arXiv:2304.01303v1 [cs.LG] 3 Apr 2023

并行回火混合时间的改进的界

Holden Lee* Zeyu Shen*

摘要

在采样算法领域，当无法直接采样时，马尔可夫链蒙特卡罗（MCMC）方法被广泛使用。然而，目标分布的多峰性往往会导致收敛和混合速度缓慢。并行回火是一种常见的解决方案。尽管它在实践中非常有效，但关于其性能的理论保证仍然有限。本文针对并行回火的谱隙提出了一个新的下界：除 $\log\ L$ 外，该下界对所有参数均呈多项式依赖，其中 $(L+1)$ 为层级数。这改进了现有最佳界；后者对模态数呈指数依赖。此外，我们还给出了谱隙的一个假设性上界，该上界对 $\log L$ 呈指数依赖，这表明从某种意义上说，我们的界是紧的。

1 引言

统计学、计算机科学和统计物理学中的一个关键问题是：在已知概率密度函数（相差一个比例常数）的情况下，如何抽取样本。由于通常难以直接从目标分布中抽取独立样本，因此人们常用马尔可夫链蒙特卡罗（MCMC）方法。然而，对于典型的 MCMC 方法，一个常见难题是：面对强多峰分布，MCMC 方法需要耗费长得不合理的时间才能达到平稳状态。并行回火是一种广泛用于从多峰分布中采样的 MCMC 算法。尽管它在实践中非常有效，但关于其性能的理论保证仍然有限。由于较大的谱隙意味着快速混合，获得混合时间上界的一种常见方法是推导谱隙下界。 [7]讨论了并行回火在从多峰分布中采样时实现快速混合的条件，并给出了一个对模态数呈指数依赖的谱隙下界。

本文给出了并行回火谱隙的一个下界：除 $\log L$ 外，它对所有参数都呈多项式依赖，其中 $(L+1)$ 是层级数。我们采用与 [7], 类似的典范路径法证明这一结果，但所用路径更加复杂。该下界对所有参数呈准多项式依赖，并且在大多数情况下优于 [7] 中现有的最佳界。作为补充，我们还给出了谱隙的一个假设性上界，该上界对 $\log L$ 也呈指数依赖，这表明从某种意义上说，我们的界是紧的。

2 预备知识

考虑一个可测空间 $(\mathcal{X}, \mathcal{F}, \lambda)$。通常， $\mathcal{X} = \mathbb{R}^d$ 和 λ 是勒贝格测度，但也可以考虑更一般的空间。假设我们想从分布 μ 中抽取样本，该分布定义在 $(\mathcal{X}, \mathcal{F})$ 上；我们通过模拟一个转移核为 P 的马尔可夫链来实现，其中 μ 是 P的平稳分布。如果 P 是 ϕ-不可约且非周期的（定义见 [5]）、非负定且 μ-可逆，那么该链会以由谱隙 [5]界定的速率依分布收敛至 μ

$$\mathbf{Gap}(P) = \inf_{f \in L_2(\mu)} \left(\frac{\mathcal{E}(f,f)}{\mathbf{Var}_\mu(f)} \right),$$

*数学系，杜克大学。电子邮箱： holden.lee@duke.edu， zeyu.shen@duke.edu。

其中 $\mathcal{E}(f,f)$ 是狄利克雷形式 $\langle f,\ (I-P)f\rangle_\mu$。特别地，对于任何相对于 μ具有密度的分布 μ_0，我们有 [4]

$$\|\mu_0 P^n-\mu\|_{\mathrm{TV}}\le\sqrt{\mathcal{X}^2\left(\mu\|\mu_0\right)}e^{-n\mathbf{Gap}(P)},\ \forall n\in\mathbb{N},$$

其中 $\|\cdot\|_{\mathrm{TV}}$ 是全变差距离。这说明，较大的谱隙意味着更快的混合 .

并行回火。并行回火是一种经典算法，旨在从多峰分布中采样时实现更快的混合。它通过加入“逆温度项” β 扩展状态空间，使分布变得平滑，从而更容易在不同模态之间转移。从分布 π 中采样时，每个温度都对应一个平稳分布 π_i；对于 π_i，一个自然的选择是 π^{β_i}，当 β_i 较小时，它具有较快的混合速度。

并行回火的详细过程如下。假设我们要从分布 π 中采样。我们选择一组 $L+1$ 逆温度 $\beta_0<\cdots<\beta_L=1$，并构建 $L+1$ 个链层级，其中 i^{th} 层级的平稳分布为 π_i。我们针对 β 的每个取值运行一条链，因此有 $L+1$ 个链层级并行运行；我们只关注 β_L 处的样本。每次迭代中，我们提议交换相邻温度层级的样本，并以 Metropolis-Hastings 比率作为接受概率。该过程在算法 1 中正式给出。

算法 1：并行回火

输入：温度 $\{\beta_i\}_{i=0}^L$，平稳分布为 $\{\pi_i\}_{i=}^L$ 的马尔可夫链 $_0$ 及其转移核 $\{T_i\}_{i=0}^L$，最大迭代次数 N；
输出：随机样本 $\{\theta_n\}_{n=1}^N$ $\subset\mathcal{X}$；
初始化：随机初始化 $\left(\theta_0^0,\ldots,\theta_L^0\right)$；
对于 $n=1,\ldots,N$ **do**
  对于 $i=0,\ldots,L$ **do**
    提出 $\widetilde{\theta}_i$ 根据转移核 T_i 从 θ_i^{n-1} 抽取；
    以概率 $\alpha\left(\widetilde{\theta}_i,\theta_i^{n-1}\right)=$ 接受 $\widetilde{\theta}_i$：$\min\left\{1,\ \frac{\pi_i(\widetilde{\theta}_i)}{\pi_i(\theta_i^{n-1})}\right\}$；
    如果接受，则
      $\theta_i^n\leftarrow\widetilde{\theta}_i$；
    else
      $\theta_i^n\leftarrow\theta_i^{n-1}$ ；
    end
  end
  对于 $i=1,\ldots,L$ **do**
    提出交换 $(i-1)^{\mathrm{st}}$ 链与 i^{th} 链；
    以概率 $\alpha(\theta_{i-1}\leftrightarrow\theta_i)=\min\left\{1,\ \frac{\pi_{i-1}(\theta_i^n)\pi_i(\theta_{i-1}^n)}{\pi_i(\theta_i)\pi_{i-1}(\theta_{i-1})}\right\}$接受交换；
    若接受，则
      $\left(\theta_{i-1}^n,\theta_i^n\right)\leftarrow\left\{\theta_i^n,\theta_{i-1}^n\right\}$；
    end
  end
  将 θ_L^n 添加到样本集；
结束

基本记号。我们用 $[n]$ 表示集合 $\{1,\ldots,n\}$。设 π 为我们希望从中采样的分布，它是一个具有 m 个模态的多峰分布。假设我们运行具有 $L+1$ 个链层级的并行回火。用 π_i 表示 π 在逆温度 β_i 下的退火版本，其中 $\pi_i\propto\pi^{\beta_i}$。用 $\mathcal{X}_{pt}=\mathcal{X}^{L+1}$ 表示并行回火链的状态空间。用 P_{pt} 表示并行回火链的转移核。此外，给定状态 $\xi=(\xi_0,\ldots,\xi_{k_1},\ldots,\xi_{k_2},\ldots,\xi_L)$，我们令 $\xi_{[i,k]}=(\xi_0,\ldots,\xi_{i-1},k,\xi_{i+1},\ldots,\xi_L)$，并且令 $(k_1,k_2)\xi=(\xi_0,\ldots,\xi_{k_2},\ldots,\xi_{k_1},\ldots,\xi_L)$。

转移核。我们用 T_i 表示 β_i处的转移核，它关于π_i可逆；并用 T 表示 $\mathcal{X}_{pt}$上的转移核。用 Q 表示交换移动的转移核。对于每次更新和交换移动，我们加入$\frac{1}{2}$自环概率，以确保其为非负定，从而使谱隙能够转化为混合时间上界。于是， T 可表示为

$$T\left(\theta, \mathrm{d}\widetilde{\theta}\right)=\frac{1}{2(L+1)} \sum_{i=0}^{L} T_i\left(\theta_i, \mathrm{d}\widetilde{\theta}_i\right) \delta_{\theta_{(-i)}}\left(\widetilde{\theta}_{(-i)}\right) \mathrm{d}\widetilde{\theta}_{(-i)}, \ \forall \theta, \widetilde{\theta} \in \mathcal{X}_{pt},$$

其中， $\theta_{(-i)}=(\theta_0, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_L)$， δ是狄拉克δ函数， T_i通常是关于 π_i的 Metropolis-Hastings 核； Q 定义为

$$Q(\theta, A)=\frac{1}{2L} \sum_{i=1}^{L} \mathbf{1}_A\Big((i-1, i)\theta\Big) \alpha(\theta_{i-1} \leftrightarrow \theta_i)+\mathbf{1}_A(\theta)\left[1-\frac{1}{2L} \sum_{i=1}^{L} \alpha(\theta_{i-1} \leftrightarrow \theta_i)\right], \ \forall A \subset \mathcal{X}_{pt}, \theta \in \mathcal{X}_{pt},$$

其中， $\alpha(\theta_{i-1} \leftrightarrow \theta_i)$是接受在 θ_{i-1} 和 θ_i之间交换的概率， $\mathbf{1}_A$ 是集合 A的指示函数。

转移核在集合上的限制。对于关于分布 μ 可逆的转移核 P，以及 P的状态空间的任意子集 A，将 P 在 A 上的限制定义为

$$P\mid_A(\theta, B)=P(\theta, B)+\mathbf{1}_B(\theta) P(\theta, A^c), \ \forall \theta \in A, B \subset A.$$

转移核的投影。取 P的状态空间的任意划分 $\mathcal{A}=\{A_k: \ k=1, \ldots, m\}$，使得对所有 k， $\pi(A_k)>0$，并将 P 相对于 $\mathcal{A}$ 的投影矩阵定义为

$$\overline{P}(k_1, k_2)=\frac{1}{\pi(A_{k_1})} \int_{A_{k_1}} \int_{A_{k_2}} P(\theta, \mathrm{d}\widetilde{\theta}) \mu(\mathrm{d}\theta), \ \forall k_1, k_2 \in [m].$$

当 π 为多峰时，通常选择 $\mathcal{A}$，使得对每个 k， $\pi|_{A_k}$ 都是单峰的。如 [3], 中所述，我们考虑由将层级分配给划分元素的所有可能方式构成的投影空间 $[m]^{L+1}$。对于 $\lambda=(\lambda_0, \ldots, \lambda_L) \in \mathcal{X}_{pt}$，将 θ 投影到 $\mathcal{A}$ 定义为

$$\mathrm{proj}_{\mathcal{A}}(\theta)=\lambda=(\lambda_0, \ldots, \lambda_L),$$

使得

$$\lambda_i=k \text{ if } \theta_i \in A_k,$$

并令 $\mathcal{X}_\lambda=\{\theta \in \mathcal{X}_{pt} | \mathrm{proj}_{\mathcal{A}}(\theta)=\lambda\}$。

两个重要量。接下来，我们定义两个重要量。第一个是

$$\phi=\min_{|i-j|=1, k \in [m]} \int \frac{\min\{\pi_i(\theta), \pi_j(\theta)\}}{\pi_i(A_k)} \lambda(\mathrm{d}\theta),$$

它表示两个相邻层级之间的最小重叠体积。直观地说， ϕ 控制相邻层级之间的温度变化速率。注意，对于任意 k_1、$k_2 \in [m]$ 和 $i \in [L]$，在平稳性下，接受在 $\theta_{i-1} \in A_{k_1}$ 和 $\theta_i \in A_{k_2}$ 之间提出的 Swap 的边缘概率为

$$\frac{\int_{\xi \in A_{k_1}} \int_{\widetilde{\xi} \in A_{k_2}} \min\left\{\pi_{i-1}(\xi) \pi_i(\widetilde{\xi}), \pi_{i-1}(\widetilde{\xi}) \pi_i(\xi)\right\} \lambda(\mathrm{d}\xi) \lambda(\mathrm{d}\widetilde{\xi})}{\pi_{i-1}(A_{k_1}) \pi_i(A_{k_2})} \geq \phi^2. \tag{1}$$

第二个是

$$B=\min_{k \in [m]} \prod_{i=1}^{L} \min\left\{1, \frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}\right\},$$

这就是交换过程中自然出现的“瓶颈比”。

用于界定谱隙的工具。我们提出典范路径法 [1], ，这是一种用于界定有限状态空间上的马尔可夫链谱隙的方法。设 P_1 和 P_2 分别是状态空间 $\mathcal{X}$上的马尔可夫链转移矩阵，满足 $|\mathcal{X}| < \infty$，且分别关于密度 π_{P_1} 和 π_{P_2}可逆。~ ~设 $\mathcal{E}_{P_1}$和 $\mathcal{E}_{P_2}$分别是 P_1和 P_2的狄利克雷形式，并令 $E_{P_1} = \{(\xi,\xi)\colon \pi_{P_1}(\xi)P_1(\xi,\xi) > 0\}$和~ ~$E_{P_2} = \{(\xi,\xi)\colon \pi_{P_2}(\xi)P_2(\xi,\xi) > 0\}$分别为 P_1和 P_2的边集。

~ ~ ~定理 1. 对于每一对 $\xi \neq \xi$，使得 $(\xi,\xi) \in E_{P_2}$，固定一条长度为$|\gamma_{\xi,\ \tilde{\xi}}| = \ell$ 的路径 $\gamma_{\xi,\ \tilde{\xi}} = (\xi = \xi_0, \xi_1, \ldots, \xi_\ell = \xi)$，使得对于 $s \in [\ell]$，$(\xi_{s-1}, \xi_s) \in E_{P_1}$。定义拥塞

$$c = \max_{(\tau,\widetilde{\tau})\in E_{P_1}} \left\{ \frac{1}{\pi_{P_1}(\tau)P_1(\tau,\widetilde{\tau})} \sum_{\gamma_{\xi,\tilde{\xi}} \ni (\tau,\widetilde{\tau})} |\gamma_{\xi,\tilde{\xi}}| \pi_{P_2}(\xi)P_2(\xi,\widetilde{\xi}) \right\}.$$

那么，我们有:

$$\mathcal{E}_{P_2} \leq c\mathcal{E}_{P_1}.$$

请注意，在定理 1 中，不同路径的长度可能不同。

3 谱隙下界

本节给出并行回火的谱隙下界。该界在定理 2 中正式表述。

定理 2. 给定 $\mathcal{X}$ 的任意划分 $\mathcal{A} = \{A_k\colon k = 1, \ldots, m\}$，使得对于所有 i 和 k，$\pi_i[A_k] > 0$，我们有

$$\mathbf{Gap}(P_{pt}) \geq \frac{\phi^2 B^{O(\log L)}}{O(m^3(L+1)^{3+2\log_2 3})} \mathbf{Gap}(\overline{T}_1) \min_{i,k} \mathbf{Gap}(T_i|_{A_k}).$$

注意， [7] 给出的界为 $\mathsf{Gap}(P_{pt}) \geq \frac{\phi^2 B^{m+3}}{2^{12} m^3 \ L+1^{()4}} \underline{\mathsf{Gap}(\overline{T}_1)\ \mathsf{min}_{i,\ k}\, \mathsf{Gap}(T_i|_{A_k})}$，这对 m 具有指数依赖；在大多数情况下，我们对 $\log L$ 的指数依赖改进了这一结果。从本质上说，我们的改进源于为典范路径法构造了一条更复杂的路径：路径的拥塞取决于必须移动到不同层级的样本数量；然而， [7] 中的路径实际上必须将每个层级中的样本都移至不同层级。为了获得更好的下界，我们实际上可以采用递归构造，使得只有 O个（$\log L$）层级中的样本需要移至不同层级。

为了证明定 理 2，我们首先注意到， $\overline{P}_{pt}$ 关于概率质量函 n

$$\overline{\pi}(\lambda) \equiv \prod_{i=0}^{L} \pi_i(A_{\lambda_i}), \quad \forall \lambda = (\lambda_0, \ldots, \lambda_L) \in [m]^{L+1}. \tag{2}$$

我们考虑如下构造的转移核 P_1： 以$1\ 2$ 的概率按照$\overline{Q}$转移，或者以$1\ 2(L+1)$ 的概率根据分布 $\{\pi_0[A_k]$ 抽取 λ_0： $k \in [m]\}$；否则保持不变。显然，P_1 关于$\overline{\pi}$也是可逆的，因此 $\overline{P}_{pt}$ 和 P_1 具有相同的平稳分布。我们借用了 [7]中的以下不等式。

引理 1.（ [7] 中的公式 (17)）间隙$(\overline{P}_{pt})$ $\geq$ 间隙(P_1)间隙$(\overline{T}_1)$ 4。

现在，我们使用典范路径法，通过与另一个 $\overline{\pi}$-可逆链进行比较，来给出间隙(P_1) 的界。定义转移矩阵 P_2： 它从 $\{0, \ldots, L\}$ 中均匀选择 i，然后根据分布 $\pi_i[A_k]$ 抽取 λ_i， $k = 1$，....., m。直观地说， P_2 可以轻松地在不同模态之间移动，因此具有较大的谱隙。通过比较 P_1 和 P_2，我们得到 $\overline{P}_{pt}$ 的谱隙下界。

在给出路径之前，我们先介绍一个递归构造路径 $\gamma_{\lambda,\ (i,j)\lambda}$ 从λ 到 $(i,j)\lambda$的过程，其中 $i < j$；即交换 i^{th} 和 j^{th} 层级上的样本。我们证明，对于路径中的任意中间状态，它与 λ 至多在 $O(\log L)$ 个层级上不同。该过程在算法 2 中正式给出。

算法 2：交换第 i^{th} 和第 j^{th} 层级上的状态的过程。

过程名称：Swap(i, j)；若 $j-i\le 1$，则交换 i^{th} 和 j^{th} 层级上的状态；否则Swap(i, $\lfloor\frac{i+j}{2}\rfloor$)；Swap($\lfloor\frac{i+j}{2}\rfloor$, j)；Swap(i, $\lfloor\frac{i+j}{2}\rfloor$)；结束

引理 2. 对于 $\gamma_{\lambda,(i,j)\lambda}$ 中的任意状态 τ，它与 λ 至多相差 $O(\log L)$ 个层级。

证明. 令 $F(\ell)$ 表示在交换第 x^{th} 层级和第 $(x+\ell)^{\text{th}}$ 层级的过程中，任意中间状态与原始状态最多相差的层级数，其中 $x+\ell\le L$。第一次调用 Swap(x, $x+\lfloor\frac{\ell}{2}\rfloor$) 时，任意中间状态与 λ 至多相差 $F(\lfloor\frac{\ell}{2}\rfloor)$ 个层级；调用 Swap($x+\lfloor\frac{\ell}{2}\rfloor$, $x+\ell$) 时，任意中间状态与 λ 至多相差 $F(\lceil\frac{\ell}{2}\rceil)$ 个 $+2$层级；第二次调用 Swap(x, $x+\lfloor\frac{\ell}{2}\rfloor$) 时，任意中间状态与 λ 至多相差 $F(\lfloor\frac{\ell}{2}\rfloor)$ 个 $+3$层级。因此，我们有

$$F(\ell)\le\max\left\{2+F\left(\left\lceil\frac{\ell}{2}\right\rceil\right),3+F\left(\left\lfloor\frac{\ell}{2}\right\rfloor\right)\right\},$$

这给出了 $F(\ell)\le O(\log\ell)$。因此，我们有：$\gamma_{\lambda,(i,j)\lambda}$ 中的任意状态与 λ之间最多相差 $O(\log(j-i))\le O(\log L)$ 个层级。 □

接下来，固定一些 $i\in\{0,\dots,L\}$ 和 $k\in[m]$，我们给出一条连接 λ 与 $\lambda_{[i,k]}$的路径 $\gamma_{\lambda,\lambda_{[i,k]}}$。令 k^* 为使 $\pi_L(A_k)$ 最大化的 k 取值。该过程在算法 3 中作了正式说明。

算法 3：构造一条连接 λ 与 $\lambda_{[i,k]}$的路径的过程。

将层级 0, λ_0处的值改为 k^*；调用 Swap(0, i)；将层级 0 处的值改为 k；调用 Swap(0, i)；将层级 0, k^*处的值改为 λ_0；

为了给 Gap(P_1) 找到上界，我们首先给出任意状态 $\tau,\widetilde{\tau}\in\gamma_{\lambda,\lambda_{[i,k]}}$下 $\overline{\pi}_{(\lambda)P_2(\lambda,\lambda_{[i,k]})}\,\overline{\pi}(\tau)P_1(\tau,\widetilde{\tau})$ 的一个上界。

引理 3. $\frac{\overline{\pi}(\lambda)P_2(\lambda,\lambda_{[i,k]})}{\overline{\pi}(\tau)P_1(\tau,\widetilde{\tau})}\le\frac{4m}{\phi^2B^{O(\log L)}}$.

证明。注意到

$$\overline{\pi}(\lambda)P_2(\lambda,\lambda_{[i,k]})=\frac{\pi_i(A_k)}{L+1}\prod_{j=0}^{L}\pi_j(A_{\lambda_j})=\frac{1}{L+1}\min\{\overline{\pi}(\lambda),\overline{\pi}(\lambda_{[i,k]})\}\max\{\pi_i(A_{\lambda_i}),\pi_i(A_k)\}.\tag{3}$$

对于任意中间状态 τ（属于 $\gamma_{\lambda,\ \lambda_{[i,k]}}$），我们可以用 $\min\{\overline{\pi}(\lambda)\ \overline{\pi}(\lambda_{[i,k]})\}$ 给出 $\overline{\pi}(\tau)$ 的一个下界。这里，我们给出以下引理。

Lem引理 4. 对于 τin $\gamma_{\lambda,\lambda_{[i,k]}}$中的任意状态，它与 $\lambda$and $\lambda_{[i,k]}$ 最多相差 $O(\log\ L)$ 个层级 .

证明。第一次调用 Swap(0, i) 时，根据引理 2，任意中间状态与 λ 最多相差$O(\log L)$ 个层级。第一次 Swap(0, i) 过程结束后，当前状态仅在 0th 和 i^{th} 层级上与 λ 不同。因此，第二次调用 Swap(0, i) 时，任意中间状态仍与 λ 最多相差 $O(\log L)$ 个层级。因此，$\gamma_{\lambda,\lambda_{[i,k]}}$ 中的任意状态与λ 最多相差 $O(\log L)$ 个层级。由于 λ 和 $\lambda_{[i,k]}$ 仅在一个层级上不同，因此我们也有 $\gamma_{\lambda,\lambda_{[i}$ 中的任意状态与 $\lambda_{[i,k]}$ 相比，最多只在 $O(\log L)$个层级上不同。 $_{,k]}$ □

从路径中固定一个状态 τ，并假设它与 λ 在 n 个层级上不同，其中 $n = O(\log L)$。这实际上意味着，最多有 n 个样本不在其原始层级中。

假设所有不在原始层级中的样本为 $\{\lambda_{\ell_1}, \ldots, \lambda_{\ell_{n-1}}, k^*\}$。对于 λ_{ℓ_i}，假设其原始层级为 ℓ_i，当前层级为 ℓ_i^*；对于 k^*，假设其当前层级为 ℓ_n^*。这里有一点需要注意：根据构造，对于所有 $i \in \{1, \ldots, n-1\}$，我们必有 $\ell_i^* < \ell_i$。因此，我们有

$$\frac{\overline{\pi}(\tau)}{\overline{\pi}(\lambda)} = \frac{\pi_{\ell_n^*}(k^*)}{\pi_0(\lambda_0)} \prod_{i=1}^{n-1} \frac{\pi_{\ell_i^*}(\lambda_{\ell_i})}{\pi_{\ell_i}(\lambda_{\ell_i})} \geq \frac{B^n}{m} = \frac{B^{O(\log L)}}{m}.$$

类似地，我们有 $\frac{\overline{\pi}(\tau)}{\overline{\pi}(\lambda_{[i,k]})} \geq \frac{B^{O(\log L)}}{m}$。因此，我们有

$$\frac{\overline{\pi}(\tau)}{\min\left(\overline{\pi}(\lambda), \overline{\pi}(\lambda_{[i,k]})\right)} \geq \frac{B^{O(\log L)}}{m}. \tag{4}$$

现在，考虑路径 $\gamma_{\lambda,\lambda_{[i,k]}}$ 上的某条边 $(\tau, \widetilde{\tau})$。如果对于某个 j，$\widetilde{\tau} = (j, j+1)\tau$，由于根据 Q 提议 Swap 的概率为 $\frac{1}{2L}$，我们有

$$\frac{\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau) P_1(\tau, \widetilde{\tau})} \leq \frac{2\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau) \overline{Q}(\tau, \widetilde{\tau})} \leq \frac{4(L+1)\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau)\phi^2} \tag{5}$$

$$= \frac{4\min\{\overline{\pi}(\lambda), \overline{\pi}(\lambda_{[i,k]})\} \max\{\pi_i(A_k), \pi_i(A_{\lambda_i})\}}{\overline{\pi}(\tau)\phi^2} \leq \frac{4m}{\phi^2 B^{O(\log L)}}. \tag{6}$$

其中，第一步是因为 P_1 按照 $\overline{Q}$ 以概率 $1/2$ 进行转移；第二步由式 (1) 得出，第三步由式 (3) 得出，第四步由式 (4) 得出。

如果对于某个 k_0，$\widetilde{\tau} = \tau_{[0,k_0]}$，那么我们必定处于算法 3 的第 1、3 或 5 行。如果处于第 1 行，则有$k_0 = k^*$ 和 $\tau = \lambda$，因此

$$\frac{\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau) P_1(\tau, \widetilde{\tau})} = \frac{2\overline{\pi}(\lambda)\pi_i(A_k)}{\overline{\pi}(\lambda)\pi_0(A_{k^*})} \leq \frac{2}{\pi_0(A_{k^*})} \leq \frac{2m}{B},$$

其中第一步根据 P_1 和 P_2的定义得出，第二步是因为 $\pi_i(A_k) \leq 1$，第三步是因为 $\pi_0(A_{k^*}) \geq \pi_L(A_{k^*})$ $B \geq \frac{B}{m}$。如果处于第 3 行，则有 $k_0 = k$和 $\tau_0 = \lambda_i$，因此

$$\frac{\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau) P_1(\tau, \widetilde{\tau})} = \frac{2\overline{\pi}(\lambda)\pi_i(A_k)}{\overline{\pi}(\tau)\pi_0(A_k)} = \frac{2\min\{\overline{\pi}(\lambda), \overline{\pi}(\lambda_{[i,k]})\} \max\{\pi_i(A_k), \pi_i(A_{\lambda_i})\}}{\min\{\overline{\pi}(\tau), \overline{\pi}(\widetilde{\tau})\} \max\{\pi_0(A_k), \pi_0(A_{\lambda_i})\}} \tag{7}$$

$$\leq \frac{2}{B} \cdot \frac{\min\{\overline{\pi}(\lambda), \overline{\pi}(\lambda_{[i,k]})\}}{\min\{\overline{\pi}(\tau), \overline{\pi}(\widetilde{\tau})\}} \leq \frac{2m}{B^{O(\log L)}}, \tag{8}$$

其中第二步由等式 (3) 得出，第三步是因为 $\pi_i(A_k) \leq \pi_0(A_k) \leq \max\{\pi_0(A_k), \pi_0(A_{\lambda_i})\}/B$ 且 $\pi_i(A_{\lambda_i}) \leq \pi_0(A_{\lambda_i}) \leq \max\{\pi_0(A_k), \pi_0(A_{\lambda_i})\}/B$，第四步由等式 (4) 得出。如果处于第 5 行，则有$\widetilde{\tau} = \lambda_{[i,k]}$ 和 $\tau_0 = k^*$，因此

$$\frac{\overline{\pi}(\lambda) P_2(\lambda, \lambda_{[i,k]})}{\overline{\pi}(\tau) P_1(\tau, \widetilde{\tau})} = \frac{2\overline{\pi}(\lambda_{[i,k]})\pi_i(A_{\lambda_i})}{\overline{\pi}(\lambda_{[i,k]})\pi_0(A_{k^*})} \leq \frac{2m}{B}$$

其中第一步是因为 P_1 和 P_2 都关于 $\overline{\pi}$ 可逆。综合以上讨论的所有情况，即完成了引理 3 的证明。

□

接下来，我们给出以下引理，它们为路径的长度和拥塞提供了一个界。

引理 5。$|\gamma_{\lambda,\ \lambda_{[i,k]}}| \leq O(L^{\log_2 3})$。

证明。令 $F(\ell)$ 表示由算法 2 产生的路径长度，该算法交换 x^{th} 层级与 $(x+\ell)^{\text{th}}$ 层级，其中 $x + \ell \leq L$。于是，我们有

$$F(\ell) = 2F\left(\left\lfloor \frac{\ell}{2} \right\rfloor\right) + F\left(\left\lceil \frac{\ell}{2} \right\rceil\right),$$

由此可得

$$F(2^t) = 3^t,\ F(\ell) \le F(2^t) = 3^t \le (2\ell)^{\log_2 3} \quad \forall t \in \mathbb{N}, 2^{t-1} \le \ell < 2^t.$$

也就是说，$F(\ell) = O(\ell^{\log_2 3})$。因此，$|\gamma_{\lambda,\ \lambda_{[i,k]}}| = 3 + 2F(i) = O(i^{\log_2 3}) \le O(L^{\log_2 3})$。 □

引理 6。对于每 条给定的边（τ，$\widetilde{\tau}$），至多有 $O(m^2(L+1)^{1+\log_2 3})$ 条路径经过这条边 .

~证明。我们首先作如下观察。固定 $i \in \{0, \ldots, L\}$，$k \in [m]$，并设 λ，λ 是两个起始状态，且对某个 $i_0 \in [L]$ 有 $\lambda_{i_0} \neq \widetilde{\lambda}_{i_0}$。令 $\gamma^s_{\lambda,\ \lambda_{[i,k]}}$ 和 $\gamma^s_{\widetilde{\lambda},\ \widetilde{\lambda}_{[i,k]}}$ 分别为每条路径中的边。注意，$|\gamma_{\lambda,\ \lambda_{[i,k]}}| = |\gamma_{\widetilde{\lambda},\ \widetilde{\lambda}_{[i,k]}}|$，因为根据算法 2 和 3，路径长度仅取决于 i。那么，主要的观察是，对于每个 $1 \le s \le |\gamma_{\lambda,\ \lambda_{[i,k]}}|$，都有 $\gamma^s_{\lambda,\ \lambda_{[i,k]}} \neq \gamma^s_{\widetilde{\lambda},\ \widetilde{\lambda}_{[i,k]}}$，这是因为 λ 和 λ 在最初的 s 步中经历完全相同的过程，从而到达 $\lambda_{[i,k]}$ 和 $\lambda_{[i,k]}$。因此，

$$\sum_{\gamma_{\lambda,\lambda_{[i,k]}} \ni (\tau,\widetilde{\tau})} 1 = \sum_{s=1}^{|\gamma_{\lambda,\lambda_{[i,k]}}|} \sum_{\gamma^s_{\lambda,\lambda_{[i,k]}} = (\tau,\widetilde{\tau})} 1 \le \sum_{s=1}^{|\gamma_{\lambda,\lambda_{[i,k]}}|} m \le O\left(mL^{\log_2 3}\right),$$

也就是说，对于固定的 i、k，每条边至多属于 $O\left(mL^{\log_2 3}\right)$ 条路径。注意，第二步成立是因为条件 $\gamma^s_{\lambda,\ \lambda_{[i,k]}} = (\tau,\widetilde{\tau})$ 唯一确定了所有 $i_0 \in [L]$ 对应的 λ_{i_0}，而 0th 层级上的样本尚未固定。由于 i 有 $L+1$ 种选择，k 有 m 种选择，因此可以得出，每条边至多属于 $O\left(m^2(L+1)^{1+\log_2 3}\right)$ 条路径。

□

现在回过头来证明定理 2。

定理 2 的证明。由引理 5 和引理 6，我们有

$$\sum_{\gamma_{\lambda,\lambda_{[i,k]}} \ni (\tau,\widetilde{\tau})} |\gamma_{\lambda,\lambda_{[i,k]}}| \le O\left(m^2(L+1)^{1+2\log_2 3}\right). \tag{9}$$

由于 P_1 和 P_2 具有相同的平稳分布 $\overline{\pi}$，应用定理 1，并结合引理 3 和公式（9），可得

$$\begin{aligned}
\mathbf{Gap}(P_2) &\le \max_{(\tau,\widetilde{\tau}) \in E_{P_1}} \left\{ \frac{1}{\pi_{P_1}(\tau) P_1(\tau,\widetilde{\tau})} \sum_{\gamma_{\lambda,\lambda_{[i,k]}} \ni (\tau,\widetilde{\tau})} |\gamma_{\lambda,\lambda_{[i,k]}}| \pi_{P_2}(\lambda) P_2(\lambda, \lambda_{[i,k]}) \right\} \mathbf{Gap}(P_1) \\
&\le \max_{(\tau,\widetilde{\tau}) \in E_{P_1}} \left\{ \frac{\pi_{P_2}(\lambda) P_2(\lambda,\lambda_{[i,k]})}{\pi_{P_1}(\tau) P_1(\tau,\widetilde{\tau})} \right\} \max_{(\tau,\widetilde{\tau}) \in E_{P_1}} \left\{ \sum_{\gamma_{\lambda,\lambda_{[i,k]}} \ni (\tau,\widetilde{\tau})} |\gamma_{\lambda,\lambda_{[i,k]}}| \right\} \mathbf{Gap}(P_1) \\
&\le \frac{O(m^3(L+1)^{1+2\log_2 3})}{\phi^2 B^{O(\log L)}} \mathbf{Gap}(P_1).
\end{aligned}$$

由引理 1，可得

$$\mathbf{Gap}(\overline{P}_{pt}) \ge \frac{\mathbf{Gap}(P_1)\mathbf{Gap}(\overline{T}_1)}{4} \ge \frac{\phi^2 B^{O(\log L)}}{O(m^3(L+1)^{2+2\log_2 3})} \mathbf{Gap}(\overline{T}_1). \tag{10}$$

这里，我们借用 [7] 中的以下不等式。

引理 7。（ [7] 中的公式（14））间隙$(P_{pt}) \ge \frac{1}{2}$间隙$(\overline{P}_{pt})$ $\min^{\lambda \in [m]^{L+1}}$间隙$(P_{pt}|_{\mathcal{X}_\lambda})$。

引理 8。（ [7] 中的方程 (15)）间隙$(P_{pt}|_{\mathcal{X}_\lambda}) \ge$ 8^{-1} ~~$(L+1)$~~ $\min_{i,\ k}$ 间隙$(T_i|_{A_k})$ $\forall \lambda \in [m]^{L+1}$.

结合引理 7、8 和方程（10）可得

$$\mathbf{Gap}(P_{pt}) \ge \frac{1}{16(L+1)} \mathbf{Gap}(\overline{P}_{pt}) \min_{i,k} \mathbf{Gap}(T_i|_{A_k}) \ge \frac{\phi^2 B^{O(\log L)}}{O(m^3(L+1)^{3+2\log_2 3})} \mathbf{Gap}(\overline{T}_1) \min_{i,k} \mathbf{Gap}(T_i|_{A_k}).$$

这就完成了定理 2 的证明。 □

4 谱隙的假设上界

在本节中，我们通过给出并行回火谱隙的一个假设性上界来补充我们的结果。 [7] 中的定理 5.2 指出，链的谱隙不大于投影链的谱隙。 [6] 讨论了并行回火出现迟滞混合的充分条件，并指出不等方差的高斯混合分布可能混合缓慢；然而，其缓慢混合本质上源于该实例的瓶颈比非常小，而且其谱隙仍可由 $O(B^{O(1)})$ 上界。接下来，我们构造一个假设性实例，并指定不同温度下各个模态的权重。我们只关注不同模态之间的转移，而不关注模态内部的转移。在该实例中，我们证明，在只允许相邻温度层级之间交换、且同一温度层级的不同模态之间不存在移动的假设下，谱隙的上界为 $O\left(B^{O(\log L)}\right)$；这表明定理 2 中的下界在某种意义上是紧的。请注意，我们可以将各个模态设置得足够远，使得同一层级的不同模态之间实际上不会发生移动。我们将类似上界能否由更自然的例子导出留待未来研究。

定理 3. 存在这样一个实例：只允许在相邻温度层级之间交换，且同一层级上的模态之间不发生移动，使得

$$\mathbf{Gap}(P_{pt}) < O\left(B^{O(\log L)}\right).$$

在给出证明之前，我们先简要说明构造思路。我们考虑一个包含m 个模态的分布，并在该分布上运行一个具有 $L+1$ 个层级的并行回火链，其中 $m = L+1$。我们构造该分布，使得在链的 i^{th} 层级上，$(i+1)^{\text{st}}$ 模态的密度占主导地位，而其他模态的密度则处于同一数量级。接着，我们根据各状态与起始状态相差多少个样本，将状态空间划分为两部分，并证明这两部分之间的转移概率很小，而每部分的密度都很大，然后利用切格不等式推导出谱隙的上界。图 1 展示了各模态的相对大小。

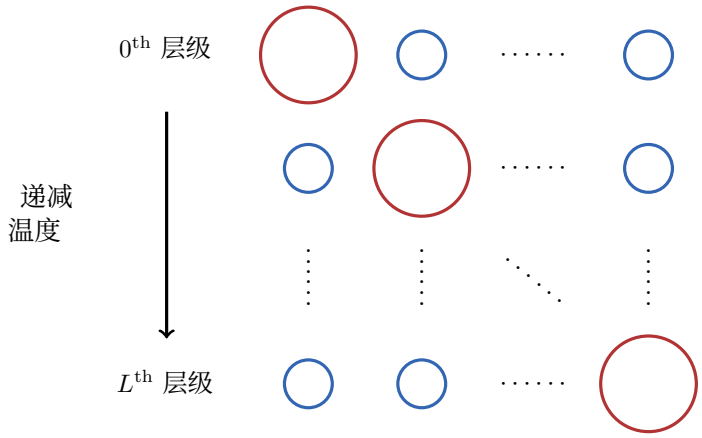

图 1：模式的相对大小示意图

为证明定理 3，我们先给出以下引理。

~ ~引理9。如果我们希望通过仅交换相邻温度层级之间的样本，将状态从 λ 变换为 λ，其中 $\lambda_L = \lambda_0$，那么存在某个中间状态 τ，它与 λ 在 $\Omega(\log L)$ 个层级上不同。

证明。对于任意状态 τ，定义 $h_\tau(i,j)$ 为 τ 中从 i^{th} 层级到 j^{th} 层级（含两者）之间与 λ 不同的层级数；即 $h_\tau(i,j) = |\{\ell\colon i \le \ell \le j, \tau_\ell \neq \lambda_\ell\}|$。为证明此引理，我们证明一个更强的命题：假设存在无限多个层级（例如 -1th 层级、$(L+1)^{\text{st}}$ 层级等，所有这些层级

都可以交换到），且 λ 长度无限；即 $\lambda \in [m]^{\mathbb{Z}}$。我们通过归纳证明，对于~ ~所有 $k \in \mathbb{Z}$，若要将状态从 λ 转换到 λ，其中 $\lambda_{k+L} = \lambda_{i_k}$ 对某个 $i_k \leq k$ 成立，则存在某个中间状态 τ，使得 $h_\tau(k+1, k+L) \geq \lfloor \log_2 L \rfloor$。

基础情形 $L = 1$ 显然成立。假设该断言对 L成立，我们证明它对 $2L$也成立。~ ~假设我们希望将一个状态从 λ 变换为 λ，其中 $\lambda_{k+2L} = \lambda_{i_k}$ 对某个 $i_k \leq k$成立。我们首先给出以下引理。

~引理 10。我们只需考虑以下情形：在将一个状态从 λ 变换为 λ的过程中，其中 $\widetilde{\lambda}_{k+2L} = \lambda_{i_k}$ 对某个 $i_k \leq k$成立，存在某个中间状态 τ^1，使得 $h_{\tau^1}(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$，并且 λ_{i_k} 被 Swap 到介于第 $(k+L+1)^{st}$ 层级与第 $(k+2L)^{th}$ 层级之间的某处；即，$\tau^1_\ell = \lambda_{i_k}$ 对某个 $k+L+1 \leq \ell \leq k+2L$成立。

证明。显然，存在一个中间状态 τ，使得 $h_\tau(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$。我们将证明，除此之外，要么可以找到这样一个中间状态 τ^1，使得 λ_{i_k} 被 Swap 到介于第 $(k+L+1)^{\text{st}}$ 层级与第 $(k+2L)^{\text{th}}$ 层级之间的某处，要么可以结束归纳。令 τ^1 为满足 $h_\tau(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$的最后一个中间状态 τ。如果此时 λ_{i_k} 不在第 $(k+L+1)^{\text{st}}$ 层级与第 $(k+2L)^{\text{th}}$ 层级之间，那么我们论证只需考虑它处于某个层级 $i'_k \leq k$的情形。实际上，我们可以观察到，如果有 $h_{\tau^1}(k+1, k+L) > 0$，那么由于 $h_{\tau^1}(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$，可得 $h_{\tau^1}(k+1, k+2L) \geq \lfloor \log_2 L \rfloor + 1 = \lfloor \log_2 2L \rfloor$，从而结束归纳。因此，我们可以假设 $h_{\tau^1}(k+1, k+L) = 0$。由于 λ_{i_k} 尚未到达第 $(k+2L)^{\text{th}}$ 层级，并且根据假设它不在第 $(k+L+1)^{\text{st}}$ 层级与第 $(k+2L)^{\text{th}}$ 层级之间，这意味着 $i'_k \leq k$。于是，根据归纳假设，我们将遇到一个较晚的中间状态 τ^2，使得 $h_{\tau^2}(k+1, k+L) \geq \lfloor \log_2 L \rfloor$。假设 λ_{i_k} 在 τ^2 中的层级 i''_k，其中 $i''_k \leq k+L$。在这里，我们只需考虑 $h_{\tau^2}(k+L+1, k+2L) = 0$的情形；否则归纳将以类似方式得出结论。再次根据归纳假设，为了将 λ_{i_k} 从层级 i''_k 向上 Swap 到第 $(k+2L)^{\text{th}}$ 层级，存在一个较晚的中间状态 τ^3，满足 $h_{\tau^3}(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$，这与 τ^1是满足 $h_\tau(k+L+1, k+2L) \geq \lfloor \log_2 L \rfloor$的最后一个中间状态 τ这一事实矛盾。引理 10 的证明完毕。

□

如果样本 s 从层级 $i_s > k+2L$ 移动到层级 $i'_s \leq k+1$，或从层级 $i_s < k+1$ 移动到层级 $i'_s \geq k+2L$，我们就称它跨越了层级（$k+1$） ～ （$k+2L$）。这里有一个观察：我们可以假设 λ_{i_k}是第一个跨越层级（$k+1$） ～ （$k+2L$）的样本。这是因为，否则我们可以取第一个跨越层级（$k+1$） ～ （$k+2L$）的样本，并应用相同的论证，必要时反转层级。以下证明均基于这一假设。

由引理 10，我们只需考虑存在某个中间状态 τ^1，使得h_{τ^1}（$k+L+1$，$k+2L$）$\geq \lfloor \log_2 L \rfloor$，且 λ_{i_k}在 τ^1 中位于（$k+L+1$）st层级与（$k+2L$）th层级之间的情形。如果h_{τ^1}（$k+1$，$k+L$）>0，则归纳得证。因此，我们可以假设 h_{τ^1}（$k+1$，$k+L$）$=0$。由于λ_{i_k}在 τ^1 中位于（$k+L+1$）st 层级与（$k+2L$）th 层级之间，且它最初处于层级 $i_k \leq k$，所以存在某个样本 s，它最初处于层级 $i_s > k+L$，现在则处于层级 $i'_s \leq k+L$。由于h_{τ^1}（$k+1$，$k+L$）$=0$，我们有 $i'_s \leq k$。此外，由于我们假设 λ_{i_k}是第一个跨越层级（$k+1$） ～（$k+2L$）的样本，我们还有 $i_s \leq k+2L$；也就是说，s最初位于（$k+L+1$）st层级与（$k+2L$）th层级之间。由于 s最终到达层级 $i'_s \leq k$，它必然曾到达（$k+1$）st层级。于是，根据层级反转后的引理 10，我们只需考虑存在某个中间状态 τ^4，使得 h_{τ^4}（$k+1$，$k+L$）$\geq \lfloor \log_2 L \rfloor$，且 s位于（$k+1$）st层级与（$k+L$）th层级之间的情形。此时，由于 s不处于其原始层级，我们有 h_{τ^4}（$k+1$，$k+2L$）$\geq \lfloor \log_2 L \rfloor + 1 = \lfloor \log_2 2L \rfloor$，从而再次完成归纳。

至此，归纳过程已经完成。为了完成引理9的证明，我们注意到，如果要将一个样本从 k^{th} 层级移动到$(k+L)^{\text{th}}$ 层级，并定义

$$f(L) = \min_{\text{all paths } \gamma} \max_{\tau \in \gamma} h_\tau(k+1, k+L),$$

显然，我们有 $f(L+1) \geq f(L)$，因为一方面，在将 λ_k 交换到 $(k+L+1)^{\text{st}}$层级的过程中，我们总会经过 $(k+L)^{\text{th}}$ 层级；另一方面，根据定义，对于任意中间状态 τ，总有 $h_\tau(k+1, k+L) \leq h_\tau(k+1, k+L+1)$。注意，在有无限多个层级的假设下，$f$仍然有定义。结合从 L到 $2L$的归纳步骤，我们可以得出，对于所有 $L \in \mathbb{N}$，$f(L) \geq \lfloor \log_2 L \rfloor$。

回到只有 $(L+1)$ 个层级时的原始命题，由于有限层级链中的路径始终也是无限层级链中的有效路径，因此原始命题成立。 □

构造。我们考虑 $L+1=m$ 和 $\beta_i=\frac{i+1}{L+1}$的情形。选取 m 个彼此间隔较远的点 $\{x_k\}_{k=1}^m$，将它们用作各模态的中心。假设我们要从 π中采样；下面给出一种构造 $\pi^{1/(L+1)}$ 的方法：

$$\pi^{1/(L+1)}=\frac{1}{Z_0}\left(w_k\pi_k+\sum_{k=1}^m\sum_{r=0}^L w_{kr}\pi_{kr}\right),$$

其中，对于每个固定的 k， π_k和 π_{kr}对于每个 $r\in\{0,\ldots,L\}$都是围绕x_k 的非归一化均匀分布，在体积为 V_k 的空间中具有单位密度，且 V_{kr}彼此不相交； Z_0 为归一化常数。假设 $\{x_k\}_{k=1}^m$ 彼此相距足够远，因此我们可以用 $\frac{1}{Z_i}\left(w_k^{i+1}\pi_k^{i+1}+\sum_{k=1}^m\sum_{r=0}^L w_{kr}^{i+1}\pi_{kr}^{i+1}\right)$近似$\pi^{(i+1)/(L+1)}$，其中 Z_i 是其归一化常数；也就是说，可以安全地忽略交叉项。考虑 $\mathcal{A}=\{A_k$的划分： $k=1$，……， $m\}$，其中 $\mathcal{X}$满足 A_k包含中心为 x_k的模态。令 $\gamma=(L+1)^3$，并且

$$w_k=\gamma^{2k},\quad V_k=\gamma^{-k^2+1},$$

$$w_{kr}=\gamma^{2r+1},\quad V_{kr}=\gamma^{-r^2-r}\quad\forall r\in\{0,\ldots,L\}.$$

令 $\overline{\pi}$ 如式 (2) 所定义。我们证明该实例给出了所需的上界。令 π_i 为 $\pi^{(i+1)/(L+1)}$的归一化形式， Z_i 为归一化常数。我们先给出以下引理，它对每个模态的密度提供了界。

引理 11. $\pi_i(A_{i+1})>1-\frac{1}{L+1}$和$1\ 2(L+1)\cancel{3<\pi_i}(A_k)<\quad{}^1\ (L\cancel{+1)^2}\ \forall i\in\{0,\ldots,L\}$和 $k\neq i+1$。

证明。固定某个层级 $i\in\{0,\ldots,L\}$。我们计算

$$\pi_i(A_k)=\frac{1}{Z_i}\left(\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^L\gamma^{(2r+1)(i+1)-r^2-r}\right)\quad\forall k\in[m].$$

由于 $2(i+1)k-k^2+1\le(i+1)^2+1$，且当 $k=i+1$时取等号，因此我们有

$$\gamma^{2(i+1)k-k^2+1}\le\gamma^{(i+1)^2+1}\text{ and achieves equality when }k=i+1,$$

又由于 $(2r+1)(i+1)-r^2-r\le(i+1)^2$，且当 $r=i$ 或 $r=i+1$时取等号，因此我们有

$$\gamma^{(2r+1)(i+1)-r^2-r}\le\gamma^{(i+1)^2}\text{ and achieves equality when }r=i\text{ or }r=i+1.$$

因此，

$$\begin{aligned}\pi_i(A_{i+1})&=\frac{\frac{1}{Z_i}\left(\gamma^{(i+1)^2+1}+\sum_{r=0}^L\gamma^{(2r+1)(i+1)-r^2-r}\right)}{\sum_{k=1}^m\left[\frac{1}{Z_i}\left(\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^L\gamma^{(2r+1)(i+1)-r^2-r}\right)\right]}\\&>\frac{\gamma^{(i+1)^2+1}}{\gamma^{(i+1)^2+1}+(L+1)^2\gamma^{(i+1)^2}}=\frac{(L+1)^3}{(L+1)^3+(L+1)^2}>1-\frac{1}{L+1},\end{aligned}$$

其中，第二步是因为

$$\gamma^{(i+1)^2+1}+\sum_{r=0}^L\gamma^{(2r+1)(i+1)-r^2-r}>\gamma^{(i+1)^2+1}$$

and

$$\sum_{k=1}^{m}\left(\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}\right)<\gamma^{(i+1)^2+1}+\sum_{k=1}^{m}\sum_{r=0}^{L}\gamma^{(i+1)^2}$$
$$=\gamma^{(i+1)^2+1}+(L+1)^2\gamma^{(i+1)^2};$$

这里，我们利用了这一事实： $m=L+1$。对于 $k\neq i+1$，同理可得

$$\pi_i(A_k)=\frac{\frac{1}{Z_i}\left(\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}\right)}{\sum_{h=1}^{m}\left[\frac{1}{Z_i}\left(\gamma^{2(i+1)h-h^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}\right)\right]}$$
$$>\frac{\gamma^{(i+1)^2}}{\gamma^{(i+1)^2+1}+(L+1)^2\gamma^{(i+1)^2}}=\frac{1}{(L+1)^3+(L+1)^2}>\frac{1}{2(L+1)^3}$$

and

$$\pi_i(A_k)<\frac{(L+1)\gamma^{(i+1)^2}}{\gamma^{(i+1)^2+1}+(L+1)\gamma^{(i+1)^2}}=\frac{L+1}{(L+1)^3+(L+1)}<\frac{1}{(L+1)^2}.$$

引理 11 的证明至此完成。 □

接下来，我们给出瓶颈比的一个下界。

引理 12. $B>\ {}^{1}\!/_{(L+1)^7}$，其中 B 是该链的瓶颈比。

证明. 固定 $i\in[L]$ 和 $k\neq i$、$i+1$ 或 $i+2$ 后，我们首先给出 $\pi_{i-1}(A_k)\,\pi_i(A_k)$ 的一个下界。我们计算

$$\pi_i(A_k)=\frac{\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}}{\sum_{h=1}^{m}\left(\gamma^{2(i+1)h-h^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}\right)}$$
$$=\frac{\gamma^{2(i+1)k-k^2+1}+\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}}{\sum_{r=1}^{L+1}\gamma^{2(i+1)r-r^2+1}+(L+1)\sum_{r=0}^{L}\gamma^{(2r+1)(i+1)-r^2-r}}$$
$$=\frac{\gamma^{1-(i+1-k)^2}+\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}}{\sum_{r=1}^{L+1}\gamma^{1-(i+1-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}},$$

and

$$\pi_{i-1}(A_k)=\frac{\gamma^{1-(i-k)^2}+\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}}{\sum_{r=1}^{L+1}\gamma^{1-(i-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}}.$$

因为当 $r=i$ 或 $i+1$ 时，$k\neq i$、$i+1$ 或 $i+2$，$\gamma^{1-(i+1-k)^2}\leq\frac{1}{\gamma}=\ {}^{1}\!/_{L+1()^3}$，且 $\gamma^{-(i+1-r)(i-r)}=1$，而当 $r=i$ 或 $i+1$ 时，$\gamma^{-(i+1-r)(i-r)}\leq\frac{1}{\gamma}=\ {}^{1}\!/_{L+1()^3}$，否则，我们有

$$\gamma^{1-(i+1-k)^2}+\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}<2+(L+1)\frac{1}{(L+1)^3}=2+\frac{1}{(L+1)^2}. \tag{11}$$

此外，因为当 $r=i$ 或 $i-1$ 时 $\gamma^{-(i-r)(i-1-r)}=1$，我们有

$$\gamma^{1-(i-k)^2}+\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}>2, \tag{12}$$

由此得到

$$\frac{\gamma^{1-(i-k)^2}+\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}}{\gamma^{1-(i+1-k)^2}+\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}}>\frac{2}{2+\frac{1}{(L+1)^2}}>1-\frac{1}{(L+1)^2}.$$

另一方面 另一方面，由于当 $r=i$ 时 $\gamma^{1-(i-r)^2}=(L+1)^3$，否则 $\gamma^{1-(i-r)^2}\le 1$，我们有 e

$$\sum_{r=1}^{L+1}\gamma^{1-(i+1-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}<(L+1)^3+L+2(L+1)+(L+1)^2\frac{1}{(L+1)^3}<(L+1)^3+3(L+1),$$

$$\sum_{r=1}^{L+1}\gamma^{1-(i-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}>(L+1)^3+2(L+1),$$

从而得到

$$\frac{\sum_{r=1}^{L+1}\gamma^{1-(i+1-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}}{\sum_{r=1}^{L+1}\gamma^{1-(i-r)^2}+(L+1)\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}}>\frac{(L+1)^3+2(L+1)}{(L+1)^3+3(L+1)}>1-\frac{1}{(L+1)^2}. \tag{13}$$

因此，我们有

$$\frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}>\left(1-\frac{1}{(L+1)^2}\right)^2>1-\frac{1}{L+1}.$$

当 $k=i$ 时，根据引理 11，我们有

$$\frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}>\frac{1-\frac{1}{L+1}}{\frac{1}{(L+1)^2}}>1.$$

当 $k=i+1$ 时，再次根据引理 11，我们有

$$\frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}>\pi_{i-1}(A_k)>\frac{1}{2(L+1)^3}.$$

当 $k=i+2$ 时，我们有 $\gamma^{1-(i+1-k)^2}=1$，因此式（11）弱化为

$$\gamma^{1-(i+1-k)^2}+\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}<3+\frac{1}{(L+1)^2},$$

同时式（12）和式（13）继续成立。因此，我们有

$$\frac{\gamma^{1-(i-k)^2}+\sum_{r=0}^{L}\gamma^{-(i-r)(i-1-r)}}{\gamma^{1-(i+1-k)^2}+\sum_{r=0}^{L}\gamma^{-(i+1-r)(i-r)}}>\frac{2}{3+\frac{1}{(L+1)^2}}>\frac{2}{3}\left(1-\frac{1}{(L+1)^2}\right),$$

and

$$\frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}>\frac{2}{3}\left(1-\frac{1}{(L+1)^2}\right)^2>\frac{2}{3}\left(1-\frac{1}{L+1}\right).$$

综合上述结果，我们可以得出

$$B=\min_{k\in[m]}\prod_{i=1}^{L}\min\left\{1,\frac{\pi_{i-1}(A_k)}{\pi_i(A_k)}\right\}>\frac{2}{3}\left(1-\frac{1}{L+1}\right)^L\frac{1}{2(L+1)^3}>\frac{1}{3e(L+1)^3}>\frac{1}{(L+1)^7},$$

其中，最后一步是因为 $L+1\ge 2$。至此，引理 12 的证明完成。 □

现在回过头来证明定理 3。

定理 3 的证明。令 $\mathcal{X}_\lambda$ 为从 $\lambda=(1,\dots,m)$ 可通过仅交换相邻层级之间的样本到达的状态集合，即，对所有 $i\in\{0,\dots,L\}$，$\lambda_i=i+1$；并令 S 为 $\mathcal{X}_\lambda$ 的子集，其中包含所有可通过仅交换相邻层级之间的样本从 λ 到达，且所有中间状态与 λ 至多在 $\lfloor\log L\rfloor-1$ 个层级上不同的状态。令 $\overline{\pi}_\lambda=\overline{\pi}_{\mathcal{X}_\lambda}$ 为限制在 $\mathcal{X}_\lambda$ 上的投影链的密度。对于 $X\subset\mathcal{X}_\lambda$，令

$$\varphi(X) = \frac{\sum_{\xi\in X,\widetilde{\xi}\in X^C} \overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi})}{\min\{\overline{\pi}_\lambda(X),\overline{\pi}_\lambda(X^C)\}},$$

其中 $X^C = \mathcal{X}_\lambda \setminus X$；令 $\varphi^* = \min_{X:\overline{\pi}_\lambda(X)\le\frac{1}{2}} \varphi(X)$。根据切格不等式，

$$\mathbf{Gap}(P_{pt}) \le 2\varphi^* \le 2\varphi(S), \tag{14}$$

接下来，我们给出 $\varphi(S)$ 的上界。我们将在以下引理中分别界定 $\varphi(S)$ 的各个部分。

引理 13. $\overline{\pi}_\lambda(S) > \frac{1}{2e}$。

证明。由引理 11，

$$\overline{\pi}_\lambda(S) \ge \overline{\pi}(\lambda) = \prod_{i=0}^{L} \pi_i(A_i) > \left(1-\frac{1}{L+1}\right)^{L+1} > \frac{1}{2e}.$$ □

引理 14. $\overline{\pi}_\lambda(\mathcal{X}_\lambda \setminus S) > \frac{1}{4e(L+1)^6}$。

证明。令 $\mathcal{P}$ 为 $\{2,\dots,m-1\}$ 的所有排列构成的集合。令 $\widetilde{S}$ 为满足 λ_0 位于第 L^{th} 层级且 λ_L 位于第 0th 层级的状态集合。由引理9，$\widetilde{S} \subset \mathcal{X}_\lambda \setminus S$，因此

$$\begin{aligned}
\overline{\pi}_\lambda(S^C) \ge \overline{\pi}(S^C) \ge \overline{\pi}(\widetilde{S}) &\ge \pi_0(A_m)\pi_L(A_1)\sum_{\sigma\in\mathcal{P}}\prod_{i=1}^{L-1}\pi_i(A_{\sigma(i+1)}) \\
&> \left(\frac{1}{2(L+1)^3}\right)^2 \prod_{i=1}^{L-1}\pi_i(A_{i+1}) \\
&> \frac{1}{4(L+1)^6}\left(1-\frac{1}{L+1}\right)^{L-1} > \frac{1}{4e(L+1)^6}.
\end{aligned}$$

其中第四步和第五步由引理 11 得出。 □

引理 15. $\sum_{\xi\in S,\ \widetilde{\xi}\in S^C} \overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi}) \le 4e\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-2}$。

证明。设 S_1 为有 $\lfloor\log L\rfloor - 1$ 个层级错位的状态集合，S_2 为有$\lfloor\log L\rfloor - 2$ 个层级错位的状态集合。注意，为了使 $\xi \in S$、$\widetilde{\xi} \in S^C$ 和 $\overline{P}_{pt}(\xi,\widetilde{\xi}) > 0$ 成立，我们必有$\xi \in S_1$ 或 $\xi \in S_2$。因此，

$$\begin{aligned}
\sum_{\xi\in S,\widetilde{\xi}\in S^C} \overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi}) &= \sum_{\xi\in S_1,\widetilde{\xi}\in S^C} \overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi}) + \sum_{\xi\in S_2,\widetilde{\xi}\in S^C} \overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi}) &(15)\\
&\le \sum_{\xi\in S_1}\overline{\pi}_\lambda(\xi) + \sum_{\xi\in S_2}\overline{\pi}_\lambda(\xi). &(16)
\end{aligned}$$

对于每个 $\xi \in S_1$，有 $\binom{L+1}{\lfloor\log L\rfloor-1}$ 种选择来确定哪些样本错位，并且错位样本有 $(\lfloor\log L\rfloor-1)!$ 种排列。此外，根据引理 11，对于所有 $i\in\{0,\dots,L\}$ 和 $k \ne i+1$，$\pi_i(A_k) < \frac{1}{(L+1)^2}$。因此，我们有

$$\sum_{\xi\in S_1}\overline{\pi}(\xi) \le \binom{L+1}{\lfloor\log L\rfloor-1}\cdot(\lfloor\log L\rfloor-1)!\cdot\left(\frac{1}{(L+1)^2}\right)^{\lfloor\log L\rfloor-1},$$

并且由于 $\overline{\pi}(\mathcal{X}_\lambda) \ge \overline{\pi}(\lambda) > \frac{1}{2e}$，

$$\begin{aligned}
\sum_{\xi\in S_1}\overline{\pi}_\lambda(\xi) = \frac{1}{\overline{\pi}(\mathcal{X}_\lambda)}\sum_{\xi\in S_1}\overline{\pi}(\xi) &< 2e\binom{L+1}{\lfloor\log L\rfloor-1}\cdot(\lfloor\log L\rfloor-1)!\cdot\left(\frac{1}{(L+1)^2}\right)^{\lfloor\log L\rfloor-1} &(17)\\
&\le 2e(L+1)^{\lfloor\log L\rfloor-1}\left(\frac{1}{(L+1)^2}\right)^{\lfloor\log L\rfloor-1} = 2e\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-1}. &(18)
\end{aligned}$$

类似地，

$$\sum_{\xi\in S_2}\overline{\pi}_\lambda(\xi)<2e\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-2}. \tag{19}$$

将式 (18) 和 (19) 代回式 (16)，得到

$$\sum_{\xi\in S,\widetilde{\xi}\in S^C}\overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi})<4e\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-2}. \qquad \square$$

结合引理 12、13、14、15 和式 (14)，得到

$$\begin{aligned}\mathbf{Gap}(P_{pt})\le 2\varphi^*\le 2\varphi(S)&=\frac{2\sum_{\xi\in S,\widetilde{\xi}\in S^C}\overline{\pi}_\lambda(\xi)\overline{P}_{pt}(\xi,\widetilde{\xi})}{\min\{\overline{\pi}_\lambda(S),\overline{\pi}_\lambda(S^C)\}}\\&<\frac{8e\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-2}}{\min\left\{\frac{1}{2e},\frac{1}{4e(L+1)^6}\right\}}=32e^2\left(\frac{1}{L+1}\right)^{\lfloor\log L\rfloor-8}<O\left(B^{O(\log L)}\right),\end{aligned}$$

这就完成了定理 3 的证明。 $\square$

5 结论与未来方向

本文的主要技术贡献是为并行回火的谱隙给出了一个改进的下界：除 $\log L$ 外，该下界对所有参数都呈多项式依赖；同时还给出了一个假设性的上界。不过，这只是理解这些算法混合时间的一小步，我们在此提出一些开放问题。首先，值得探究是否存在一个自然的例子，使得谱隙的上界仍会随 $\log L$ 呈指数级变化。其次，鉴于 [2] 为模拟退火提供了谱隙下界，而模拟退火是另一种有效的多峰分布采样算法，并且对广泛的一类分布，该下界对所有参数都呈多项式依赖，因此值得研究是否存在这样的实例：模拟退火表现良好，而并行回火却不然。此外，还可以探究如何将我们的界应用于实际分布，以获得更好的混合时间保证。请注意，$\log L$ 仅为一个下界；在加入更多假设后，该界可能会更好。

参考文献

[1] Persi Diaconis 和 Daniel Stroock。马尔可夫链特征值的几何界。《应用概率年刊》，1(none)，1991。[2] Rong Ge、Holden Lee 和 Andrej Risteski。超越对数凹性：使用模拟退火朗之万蒙特卡洛对多峰分布进行采样的可证明保证，2017。[3] Neal Madras 和 Zhongrong Zheng。论交换算法。《随机结构与算法》，22:66–97，2003年1月。[4] Gareth Roberts 和 Richard Tweedie。可逆马尔可夫链的几何 l2 收敛与 l1 收敛等价。《应用概率杂志》，38A，2000年5月。[5] Gareth O. Roberts 和 Jeffrey S. Rosenthal。一般状态空间马尔可夫链与马尔可夫链蒙特卡罗算法。《概率调查》，1(none)，2004年1月。[6] Dawn Woodard、Scott Schmidler 和 Mark Huber。并行回火与模拟退火迟滞混合的充分条件。《电子概率杂志》 *[仅电子版]*, 14, 2009年8月。[7] Dawn B. Woodard、Scott C. Schmidler 和 Mark Huber。多峰分布上并行回火与模拟退火快速混合的条件。《应用概率年刊》，19(2)，2009年4月。