \exp(c_2 d)\, L\, \kappa^{d/2}}{w_{\min}},$$

for some fixed constants $c_1, c_2 > 0$.

Proof. The $\mathcal{L}^2$-norm of f_0 is given by

$$\|f_0\|^2_{\mathcal{L}^2(P)} = \sum_{i=1}^{L} \int p(i,x)\, |f_0(i,x)|^2\, dx = \frac{1}{r_1} \int \frac{(p^0(1,x))^2}{p_1^*(x)}\, dx, \tag{37}$$

where the second equality follows from the fact that P^0 is supported only on $i = 1$. By Lemma 7, for all $x \in \mathbb{R}^d$, we have

$$p_1^*(x) \;\geq\; w_{\min}\, \widetilde{p}_1(x) = w_{\min} \sum_{j=1}^{n} w_j \widetilde{p}_{(1,j)}(x). \tag{38}$$

where $\widetilde{p}_1$ is defined in Equation (9) and $\widetilde{p}_{(1,j)}$ is defined in Equation (16). Substituting (38) into (37), we obtain

$$\|f_0\|^2_{\mathcal{L}^2(P)} \leq \frac{1}{r_1 w_{\min}} \int \frac{(p^0(1,x))^2}{\sum_{j=1}^n w_j \widetilde{p}_{(1,j)}(x)}\, dx.$$

Using the convexity of the χ^2-divergence, we further bound this as

$$\|f_0\|^2_{\mathcal{L}^2(P)} \leq \frac{1}{r_1 w_{\min}} \sum_{j=1}^{n} w_j \int \frac{(p^0(1,x))^2}{\widetilde{p}_{(1,j)}(x)}\, dx. \tag{39}$$

For each $j \in [n]$, the density $\widetilde{p}_{(1,j)}(x)$ can be lower bounded as

$$\widetilde{p}_{(1,j)}(x) \geq \left(\frac{\beta_1}{2\pi\gamma_{\max}}\right)^{d/2} \exp\left(-\frac{\beta_1}{2\gamma_{\min}}\|x-\mu_j\|^2\right). \tag{40}$$

Since $p^0(1,\cdot) \sim \mathcal{N}(0, \sigma_0^2 I)$, we have

$$(p^0(1,x))^2 = \left(\frac{1}{2\pi\sigma_0^2}\right)^{d} \exp\left(-\frac{1}{\sigma_0^2}\|x\|^2\right).$$

From Equation (30), we have fixed constants $0 < s_1, s_2 < 2$ such that

$$s_1 \frac{\gamma_{\min}}{\beta_1} \leq \sigma_0^2 \leq s_2 \frac{\gamma_{\min}}{\beta_1}.$$

This gives

$$(p^0(1,x))^2 \leq \left(\frac{\beta_1}{2\pi s_1 \gamma_{\min}}\right)^{d} \exp\left(-\frac{\beta_1}{s_2\gamma_{\min}}\|x\|^2\right). \tag{41}$$

Substituting Equations (40) and (41) into Equation (39), we obtain

$$\begin{aligned}
&\|f_0\|^2_{\mathcal{L}^2(P)} \\
&\leq \frac{\kappa^{d/2}}{s_1^d\, r_1 w_{\min}} \sum_{j=1}^{n} w_j \left(\frac{\beta_1}{2\pi\gamma_{\min}}\right)^{d/2} \exp\left(\frac{\beta_1}{(2-s_2)\gamma_{\min}}\|\mu_j\|^2\right) \int \exp\left(-\frac{\beta_1(2-s_2)}{2\gamma_{\min} s_2}\left\|x + \frac{s_2\,\mu_j}{2-s_2}\right\|^2\right) \mathrm{d}x \\
&= \frac{\kappa^{d/2}\, s_2^{d/2}}{s_1^d\,(2-s_2)^{d/2}\, r_1 w_{\min}} \sum_{j=1}^{n} w_j \exp\left(\frac{\beta_1}{(2-s_2)\gamma_{\min}}\|\mu_j\|^2\right).
\end{aligned}$$

From Equation (27), we have a fixed constant $s_3 > 0$ such that $\beta_1 \leq s_3\gamma_{\min}/D^2$. Substituting this, we get

$$\|f_0\|^2_{\mathcal{L}^2(P)} \leq \frac{\kappa^{d/2}\, s_2^{d/2}}{s_1^d\,(2-s_2)^{d/2}\, r_1 w_{\min}} \exp\left(\frac{s_3}{2-s_2}\right).$$

By Lemma 19, we have $r_1 \geq 1/(e^2 L)$. Substituting this into the above bound on $\|f_0\|^2_{\mathcal{L}^2(P)}$ proves the lemma. $\square$

Lemma 21. *Let X follow the d-dimensional Gaussian distribution with mean μ and covariance matrix Σ. Denote the largest eigenvalue of Σ by $\|\Sigma\|$. Then,*

$$\mathbb{P}\left(\|X\| \leq \|\mu\| + \sqrt{d\|\Sigma\|} + \sqrt{2\|\Sigma\|\log(1/\varepsilon)}\right) \geq \varepsilon.$$

Proof. Using the standard concentration inequality for Lipschitz functions of Gaussian random vectors, we get

$$\mathbb{P}\left(\|X-\mu\| \geq \mathbb{E}(\|X-\mu\|)+t\right) \leq e^{-t^2/2\|\Sigma\|}.$$

Since $\mathbb{E}(\|X-\mu\|) \leq \sqrt{d\|\Sigma\|}$, we get

$$\mathbb{P}\left(\|X-\mu\| \geq \sqrt{d\|\Sigma\|}+t\right) \leq e^{-t^2/2\|\Sigma\|}.$$

Letting $t=\sqrt{2\gamma_{\max}\log(1/\varepsilon)}$ and applying triangle inequality, we get the asserted bound. □

Lemma 22. *Suppose $1 \leq \ell \leq L$, and let Algorithm 1 be run with the potential function $f(x)$ defined in Equation* (7)*, inverse temperatures $\beta_1 < \cdots < \beta_\ell$, and using the parameters specified in Equations* (27)*,* (28)*,* (29)*,* (30)*,* (31) *and* (32)*. Assume that the partition function estimates $\widehat{Z}_1, \ldots, \widehat{Z}_\ell$ satisfy Equation* (33)*. Let P^N denote the distribution obtained after running Algorithm 1 for N steps, and let P denote its stationary distribution. Then the total variation distance between P and P^N satisfies*

$$\|P-P^N\|_{\mathrm{tv}} \;\leq\; \varepsilon.$$

Proof. Under the assumptions of the lemma, and by Lemmas 18 and 19, the $[\ell]\times\mathcal{X}^0$-restricted spectral gap of M^* satisfies

$$\operatorname{SpecGap}_{[\ell]\times\mathcal{X}^0}(M^*) \;\geq\; \Omega\left(\frac{w_{\min}^5\,\gamma_{\min}^{d/2}}{R^d\,\ell^4\,\kappa^{d/2}\,\exp(cd)}\right),$$

where $c>0$ is a fixed constant. Moreover, from Lemma 20, $\|f_0\|^2_{\mathcal{L}^2(P)}$ is bounded above by B, where

$$B=\frac{c_1\,\exp(c_2 d)\ell\kappa^{d/2}}{w_{\min}},$$

and $c_1, c_2>0$ are fixed constants. Applying Lemma 6 with the above parameters yields the desired total variation bound. □

Lemma 23. *Assume the same conditions and notations as in Lemma 22. Let P_ℓ^* denote the marginal stationary distribution at temperature level ℓ, with density $p_\ell^*(x) \propto \exp(-\beta_\ell f(x))$, and let P_ℓ^N denote the marginal distribution at level ℓ after running Algorithm 1 for N steps, with density $p_\ell^N(x) \propto P^N(\ell,x)$. Then the total variation distance between P_ℓ^* and P_ℓ^N is bounded by*

$$\|P_\ell^*-P_\ell^N\|_{\mathrm{tv}} \;\leq\; \frac{3e^2\ell}{2}\,\varepsilon.$$

Proof. By Lemma 19, we have $\min_{i\in[\ell]} r_i \geq 1/(e^2\ell)$. The proof now follows directly from Lemmas 2 and 22. □

In the following lemma, we analyze how many times Algorithm 1 must be re-run, with a fixed number of steps N, in order to obtain a sample from the desired temperature level.

Lemma 24. *Suppose the partition function estimates $\widehat{Z}_1, \ldots, \widehat{Z}_\ell$ satisfy Equation* (33)*. Let $I_N \in [\ell]$ denote the temperature index of the state returned after running Algorithm 1 for N steps. Suppose the algorithm is run independently T times, each for N steps. Then, for any fixed temperature level $k \in [\ell]$, if*

$$T \;\geq\; e^2\ell\log\left(\tfrac{1}{\alpha}\right), \qquad \alpha\in(0,1),$$

the probability that at least one of the T runs returns a sample from level k satisfies

$$\mathsf{P}\left(\exists\, t\in[T] \text{ such that } I_N^{(t)}=k\right) \;\geq\; 1-\alpha,$$

where $I_N^{(t)}$ is the temperature level returned in the t-th run.

Proof. Let $k \in [\ell]$. From Lemma 19, we have

$$\mathsf{P}(I_N \neq k) \;=\; 1 - \mathsf{P}(I_N = k) \;\leq\; 1 - \frac{1}{e^2\,\ell}.$$

Hence,

$$\mathsf{P}\left(\nexists\, t \in [T] \text{ such that } I_N^{(t)} = k\right) \leq \left(1 - \tfrac{1}{e^2\,\ell}\right)^T \;\leq\; \exp\!\left(-\tfrac{T}{e^2\,\ell}\right).$$

Setting this upper bound no larger than δ, and solving for T, completes the proof of the lemma. □

Assuming the partition function estimates satisfy Equation (33), we have shown that the algorithm reaches total variation distance at most ε within the time complexity specified in Equation (29). We now show that partition function estimates satisfy Equation (33). By combining these two components, we establish the overall time complexity for the complete algorithm.

Lemma 25. *Let $\delta \in (0,1)$ and $1 \leq \ell \leq L$. Suppose the parameters satisfy Equations* (26), (27), (28), (30), (31), *and* (32), *and assume that the partition function estimates $\widehat{Z}_1, \ldots, \widehat{Z}_\ell$ satisfy Equation* (33). *Let $s = L^2 \log(1/\delta)$. Collect s samples from Algorithm 1, denoted by $(x_j)_{j=1}^s$. Define the next partition function estimate $\widehat{Z}_{\ell+1}$ as*

$$\widehat{Z}_{\ell+1} := \overline{r}\,\widehat{Z}_\ell, \quad \textit{where} \quad \overline{r} := \frac{1}{s}\sum_{j=1}^{s} \exp\!\big(-(\beta_{\ell+1} - \beta_\ell) f(x_j)\big).$$

Then, with probability at least $1-\delta$, the estimate $\widehat{Z}_{\ell+1}$ also satisfies Equation (33). *In particular,*

$$\left|\frac{\widehat{Z}_{\ell+1}/Z_{\ell+1}}{\widehat{Z}_1/Z_1}\right| \in \left[\left(1 - \frac{1}{L}\right)^{\ell}, \left(1 + \frac{1}{L}\right)^{\ell}\right].$$

The proof of Lemma 25 requires the following results.

Lemma 26 (Lemma 9.1 of Ge et al. [2018])**.** *Suppose that P_1 and P_2 are probability measures on Ω with density functions (with respect to a reference measure)*

$$p_1(x) = \frac{g_1(x)}{Z_1}, \quad \textit{and} \quad p_2(x) = \frac{g_2(x)}{Z_2}.$$

Suppose $\widetilde{P}_1$ is a measure such that $\|\widetilde{P}_1 - P_1\|_{tv} < c/2C^2$, and $g_2(x)/g_1(x) \in [0, C]$ for all $x \in \Omega$. Given n samples $x_1, \ldots, x_n$ from $\widetilde{P}_1$, define the random variable

$$\overline{r} = \frac{1}{n}\sum_{i=1}^{n} \frac{g_2(x_i)}{g_1(x_i)}.$$

Let

$$r = \mathsf{E}_{x\sim P_1} \frac{g_2(x)}{g_1(x)} = \frac{Z_2}{Z_1}.$$

and suppose $r \geq 1/C$. Then with probability at least $1 - e^{-nc^2/(2C^4)}$,

$$\left|\frac{\overline{r}}{r} - 1\right| \leq c.$$

Lemma 27 (Lemma G.16 of Ge et al. [2018])**.** *Suppose that $f(x) = -\log\left[\sum_{i=1}^n w_i\, e^{-f_i(x)}\right]$, where $f_i(x) = f_0(x - \mu_i)$, and $f_0 \colon \mathbb{R}^d \to \mathbb{R}$ is a κ-strongly convex and K-smooth function. For any $a > 0$, let P_a denote the probability measure with density $p_a(x) \propto e^{-af(x)}$. Let Z_a be the corresponding normalization constant, given by $Z_a = \int_{\mathbb{R}^d} e^{-af(x)}\,\mathrm{d}x$. Suppose that $\|\mu_i\| \leq D$ for all $i \in [n]$, and let $\alpha, \beta > 0$. Let*

$$A = D + \frac{1}{\sqrt{\alpha\kappa}}\left(\sqrt{d} + \sqrt{d\log\left(\frac{K}{\kappa}\right) + 2\log\left(\frac{2}{w_{\min}}\right)}\right).$$

If $\alpha < \beta$, then

$$\min_{x\in\mathbb{R}^d} \frac{p_\alpha(x)}{p_\beta(x)} \geq \frac{Z_\beta}{Z_\alpha} \qquad \textit{and} \qquad \frac{Z_\beta}{Z_\alpha} \in \left[\frac{1}{2}e^{-\frac{1}{2}(\beta-\alpha)KA^2}, 1\right].$$

Proof of Lemma 25. By Equation (27) and Lemma 27, we have

$$\frac{\exp(-\beta_{\ell+1} f(x))}{\exp(-\beta_\ell f(x))} = \exp\left(-(\beta_{\ell+1} - \beta_\ell) f(x)\right) \in [0, 1/(2\,e)]$$

for all $\ell \in [L-1]$. Moreover, by substituting $\varepsilon = 4/(3\ell L)$ into Lemma 23, we obtain

$$\|P_\ell^* - P_\ell^{\widetilde{N}}\|_{\mathrm{tv}} \;\le\; \frac{2e^2}{L},$$

when

$$\widetilde{N} \ge \frac{C' L^4 R^d \kappa^{d/2} \exp(c'd)}{\gamma_{\min}^{d/2} w_{\min}^5} \log\left(\frac{L^4 \kappa^d}{w_{\min}^2}\right),$$

where $C', c' > 0$ are fixed constants. Next, by applying Lemma 26 with constants $C = 1/2e$ and $c = 1/L$, we obtain the following bound

$$\left|\frac{\widehat{Z}_{\ell+1}/Z_{\ell+1}}{\widehat{Z}_\ell/Z_\ell}\right| \in \left[1 - \frac{1}{L},\ 1 + \frac{1}{L}\right].$$

The lemma then follows by induction on ℓ. □

C.4.3 Proof of Theorem 2

Proof of Theorem 2. Let L denote the number of temperature levels defined in Equation (26). By applying Lemma 25 inductively with $\delta = \varepsilon/(4L)$, we obtain that, with probability at least $1 - \varepsilon/4$, the following bound holds

$$\frac{\widehat{Z}_\ell}{Z_\ell} \in \left[\left(1 - \frac{1}{L}\right)^{\ell-1}, \left(1 + \frac{1}{L}\right)^{\ell-1}\right] \cdot \frac{\widehat{Z}_1}{Z_1} \qquad \text{for all } \ell \in [L].$$

To ensure this guarantee, it suffices to generate $s = L^2 \log(4L/\varepsilon)$ samples from each temperature level $i \in [L]$, resulting in a total of $sL = L^3 \log(4L/\varepsilon)$ samples from Algorithm 1. Applying Lemma 24 with $\alpha = \varepsilon/(4L^4 \log(4L/\varepsilon))$, we obtain that, with probability at least $1 - \varepsilon/4$, we obtain s samples from each temperature level $i \in [L]$ by running Algorithm 1 for N steps (as defined in Equation (29)) and repeating this process independently T times, where

$$T = sL \cdot e^2 L \log\left(\frac{1}{\alpha}\right) = e^2 L^4 \log\left(\frac{4L}{\varepsilon}\right) \log\left(\frac{4L^4}{\varepsilon} \log\left(\frac{4L}{\varepsilon}\right)\right).$$

Hence, the total time complexity for getting partition function estimates is

$$T_{\text{partition}} = T \cdot N = \frac{C'\, L^8 R^d \kappa^{d/2} \exp(c'd)}{\gamma_{\min}^{d/2} w_{\min}^5} \log^3\left(\frac{L\kappa}{\varepsilon w_{\min}}\right),$$

where $c', C' > 0$ are fixed constants. By applying Lemma 23 and Lemma 24, we conclude that, with probability at least $1 - \varepsilon/4$, Algorithm 1 produces a sample from a distribution that is within total variation distance $\varepsilon/4$ of the target distribution P^* in time T_{sampling}, where

$$\begin{aligned} T_{\text{sampling}} &= e^2\, L\, \log\left(\frac{4}{\varepsilon}\right) \frac{C'' L^4 R^d \kappa^{d/2} \exp(c''d)}{\gamma_{\min}^{d/2} w_{\min}^5} \log\left(\frac{L^2 \kappa^d}{\varepsilon^2 w_{\min}^2}\right) \\ &= \frac{C' L^5 R^d \kappa^{d/2} \exp(c'd)}{\gamma_{\min}^{d/2} w_{\min}^5} \log^2\left(\frac{L\kappa}{\varepsilon w_{\min}}\right), \end{aligned}$$

where $c', C', c'', C'' > 0$ are fixed constants. The overall time complexity T consists of two components: the time to get partition function estimates, and the time to generate sample from the target distribution

$$T = T_{\text{partition}} + T_{\text{sampling}}.$$

This completes the proof of the theorem.

□