POLYNOMIAL COMPLEXITY SAMPLING FROM MULTIMODAL DISTRIBUTIONS USING SEQUENTIAL MONTE CARLO

RUIYU HAN, GAUTAM IYER, AND DEJAN SLEPČEV

Abstract. We study a sequential Monte Carlo algorithm to sample from the Gibbs measure with a non-convex energy function at a low temperature. We use the practical and popular geometric annealing schedule, and use a Langevin diffusion at each temperature level. The Langevin diffusion only needs to run for a time that is long enough to ensure local mixing within energy valleys, which is much shorter than the time required for global mixing. Our main result shows convergence of Monte Carlo estimators with time complexity that, approximately, scales like the forth power of the inverse temperature, and the square of the inverse allowed error. We also study this algorithm in an illustrative model scenario where more explicit estimates can be given.

1. Introduction

We show that under general non-degeneracy conditions, the Annealed Sequential Monte Carlo algorithm (detailed in Algorithm 1) produces samples from multimodal distributions with time complexity that is a polynomial in the inverse temperature, with a precise dimension independent degree. We begin (Section 1.1) with an informal description of the algorithm, and our results. Following this we survey (Section 1.2) the literature, provide a gentle introduction to the area, and place our work in the context of existing results. Our main results are stated precisely (Section 2) below, and the remainder of this paper is devoted to the proofs.

1.1. Informal statement of main results.

Let $U\colon \mathcal{X} \to \mathbb{R}$ be an energy function defined on a configuration space $\mathcal{X}$. Consider the Gibbs distribution π_ε whose density is given by

$$\pi_\varepsilon(x) = \frac{1}{Z_\varepsilon}\tilde{\pi}_\varepsilon(x), \quad \text{where} \quad \tilde{\pi}_\varepsilon(x) \stackrel{\text{def}}{=} e^{-U(x)/\varepsilon} \text{ and } Z_\varepsilon \stackrel{\text{def}}{=} \int_{\mathcal{X}} \tilde{\pi}_\varepsilon(y)\,dy. \tag{1.1}$$

where dy denotes some fixed measure on the configuration space $\mathcal{X}$. In many applications arising in physics, the parameter $\varepsilon > 0$ is proportional to the absolute temperature. We adopt (and abbreviate) this terminology and will subsequently refer to the parameter ε as the temperature. In this paper the space $\mathcal{X}$ will typically be the d-dimensional Euclidean space $\mathbb{R}^d$, or the torus $\mathbb{T}^d$.

Our aim is to study convergence of an *Annealed Sequential Monte Carlo (ASMC)* algorithm. This is a *Sequential Monte Carlo (SMC)* algorithm (see for instance [CP20,

---

2020 *Mathematics Subject Classification.* Primary: 60J22, Secondary: 65C05, 65C40, 60J05, 60K35.

Key words and phrases. Markov Chain Monte Carlo, sequential Monte Carlo, annealing, multimodal distributions, high dimensional sampling.

This work has been partially supported by the National Science Foundation under grants DMS-220606, DMS-2342349, DMS-2406853, and the Center for Nonlinear Analysis.

Chapters 3.3, 17], or [Liu08, Chapter 3.4]), where particles are moved through a sequence of interpolating measures obtained by gradually reducing the temperature according to a specified annealing schedule. We use the practical and popular geometric annealing schedule where the inverse temperatures are linearly spaced [SBCCD24]. Our main result shows convergence of Monte Carlo estimators using ASMC with time complexity that, approximately, scales like the forth power of the inverse temperature, and the square of the inverse allowed error.

Before stating our main result, we briefly recall the ASMC algorithm.

1. Choose a finite sequence of temperatures $\eta_1 > \eta_2 \cdots > \eta_M$ (called an *annealing schedule*) so that π_{η_1} is easy to sample from and $\eta_M = \eta$ is the desired final temperature.
2. Choose a family of Markov processes $\{Y_{\varepsilon,\cdot}\}_{\varepsilon>0}$ so that for every $\varepsilon > 0$ the stationary distribution of $Y_{\varepsilon,\cdot}$ is π_ε, and fix a running time $T > 0$.
3. Choose arbitrary initial points $y_1^1, \ldots, y_N^1$.
4. For each $i \in \{1, \ldots, N\}$, run (independent) realizations of $Y_{\eta_1,\cdot}$ for time T, starting from y_1^i, to obtain x_1^i.
5. Assign each point x_1^i the weight $\tilde{\pi}_{\eta_2}(x_1^i)/\tilde{\pi}_{\eta_1}(x_1^i)$. Choose $(y_2^1, \ldots, y_2^N)$ to be a *resampling* of the points $(x_1^1, \ldots, x_1^N)$ from the multinomial distribution with probabilities proportional to the assigned weights.
6. Repeat the previous two steps, reducing the temperature until the final temperature is reached.

This is stated more precisely as Algorithm 1 in Section 2.1, below. Clearly if we choose T larger than the *mixing time* of $Y_{\eta,\cdot}$ at the final temperature $\eta = \eta_M$, then the above procedure will produce good samples from π_η. This, however, is not practical – when U is not convex the mixing time of $Y_{\eta,\cdot}$ grows exponentially with $1/\eta$. When η is small waiting for the mixing time of $Y_{\eta,\cdot}$ at the desired final temperature computationally infeasible. We will instead show that we only need to choose T to be larger than the mixing time of $Y_{\eta_1,\cdot}$ at the *initial temperature* η_1. Since η_1 is large, this is computationally tractable. The price we pay is *only polynomially many* temperature levels M, provided we use the popular *geometric annealing schedule*, [VCK25], where the inverse temperatures are linearly spaced (and hence the densities form a geometric sequence).

Roughly speaking, our main result is as follows.

Theorem 1.1. *Suppose $U\colon \mathbb{T}^d \to \mathbb{R}$ is a non-degenerate double-well function with wells of equal depth (but not necessarily the same shape). For $\varepsilon > 0$ let $Y_{\varepsilon,\cdot}$ be a solution to the overdamped Langevin equation*

$$dY_{\varepsilon,t} = -\nabla U(Y_{\varepsilon,t})\,dt + \sqrt{2\varepsilon}\,dW_t, \tag{1.2}$$

where W is a standard d-dimensional Brownian motion on the torus. There exists constants C_N, C_T, depending on U and d, such that the following holds. For any $\delta > 0$, $\eta > 0$, choose M, N, T according to

$$M = \left\lceil \frac{1}{\eta} \right\rceil, \quad N = \frac{C_N M^2}{\delta^2}, \quad \textit{and} \quad T \geqslant C_T\Big(M^2 + \log\Big(\frac{1}{\delta}\Big) + \frac{1}{\eta}\Big)$$

and a suitable geometric annealing schedule $\{1/\eta_k\}_{k=1,\ldots,M}$ so that η_1 is sufficiently large, and $\eta_M = \eta$. Then the points $x^1, \ldots, x^N$ obtained from ASMC (with the *parameters above) are such that for any bounded test function h we have*

$$\boldsymbol{E}\Big(\frac{1}{N}\sum_{i=1}^{N} h(x^i) - \int_{\mathbb{T}^d} h(x)\pi_\eta(x)\,dx\Big)^2 < \|h\|_{\mathrm{osc}}^2 \delta^2.$$

Theorem 1.1 shows that the time complexity of obtaining good samples from π_η using ASMC is polynomial in $1/\eta$, with a degree independent of dimension. We note that the drift in (1.2) is independent of temperature ε, and so computational complexity of ASMC is proportional to MNT. In contrast, the time complexity of obtaining good samples by directly simulating the process $Y_{\eta,\cdot}$ is $e^{O(1/\eta)}$.

The assumption that U has a double-well structure is mainly to simplify the technical presentation. Our proof will generalize without difficulty to the situation where U has more than two wells, at the expense of several technicalities that further obscure the heart of the matter. As a result we only present the proof of Theorem 1.1 in the double-well setting.

We assumed that the wells have equal depth above only for simplicity. Our main result (Theorem 2.8) will generalize of Theorem 1.1 so that it applies to a large class of double-well energy functions, where the low temperature sampling problem we study is a nondegenerate in the sense that each well has a non-negligible fraction of the total mass. The precise assumptions required are laid out in Section 4.1. In particular, Lemma 4.4 shows that if the wells have *nearly equal depth*, then the problem is nondegenerate and Theorem 2.8 applies.

We also remark that Theorem 1.1 requires no prior knowledge of the location or the depth of the wells. In particular, if the target distribution is a mixture, we require no knowledge of the decomposition of the domain into components of a mixture, and only require access to the energy and its gradient.

The main tool used in the proof is a spectral decomposition. This decomposes any initial distribution into components corresponding to the (target) stationary distribution, a mass imbalance between wells, and higher order terms. The higher order terms decay exponentially and do not present a problem. The term corresponding to the mass imbalance extremely slowly (at a rate that is exponentially small in the inverse temperature), and is the bottleneck.

This, precisely, is the term that can be eliminated using ASMC. At high temperatures, all terms converge rapidly, and it is easy to obtain samples with a small mass imbalance. The resampling step used to move between temperature levels does not disturb this much, resulting in a distribution that has a small mass imbalance at a lower temperature. Iterating this should, in principle, yield samples from distributions that have a small mass imbalance at every temperature level. This is the main idea behind of the proof of Theorem 1.1, and is presented in Section 4, below.

The details of the proof, however, are somewhat involved. To precisely quantify the error at each level, we require precise bounds on the shape of the eigenfunctions, and how they change with temperature. In particular, the proof relies on bounding the inner-product between the normalized eigenfunctions at successive temperature levels, which involves dimensional constants that are not explicit. As a result, the constants C_N, C_T in Theorem 1.1 are not explicit. Moreover, the assumption that the state space is the compact torus $\mathbb{T}^d$, while inconsequential to issues arising from of multimodality, is required to ensure the validity of some of our spectral estimates.

To obtain a better understanding of the dynamics, and a more explicit constants, we also study ASMC in an idealized scenario. In this idealized scenario, we assume the domain is divided into J energy valleys, and we have access to a Markov process that mixes quickly in each valley but very slowly globally. In this case (Theorem 2.2, in Section 2.2, below) we also prove polynomial time complexity bounds, but obtain more explicit constants, and control their dimensional dependence.

We numerically illustrate some aspects of the performance of ASMC algorithm in Section 2.4. In particular we highlight how the algorithm adjusts the mass in each valley, which can change as temperature changes and we investigate how the accuracy of the algorithm depends on the number of levels M, under fixed computational budget. A reference implementation is provided in [HIS25].

1.2. **Literature review.** We first recall the widely used sampling techniques for nice (e.g. log-concave) measures and then discuss the literature on sampling multimodal distributions.

1.2.1. *General sampling algorithms.* Perhaps the simplest practical technique for drawing samples of the target distribution $\pi \propto \exp(-U)$ is based on rejection sampling. One first draws samples of distribution μ from which exact samples can be obtained easily (say a Gaussian or a Lebesgue measure on a square), and is such that π is absolutely continuous with respect to μ. One then accepts these samples with probability proportional to $\frac{d\pi}{d\mu}$. If the measure π is much more concentrated than μ the acceptance probability becomes very small. For a Gibbs distribution in d-dimensions at temperature η, the acceptance rate is typically proportional to $1/\eta^d$, making the cost of this method prohibitively expensive.

Thus in high-dimensions, a different approach is needed. Most of the widely used methods are based on a stochastic process whose invariant measure is π. The largest class of these are Markov Chain Monte Carlo (MCMC) methods which include the seminal Metropolis-Hastings algorithm, Langevin Monte Carlo, Metropolis adjusted Langevin Algorithm (MALA), Hamiltonian Monte Carlo (HMC) and others [SAAG24]. We now briefly recall some of the main algorithms, as any of these can be used in the step 2 of ASMC (the Markov transition step), provided it rapidly mixes within the modes of the distribution.

The Langevin Monte Carlo (LMC) algorithm relies on updating individual particles following the overdamped Langevin equation (1.2). The law of the solution, denoted by μ_t^ε, satisfies the Fokker–Planck equation and converges to the stationary distribution π_ε exponentially as $t \to \infty$. If the energy function U is uniformly convex, and satisfies $\alpha I \leqslant \operatorname{Hess} U$, then it is known that the 2-Wasserstein distance converges exponentially with rate α (i.e. $W_2(\mu_t^\varepsilon, \pi_\varepsilon) \leqslant \exp(-\alpha t) W_2(\mu_0, \pi_\varepsilon)$). To use this algorithm in practice, one needs to discretize the SDE, which is often done using the explicit Euler–Maruyama scheme. Convergence of the time discretized SDE were proved in [VW19] using KL divergence, and in [Che23] using W_2.

The general Langevin dynamics allows for inertial effects and is modeled by a system of an SDE for momentum and ODE for position. This property is the foundation of popular Hamiltonian Monte Carlo (HMC) algorithm, which extends the configuration space to include the momentum variable p, and considers Hamiltonian dynamics whose invariant measure is $\pi^H \propto \exp\big(-U(x) - \frac{1}{2}|p|^2\big)$. Observe that the first marginal of π^H is exactly the target Gibbs measure π_1 (with temperature $\varepsilon = 1$). To numerically obtain samples from π^H, the HMC algorithm alternates between the flow of the Hamiltonian dynamics in the phase space, and drawing a new random momentum, whose marginal distribution is a standard Gaussian. The optimal convergence rate for the idealized (i.e. one with an exact Hamiltonian dynamics solver) HMC was proved by Chen and Vempala in [CV22].

While the dynamics above have π_1 as the invariant measure at the continuum level, this is not preserved at the level of numerical schemes resulting in bias that the estimates above control. The original Metropolis–Hastings algorithm [MRR+53, Has70] offers an algorithm where the target measure π_1 is the invariant measure at the discrete level. The algorithm proposes a new sample from a (simple) proposal distribution, and then accepts/rejects the proposal in a manner that ensures the desired target distribution is the invariant measure. This accept/reject step can be combined with several other methods. In particular when added to LMC one obtains the popular MALA algorithm. Other algorithms in this direction include the Proximal sampler, both of which are studied in [Che23].

1.2.2. *Sampling from multimodal distributions.* Sampling from multimodal distributions, is challenge that none of the algorithms like LMC, MALA, or HMC can effectively overcome as their convergence rate becomes extremely slow as the separation between the modes increases. As a result, there is a broad spectrum of works studying algorithms that are suitable for sampling multimodal distributions.

Annealed importance sampling, introduced by Neal [Nea01], involves drawing samples from a sequence of auxiliary distributions starting from starting from one that is easy to sample from, and ending with the target distribution. Samples are moved from one distribution to the next by a reweighting procedure, and then improved by iterating a Markov chain. The algorithm outputs a set of weighted sample points representing the target distribution. While this is extremely popular and versatile, one drawback is that the variance of the weights can become extremely large, and with most of the mass being distributed over only a few points [CP20, Chapter 9].

Sequential Monte Carlo (SMC) was first developed to study of the average extension of molecular chains [HM54, RR55]. Its use in sampling [DdFG01, CP20, SBCCD24] can be seen as generalization of AIS, with the addition of a key resampling step that leads to balanced particle weights. SMC algorithms are enormously popular in a variety of applications and numerous modifications have been developed.

There are a number of works that consider convergence of SMC including obtaining central limit theorems [Cho04, CP20]. As remarked in Section 11.2.4 of [CP20], the variance of the error typically grows exponentially with the number of levels. The variance can be controlled [CP20, Section 11.4] under restrictive conditions that do not apply to multimodal distributions.

The works of Schweitzer [Sch12] and Paulin, Jasra, and Thiery [PJT19] are the first to rigorously consider the convergence of SMC for multimodal distributions and prove bounds on the variance of the error. However their assumptions require a strong stability condition on the underlying Markov kernels, which can not be used to in the context of Theorem 1.1. Building on the coupling technique developed in [MMS23], Matthews and Schmidler [MS24] prove finite sample error bounds for SMC in multimodal setting. Their assumption on the underlying Markov kernels is restrictive requiring knowledge of partition of the domain corresponding to the modes, and also can not be used in the context of Theorem 1.1.

The recent work of of Lee and Santana-Gijzen [LSG24] takes a similar angle as our work in that it shows convergence results for SMC under assumptions of local mixing within the wells and boundedness of ratios of the densities of consecutive levels. While these assumptions resemble the assumptions we make, there is a key difference: they require a sequence of interpolating measures where the mass in each component of the mixture is known and remains constant. Devising such an interpolation sequence requires knowing the components of the mixture, which is not available in many practical problems and in particular precludes using interpolations based on adjusting the temperature in the Gibbs measure, such as the geometric annealing we study.

The main differentiating factor between our work, and the SMC papers mentioned above, is that we do not require structural assumptions on the underlying Markov kernels, and do not require any prior knowledge of the mixture components. As such, our result, stated in Theorem 2.8, is the first to provide polynomial time complexity bounds for ASMC using Langevin diffusions and a geometric annealing schedule.

Parallel, simulated, and related tempering methods. Parallel tempering was introduced in a form by Swendsen and Wang [SW86] and developed by Geyer in [Gey91]. Simulated tempering introduced by Marinari and Parisi [MP92] and developed further by Geyer [GT95]. These algorithms rely on Markov chains that run on a product space of the desired configuration space and various levels of the temperature. Samples drawn at a particular value of the temperature may be modified into samples from either a higher, or a lower temperature. At the lowest temperature the marginal of the invariant measure on the product space is the target measure, while at the highest it is a measure where the Markov chain mixes rapidly.

There are notable results on rigorously showing convergence of parallel and simulated tempering. In particular, Woodard, Schmidler, and Huber [WSH09a] obtain conditions under which tempering methods are rapidly mixing. When applied to sampling multimodal distributions the authors considered distributions which have separated modes, but require the variance near each mode to be of size one. Thus their results do not address the low temperature regime that Theorem 1.1 applies to. In [WSH09b] the authors prove that the mixing of these tempering approaches slows exponentially with dimension if components of multimodal measures have different variances. If all the modes have the same shape, Ge, Lee, and Risteski [GLR18, GLR20] show the convergence in TV norm of simulated tempering with error rates that are polynomial in inverse temperature and dimension, provided we have initial estimate on the ratio of the normalizing constants. The precise degree of the polynomial, however, is not explicitly identified.

Further tempering methods in this family include tempered transitions introduced by Neal in [Nea96], which rely on compositions of transitions steps that result in jumps at the lowest temperature and tempered Hamiltonian Monte Carlo [Nea11]. Though, to the best of our knowledge, there are no results that apply in the setting of Theorem 1.1 and provide polynomial time complexity bounds.

Annealing without reweighting or resampling. There are a number of annealing approaches that evolve a measure from one that is easy to sample from, to the desired target distribution. In particular, the annealed Langevin Monte Carlo considers Langevin dynamics with slowly changing stationary measure [GTC25,VCK25]. These papers show rigorous convergence results for target measures satisfying restrictive structure conditions. In general the annealed LMC lacks a way to easily adjust the mass within a well at low temperatures. As a result, the convergence rate is exponentially small in the inverse temperature, and this was rigorously shown for geometric tempering schedule in [VCK25].

Further approaches. Some recent papers explore new avenues to sampling multimodal distributions. These include approaches based on exploring ideas from diffusion models [VCK25] where the authors show rigorous complexity bounds. This method, however, suffers from the curse of dimensionality and the error bounds scale like δ^d, where δ is allowable error and d is dimension.

The work [PHLa20] proposed a framework of MCMC algorithms for multimodal sampling, which combines an optimization step to find the modes with Markov transition steps. They showed the weak law of large numbers of Monte Carlo integral using samples generated by the Auxiliary Variable Adaptive MCMC algorithm.

Another direction explored is to use ensemble methods that involve Markov Chains whose jump rates use the estimating the density of the measure represented by the particle configuration [LLN19, LWZ22, LSW23]. These approaches can be seen as particle approximations of gradient flows of KL divergence in spherical Hellinger metric, which converge exponentially fast with rate that is independent of the height of the barrier. However this method also suffers from the curse of dimensionality, as the kernel density estimation used to estimate density based on the configuration of particles introduces bias that becomes large in high dimensions.

A few methods modify (1.2) in a manner that allows particles to move between modes faster. The authors of [ERY24] do this by modifying the diffusion, and the authors of [RBS15, DFY20, CFIN23] do this by introducing an additional drift term. In both cases the modified equation has terms that grow exponentially with the inverse temperature, and a numerical implementation is computationally expensive.

Plan of the paper. In Section 2 we precisely state our algorithm, and state results guaranteeing convergence both for ASMC in an idealized scenario (Theorem 2.2), and for a double-well energy function (Theorem 2.8, which generalizes Theorem 1.1). For the idealized scenario we are able to obtain explicit constants, and track the dimensional dependence (Proposition 3.1). Numerical simulations illustrating relevant aspects of the performance of ASMC in model situations are shown in Section 2.4. We prove Theorems 2.2 and 2.8 in Sections 3 and 4 respectively. The proof of Theorem 2.2 relies on a few lemmas which are proved in Section 5. The proof of Theorem 2.8 is a little more involved and the required lemmas are proved in Sections 6, 7 and 8 respectively. Finally in Section 9 we show that that regular enough energy functions satisfy the assumptions required for Theorems 2.2 and 2.8, and obtain the dimension independent stated in Proposition 3.1.

2. Main results

2.1. **Annealed Sequential Monte Carlo (ASMC).** We now briefly introduce the ASMC algorithm, which is stated precisely as Algorithm 1, below. In many situations of interest, the configuration space $\mathcal{X}$ admits a decomposition into *energy valleys.* MCMC samplers (such as (1.2)) are typically confined to an energy valley for time $e^{O(1/\varepsilon)}$ before moving to a different valley (see for instance [Arr89]). Of course, waiting time $e^{O(1/\varepsilon)}$ to explore the state space is practically infeasible, and directly using an MCMC sampler is prohibitively slow at low temperatures.

Annealing and tempering (both terms having origin in metallurgy and describing heat treatment of metals) based algorithms, in particular ASMC we study, have been introduced to overcome the issue of slow global mixing of the MCMC algorithms. ASMC a special case of a *sequential Monte Carlo* algorithm, as samples are drawn in sequence from an auxiliary family of distributions, starting from one that is easy to sample from and ending with the target distribution. The name ASMC stems from the fact that the auxiliary family of distributions used are obtained by starting from the Gibbs distribution at a high temperature, and then gradually lowering the temperature until the desired temperature is reached.

To use ASMC, we choose an *annealing schedule*, which is a sequence of temperatures $\eta_1 > \eta_2 \cdots > \eta_M$, chosen so that the MCMC sampler converges fast at temperature η_1, the desired final temperature is $\eta_M = \eta$. Samples at temperature η_k are transformed to samples at temperature η_{k+1} by reweighting them with the ratio of densities $\pi_{\eta_{k+1}}/\pi_{\eta_k}$. To ensure the mass is spread across sample points, the weights are redistributed using a *resampling* process. The samples are then improved by iterating an MCMC sampler for a fixed amount of time, and then the above processes is repeated at the next temperature until the final temperature is reached.

It is important to note that for the reweighting step, one *does not* have access to the normalized densities π_{η_k} in practice, as the normalization constants are not known and are hard to compute. However, using weights proportional to the ratio of the normalized densities $\pi_{\eta_{k+1}}/\pi_{\eta_k}$ is *equivalent* to using weights proportional to the ratio of the unnormalized densities $\tilde{\pi}_{\eta_{k+1}}/\tilde{\pi}_{\eta_k}$. The unnormalized densities are known, and are used in the reweighting step instead of the normalized densities.

We now describe the *resampling* step: given points $x_k^1, \dots, x_k^N$ which are (approximate) samples from π_{η_k}, we obtain $y_{k+1}^1, \dots, y_{k+1}^N$ by resampling from the points $\{x_k^1, \dots, x_k^N\}$ using the multinomial distribution with probabilities

$$\boldsymbol{P}(y_{k+1}^i = x_k^j) = \frac{\tilde{r}_k(x_k^j)}{\sum_{n=1}^N \tilde{r}_k(x_k^n)}, \quad \text{where} \quad \tilde{r}_k \stackrel{\text{def}}{=} \frac{\tilde{\pi}_{\eta_{k+1}}}{\tilde{\pi}_{\eta_k}}. \tag{2.1}$$

Some points may be repeated or lost. Nevertheless, an elementary heuristic (explained in Section 3.3, after (3.12), below) suggests that the new points $y_{k+1}^1, \dots, y_{k+1}^N$ should be good samples from $\pi_{\eta_{k+1}}$.

Remark 2.1. Instead of resampling at every step, modern, practical algorithms typically control the variance of the weights using more sophisticated resampling procedures. A popular approach is to introduce a measure of the quality of the weight distribution and only resample when the quality becomes lower than a desired threshold, which is called *adaptive resampling*. For this and other resampling approaches see, for instance, the books [CP20, Sections 10.2] or [Liu08, Chapter 3.4].

We now provide a brief heuristic explanation as to why one may be able to obtain good quality samples in polynomial time using Algorithm 1. First, since η_1 is large and the process $Y_{\eta_1,\cdot}$ mixes quickly, and so the distribution of $x_1^1, x_1^2, \dots, x_1^N$ will be close to the Gibbs measure π_{η_1}. Now the resampling step may produce degenerate samples with several repeated points. However the fraction of points in each energy valley will be comparable to the π_{η_2}-mass of the same valley. In the situation we consider, the main bottleneck to fast mixing is moving mass between valleys. Since

Algorithm 1 Annealed Sequential Monte Carl (ASMC) to sample from π_η.

Require: Temperature η, energy function U, and Markov processes $\{Y_{\varepsilon,\cdot}\}_{\varepsilon \geqslant \eta}$ so that the stationary distribution of $Y_{\varepsilon,\cdot}$ is π_ε.

Tunable parameters:

(1) Number of levels $M \in \mathbb{N}$, and annealing schedule $\eta_1 > \cdots > \eta_M = \eta$.
(2) Number of realizations $N \in \mathbb{N}$, and initial points $y_1^1, \ldots, y_1^N \in \mathcal{X}$.
(3) Level running time $T > 0$.

for $k \in \{1, \ldots, M-1\}$ **do**
  For each $i \in \{1, \ldots N\}$, simulate $Y_{\eta_k,\cdot}$ for time T starting at y_k^i to obtain x_k^i.
  Choose $(y_{k+1}^1, \ldots, y_{k+1}^N)$ by resampling from $\{x_k^1, \ldots, x_k^N\}$ using the multinomial distribution with probabilities given by (2.1).
end for
For each $i \in \{1, \ldots N\}$, simulate $Y_{\eta_M,\cdot}$ for time T starting at y_M^i to obtain x^i.
return $(x^1, \ldots, x^N)$.

the samples at temperature η_2 have approximately the right fraction of mass in each energy valley, the distribution after running $Y_{\eta_2,\cdot}$ for time T will be close to the Gibbs distribution π_{η_2}. Repeating this argument should iteratively yield good samples at the desired final temperature η_M.

A rigorous proof of the above quantifying the convergence rate, however, requires some care. The number of levels M is large (grows linearly in the inverse temperature), and the error going from level k to $k+1$ accumulates *multiplicatively*. Nevertheless, we will show that if $\eta_1, \ldots, \eta_M$ according to the geometric annealing schedule, then the total error accumulates slowly enough that Algorithm 1 produces good samples in time that is polynomial in $1/\eta$. Carrying out the details of this heuristic for a double-well energy function using Langevin diffusions as the MCMC sampler (as described in Theorem 1.1) is technical, and requires several model specific bounds that distract from the main idea. Thus, we first consider an illustrative model problem where we can study Algorithm 1, and then revisit it in the context of Theorem 1.1.

2.2. **ASMC for a Local Mixing Model.** We now present an idealized scenario where we can analyze Algorithm 1 quantitatively, and obtain explicit constants in our error estimates. Suppose the number of components of the multimodal measure, $J \geqslant 2$, and the domain $\mathcal{X}$ can be partitioned into J domains $\Omega_1, \ldots, \Omega_J$. We are interested situations where we have access to a process $Y_{\varepsilon,\cdot}$ that mixes quickly in each domain Ω_j, however, transitions very slowly between domains and hence mixes slowly overall.

To model this behavior, for every $\varepsilon > 0$ let $\chi_\varepsilon \in (0,1)$ denote probability of staying in the same domain after time 1. Let $Y_{\varepsilon,\cdot}$ be the discrete time Markov process defined as follows. At time $n \in \mathbb{N}$, let j be the unique element of $\{1, \ldots, J\}$ such that $Y_{\varepsilon,n} \in \Omega_j$. Flip an independent coin that lands heads with probability χ_ε and tails with probability $1 - \chi_\varepsilon$. If the coin lands heads, we choose $Y_{\varepsilon,n+1} \in \mathcal{X}$ independently from the distribution π_ε. If the coin landed tails, we choose $Y_{\varepsilon,n+1} \in \Omega_j$ independently from the distribution with density

$$\frac{\pi_\varepsilon \mathbf{1}_{\Omega_j}}{\pi_\varepsilon(\Omega_j)}\,.$$

In other words, $Y_{\varepsilon,\cdot}$ is the Markov process whose one step transition density is

$$p_1^\varepsilon(x,y) = (1-\chi_\varepsilon)\pi_\varepsilon(y) + \chi_\varepsilon \sum_{j=1}^{J} \mathbf{1}_{\{x,y\in\Omega_j\}} \frac{\pi_\varepsilon(y)}{\pi_\varepsilon(\Omega_j)}\,. \tag{2.2}$$

Notice that the expected transition time between domains is the bottleneck to mixing, and is of order $1/(1-\chi_\varepsilon)$. One situation of interest, is when χ_ε is extremely close to 1 (for instance $\chi_\varepsilon \approx \exp(e^{-O(1/\varepsilon)})$). This models the behavior that arises several applications of interest, including Langevin dynamics driven by the gradient of an energy function with multiple wells, and this is studied in detail in Section 2.3, below. In such situations waiting for time $1/(1-\chi_\varepsilon)$ is prohibitively expensive when ε is small, and can not be done in practice.

Suppose now we are interested in computing Monte Carlo integrals with respect to the Gibbs distribution π_η for some small temperature $\eta > 0$. A direct Monte Carlo approach simulating $Y_{\eta,\cdot}$ is unfeasible as it requires simulating $Y_{\eta,\cdot}$ for time $O(1/(1-\chi_\eta))$, which very long when η is small. We now show that Algorithm 1, with a judicious choice of parameters, makes this time an order of magnitude smaller.

Theorem 2.2. *Suppose for some* $0 \leqslant \eta_{\min} < \eta_{\max} \leqslant \infty$ *we have*

$$C_{\mathrm{LBV}} \stackrel{\text{def}}{=} \sum_{j=1}^{J} \int_{\eta_{\min}}^{\eta_{\max}} |\partial_\varepsilon \ln \pi_\varepsilon(\Omega_j)| \, d\varepsilon < \infty. \tag{2.3}$$

For any finite $\eta_1 \in (\eta_{\min}, \eta_{\max}]$, $\delta, \eta, \nu > 0$ *with* $\eta \in [\eta_{\min}, \eta_1)$, *and constants* C_T, $C_N > 0$ *choose* $M, N, T \in \mathbb{N}$ *so that*[1]

$$M \geqslant \Big\lceil \frac{1}{\nu\eta} \Big\rceil, \qquad N \geqslant \frac{C_N M^2}{\delta^2}, \qquad \textit{and} \qquad T \geqslant t_{\mathrm{mix}}\Big(Y_{\eta_1,\cdot}, \frac{\delta}{C_T}\Big), \tag{2.4}$$

and choose $\eta_2, \dots, \eta_M$ *so that* $\eta_M = \eta$ *and* $1/\eta_1, \dots, 1/\eta_M$ *are linearly spaced.*

For every $\delta, \nu > 0$, *there exists (explicit) constants* $C_N = C_N(U/\eta_1, J, \nu)$, *and* $C_T = C_T(U/\eta_1, J, \nu)$ *such that if the process* $Y_{\varepsilon,\cdot}$ *in Algorithm 1 have transition density* (2.2), *and if the parameters to Algorithm 1 are chosen as in* (2.4), *then for every bounded test function* h, *and arbitrary initial data* $\{x_0^i\}$, *the points* $(x^1, \dots, x^N)$ *returned by Algorithm 1 satisfy*

$$\Big\| \frac{1}{N} \sum_{i=1}^{N} h(x^i) - \int_{\mathcal{X}} h(x) \pi_\eta(x)\, dx \Big\|_{L^2(\boldsymbol{P})} < \|h\|_{\mathrm{osc}} \delta. \tag{2.5}$$

We prove Theorem 2.2 in Section 3, below.

Remark 2.3. In (2.5) above, we clarify that the Monte Carlo sum $\frac{1}{N}\sum_1^N h(x^i)$ is a random variable, as the points x^i are random, and the notation $\|\cdot\|_{L^2(\boldsymbol{P})}$ denotes the $L^2(\boldsymbol{P})$ norm with respect to the underlying probability measure $\boldsymbol{P}$. Explicitly, if X is a random variable, then $\|X\|_{L^2(\boldsymbol{P})} = (\boldsymbol{E}X^2)^{1/2}$.

[1] Here $t_{\mathrm{mix}}(Y_{\varepsilon,\cdot}, \delta)$ denotes the δ-mixing time of the process $Y_{\varepsilon,\cdot}$ (see for instance [LP17]), and measures the TV-rate of convergence of $Y_{\varepsilon,\cdot}$ to the stationary distribution π_ε. Explicitly, if p_n^ε denotes the n-step transition density of $Y_{\varepsilon,\cdot}$, then the δ-mixing time is given by

$$t_{\mathrm{mix}}(Y_{\varepsilon,\cdot}, \delta) \stackrel{\text{def}}{=} \inf\Big\{ n \in \mathbb{N} \;\Big|\; \sup_{x\in\mathcal{X}} \|p_n^\varepsilon(x,\cdot) - \pi_\varepsilon(\cdot)\|_{L^1} \leqslant 2\delta \Big\}.$$

Remark 2.4. The constants C_N and C_T can be computed explicitly in terms of the energy function U, the number of domains J and the parameter ν and their values are stated precisely in Section 3.2, below.

Remark 2.5. We will show (Corollary 9.1, below) that the finiteness condition (2.3) will be satisfied for a large class of double-well energy functions where the sampling at low temperature is a nondegenerate problem and each well has a non-negligible fraction of the total mass. In particular, (2.3) will hold for double-well functions with wells of *nearly equal depth*. More precisely, if the difference between the energies at local minima is comparable to the minimum temperature $\eta_{\min}$, then C_{LBV} can be bounded independent of $\eta_{\min}$.

Remark 2.6. If the difference between the energies at local minima is much larger than the minimum temperature $\eta_{\min}$, then the as $\varepsilon \to \eta_{\min}$ some of the domains Ω_j will contain a fraction of the total mass which is exponentially small in the inverse temperature. Hence the multimodal nature of the target distribution degenerates, and the sampling from this distribution requires the simulation of rare events. This goes beyond the scope of the present work and Theorem 2.2 does not apply.

Remark 2.7. Theorem 2.2 shows that the *averaged* empirical measure is TV close to the Gibbs distribution. Explicitly, the averaged empirical measure μ is defined by

$$\mu(A) \stackrel{\text{def}}{=} \frac{1}{N}\boldsymbol{E}\sum_{i=1}^{N}\delta_{x^i}(A) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{P}(x^i \in A),$$

where $x^1, \dots, x^N$ are the points returned by Algorithm 1. Now Theorem 2.2 and Jensen's inequality immediately imply $|\mu(A) - \pi_\eta(A)| \leqslant \delta$ for every Borel set A, and hence

$$\|\mu - \pi_\eta\|_{\text{TV}} \leqslant \delta.$$

Computational Complexity. We now estimate the computational cost of Monte Carlo integration using Theorem 2.2 and compare it to the direct approach using the process $Y_{\eta,\cdot}$. In this idealized situation, we assume the cost of simulating the process Y^ε for time T is $O(T)$. Using an alias method [Vos91] one can perform the resampling step in time $O(N)$, which makes the computational cost of Algorithm 1 of order MNT. To estimate T, we need to estimate the δ-mixing time of the process $Y_{\varepsilon,\cdot}$ for $\varepsilon = \eta_1$. For this, we use (2.2) to deduce that the n-step transition density of $Y_{\varepsilon,\cdot}$ is

$$p_n^\varepsilon(x,y) = (1-\chi_\varepsilon^n)\pi_\varepsilon(y) + \chi_\varepsilon^n\sum_{j=1}^{J}\mathbf{1}_{\{x,y\in\Omega_j\}}\frac{\pi_\varepsilon(y)}{\pi_\varepsilon(\Omega_j)}\,.$$

This immediately implies

$$t_{\text{mix}}(Y_{\varepsilon,\cdot},\delta) \leqslant \frac{\ln(\delta/2)}{\ln\chi_\varepsilon} \approx \frac{|\ln\delta|}{1-\chi_\varepsilon}.$$

Thus Theorem 2.2 implies the computational cost of running Algorithm 1 to achieve the Monte Carlo error (2.5) is

$$\text{cost}(\text{Algorithm 2.2}) = O(MNT) \leqslant \frac{C(U)|\ln\delta|}{\eta^3\delta^2|\ln\chi_{\eta_1}|}, \tag{2.6}$$

for some U-dependent constant $C(U)$.

On the other hand, achieving the same Monte Carlo error by simulating independent realizations of $Y_{\eta,\cdot}$ has a computational cost of

$$\text{cost(Direct Monte Carlo)} = O\Big(\frac{\ln\delta}{\delta^2|\ln\chi_\eta|}\Big). \tag{2.7}$$

We note that the cost of ASMC in (2.6) involves a polynomial in the final temperature η, and the mixing time at the initial temperature η_1, which is small. In contrast, the direct Monte Carlo cost (2.7) involves the mixing time at the *final* temperature η, which is typically exponential in $1/\eta$.

2.3. **ASMC for a double-well energy function.** We now study Algorithm 1 when the configuration space $\mathcal{X}$ is the d-dimensional torus $\mathbb{T}^d$. Here the Gibbs measure π_ε arises naturally as the stationary distribution of the *overdamped Langevin equation* (1.2).

When U is convex, the process Y mixes quickly even in high dimensions [BGL14], and provides a very efficient way to sample from the Gibbs distribution π_ε. When U is not convex, however, the process $Y_{\varepsilon,\cdot}$ mixes extremely slowly. In fact, the well known Arrhenius law [Arr89] states that in general it takes time $t \approx e^{C/\varepsilon}$ before the distribution of $Y_{\varepsilon,\cdot}$ becomes close to the Gibbs distribution π_ε. At low temperatures, this is too long to be practical.

The reason Langevin dynamics mixes so slowly is because the drift in (1.2) pulls trajectories towards local minima of U. In order to escape an energy valley, the noise term in (1.2) has to go against the drift for an $O(1)$ amount of time, which happens with *exponentially small* probability. In each energy valley, however, the energy function U is essentially convex which makes the process $Y_{\varepsilon,\cdot}$ mix quickly in valleys. We also note that the situation considered in Section 2.2 is an idealized model for the dynamics of (1.2).

We study Algorithm 1 for target distributions corresponding to double-well energy functions and show that appropriate choice of parameters allows one to compute integrals with respect to the Gibbs distribution, with time complexity that is polynomial in the inverse temperature. We again remark that the assumption that U is a double-well energy function is mainly to simplify the presentation, and the generalization to energy functions with more wells is straightforward.

Theorem 2.8. *Suppose for some $0 \leqslant \eta_{\min} < \eta_{\max} \leqslant \infty$, the function U is a double-well function that satisfies Assumptions 4.1, 4.2 and 4.3 in Section 4 below. Let $\hat\gamma_r \geqslant 1$ be the ratio of the saddle height to the energy barrier, defined precisely in (4.3), below. Given $\eta_1 \in (\eta_{\min}, \eta_{\max}]$ finite, $\alpha, \delta, \eta, \nu > 0$ with $\eta \in [\eta_{\min}, \eta_1)$, and constants C_T, $C_N > 0$ choose $M, N \in \mathbb{N}$, and $T \in \mathbb{R}$ so that*

$$M \geqslant \Big\lceil \frac{1}{\nu\eta} \Big\rceil, \quad T \geqslant C_T\Big(M^{(1+\alpha)\hat\gamma_r} + \log\Big(\frac{1}{\delta}\Big) + \frac{1}{\eta}\Big) \quad \text{and} \quad N \geqslant \frac{C_N M^2}{\delta^2}, \tag{2.8}$$

and choose $\eta_2, \dots, \eta_M$ so that $\eta_M = \eta$ and $1/\eta_1, \dots, 1/\eta_M$ are linearly spaced.

For every $\alpha, \delta, \nu > 0$, there exist constants $C_T = C_T(\alpha, \nu, U/\eta_1)$ and $C_N(\nu, U/\eta_1)$ such that if the process $Y_{\varepsilon,\cdot}$ in Algorithm 1 is given by (1.2), and the parameters to Algorithm 1 are chosen as in (2.8), then for every bounded test function h, and arbitrary initial data $\{x_0^i\}$, the points $(x^1, \dots, x^N)$ returned by Algorithm 1 satisfy

$$\Big\| \frac{1}{N}\sum_{i=1}^N h(x^i) - \int_{\mathbb{T}^d} h(x)\pi_\eta(x)\,dx \Big\|_{L^2(\boldsymbol{P})} < \|h\|_{\mathrm{osc}}\delta. \tag{2.9}$$

We remark that Assumptions 4.1–4.3 are nondegeneracy assumptions, and do not require symmetry, or similarity of the shape of the wells. The proof of Theorem 2.8 follows the same general strategy as that of Theorem 2.2, however the details more technically involved. In Theorem 2.2 the main idea is to show that if at level k, the initial mass distributions in the domains $\Omega_1, \ldots, \Omega_J$ is distributed according to π_{η_k}, then the process $Y_{\eta_k,\cdot}$ will correct the shape and quickly give a distribution that is close to the Gibbs distribution π_{η_k}. To show this in the context of (1.2), we consider a spectral decomposition based on eigenvalues of the generator of (1.2). We will show that if the projection of the initial distribution onto the *second* eigenspace is small, then the Langevin dynamics will quickly correct the shape and yield a distribution close to the Gibbs measure. The proof of this involves several technical lemmas controlling the shape of the eigenfunctions and introduces dimensional pre-factors that are not explicit. This takes up the bulk of the paper and begins in Section 4, below.

Time and computational complexity. We now briefly discuss the computational cost of integration using Theorem 2.8. Suppose U is a double-well function with wells of equal depth, so that $\hat{\gamma}_r$ is exactly 1. As mentioned earlier, the resampling step costs $O(N)$ and so the time complexity of running Algorithm 1 (for $\eta < 1$ with $\nu = 1$) to achieve the error tolerance (2.9) is

$$O(MTN) \leqslant \frac{\tilde{C}_d}{\eta^3 \delta^2} T \leqslant \frac{C_d}{\eta^3 \delta^2} \Big(\frac{1}{\eta^{1+\alpha}} + \log\Big(\frac{1}{\delta}\Big)\Big), \tag{2.10}$$

for some dimensional constants $\tilde{C}_d, C_d$. The second inequality gives us the precise, polynomial, time complexity of the algorithm. The significance of the first inequality is that the computational complexity of the algorithm is, up to dimensional constant, $\frac{1}{\eta^3\delta^2}$ times the computational complexity of the numerical algorithm which mixes the distribution sufficiently well within the wells.

In order to use Algorithm 1 in practice, one has to time discretize (1.2) and consider the bias induced by this discretization. Obtaining the precise errors for numerical discretizations of LMC and other algorithms is an active area of research, and we refer the reader to the notes by Chewi [Che23] for comprehensive overview. Obtaining rigorous computational complexity of ASMC is a challenging open problem, as the wells are not exactly log-concave and one would need to control various terms in our proof up to discretization error. We remark, however, that in our formulation the drift in (1.2) is independent of the temperature ε, and so for our purposes, the number of iterations required to simulate (1.2) for a given length of time T is proportional to time T and independent of the temperature ε. Let us also remark that, assuming the scaling for smooth log-concave wells can be reached the estimates of Chewi [Che23, Theorem 4.1.2] suggest that the total computational complexity of the algorithm, in terms of the number of evaluations of ∇U, would be $c_U d \frac{1}{\delta^2}$ times the time complexity, when applied to integrating bounded, Lipschitz continuous functions. The restriction to a smaller class of test functions is needed since the Wasserstein error controlled in Theorem [Che23, Theorem 4.1.2] needs to control the integration error. We remark that better error bounds can be obtained by using different discretizations of LMC and by using MALA (see [Che23]).

For comparison we note the cost of using rejection sampling to achieve a comparable error is C_d/η^d, which is huge when η is small and the dimension is large. Also, the cost of using LMC requires simulating $1/\delta^2$ realizations of (1.2) for a time that is comparable to the δ-mixing time. By the Arrhenius law [Arr89, BGK05] this is $e^{O(1/\eta)}$, which is much larger than (2.10) when η is small.

2.4. **Numerical experiments.** In practice, one typically needs to sample from the target distribution with density proportional to e^{-V}, for some given energy function V. In situations of interest the energy V has deep valleys and the associated Gibbs measure has several components (modes). To apply ASMC, we choose a temperature $\eta > 0$, which is small enough so that the Gibbs measure with energy function

$$U \stackrel{\text{def}}{=} \eta V,$$

is easy to sample from. Then we run Algorithm 1 with the energy function U, with initial temperature $\eta_1 = 1$, and final temperature η to deliver samples from the Gibbs measure with density proportional to e^{-V}. A reference implementation is provided in [HIS25].

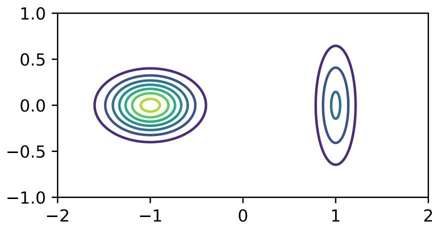

FIGURE 1. Contour plot of the anisotropic Gaussian mixture in $\mathbb{R}^2$, defined in (2.11), and used in experiments for Figure 2.

For the first illustration, we consider a two-dimensional distribution. In this case integrals with respect to the Gibbs distribution can also be effectively computed using quadrature, and can be used as a reference for our numerical simulations. We choose the Gibbs measure to be a mixture of two dimensional, anisotropic Gaussians given by

$$\pi = \sum_{i=1}^{2} a_i G_{\mu_i, \Sigma_i}. \tag{2.11}$$

Here $G_{\mu,\Sigma}$ is the PDF of the two dimensional Gaussian with mean μ and covariance matrix Σ. We choose parameters $a_1 = 0.7$, $a_2 = 0.3$, $\mu_1 = -e_1$, $\mu_2 = e_1$ and

$$\Sigma_1 = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.04 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.18 \end{pmatrix}$$

A contour plot of π is shown on the left of Figure 1. The left panel of Figure 2 shows the results of numerical simulations computing the Monte Carlo integral of the indicator function of a separating hyperplane using samples from Algorithm 1. For comparison, we also show the results of computing the same integral using direct LMC, and using quadrature. To generate this plot we used $N = 10^4$, time step 0.0025, $M = 5$, $T = 500$. For confirmation, we verify the mean error and standard deviation decrease like $1/\sqrt{N}$, and show our results in Figure 2.

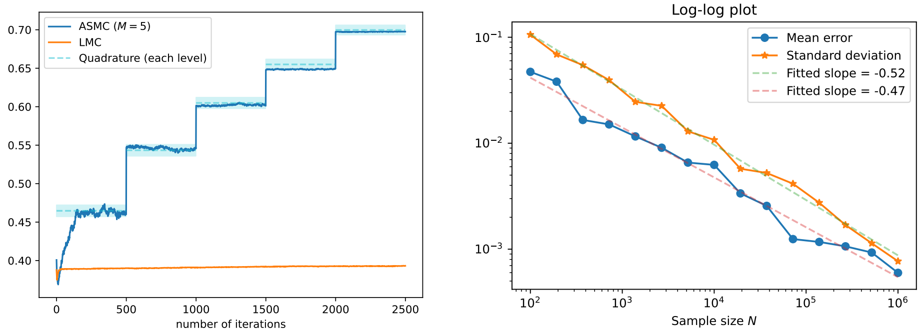

FIGURE 2. Left: A Monte Carlo integral computed using ASMC, LMC, and quadrature in 2D. Right: A Log-log plot of the mean error and standard deviation using ASMC as the number of particles varies.

Our next experiment, illustrated in Figure 3, focuses on the trade-off between increasing the number of levels and the number of time steps under a fixed computational budget. We use samples obtained by Algorithm 1 to compute a Monte Carlo integral in dimension 20. The target measure is a mixture of Gaussians given by (2.11) with parameters $a_1 = 0.2$, $a_2 = 0.8$, $\mu_1 = -e_1$, $\mu_2 = e_1$ and

$$\Sigma_1 = \frac{1}{16} I_d, \quad \Sigma_2 = \frac{1}{25} I_d, \quad d = 20.$$

We vary the number of levels M and the level running time T, while keeping the total number of iterations MT constant. To generate the plots we used a total of 5000 iterations per run, sample size $N = 10^4$ and time step 0.001, and 100 independent Monte Carlo runs per choice of M and T. We observe that ASMC produces good results for intermediate values of M, but performs poorly when M is too large or too small when compared to d. We note that this is not surprising. When $M = 1$ ASMC becomes the rejection sampler, and with M being small it is closely approximating a rejection sampler with a few intermediate levels. Since the jumps in temperature are large the resulting bias is large. When M is very large and T is quite small the Markov transitions do not have a chance to mix even within the wells. Thus the procedure basically only involves importance reweighting and resampling, thus leading to most of the mass concentrated at few nodes, and large error.

3. Error Estimates for the Local Mixing Model (Theorem 2.2)

3.1. **Notation and convention.** Before delving into the proof of Theorem 2.2 we briefly list notational conventions that will be used throughout this paper.

(i) We will always assume $C > 0$ is a finite constant that can increase from line to line, provided it does not depend on the temperature η.

(ii) We use the convention that the expectation operator $\boldsymbol{E}$ has lower precedence than multiplication. That is $\boldsymbol{E}XY$ denotes the expectation of the product $\boldsymbol{E}[XY]$, and $\boldsymbol{E}X^2$, denotes the expectation of the square $\boldsymbol{E}[X^2]$

(iii) When taking expectations and probabilities, a subscript will denote the conditional expectation / conditional probability. That is $\boldsymbol{E}_X Y = \boldsymbol{E}(Y \mid X)$ denotes the conditional expectation of Y given the σ-algebra generated by X.

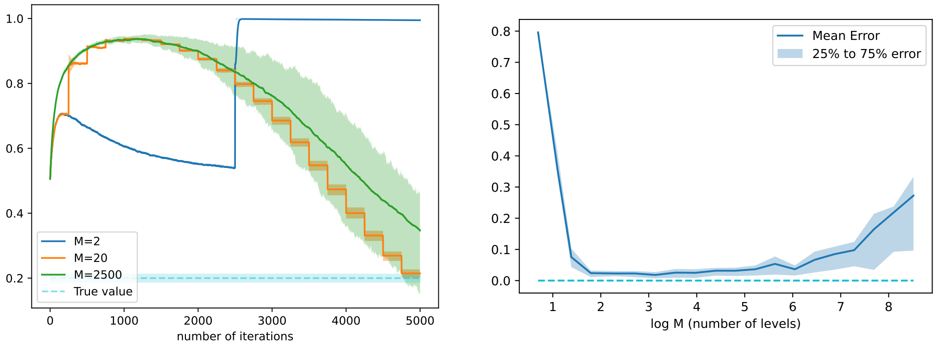

FIGURE 3. Mean error of an integral in dimension $d = 20$ computed using ASMC as M, T vary, while holding MT constant. Shaded regions indicate the 25%-75% quintile range. Left: A plot of the Monte Carlo integral vs the number of iterations for a few values of M. Right: A plot of the mean error vs $\log M$.

(iv) When averaging functions of Markov processes, a superscript will denote the initial distribution. That is $\boldsymbol{E}^\mu f(Y_t)$ denotes $\boldsymbol{E} f(Y_t)$ given $Y_0 \sim \mu$. When $\mu = \delta_y$ is the Dirac δ-measure supported at y, we will use $\boldsymbol{E}^y$ to denote $\boldsymbol{E}^{\delta_y}$.

(v) We interchangeably use π_ε to denote the measure and the density. That is for $x \in \mathbb{T}^d$, $\pi_\varepsilon(x)$ is given by (1.1), however for Borel sets A, $\pi_\varepsilon(A)$ denotes $\int_A \pi_\varepsilon(x)\,dx$.

3.2. **Description of the Constants in Theorem 2.2.** As remarked earlier, the constants C_N and C_T in Theorem 2.2 are explicit. Since these determine the efficiency of Algorithm 1, we state them precisely before embarking on the proof of Theorem 2.2. First we need auxiliary constant $C_r = C_r(U/\eta_1, \nu)$ that will be used to bound the ratio of the densities at each level. For $k \in \{1, \dots, M\}$, by a slight abuse of notation we define

$$\pi_k \stackrel{\text{def}}{=} \pi_{\eta_k}, \quad \tilde{\pi}_k \stackrel{\text{def}}{=} \tilde{\pi}_{\eta_k} \quad \text{and} \quad Z_k \stackrel{\text{def}}{=} Z_{\eta_k}$$

where π_{η_k}, $\tilde{\pi}_{\eta_k}$ and Z_{η_k} are defined by (1.1) with $\varepsilon = \eta_k$. Next we define

$$r_k \stackrel{\text{def}}{=} \frac{\pi_{k+1}}{\pi_k} \tag{3.1}$$

to be the ratio of *normalized* densities at levels $k+1$ and k. In practice, we do not have access to r_k as we do not have access to the normalization constants Z_k. This is why Algorithm 1 is formulated using the ratio of unnormalized densities $\tilde{r}_k$ defined in (2.1). The auxiliary constant C_r mentioned above is defined by

$$C_r \stackrel{\text{def}}{=} \max_{1 \leqslant k \leqslant M-1} \|r_k\|_{L^\infty(\mathcal{X})}. \tag{3.2}$$

Clearly $C_r \to 1$ as $\nu \to 0$. However, choosing ν very small increases the number of levels M and hence the computational cost of Algorithm 1. A bound for C_r, which may be easier to check in practice, is

$$C_r \leqslant \inf_{c>0} (1 + s_c) \exp(c\nu) \tag{3.3}$$

where

$$s_c \stackrel{\text{def}}{=} \frac{\int_{\{U_0>c\}} e^{-U_0}\,dx}{\int_{\{U_0\leqslant c\}} e^{-U_0}\,dx} < \infty \quad \text{and} \quad U_0 \stackrel{\text{def}}{=} \frac{U - \inf U}{\eta_1}. \tag{3.4}$$

We prove (3.3) in Lemma 9.2, below.

Now, the proof of Theorem 2.2 will show that constants C_T and C_N are given by

$$C_T \stackrel{\text{def}}{=} 4JC_r(2C_\beta+1), \quad C_N \stackrel{\text{def}}{=} J^2(2C_\beta+1)^2(1+C_r)^2. \tag{3.5}$$

where

$$C_\beta \stackrel{\text{def}}{=} \exp(2C_rC_{\mathrm{LBV}}). \tag{3.6}$$

Dimensional dependence. Suppose now $\mathcal{X} = \mathbb{R}^d$. For a certain class of energy functions, it is possible to make the constants C_T, C_N independent of d by choosing a geometric annealing schedule with M linear in d. One such class of energies are those which can be separated into a sum of two functions – one which depends on the first $\tilde{d}$ coordinates and may have multiple local minima; and the other only depends on the last $d-\tilde{d}$ coordinates and is convex.

Explicitly, suppose there exists an integer $\tilde{d} \leqslant d$ such that the function U is of the form

$$U_0(x) = \tilde{U}_0(x_1,\dots,x_{\tilde{d}}) + V_0(x_{\tilde{d}+1},\dots,x_d). \tag{3.7}$$

Here $\tilde{U}_0$ is an any function for which $e^{-\tilde{U}_0}$ is integrable, and may have several local minima. The function V_0 is assumed to be a convex function for which there exist constants $\alpha_0 > 0, k_0 > 1$, $\alpha_u, \alpha_b \in \mathbb{R}$ and a point $x_0 \in \mathbb{R}^{d-\tilde{d}}$ such that for all $x \in \mathbb{R}^{d-\tilde{d}}$, we have

$$\alpha_0|x-x_0|^{k_0} + \alpha_b \leqslant V_0(x) \leqslant \alpha_0|x-x_0|^{k_0} + \alpha_u. \tag{3.8}$$

One class of functions that have this structure are Gaussian mixtures of points that are located on a $\tilde{d}$ dimensional hyperplane, and whose covariance matrices in the perpendicular direction are all equal. For such energies we have the following dimension independent bounds.

Proposition 3.1. *Assume that U_0 satisfies and (3.7) and (3.8). Choose*

$$M \geqslant \left\lceil \frac{d}{\eta} \right\rceil$$

and η_k such that $1/\eta_1, \dots, 1/\eta_M$ are linearly spaced. Then C_T and C_N in (3.5) can be bounded above in terms of α_0, α_b, α_u, k_0, U_0, but independent *of d.*

Proposition 3.1 can be proved using asymptotics for the incomplete gamma function and is presented in Section 9, below.

3.3. **Proof of Theorem 2.2.** In order to prove Theorem 2.2, we note that Algorithm 1 consists of repeating two steps: (local) exploration using using the process $Y_{\varepsilon,\cdot}$ (Algorithm 1, step 2), and then resampling (Algorithm 1, step 3). We now state lemmas for the errors accumulated in each of these steps.

To quantify the Monte Carlo error made by running the process $Y_{\varepsilon,\cdot}$ in the (local) exploration step, we introduce the following notation. Given $\varepsilon, t > 0$ and a bounded test function h, define the Monte Carlo error $\operatorname{Err}_{\varepsilon,t}(h)$ by

$$\operatorname{Err}_{\varepsilon,t}(h) \stackrel{\text{def}}{=} \Big\| \frac{1}{N} \sum_{i=1}^{N} h(Y^i_{\varepsilon,t}) - \int_{\mathbb{T}^d} h \pi_\varepsilon \, dx \Big\|_{L^2(\boldsymbol{P})},$$

where $Y^i_{\varepsilon,\cdot}$ are N independent realizations of a Markov process with transition density (2.2).

Lemma 3.2. *Given N (random) points $y^1, \dots, y^N$, let $Y^i_{\varepsilon,\cdot}$ be N independent realizations of the Markov process with transition density* (2.2) *and initial distribution $Y^i_{\varepsilon,0} = y^i$. Then for any bounded test function h, and any $T \in \mathbb{N}$ we have*

$$\operatorname{Err}_{\varepsilon,T}(h) \leqslant \chi_\varepsilon^T \Big\| \sum_{j=1}^{J} \Big(1 - \frac{\mu_0(\Omega_j)}{\pi_\varepsilon(\Omega_j)}\Big) \int_{\Omega_j} h \pi_\varepsilon \, dx \Big\|_{L^2(\boldsymbol{P})} + \frac{\|h\|_{\mathrm{osc}}}{2\sqrt{N}}, \tag{3.9}$$

where μ_0 is the empirical measure

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{N} \delta_{y^i}. \tag{3.10}$$

Consequently,

$$\operatorname{Err}_{\varepsilon,T}(h) \leqslant \frac{\|h\|_{\mathrm{osc}}}{2} \Big(\chi_\varepsilon^T \sum_{j=1}^{J} \operatorname{Err}_{\varepsilon,0}(\mathbf{1}_{\Omega_j}) + \frac{1}{\sqrt{N}} \Big). \tag{3.11}$$

We clarify that $\mu_0(\Omega_j)$ is random as the initial points y^i are themselves random. The second term on the right of (3.11) is the standard Monte Carlo error which can be made small by making N large. To make the first term small, we have two options: The first option is to wait for the mixing time of $Y_{\varepsilon,\cdot}$, and obtain smallness from the χ_ε^T factor. The second is to ensure $\sum_j \operatorname{Err}_{\varepsilon,0}(\mathbf{1}_{\Omega_j})$ is small. In our situation the first option is undesirable as it requires $T \gg 1/|\ln \chi_\varepsilon|$, which is too large to be practical. Instead we use the second option, and make $\sum_j \operatorname{Err}_{\varepsilon,0}(\mathbf{1}_{\Omega_j})$ small by ensuring the fraction of initial points in each domain Ω_j is close to $\pi_\varepsilon(\Omega_j)$.

We now turn to the resampling step. Suppose we have N i.i.d. samples $x^1, \dots, x^N$ from a distribution with an unnormalized probability density function $\tilde{p}$. Let $\tilde{q}$ be another unnormalized probability density function, such that $\{\tilde{q} > 0\} \subseteq \{p > 0\}$. Choose $(y^1, \dots, y^N)$ to be a *resampling* of the points $(x^1, \dots, x^N)$ using the multinomial distribution with probability

$$\boldsymbol{P}(y^i = x^j) = \frac{\tilde{r}(x^j)}{\sum_{i=1}^{N} \tilde{r}(x^i)}, \quad \text{where} \quad \tilde{r} \stackrel{\text{def}}{=} \frac{\tilde{q}}{\tilde{p}}. \tag{3.12}$$

Of course, some of the points x^i may be chosen multiple times and the points $y^1, \dots, y_N$ may not be distinct. Nevertheless, a simple heuristic argument suggests that when N is large the distribution of each of the points y^i will have a density proportional to $\tilde{q}$. Indeed, suppose $\mathcal{X}$ is finite, $N \gg |\mathcal{X}|$ and p, q are the normalized probability distributions corresponding to p, q respectively. Then each $x \in \mathcal{X}$ occurs amongst the points $\{x^1, \dots, x^N\}$ roughly $Np(x)$ times, and so

$$\boldsymbol{P}(y^i = x) \approx \frac{\tilde{r}(x) N p(x)}{\sum_{x' \in \mathcal{X}} \tilde{r}(x') N p(x')} = \frac{\tilde{q}(x)}{\sum_{x' \in \mathcal{X}} \tilde{q}(x')} \approx q(x).$$

To make the above quantitative, and usable in our context, some care has to be taken. The points y^i are only conditionally independent given $x^1, \dots, x^N$; they are not unconditionally independent, and it is hard to estimate the unconditional joint distribution. We will instead obtain a Monte Carlo estimate which both quantifies the error and is sufficient for our purposes.

Lemma 3.3. *Suppose $x^1, \dots, x^N$ are N (not necessarily i.i.d.) random points in $\mathcal{X}$. Let $\tilde{p}$, $\tilde{q}\colon \mathcal{X} \to [0, \infty)$ be two unnormalized probability density functions, and choose $y^1, \dots, y^N$ independently from $\{x^1, \dots, x^N\}$ according to (3.12). Then for any test function $h \in L^\infty(\mathcal{X})$, we have*

$$\begin{aligned}
\Big\| \frac{1}{N} \sum_{i=1}^N h(y^i) - \int_{\mathcal{X}} hq\,dx \Big\|_{L^2(\boldsymbol{P})} &\leqslant \frac{1}{\sqrt{N}} \Big\| h - \int_{\mathcal{X}} hq\,dx \Big\|_{L^\infty} \\
&\quad + \Big\| h - \int_{\mathcal{X}} hq\,dx \Big\|_{L^\infty} \Big\| 1 - \frac{1}{N} \sum_{i=1}^N r(x^i) \Big\|_{L^2(\boldsymbol{P})} \\
&\quad + \Big\| \frac{1}{N} \sum_{i=1}^N r(x^i) \Big(h(x^i) - \int_{\mathcal{X}} hq\,dx \Big) \Big\|_{L^2(\boldsymbol{P})}.
\end{aligned} \tag{3.13}$$

Here r is the ratio

$$r \stackrel{\text{def}}{=} \frac{q}{p}, \quad \text{where} \quad p = \frac{\tilde{p}}{\int_{\mathcal{X}} \tilde{p}\,dx} \quad \text{and} \quad q = \frac{\tilde{q}}{\int_{\mathcal{X}} \tilde{q}\,dx}. \tag{3.14}$$

Note Lemma 3.3 does not assume $x^1, \dots, x^N$ are independent, or even that they have distribution p. If, however, the points $x^1, \dots, x^N$ give good Monte Carlo estimates for integrals with respect to p, then the right hand side of (3.13) will be small. Explicitly, in the typical situation where $x^i \sim p$ are i.i.d, we will have

$$\Big\| \frac{1}{N} \sum_{i=1}^N g(x^i) - \int_{\mathcal{X}} gp\,dx \Big\|_{L^2(\boldsymbol{P})}^2 \leqslant \frac{C \operatorname{Var}(g)}{N}.$$

for any bounded test function g. Combined with the fact that

$$\int_{\mathcal{X}} rp\,dx = \int_{\mathcal{X}} q\,dx = 1 \quad \text{and} \quad \int_{\mathcal{X}} hrp\,dx = \int_{\mathcal{X}} hq\,dx,$$

this shows the right hand side of (3.13) is $O(1/\sqrt{N})$.

We now use Lemma 3.2 and Lemma 3.3 to derive a recurrence relation for the Monte Carlo error between levels k and $k+1$ in Algorithm 1.

Lemma 3.4. *For each $k = 1, \dots, M-1$,*

$$\begin{aligned}
\max_{1 \leqslant \ell \leqslant J} \operatorname{Err}_{k+1,0}(\mathbf{1}_{\Omega_\ell}) &\leqslant \frac{1 + \|r_k\|_{\text{osc}}}{\sqrt{N}} \\
&\quad + \Big(1 + 2 \sum_{j=1}^J \Big| \frac{\pi_{k+1}(\Omega_j)}{\pi_k(\Omega_j)} - 1 \Big| \Big) \cdot \max_{1 \leqslant \ell \leqslant J} \operatorname{Err}_{k,0}(\mathbf{1}_{\Omega_\ell}).
\end{aligned} \tag{3.15}$$

Here r_k is the ratio of the normalized densities defined in (3.1), and by a slight abuse of notation we use $\operatorname{Err}_{k,\cdot}$ to denote $\operatorname{Err}_{\eta_k,\cdot}$.

The proof of Theorem 2.2 now reduces to solving the recurrence relation (3.15) and using Lemma 3.2. Notice that (3.15) involves a bound on $\|r_k\|_{\text{osc}}$, and the maximum of this as k varies is precisely the constant C_r defined in (3.2). A bound on C_r that may be easier to obtain in practice is (3.3), which we prove in Lemma 9.2, below. Momentarily postponing the proofs of the above lemmas, we prove Theorem 2.2.

Proof of Theorem 2.2. Applying (3.11) with $\varepsilon = \eta_M$ gives that

$$\operatorname{Err}_{M,T}(h) \leqslant \frac{\|h\|_{\mathrm{osc}}}{2}\Big(J \max_{j=1,\dots,J} \operatorname{Err}_{M,0}(\mathbf{1}_{\Omega_j}) + \frac{1}{\sqrt{N}}\Big). \tag{3.16}$$

We will show that the right hand side of (3.16) is bounded above by $\delta\|h\|_{\mathrm{osc}}$. For the first term, a direct calculation using (3.15) immediately shows that

$$\begin{aligned}\max_{j=1,\dots,J} \operatorname{Err}_{M,0}(\mathbf{1}_{\Omega_j}) \leqslant \Big(\prod_{\ell=2}^{M-1} \Theta(\ell,\ell+1)\Big) \max_{j=1,\dots,J} \operatorname{Err}_{2,0}(\mathbf{1}_{\Omega_j})\\ + \sum_{k=2}^{M-1} \frac{1+\|r_k\|_{\mathrm{osc}}}{\sqrt{N}} \prod_{\ell=k+1}^{M-1} \Theta(\ell,\ell+1)\end{aligned} \tag{3.17}$$

where

$$\Theta(\ell,\ell+1) \stackrel{\text{def}}{=} 1 + 2\sum_{j=1}^{J}\Big|\frac{\pi_{\ell+1}(\Omega_j)}{\pi_\ell(\Omega_j)} - 1\Big|.$$

To finish the proof, we now need to estimate the terms $\prod_{\ell=k}^{M-1}\Theta(\ell,\ell+1)$ and $\max_{1\leqslant j\leqslant J}\operatorname{Err}_{2,0}(\mathbf{1}_{\Omega_j})$.

Step 1: Estimating $\prod_{\ell=k}^{M-1}\Theta(\ell,\ell+1)$*.* Notice that for every $j = 1,\dots,J$, and every $k = 1,\dots,M$, we have

$$0 < \frac{\pi_{k+1}(\Omega_j)}{\pi_k(\Omega_j)} \leqslant \|r_k\|_{L^\infty} \stackrel{(3.2)}{\leqslant} C_r. \tag{3.18}$$

Using the fact that

$$|y-1| \leqslant (1\vee y)|\ln y|, \tag{3.19}$$

for any $k = 1,\dots,M-1$, we obtain

$$\begin{aligned}\prod_{\ell=k}^{M-1}\Theta(\ell,\ell+1) &\stackrel{\text{AM-GM}}{\leqslant} \Big(1 + \frac{2}{M-k}\sum_{j=1}^{J}\sum_{\ell=k}^{M-1}\Big|\frac{\pi_{\ell+1}(\Omega_j)}{\pi_\ell(\Omega_j)} - 1\Big|\Big)^{M-k}\\ &\stackrel{(3.18),(3.19)}{\leqslant}\Big(1 + \frac{2}{M-k}\sum_{j=1}^{J}\sum_{\ell=k}^{M-1} C_r\Big|\log\Big(\frac{\pi_{\ell+1}(\Omega_j)}{\pi_\ell(\Omega_j)}\Big)\Big|\Big)^{M-k}\\ &= \Big(1 + \frac{2}{M-k}\sum_{j=1}^{J}\sum_{\ell=k}^{M-1} C_r\Big|\int_{\eta_{\ell+1}}^{\eta_\ell}\partial_\varepsilon \ln\pi_\varepsilon(\Omega_j)\,d\varepsilon\Big|\Big)^{M-k}\\ &\stackrel{(2.3)}{\leqslant} \exp(2C_rC_{\mathrm{LBV}}) \stackrel{(3.6)}{=} C_\beta.\end{aligned} \tag{3.20}$$

Step 2: Estimating $\operatorname{Err}_{2,0}(\mathbf{1}_{\Omega_j})$*.* Applying Lemma 3.3 with $p = \pi_1$, $q = \pi_2$, $h = \mathbf{1}_{\Omega_j}$, and $x^i = Y^i_{1,T}$, to obtain

$$\operatorname{Err}_{2,0}(\mathbf{1}_{\Omega_j}) \leqslant \frac{1}{\sqrt{N}} + \operatorname{Err}_{1,T}(r_1) + \operatorname{Err}_{1,T}(r_1(\mathbf{1}_{\Omega_j} - \pi_2(\Omega_j))). \tag{3.21}$$

Now the processes $Y^1_{1,\cdot}, \ldots, Y^N_{1,\cdot}$ are all independent.[2] Thus for any bounded test function h,

$$\begin{aligned}
(\mathrm{Err}_{1,T}(h))^2 &= \boldsymbol{E}\Big(\frac{1}{N}\sum_1^N h(Y^i_{1,T}) - \int_{\mathcal{X}} h\,\pi_1\,dx\Big)^2 \\
&= \boldsymbol{E}\bigg(\frac{1}{N}\sum_1^N \big(h(Y^i_{1,T}) - \boldsymbol{E}h(Y^i_{1,T})\big) + \frac{1}{N}\sum_1^N \boldsymbol{E}h(Y^i_{1,T}) - \int_{\mathcal{X}} h\,\pi_1\,dx\bigg)^2 \\
&\leqslant \frac{1}{N}\|h\|^2_{L^\infty} + \Big(\frac{1}{N}\sum_{i=1}^N \boldsymbol{E}h(Y^i_{1,T}) - \int_{\mathcal{X}} h\,\pi_1\,dx\Big)^2 \\
&\leqslant \frac{1}{N}\|h\|^2_{L^\infty} + \|h\|^2_{L^\infty}\Big(\frac{1}{N}\sum_{i=1}^N \|p^1_T(y^i_1,\cdot) - \pi_1\|_{L^1}\Big)^2,
\end{aligned}$$

and hence

$$\mathrm{Err}_{1,T}(h) \leqslant \frac{1}{\sqrt{N}}\|h\|_{L^\infty} + \|h\|_{L^\infty}\frac{1}{N}\sum_{i=1}^N \|p^1_T(y^i_1,\cdot) - \pi_1\|_{L^1}. \tag{3.22}$$

Notice that the choice of C_T and C_N in (3.5), implies

$$T \geqslant t_{\mathrm{mix}}\Big(Y_{\eta_1,\cdot}, \frac{\tilde\delta}{4C_r}\Big), \quad \text{and} \quad \frac{1}{\sqrt{N}} \leqslant \frac{1+C_r}{\sqrt{N}} \overset{(2.4)}{\leqslant} \frac{\tilde\delta}{M} \tag{3.23}$$

where

$$\tilde\delta = \frac{\delta}{J(2C_\beta + 1)}. \tag{3.24}$$

Using (3.23) in (3.22) with $h = r_1$ gives

$$\mathrm{Err}_{1,T}(r_1) \leqslant \frac{1}{\sqrt{N}}\|r_1\|_{L^\infty} + \frac{\tilde\delta}{2C_r}\|r_1\|_{L^\infty} \overset{(3.2),(3.23)}{\leqslant} \frac{\tilde\delta}{M} + \frac{\tilde\delta}{2}. \tag{3.25}$$

Similarly,

$$\mathrm{Err}_{1,T}(r_1(\mathbf{1}_{\Omega_j} - \pi_2(\Omega_j))) \leqslant \frac{\tilde\delta}{M} + \frac{\tilde\delta}{2}. \tag{3.26}$$

Plugging (3.25),(3.26) and (3.23) into (3.21) yields

$$\max_{j=1,\ldots,J} \mathrm{Err}_{2,0}(\mathbf{1}_{\Omega_j}) \leqslant \frac{3\tilde\delta}{M} + \tilde\delta. \tag{3.27}$$

Now using (3.17), we obtain

$$\begin{aligned}
\max_{j=1,\ldots,J} \mathrm{Err}_{M,0}(\mathbf{1}_{\Omega_j}) &\overset{(3.20),\ (3.27)}{\leqslant} C_\beta\Big(\frac{3\tilde\delta}{M} + \tilde\delta\Big) + \sum_{k=2}^{M-2} C_\beta\frac{\tilde\delta}{M} + \frac{\tilde\delta}{M} \\
&= \Big(2C_\beta + \frac{1}{M}\Big)\tilde\delta \overset{(3.24)}{\leqslant} \frac{\delta}{J}.
\end{aligned} \tag{3.28}$$

Using (3.23) and (3.28) in (3.16) implies

$$\mathrm{Err}_{M,T}(h) \leqslant \frac{\|h\|_{\mathrm{osc}}}{2}\Big(\frac{J\delta}{J} + \frac{\tilde\delta}{M}\Big) \overset{(3.24)}{<} \delta\|h\|_{\mathrm{osc}}.$$

[2]For $k \geqslant 2$ the processes $Y^1_{k,\cdot}, \ldots, Y^N_{k,\cdot}$ are no longer independent as the initial distributions are not independent.

This proves (2.5), concluding the proof. □

4. **Error Estimates for a double-well energy (Theorem 2.8).**

The aim of this section is to prove Theorem 2.8 and obtain error estimates when ASMC is used to sample from a double-well energy function on a d-dimensional torus.

4.1. **Assumptions and Notation.**

We begin by precisely stating the assumptions that were used in Theorem 2.8. The first assumption requires U to be a regular, double-well function with nondegenerate critical points.

Assumption 4.1. *The function $U \in C^6(\mathbb{T}^d, \mathbb{R})$, has a nondegenerate Hessian at all critical points, and has exactly two local minima located at $x_{\min,1}$ and $x_{\min,2}$. We normalize U so that*

$$0 = U(x_{\min,1}) \leqslant U(x_{\min,2}).$$

Our next assumption concerns the saddle between the local minima $x_{\min,1}$ and $x_{\min,2}$. Define the saddle height between $x_{\min,1}$ and $x_{\min,2}$ to be the minimum amount of energy needed to go from the global minimum $x_{\min,1}$ to $x_{\min,2}$, and is given by

$$\hat{U} = \hat{U}(x_{\min,1}, x_{\min,2}) \stackrel{\text{def}}{=} \inf_{\omega} \sup_{t \in [0,1]} U(\omega(t)). \tag{4.1}$$

Here the infimum above is taken over all continuous paths $\omega \in C([0,1]; \mathbb{T}^d)$ such that $\omega(0) = x_{\min,1}$, $\omega(1) = x_{\min,2}$. To prove Theorem 2.8 we need to assume a nondegeneracy condition on the saddle.

Assumption 4.2. *The saddle height between $x_{\min,1}$ and $x_{\min,2}$ is attained at a unique critical point $s_{1,2}$ of index one. That is, the first eigenvalue of* $\operatorname{Hess} U(s_{1,2})$ *is negative and the others are positive.*

We can now define the ratio $\hat{\gamma}_r$ that appeared in (2.8), above. The *energy barrier*, denoted by $\hat{\gamma}$, is defined to be the minimum amount of energy needed to go from the (possibly local) minimum $x_{\min,2}$ to the global minimum $x_{\min,1}$. In terms of $s_{1,2}$, the energy barrier $\hat{\gamma}$ and the saddle height are given by

$$\hat{\gamma} \stackrel{\text{def}}{=} U(s_{1,2}) - U(x_{\min,2}), \quad \text{and} \quad \hat{U} = U(s_{1,2}). \tag{4.2}$$

The ratio $\hat{\gamma}_r$ is the ratio of the saddle height $\hat{U}$ to the energy barrier $\hat{\gamma}$, given by

$$\hat{\gamma}_r \stackrel{\text{def}}{=} \frac{\hat{U}}{\hat{\gamma}}. \tag{4.3}$$

Finally, we require the distribution π_η to be truly multimodal in the temperature range of interest. That is, we require the mass in the *basins of attraction* around each of the local minima $x_{\min,1}$ and $x_{\min,2}$ to be bounded away from 0. We recall the basin of attraction around $x_{\min,i}$, denoted by Ω_i, is the set of all initial points for which the gradient flow of U eventually reaches $x_{\min,i}$. Precisely, Ω_i is defined by

$$\Omega_i \stackrel{\text{def}}{=} \Big\{ y \in \mathbb{T}^d \;\Big|\; \lim_{t \to \infty} y_t = x_{\min,i}, \text{ where } \dot{y}_t = -\nabla U(y_t) \text{ with } y_0 = y \Big\},$$

and our multimodality condition is as follows.

Assumption 4.3. *There exists* $0 \leqslant \eta_{\min} < \eta_{\max} \leqslant \infty$*, a constant* C_m *such that*

$$\inf_{\substack{\varepsilon \in [\eta_{\min}, \eta_{\max}] \\ 0<\varepsilon<\infty}} \pi_\varepsilon(\Omega_i) \geqslant \frac{1}{C_m^2}. \tag{4.4}$$

We will show (Lemma 4.4, below) that (4.4) is satisfied if the wells have *nearly equal depth.* That is, if $U(x_{\min,2}) - U(x_{\min,1}) \leqslant O(\eta_{\min})$, then one can show (4.4) holds for some constant C_m that is independent of $\eta_{\min}$. We state this precisely as the following lemma.

Lemma 4.4. *Suppose* U *satisfies Assumptions 4.1, 4.2, and there exists a temperature* $\eta_{\min} \geqslant 0$ *and constant* $C_\ell > 0$ *such that*

$$U(x_{\min,2}) - U(x_{\min,1}) \leqslant C_\ell \eta_{\min}. \tag{4.5}$$

Then for any finite $\eta_{\max} > \eta_{\min}$ *there exists a constant* $C_m = C_m(U, \eta_{\max}, C_\ell)$*, independent of* $\eta_{\min}$ *such that* (4.4) *holds.*

Remark 4.5. We note that the condition (4.5) implies the finiteness condition (2.3) that was used in Theorem 2.2. This is shown in Corollary 9.1, below, and was previously referred to in Remark 2.5.

4.2. **Proof of Theorem 2.8.** In this subsection, we explain the main idea behind the proof of Theorem 2.8. For simplicity and without loss of generality we assume $\eta_1 = 1$. We begin by rewriting our algorithm in a manner that that is convenient for the proof. Fix $T > 0$ and $N \in \mathbb{N}$ that will be chosen later.

Step 1: We start with N arbitrary points $y_1^1, \ldots, y_1^N$.

Step 2: Langevin step. For each $k \in \{1, \ldots, M\}$, and $i \in \{1, \ldots N\}$, let $X_{k,\cdot}^i$ be the solution to the overdamped Langevin equation (1.2) with initial data $X_{k,0}^i = y_k^i$, driven by independent Brownian motions.

Step 3: Resampling step. Given the processes $\{X_{k,\cdot}^i \mid i \leqslant N, k \leqslant M-1\}$ we choose the points $\{y_{k+1}^1, \ldots, y_{k+1}^N\}$ independently from $\{X_{k,T}^1, \ldots, X_{k,T}^N\}$ so that

$$\boldsymbol{P}(y_{k+1}^i = X_{k,T}^j) = \frac{\tilde{r}_k(X_{k,T}^j)}{\sum_{i=1}^N \tilde{r}_k(X_{k,T}^i)}.$$

Here $\tilde{r}_k$ is the ratio defined by (2.1).

We now briefly recall a few standard facts about the overdamped Langevin dynamics (1.2) that will be used in the proof. Let L_ε be the generator of (1.2), whose action on smooth test functions is defined by

$$L_\varepsilon f \stackrel{\text{def}}{=} -\varepsilon \Delta f + \nabla U \cdot \nabla f. \tag{4.6}$$

Let L_ε^* be the dual operator defined by

$$L_\varepsilon^* f = -\nabla \cdot (\nabla U f) - \varepsilon \Delta f\,. \tag{4.7}$$

It is well known [Øks03, Chapter 8] that if $Y_{\varepsilon,\cdot}$ solves (1.2) then its density $f_t \stackrel{\text{def}}{=} \mathrm{PDF}(Y_{\varepsilon,t})$ satisfies the *Fokker-Planck equation*, a.k.a. the *Kolmogorov forward equation*

$$\partial_t f + L_\varepsilon^* f = 0. \tag{4.8}$$

One can readily check that the Gibbs distribution π_ε is a stationary solution of (4.8), and hence must be the stationary distribution of (1.2). A direct calculation shows that

$$\partial_t\Big(\frac{f}{\pi_\varepsilon}\Big) + L_\varepsilon\Big(\frac{f}{\pi_\varepsilon}\Big) = 0. \tag{4.9}$$

The mixing properties of Langevin dynamics can be deduced directly from the spectral properties of the operator L_ε, as we now explain. It is well known (see for instance [Kol00, Chapter 8]) that on the weighted space $L^2(\pi_\varepsilon)$ the operator L_ε is self-adjoint and has a discrete spectrum with eigenvalues

$$0 = \lambda_{1,\varepsilon} < \lambda_{2,\varepsilon} \leqslant \lambda_{3,\varepsilon} \cdots$$

with corresponding $L^2(\pi_\varepsilon)$ normalized eigenfunctions $\psi_{1,\varepsilon}$, $\psi_{2,\varepsilon}$, etc. The first eigenvalue $\lambda_{1,\varepsilon} = 0$ corresponds to the constant eigenfunction $\psi_{1,\varepsilon} \equiv 1$. In our situation, because U has two wells, it is well known that (see for instance Propositions 2.1, 2.2 in Chapter 8 of [Kol00]) for every $\gamma < \hat\gamma$ there exists constants C_γ and Λ (independent of ε) such that

$$\lambda_{2,\varepsilon} \leqslant C_\gamma \exp\Big(-\frac{\gamma}{\varepsilon}\Big) \quad \text{and} \quad \lambda_{i,\varepsilon} \geqslant \Lambda, \quad \forall i \geqslant 3. \tag{4.10}$$

As a result, equation (4.9) implies

$$\begin{aligned}\Big\|\frac{f_t}{\pi_\varepsilon} - 1\Big\|^2_{L^2(\pi_\varepsilon)} &= \Big\|e^{-L_\varepsilon t}\Big(\frac{f_0}{\pi_\varepsilon}\Big) - 1\Big\|^2_{L^2(\pi_\varepsilon)} \\ &\leqslant \exp\big(-2tC_\gamma e^{-\gamma/\varepsilon}\big)\langle f_0, \psi_{2,\varepsilon}\rangle^2_{L^2} + e^{-\Lambda t}\Big\|\frac{f_0}{\pi_\varepsilon} - 1\Big\|^2_{L^2(\pi_\varepsilon)}.\end{aligned}$$

The second term on the right decays fast with t, and the metastability phenomenon described above is due to the slow decay of the first term. While the first term decays slowly with t, we can make it small by ensuring

$$\langle f_0, \psi_{2,\varepsilon}\rangle^2_{L^2} = \Big(\int_{\mathbb{T}^d} f_0\, \psi_{2,\varepsilon}\, dx\Big)^2$$

is small. We note that $\langle f_0, \psi_{2,\varepsilon}\rangle_{L^2}$ measures the difference in the f_0 and the π_ε-mass distribution in each well. This confirms our previous statement that the Langevin dynamics mixes quickly when the mass of the initial distribution in each well is close to the π_k-mass of the same well.

To use this quantitatively in our situation, we need an estimate on the Monte Carlo error when using N independent realizations to compute the integral of a test function. We recall the standard Langevin Monte Carlo algorithm (LMC) approximates the integral of a test function h with respect to the Gibbs measure π_ε by

$$\int_{\mathbb{T}^d} h\pi_\varepsilon\, dx \approx \frac{1}{N}\sum_{i=1}^N h(Y^i_{\varepsilon,t}),$$

where $Y^1_{\varepsilon,\cdot}, \ldots, Y^N_{\varepsilon,\cdot}$ are N independent solutions to (1.2). The right hand side approaches the left hand side as $N, t \to \infty$. Our first lemma controlling the error is as follows.

Lemma 4.6. *Assume that for each $i \in \{1, \dots, N\}$, $\mathrm{PDF}(Y^i_{\varepsilon,0}) = q^i_{\varepsilon,0}$. Then for any bounded test function h,*

$$\mathrm{Err}_{\varepsilon,T}(h) \leqslant e^{-\lambda_{2,\varepsilon}T}\Big|\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon \, dx\Big| \, \mathrm{Err}_{\varepsilon,0}(\psi_{2,\varepsilon}) + \frac{1}{2\sqrt{N}}\|h\|_{\mathrm{osc}} + \mathcal{E}_{\varepsilon,T}(h) \tag{4.11}$$

where

$$\mathcal{E}_{\varepsilon,T}(h) \stackrel{\text{def}}{=} \|h\|_{\mathrm{osc}} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\|\frac{q^i_{\varepsilon,0}}{\pi_\varepsilon}\Big\|^{\frac{1}{2}}_{L^\infty(\pi_\varepsilon)}. \tag{4.12}$$

We will now use Lemma 3.3 and Lemma 4.6 to derive Monte Carlo error estimates between levels k and $k+1$ in Algorithm 1. Recall in Algorithm 1, M is chosen according to (2.8), $\eta_1 = 1$, $\eta_M = \eta$, and the reciprocals $1/\eta_1, \dots, 1/\eta_M$ are linearly spaced. That is η_k is chosen according to

$$\eta_k \stackrel{\text{def}}{=} \frac{(M-1)\eta}{(M-1)\eta + (k-1)(1-\eta)}. \tag{4.13}$$

For simplicity of notation, we use a subscript of k on the error, eigenvalue and eigenfunction to denote the corresponding quantities at $\varepsilon = \eta_k$. Explicitly, we write

$$\lambda_{2,k} \stackrel{\text{def}}{=} \lambda_{2,\eta_k}, \quad \psi_{2,k} \stackrel{\text{def}}{=} \psi_{2,\eta_k}, \quad \text{and} \quad \mathrm{Err}_{k,0}(\psi_{2,k}) \stackrel{\text{def}}{=} \mathrm{Err}_{\eta_k,0}(\psi_{2,k}).$$

The main idea behind the proof of Theorem 2.8 is to first estimate $\mathrm{Err}_{k+1,0}(\psi_{2,k+1})$ in terms of $\mathrm{Err}_{k,0}(\psi_{2,k})$, and then use Lemma 4.6 to obtain (2.9). Obtaining this recurrence relation, however, requires a fair amount of technical work. We state this in the next lemma.

Lemma 4.7. *Choose M as in (2.8) and η_k as in (4.13). For any $\alpha > 0$, there exist constants $C_\alpha = C_\alpha(\alpha, U) > 0$ (depending on α) and $\tilde{C}_N = \tilde{C}_N(U) > 1$ (independent of α) such that for any $\delta > 0$, if*

$$T \geqslant C_\alpha\Big(M^{(1+\alpha)\hat{\gamma}_r} + \log\Big(\frac{1}{\delta}\Big) + \frac{1}{\eta}\Big), \quad N \geqslant \frac{\tilde{C}_N M^2}{\delta^2},$$

then for each $2 \leqslant k \leqslant M-1$, we have

$$\mathrm{Err}_{k+1,0}(\psi_{2,k+1}) \leqslant \beta_k \, \mathrm{Err}_{k,0}(\psi_{2,k}) + c_k. \tag{4.14}$$

Here the constants β_k, c_k are such that for every $k \in \{2, \dots, M-1\}$ we have

$$\prod_{j=k}^{M-1} \beta_j \leqslant C_\beta \quad \text{and} \quad c_k \leqslant \frac{\delta}{M}, \tag{4.15}$$

for some dimensional constant $C_\beta > 1$ (independent of α, δ).

The proof of Theorem 2.8 now reduces to solving the recurrence relation (4.14) and using Lemma 4.6. Notice that in order to use Lemma 4.6 and Lemma 4.9, we need an estimate for $\|q_{k,0}/\pi_k\|_{L^\infty(\pi_k)}$. This is addressed in the following lemma.

Lemma 4.8. *For every $2 \leqslant k \leqslant M$, $1 \leqslant i \leqslant N$, let $q^i_{k,0}$ be the probability density function of $X^i_{k,0}$. For any $T_0 > 0$, there exists a constant $C_q = C_q(U, T_0)$ such that if $T \geqslant T_0$, then*

$$\max_{i=1,\dots,N} \Big\|\frac{q^i_{k,0}}{\pi_k}\Big\|_{L^\infty(\pi_k)} \leqslant C_q \exp\Big(\|U\|_{\mathrm{osc}}\Big(\frac{1}{\eta_k} - 1\Big)\Big). \tag{4.16}$$

While the right hand side of (4.16) is exponentially large, it will only be used in (4.12) which has an exponentially small $e^{-\Lambda T}$ factor. Choosing $T \geqslant O(1/\eta_k)$ will allow us to control it.

Finally, we will need an estimate for $\mathrm{Err}_{2,0}(\psi_{2,2})$ which we obtain as the mixing time when $k = 1$ is of order 1.

Lemma 4.9. *There exists a constant $C_1 = C_1(U)$ such that for any $\delta > 0$, we have*

$$\mathrm{Err}_{2,0}(\psi_{2,2}) \leqslant \delta \quad \textit{provided} \quad T \geqslant C_1\Big(\log\Big(\frac{1}{\delta}\Big) + 1\Big), \quad N \geqslant \frac{\tilde{C}_N}{\delta^2}. \tag{4.17}$$

Here $\tilde{C}_N$ is the same constant as in Lemma 4.7.

We are now well-equipped to prove Theorem 2.8.

Proof of Theorem 2.8. Fix $\alpha, \delta > 0$, and define

$$\tilde{\delta} = \frac{\delta}{4C_\beta}, \tag{4.18}$$

where C_β is the constant in Lemma 4.7. Choose M as in (2.8) and define

$$\begin{aligned} T \overset{\text{def}}{=} \max\Big\{ & C_\alpha\Big(M^{(1+\alpha)\hat{\gamma}_r} + \log\Big(\frac{1}{\tilde{\delta}}\Big) + \frac{1}{\eta}\Big),\ C_1\Big(\log\Big(\frac{1}{\tilde{\delta}}\Big) + 1\Big), \\ & \frac{1}{\Lambda}\Big(\log\Big(\frac{1}{\tilde{\delta}}\Big) + \frac{\|U\|_{\mathrm{osc}}}{2\eta} + \frac{1}{2}\log(C_q)\Big), 1\Big\}, \end{aligned} \tag{4.19}$$

$$N \overset{\text{def}}{=} \frac{\tilde{C}_N M^2}{\tilde{\delta}^2}. \tag{4.20}$$

Here $C_q = C_q(U, 1)$ is the constant in Lemma 4.8 with $T_0 = 1$, and C_α, C_1, Λ and are the constants in Lemma 4.7, Lemma 4.9, and (4.10) respectively.

Notice that if T, N are chosen according to (4.19) and (4.20), then we can find constants $C_T = C_T(\alpha, U) > 0$ and $C_N = C_N(U)$ so that this choice is consistent with the choice in (2.8). We will now show that (2.9) holds for any bounded test function $h \in L^\infty(\mathbb{T}^d)$.

Using Lemma 4.6, we obtain

$$\begin{aligned} \mathrm{Err}_{M,T}(h) \leqslant & \Big|\int_{\mathbb{T}^d} h\psi_{2,k}\pi_k\,dx\Big| e^{-\lambda_{2,k}T}\,\mathrm{Err}_{M,0}(\psi_{2,M}) \\ & + \frac{1}{2\sqrt{N}}\|h\|_{\mathrm{osc}} + \mathcal{E}_{\eta,T}(h). \end{aligned} \tag{4.21}$$

We will now show that the right hand side of (4.21) is bounded above by $\delta\|h\|_{\mathrm{osc}}$. For the first term, a direct calculation using (4.14) immediately shows that for T, N as in (4.19), (4.20) we have

$$\begin{aligned} \mathrm{Err}_{M,0}(\psi_{2,M}) &\leqslant \Big(\prod_{j=2}^{M-1}\beta_j\Big)\mathrm{Err}_{2,0}(\psi_{2,2}) + \sum_{k=2}^{M-2} c_k\Big(\prod_{j=k+1}^{M-1}\beta_j\Big) + c_{M-1} \\ &\overset{(4.15),\,(4.17)}{\leqslant} C_\beta\tilde{\delta} + \sum_{k=2}^{M-2} C_\beta\frac{\tilde{\delta}}{M} + \frac{\tilde{\delta}}{M} \leqslant 2C_\beta\tilde{\delta} \overset{(4.18)}{=} \frac{\delta}{2}. \end{aligned} \tag{4.22}$$

Next, we see

$$\frac{1}{\sqrt{N}} \overset{(4.20)}{\leqslant} \frac{\tilde{\delta}}{M} \overset{(4.18)}{\leqslant} \frac{\delta}{4}. \tag{4.23}$$

Finally,

$$
\begin{aligned}
\mathcal{E}_{\eta,T}(h) &\overset{(4.12)}{=} \|h\|_{\mathrm{osc}} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\| \frac{q^i_{M,0}}{\pi_M} \Big\|^{\frac12}_{L^\infty(\pi_M)} \\
&\overset{(4.16)}{\leqslant} C_q^{\frac12} \|h\|_{\mathrm{osc}} \exp\Big(\frac{\|U\|_{\mathrm{osc}}}{2\eta}\Big) e^{-\Lambda T} \\
&\overset{(4.19)}{\leqslant} \|h\|_{\mathrm{osc}} \tilde\delta < \|h\|_{\mathrm{osc}} \frac{\delta}{4}. \qquad (4.24)
\end{aligned}
$$

Using (4.22), (4.23) and (4.24) in (4.21) implies

$$\mathrm{Err}_{M,T}(h) \leqslant \delta \|h\|_{\mathrm{osc}}.$$

This proves (2.9), concluding the proof. □

It remains to prove Lemmas 4.6, 4.7, 4.8 and 4.9, which will be done in subsequent sections.

5. Proof of Lemmas for the Local Mixing Model.

In this section, we prove Lemmas 3.2, 3.3 and 3.4 that were used Section 3 to prove Theorem 2.2. We also prove the bound for $\|r_k\|_{L^\infty}$ stated in (3.3), that may be easier to use in practice. Since the ideas used in Lemma 3.2 and 3.4 are related, we prove Lemma 3.3 first.

5.1. The Resampling Error (Lemma 3.3).

Notice that the points $y^1, \dots, y_N$ chosen according to (3.12) are identically distributed, but need not be independent. However, given the points $x^1, \dots, x_N$, the points $y^1, \dots, y_N$ are (conditionally) independent. The main idea behind the proof of Lemma 3.3 is to split the error into the sum of a conditional mean, and a conditional standard deviation, and use conditional independence of $y^1, \dots, y_N$.

Proof of Lemma 3.3. For simplicity of notation, let

$$\mathbf{x} \overset{\text{def}}{=} \{x^1, ..., x^N\}, \quad \text{and} \quad \tilde h \overset{\text{def}}{=} h - \int_{\mathcal{X}} hq\,dx, \qquad (5.1)$$

and let $\boldsymbol{E}_{\mathbf{x}}$ denote the conditional expectation given the σ-algebra generated by $\mathbf{x}$. By the tower property,

$$
\begin{aligned}
\boldsymbol{E}\Big(\frac1N \sum_1^N h(y^i) - \int_{\mathcal{X}} hq\,dx\Big)^2 &= \boldsymbol{E}\Big(\frac1N \sum_{i=1}^N \tilde h(y^i)\Big)^2 \\
&= \boldsymbol{E}\Big[\boldsymbol{E}_{\mathbf{x}}\Big(\frac1N\sum_{i=1}^N \tilde h(y^i)\Big)^2\Big]. \qquad (5.2)
\end{aligned}
$$

We write

$$\boldsymbol{E}_{\mathbf{x}}\Big(\frac1N \sum_{\ell=1}^N \tilde h(y^i)\Big)^2 = J_1 + J_2,$$

where

$$J_1 \overset{\text{def}}{=} \Big(\frac1N\sum_{\ell=1}^N \boldsymbol{E}_{\mathbf{x}} \tilde h(y^i)\Big)^2 \quad \text{and} \quad J_2 \overset{\text{def}}{=} \boldsymbol{E}_{\mathbf{x}}\Big(\frac1N \sum_{\ell=1}^N \tilde h(y^i) - \frac1N\sum_{\ell=1}^N \boldsymbol{E}_{\mathbf{x}}\tilde h(y^i)\Big)^2.$$

Notice that the points $y^1, \dots, y^N$ are not independent; however, when conditioned on $\mathbf{x}$, the points y^i are independent and identically distributed. Thus

$$J_2 \leqslant \frac{1}{N^2} \sum_{\ell=1}^{N} \operatorname{Var}_{\mathbf{x}}(\tilde{h}(y^i)) \leqslant \frac{1}{N} \|\tilde{h}\|_{L^\infty}^2. \tag{5.3}$$

To bound J_1, we note

$$\begin{aligned}\frac{1}{N} \sum_{\ell=1}^{N} \boldsymbol{E}_{\mathbf{x}} \tilde{h}(y^i) = \boldsymbol{E}_{\mathbf{x}} \tilde{h}(y_1) &= \frac{\sum_{i=1}^{N} \tilde{h}(x^i) \tilde{r}(x^i)}{\sum_{i=1}^{N} \tilde{r}(x^i)} \overset{(3.14)}{=} \frac{\sum_{i=1}^{N} \tilde{h}(x^i) r(x^i)}{\sum_{i=1}^{N} r(x^i)} \\ &= J_3 + J_4,\end{aligned} \tag{5.4}$$

where

$$J_3 = \frac{\sum_{i=1}^{N} \tilde{h}(x^i) r(x^i)}{\sum_{i=1}^{N} r(x^i)} \Big(1 - \frac{1}{N} \sum_{i=1}^{N} r(x^i) \Big), \quad \text{and} \quad J_4 = \frac{1}{N} \sum_{i=1}^{N} \tilde{h}(x^i) r(x^i).$$

Clearly

$$|J_3| \leqslant \|\tilde{h}\|_{L^\infty} \Big(1 - \frac{1}{N} \sum_{i=1}^{N} r(x^i) \Big).$$

Thus

$$\begin{aligned}\Big\| \frac{1}{N} \sum_{1}^{N} h(y^i) - \int_{\mathcal{X}} hq \, dx \Big\|_{L^2(\boldsymbol{P})} &\overset{(5.2)}{\leqslant} (\boldsymbol{E} J_1)^{1/2} + (\boldsymbol{E} J_2)^{1/2} \\ &\overset{(5.4),\ (5.3)}{\leqslant} (\boldsymbol{E} J_3^2)^{1/2} + (\boldsymbol{E} J_4^2)^{1/2} + \frac{\|\tilde{h}\|_{L^\infty}}{\sqrt{N}}.\end{aligned}$$

Using the definition of $\tilde{h}$ in (5.1) we obtain (3.13) as desired. □

5.2. **The Monte Carlo error (Lemma 3.2).** In this section, we prove Lemma 3.2 which provides an estimate for the error when using independent realizations of the process $Y_{\varepsilon,\cdot}$ to compute Monte Carlo integrals.

Proof of Lemma 3.2. Since

$$\sum_{j=1}^{J} \Big(1 - \frac{\mu_0(\Omega_j)}{\pi_\varepsilon(\Omega_j)} \Big) \pi_\varepsilon(\Omega_j) = 0,$$

both sides of (3.9) remain unchanged when a constant is added to the function h. Thus without loss of generality we may replace h with $h - \inf h + \frac{1}{2}\|h\|_{\text{osc}}$, and assume $\|h\|_{L^\infty} = \frac{1}{2}\|h\|_{\text{osc}}$. Next we write

$$\frac{1}{N} \sum_{1}^{N} h(Y_{\varepsilon,T}^i) - \int_{\mathcal{X}} h \, \pi_\varepsilon \, dx = I_1 + I_2, \tag{5.5}$$

where

$$I_1 \overset{\text{def}}{=} \frac{1}{N} \sum_{1}^{N} \big(h(Y_{\varepsilon,T}^i) - \boldsymbol{E}_0 h(Y_{\varepsilon,T}^i) \big), \qquad I_2 \overset{\text{def}}{=} \frac{1}{N} \boldsymbol{E}_0 \sum_{1}^{N} h(Y_{\varepsilon,T}^i) - \int_{\mathcal{X}} h \, \pi_\varepsilon \, dx,$$

and $\boldsymbol{E}_0$ denote the conditional expectation with respect to $\sigma(y^1, \dots, y^N)$.

Now since $Y^i_{\varepsilon,\cdot}$ are conditionally independent given $y^1, \dots, y^N$,

$$\boldsymbol{E}_0 I_1^2 = \frac{1}{N}\sum_1^N \boldsymbol{E}_0\big(h(Y^i_{\varepsilon,T}) - \boldsymbol{E}_0 h(Y^i_{\varepsilon,T})\big)^2 \leqslant \frac{1}{N}\|h\|^2_{L^\infty} = \frac{1}{4N}\|h\|^2_{\mathrm{osc}}. \tag{5.6}$$

Next using (2.2) and (3.10) we compute

$$\begin{aligned}\frac{1}{N}\boldsymbol{E}_0\sum_1^N h(Y^i_{\varepsilon,T}) &= \frac{1}{N}\sum_{i=1}^N \int p^\varepsilon_T(y^i, z)h(z)\,dz \\ &= (1-\chi^T_\varepsilon)\int_{\mathcal{X}} h\pi_\varepsilon\,dx + \chi^T_\varepsilon\sum_{j=1}^J \frac{\mu_0(\Omega_j)}{\pi_\varepsilon(\Omega_j)}\int_{\Omega_j} h\pi_\varepsilon\,dx,\end{aligned}$$

and hence

$$I_2 = \chi^T_\varepsilon \sum_{j=1}^J\Big(\frac{\mu_0(\Omega_j)}{\pi_\varepsilon(\Omega_j)} - 1\Big)\int_{\Omega_j} h\pi_\varepsilon\,dx. \tag{5.7}$$

Using (5.6) in (5.5) implies

$$\mathrm{Err}_{\varepsilon,T}(h) = \|I_1 + I_2\|_{L^2(\boldsymbol{P})} \leqslant \|I_1\|_{L^2(\boldsymbol{P})} + \|I_2\|_{L^2(\boldsymbol{P})} \leqslant \frac{\|h\|_{\mathrm{osc}}}{2\sqrt{N}} + \|I_2\|_{L^2(\boldsymbol{P})}.$$

Using (5.7) in the above implies (3.9) as desired.

$$c_\varepsilon(j,T) \stackrel{\text{def}}{=} \chi^T_\varepsilon \frac{\mu_0(\Omega_j)}{\pi_\varepsilon(\Omega_j)} + (1-\chi^T_\varepsilon).$$

Finally, (3.11) follows immediately from (3.9) and the fact that for every $j \in \{1,\dots,J\}$ we have

$$\Big|\int_{\Omega_j} h\pi_\varepsilon\,dx\Big| \leqslant \|h\|_\infty \int_{\Omega_j}\pi_\varepsilon\,dx = \frac{\|h\|_{\mathrm{osc}}}{2}\pi_\varepsilon(\Omega_j). \qquad \square$$

5.3. **A recurrence relation for the error (Lemma 3.4).** We now prove Lemma 3.4, which obtains a recurrence relation for the Monte Carlo error between levels k and $k+1$ in Algorithm 1.

Proof of Lemma 3.4. Fix $k \in \{1,\dots,M-1\}$, and $\ell \in \{1,\dots,J\}$. Applying Lemma 3.3 with $p = \pi_k$, $q = \pi_{k+1}$, $h = \mathbf{1}_{\Omega_\ell}$, and $x^i = Y^i_{k,T}$ gives

$$\mathrm{Err}_{k+1,0}(\mathbf{1}_{\Omega_\ell}) \leqslant \frac{1}{\sqrt{N}} + \mathrm{Err}_{k,T}(r_k) + \mathrm{Err}_{k,T}\big(r_k(\mathbf{1}_{\Omega_\ell} - \pi_{k+1}(\Omega_\ell))\big). \tag{5.8}$$

We now bound the last two terms on the right hand side of (5.8). Applying Lemma 3.2 with $h = r_k$, $\varepsilon = \eta_k$ and using (3.1) gives

$$\begin{aligned}\mathrm{Err}_{k,T}(r_k) &\leqslant \frac{\|r_k\|_{\mathrm{osc}}}{2\sqrt{N}} + \chi^T_k\Big\|\sum_{j=1}^J\Big(1 - \frac{\mu_0(\Omega_j)}{\pi_k(\Omega_j)}\Big)\pi_{k+1}(\Omega_j)\Big\|_{L^2(\boldsymbol{P})} \\ &= \frac{\|r_k\|_{\mathrm{osc}}}{2\sqrt{N}} + \chi^T_k\Big\|\sum_{j=1}^J\Big(1 - \frac{\mu_0(\Omega_j)}{\pi_k(\Omega_j)}\Big)(\pi_{k+1}(\Omega_j) - \pi_k(\Omega_j)\Big\|_{L^2(\boldsymbol{P})} \\ &\leqslant \frac{\|r_k\|_{\mathrm{osc}}}{2\sqrt{N}} + \chi^T_k\Big(\sum_{j=1}^J\Big|\frac{\pi_{k+1}(\Omega_j)}{\pi_k(\Omega_j)} - 1\Big|\Big)\max_{1\leqslant j\leqslant J}\mathrm{Err}_{k,0}(\mathbf{1}_{\Omega_j}). \qquad (5.9)\end{aligned}$$

Here μ_0 is defined by (3.10) with $y^i = Y^i_{k,0}$, and the second equality above is true because $\sum_j \mu_0(\Omega_j) = 1$.

For the last term on the right of (5.8) again apply Lemma 3.2 with $\varepsilon = \eta_k$ and $h = r_k(\mathbf{1}_{\Omega_\ell} - \pi_{k+1}(\Omega_\ell))$ to obtain

$$\operatorname{Err}_{k,T}\bigl(r_k(\mathbf{1}_{\Omega_\ell} - \pi_{k+1}(\Omega_\ell))\bigr) \leqslant \frac{\|r_k\|_{\mathrm{osc}}}{2\sqrt{N}} + \chi^T_\varepsilon J_1 \tag{5.10}$$

where

$$J_1 \stackrel{\text{def}}{=} \Bigl\| \sum_{j=1}^{J} \Bigl(1 - \frac{\mu_0(\Omega_j)}{\pi_k(\Omega_j)}\Bigr) \pi_{k+1}(\Omega_\ell)(\delta_{j,\ell} - \pi_{k+1}(\Omega_j)) \Bigr\|_{L^2(\boldsymbol{P})} .$$

Since $\sum_j \mu_0(\Omega_j) = 1$ we note

$$\begin{aligned} J_1 &= \Bigl\| \sum_{j=1}^{J} \Bigl(1 - \frac{\mu_0(\Omega_j)}{\pi_k(\Omega_j)}\Bigr) \pi_{k+1}(\Omega_\ell)(\delta_{j,\ell} - \pi_{k+1}(\Omega_j) + \pi_k(\Omega_j)) \Bigr\|_{L^2(\boldsymbol{P})} \\ &\leqslant \pi_{k+1}(\Omega_\ell) \sum_{j=1}^{J} \Bigl| \frac{\delta_{j,\ell}}{\pi_k(\Omega_j)} + 1 - \frac{\pi_{k+1}(\Omega_j)}{\pi_k(\Omega_j)} \Bigr| \max_{1 \leqslant j \leqslant J} \operatorname{Err}_{k,0}(\mathbf{1}_{\Omega_j}) \\ &\leqslant \Bigl(1 + \pi(\Omega_\ell) \sum_{j=1}^{J} \Bigl| 1 - \frac{\pi_{k+1}(\Omega_j)}{\pi_k(\Omega_j)} \Bigr| \Bigr) \max_{1 \leqslant j \leqslant J} \operatorname{Err}_{k,0}(\mathbf{1}_{\Omega_j}) . \qquad (5.11) \end{aligned}$$

Using (5.9), (5.10) and (5.11) in (5.8) yields (3.15) as desired. ☐

6. Error estimates for the Langevin Dynamics (Lemma 4.6 and 4.8).

In this section we prove Lemmas 4.6 and 4.8. The proof of Lemma 4.6 is based on a spectral decomposition, and is presented in Section 6.1, below. The proof of Lemma 4.16 is based on the maximum principle and is presented in Section 6.2, below.

6.1. The Monte Carlo error in the Langevin Step (Lemma 4.6).

The proof of Lemma 4.6 has three main steps. First, we separate the error into the sum of the conditional mean and the conditional standard deviation. The conditional standard deviation is $O(1/\sqrt{N})$ and is easily bounded. Then we decompose the conditional mean as the sum of $\operatorname{Err}_{\varepsilon,0}(\psi_{2,\varepsilon})$ and a remainder using the spectral decomposition of L_ε, and bound the remainder terms.

Proof of Lemma 4.6. Since $\int_{\mathbb{T}^d} \psi_{2,\varepsilon} \pi_\varepsilon \, dx = 0$, both sides of (4.11) remain unchanged when a constant is added to the function h. Thus without loss of generality we may again replace h with $h - \inf h + \frac{1}{2}\|h\|_{\mathrm{osc}}$, and assume $\|h\|_{L^\infty} = \frac{1}{2}\|h\|_{\mathrm{osc}}$. As before, let $\boldsymbol{E}_0$ denote the conditional expectation given the σ-algebra generated by $\{Y^1_{\varepsilon,0}, \ldots, Y^N_{\varepsilon,0}\}$.

Step 1: By the tower property of conditional expectation, we have

$$(\operatorname{Err}_{\varepsilon,T}(h))^2 = \boldsymbol{E}\Bigl[\boldsymbol{E}_0\Bigl(\frac{1}{N}\sum_{i=1}^{N} h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx\Bigr)^2\Bigr]. \tag{6.1}$$

Observe that

$$\frac{1}{N}\sum_{i=1}^{N} h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx = I_1 + I_2,$$

where

$$I_1 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} h(Y^i_{\varepsilon,T}) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{E}_0 h(Y^i_{\varepsilon,T}),$$
$$I_2 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{E}_0 h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx.$$

Notice $\boldsymbol{E}_0 I_1 = 0$, and I_2 is $\sigma(Y^1_{\varepsilon,0}, ..., Y^N_{\varepsilon,0})$-measurable. Hence

$$\boldsymbol{E}_0\Big(\frac{1}{N}\sum_{i=1}^{N} h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx\Big)^2 = \boldsymbol{E}_0 I_1^2 + I_2^2. \tag{6.2}$$

Next, notice that after conditioning on $\{Y^1_{\varepsilon,0}, ..., Y^N_{\varepsilon,0}\}$, the random variables $Y^i_{\varepsilon,T}$ are independent. Hence

$$\boldsymbol{E}_0 I_1^2 = \frac{1}{N^2}\sum_{i=1}^{N} \operatorname{Var}_0(h(Y^i_{\varepsilon,T})) \leqslant \frac{1}{N}\|h\|^2_{L^\infty}. \tag{6.3}$$

Thus, we conclude

$$\begin{aligned}(\operatorname{Err}_{\varepsilon,T}(h))^2 &\stackrel{(6.1)}{=} \boldsymbol{E}\Big[\boldsymbol{E}_0\Big(\frac{1}{N}\sum_{i=1}^{N} h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx\Big)^2\Big] \\ &\stackrel{(6.2)}{=} \boldsymbol{E}\boldsymbol{E}_0 I_1^2 + \boldsymbol{E} I_2^2 \stackrel{(6.3)}{\leqslant} \frac{1}{N}\|h\|^2_{L^\infty} + \boldsymbol{E} I_2^2.\end{aligned} \tag{6.4}$$

Step 2: In this step, we use a spectral decomposition to rewrite I_2. Notice that h can be decomposed into components along the subspace spanned by $\{1, \psi_{2,\varepsilon}\}$ and its orthogonal complement. This decomposition gives

$$h = \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx + f_0 + f_0^\perp,$$

where

$$f_0(y) \stackrel{\text{def}}{=} \Big(\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon \, dx\Big)\psi_{2,\varepsilon}(y) \tag{6.5}$$

$$f_0^\perp(y) \stackrel{\text{def}}{=} h(y) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx - \Big(\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon \, dx\Big)\psi_{2,\varepsilon}(y). \tag{6.6}$$

Therefore,

$$I_2 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{E}_0 h(Y^i_{\varepsilon,T}) - \int_{\mathbb{T}^d} h\pi_\varepsilon \, dx = I_3 + I_4, \tag{6.7}$$

where

$$I_3 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{E}_0 f_0(Y^i_{\varepsilon,T}), \quad \text{and} \quad I_4 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{E}_0 f_0^\perp(Y^i_{\varepsilon,T}).$$

Step 2.1: I_3 bound. To bound bound I_3, we first claim

$$\boldsymbol{E}_0\psi_{2,\varepsilon}(Y^i_{\varepsilon,T}) = e^{-\lambda_{2,\varepsilon}T}\psi_{2,\varepsilon}(Y^i_{\varepsilon,0})\,. \tag{6.8}$$

To see this, recall that for any $g_0 \in L^\infty(\mathbb{T}^d)$ the function g defined by

$$g_t(y) = \boldsymbol{E}^y g_0(Y_{\varepsilon,t}),$$

solves the Kolmogorov backward equation

$$\partial_t g + L_\varepsilon g = 0, \tag{6.9}$$

with initial data g_0. Here $Y_{\varepsilon,\cdot}$ is a solution to the Langevin equation (1.2). Since $\psi_{2,\varepsilon}$ is the second eigenfunction of the operator L_ε (defined in (4.6)), we see

$$\boldsymbol{E}^y\psi_{2,\varepsilon}(Y_{\varepsilon,t}) = e^{-\lambda_{2,\varepsilon}t}\psi_{2,\varepsilon}(y),$$

which immediately implies (6.8).

Now (6.5) and (6.8) imply

$$\begin{aligned} I_3 =& \frac{1}{N}\sum_{i=1}^N \boldsymbol{E}_0\Big[\Big(\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\,dx\Big)\psi_{2,\varepsilon}(Y^i_{\varepsilon,T})\Big] \\ =& \Big(\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\,dx\Big)\frac{1}{N}\sum_{i=1}^N e^{-\lambda_{2,\varepsilon}T}\psi_{2,\varepsilon}(Y^i_{\varepsilon,0}), \end{aligned}$$

and hence

$$\begin{aligned} (\boldsymbol{E}I_3^2)^{\frac12} &= \Big|\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\,dx\Big| e^{-\lambda_{2,\varepsilon}T}\Big\|\frac{1}{N}\sum_{i=1}^N\psi_{2,\varepsilon}(Y^i_{\varepsilon,0})\Big\|_{L^2(\boldsymbol{P})} \\ &= \Big|\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\,dx\Big| e^{-\lambda_{2,\varepsilon}T}\,\mathrm{Err}_{\varepsilon,0}(\psi_{2,\varepsilon}). \end{aligned} \tag{6.10}$$

Step 2.2: I_4 bound. To bound I_4 we note

$$I_4 = \frac{1}{N}\sum_{i=1}^N \boldsymbol{E}_0 f_0^\perp(Y^i_{\varepsilon,T}) = \frac{1}{N}\sum_{i=1}^N f_T^\perp(Y^i_{\varepsilon,0}),$$

where $f_t^\perp(y) = \boldsymbol{E}^y f_0^\perp(Y_{\varepsilon,t})$ and, as before, $Y_{\varepsilon,\cdot}$ is a solution of (1.2). Observe that,

$$\begin{aligned} \boldsymbol{E}I_4^2 &= \boldsymbol{E}\Big(\frac{1}{N}\sum_{i=1}^N f_T^\perp(Y^i_{\varepsilon,0})\Big)^2 \leqslant \frac{1}{N}\sum_{i=1}^N \boldsymbol{E} f_T^\perp(Y^i_{\varepsilon,0})^2 = \frac{1}{N}\sum_{i=1}^N\int_{\mathbb{T}^d}(f_T^\perp)^2 q^i_{\varepsilon,0}\,dx \\ &= \frac{1}{N}\sum_{i=1}^N\int_{\mathbb{T}^d}(f_T^\perp)^2\frac{q^i_{\varepsilon,0}}{\pi_\varepsilon}\pi_\varepsilon\,dx \leqslant \|f_T^\perp\|^2_{L^2(\pi_\varepsilon)}\max_{i=1,\dots,N}\Big\|\frac{q^i_{\varepsilon,0}}{\pi_\varepsilon}\Big\|_{L^\infty(\pi_\varepsilon)}. \end{aligned} \tag{6.11}$$

To bound $\|f_T^\perp\|^2_{L^2(\pi_\varepsilon)}$, we note that $f^\perp$ solves the Kolmogorov backward equation (6.9), and hence we have the spectral decomposition

$$\|f_T^\perp\|^2_{L^2(\pi_\varepsilon)} = \sum_{i=1}^\infty e^{-2\lambda_{i,\varepsilon}T}\Big|\int f_0^\perp\psi_{i,\varepsilon}\,\pi_\varepsilon\,dx\Big|^2$$

Using (6.5) and (6.6) the first two terms on the right vanish, and hence the spectral decomposition gives

$$\|f_T^\perp\|^2_{L^2(\pi_\varepsilon)} = \sum_{i=3}^\infty e^{-2\lambda_{i,\varepsilon}T}\Big|\int f_0^\perp\psi_{i,\varepsilon}\,\pi_\varepsilon\,dx\Big|^2 \overset{(4.10)}{\leqslant} \|f_0^\perp\|^2_{L^2(\pi_\varepsilon)}e^{-2\Lambda T}. \tag{6.12}$$

We will now bound $\|f_0^\perp\|_{L^2(\pi_\varepsilon)}$. Notice

$$\begin{aligned}\|f_0^\perp\|_{L^2(\pi_\varepsilon)} &\leqslant \Big\|h - \int_{\mathbb{T}^d} h\pi_\varepsilon\Big\|_{L^2(\pi_\varepsilon)} + \Big\|\Big(\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\Big)\psi_{2,\varepsilon}\Big\|_{L^2(\pi_\varepsilon)} \\ &\leqslant \|h\|_{L^2(\pi_\varepsilon)} + \Big|\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\Big| \leqslant 2\|h\|_{L^2(\pi_\varepsilon)} \leqslant 2\|h\|_{L^\infty}.\end{aligned}$$

Together with (6.12) this gives

$$\|f_T^\perp\|_{L^2(\pi_\varepsilon)} \leqslant 2\|h\|_{L^\infty} e^{-\Lambda T}. \tag{6.13}$$

Therefore, plugging (6.13) into (6.11) yields

$$(\boldsymbol{E} I_4^2)^{\frac{1}{2}} \leqslant 2\|h\|_{L^\infty(\pi_\varepsilon)} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\|\frac{q_{\varepsilon,0}^i}{\pi_\varepsilon}\Big\|_{L^\infty(\pi_\varepsilon)}^{\frac{1}{2}} = \mathcal{E}_{\varepsilon,T}(h). \tag{6.14}$$

Step 3: Based on the previous steps,

$$\begin{aligned}(\mathrm{Err}_{\varepsilon,T}(h))^2 &\overset{(6.4)}{\leqslant} \frac{1}{N}\|h\|_{L^\infty}^2 + \boldsymbol{E} I_2^2 \overset{(6.7)}{\leqslant} \frac{1}{N}\|h\|_{L^\infty}^2 + \Big((\boldsymbol{E} I_3^2)^{\frac{1}{2}} + \boldsymbol{E}(I_4^2)^{\frac{1}{2}}\Big)^2 \\ &\overset{(6.10),(6.14)}{\leqslant} \frac{1}{N}\|h\|_{L^\infty}^2 + \Big(\Big|\int_{\mathbb{T}^d} h\psi_{2,\varepsilon}\pi_\varepsilon\, dx\Big| e^{-\lambda_{2,\varepsilon}T}\, \mathrm{Err}_{\varepsilon,0}(\psi_{2,\varepsilon}) + \mathcal{E}_{\varepsilon,T}(h)\Big)^2.\end{aligned}$$

Taking square root on both sides and using $\|h\|_{L^\infty} = \frac{1}{2}\|h\|_{\mathrm{osc}}$ finish the proof. ☐

6.2. **Growth of $\|q_{k,0}^i/\pi_k\|_{L^\infty}$ (Lemma 4.8).** In this section, we prove Lemma 4.8 which will be used in the proof of Lemma 4.7, and was also used in the proof of Theorem 2.8 to obtain (4.24). Let $q_{k,t}^i$ denote the probability density of $X_{k,t}^i$. The proof Lemma 4.8 involves controlling the growth of $\|q_{k,t}^i/\pi_k\|_{L^\infty}$ in the Langevin step, and in the resampling step. In the Langevin step, $\|q_{k,t}^i/\pi_k\|_{L^\infty}$ is nonincreasing due to maximum principle. In the resampling step, the growth of $\|q_{k,0}/\pi_k\|_{L^\infty(\pi_k)}$ between levels is tracked using duality.

Proof of Lemma 4.8. The proof contains three steps. Fix $T_0 > 0$. In the first step, there exists a constant $C_q = C_q(U, T_0)$ such that for any $T > T_0$ we have

$$\max_{i=1,\dots,N} \Big\|\frac{q_{1,T}^i}{\pi_1}\Big\|_{L^\infty} \leqslant C_q. \tag{6.15}$$

Next we will show that for every $k \geqslant 2$,

$$\max_{i=1,\dots,N} \Big\|\frac{q_{k,0}^i}{\pi_k}\Big\|_{L^\infty} \leqslant \frac{1}{\prod_{\ell=1}^{k-1} \min r_\ell} \max_{i=1,\dots,N} \Big\|\frac{q_{1,T}^i}{\pi_1}\Big\|_{L^\infty}, \tag{6.16}$$

where r_k is defined in (3.1).

Finally we show by direct computation that

$$\prod_{\ell=1}^{k-1} \min r_\ell \geqslant \exp\Big(\Big(1 - \frac{1}{\eta}\Big)\|U\|_{\mathrm{osc}}\Big). \tag{6.17}$$

Combining (6.15), (6.16) and (6.17) completes the proof. We will now each of the above inequalities.

Step 1: Proof of (6.15). Since $X_{1,\cdot}^i$ solves (1.2) with $\varepsilon = \eta_1 = 1$, it's density, denoted by $q_{1,t}^i$ must solve (4.8) (with $\varepsilon = \eta_1$). Since L_1^* (defined by (4.7), with $\varepsilon = 1$) is nondegenerate, parabolic regularity implies there exists a constant $C = C(U, T_0)$ such that

$$\max_{i=1,\dots,N} \|q^i_{1,T_0}\|_{L^\infty} \leqslant C \|q^i_{1,T_0/2}\|_{L^1} = C\,.$$

The last equality above followed because $q^i_{1,t}$ is a probability density and so $\|q^i_{1,t}\|_{L^1} = 1$ for every $t > 0$. Thus

$$\Big\| \frac{q^i_{1,T_0}}{\pi_1} \Big\|_{L^\infty} \leqslant \frac{\|q^i_{1,T_0}\|_{L^\infty}}{\min \pi_1} \leqslant \frac{C}{\min \pi_1} \tag{6.18}$$

and we choose $C_q = C_q(U, T_0)$ to be the right hand side of the above.

Since $q^i_{1,t}/\pi_1$ solves the backward equation (6.9) with $\varepsilon = \eta_1 = 1$, the maximum principle and (6.18) imply (6.15) for all $T \geqslant T_0$.

Step 2: Proof of (6.16). We claim that for all $k \in \{1, \dots, M-1\}$, we have

$$\max_{i=1,\dots,N} \Big\| \frac{q^i_{k+1,0}}{\pi_{k+1}} \Big\|_{L^\infty} \leqslant \frac{1}{\min r_k} \max_{i=1,\dots,N} \Big\| \frac{q^i_{k,T}}{\pi_k} \Big\|_{L^\infty}. \tag{6.19}$$

Next we note that $q^i_{k,t}/\pi_k$ satisfies (6.9) with $\varepsilon = \eta_k$. Thus, by the maximum principle we have

$$\Big\| \frac{q^i_{k,T}}{\pi_k} \Big\|_{L^\infty} \leqslant \Big\| \frac{q^i_{k,0}}{\pi_k} \Big\|_{L^\infty}, \tag{6.20}$$

for every $k \in \{2, \dots, M\}$ and every $i \in \{1, \dots, N\}$. The bound (6.16) immediately follows from (6.19) and (6.20). Thus it only remains to prove (6.19).

We note that if $X^1_{k,T}$, ..., $X^N_{k,T}$ were i.i.d. then one has an explicit formula for $q^i_{k+1,0}$, from which (6.19) follows immediately. In our situation these processes are not independent, and so we prove (6.19) using duality, and without relying on an explicit formula.

For any test function $h \in L^1(\pi_{k+1})$ be a test function we have

$$\begin{aligned}
\int_{\mathbb{T}^d} h(x) q^i_{k+1,0}(x)\,dx &= \boldsymbol{E} h(X^i_{k+1,0}) \\
&= \boldsymbol{E}\boldsymbol{E}\big(h(X^i_{k+1,0}) \,\big|\, X^k_{1,T}, \dots, X^N_{k,T}\big) = \boldsymbol{E}\Big(\frac{\sum_j h(X^j_{k,T}) r_k(X^j_{k,T})}{\sum_j r_k(X^j_{k,T})} \Big) \\
&\leqslant \frac{1}{N \min r_k} \sum_{j=1}^N \boldsymbol{E}\big| h(X^j_{k,T}) r_k(X^j_{k,T}) \big|. \qquad (6.21)
\end{aligned}$$

Next, we note that for every $j = 1, \dots, N$,

$$\begin{aligned}
\boldsymbol{E}\big| h(X^j_{k,T}) r_k(X^j_{k,T}) \big| &= \int_{\mathbb{T}^d} |h| r_k q^j_{k,T}\,dx \overset{(3.1)}{=} \int_{\mathbb{T}^d} |h| \frac{r_k q^j_{k,T}}{r_k \pi_k} \pi_{k+1}\,dx \\
&\leqslant \|h\|_{L^1(\pi_{k+1})} \Big\| \frac{q^j_{k,T}}{\pi_k} \Big\|_{L^\infty}. \qquad (6.22)
\end{aligned}$$

Thus (6.21) and (6.22) imply

$$\Big| \int_{\mathbb{T}^d} h(x) q^i_{k+1,0}(x)\,dx \Big| \leqslant \frac{\|h\|_{L^1(\pi_{k+1})}}{\min r_k} \max_{j=1,\dots,N} \Big\| \frac{q^j_{k,T}}{\pi_k} \Big\|_{L^\infty},$$

from which (6.19) follows by duality.

Step 3: Proof of (6.17). Observe that

$$r_k(x) = \frac{Z_k}{Z_{k+1}} \exp\Big(-\Big(\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k}\Big)U(x)\Big) \tag{6.23}$$

where $Z_k \stackrel{\text{def}}{=} Z_{\eta_k}$, and Z_{η_k} is the normalization constant in (1.1). Hence, for all $k \in \{1, \dots, M-1\}$ the minimum of r_k is attained at the same point, which we denote by x^*. Thus,

$$\prod_{\ell=1}^{k-1} \min r_\ell = \prod_{\ell=1}^{k-1} r_\ell(x^*) = \frac{\pi_k(x^*)}{\pi_1(x^*)} = \frac{Z_1}{Z_k} \exp\Big(\Big(1 - \frac{1}{\eta_k}\Big)U(x^*)\Big). \tag{6.24}$$

Since $U \geqslant 0$ by assumption, and $\eta_k < \eta_1 = 1$, we must have $Z_1 \geqslant Z_k$. Using this in (6.24) immediately implies (6.17) as desired. This completes the proof of Lemma 4.8. $\square$

7. Iterating error estimates (Lemmas 4.7 and 4.9)

Lemma 4.7 consists of two main parts: the derivation of recurrence relation (4.14), and obtaining the estimate (4.15) for β_k, c_k. We do each of these steps in Sections 7.1, 7.2 and 7.3. We combine these and prove Lemma 4.7 in Section 7.4.

7.1. **Recurrence relation.** We will now prove (4.14) by combining the estimate for the Monte Carlo error (Lemma 4.6) and the resampling error (Lemma 3.3). For clarity, we state this as a new lemma and give explicit formulae for the constants β_k and c_k appearing in (4.14).

Lemma 7.1. *For each $2 \leqslant k \leqslant M-1$, the inequality* (4.14) *holds with β_k and c_k given by*

$$\begin{aligned}\beta_k \stackrel{\text{def}}{=} e^{-\lambda_{2,k}T}\Big(\Big|\int_{\mathbb{T}^d} r_k \psi_{2,k} \pi_k \, dx\Big| \cdot \|\psi_{2,k+1}\|_{L^\infty} \\ + \Big|\int_{\mathbb{T}^d} \psi_{2,k+1}\psi_{2,k}\pi_{k+1} \, dx\Big|\Big)\end{aligned} \tag{7.1}$$

$$\begin{aligned}c_k \stackrel{\text{def}}{=} 3\|\psi_{2,k+1}\|_{L^\infty}\|r_k\|_{L^\infty}\Big(\frac{1}{2\sqrt{N}} + e^{-\Lambda T} \max_{i=1,\dots,N} \Big\|\frac{q^i_{k,0}}{\pi_k}\Big\|^{\frac12}_{L^\infty(\pi_k)}\Big) \\ + \frac{1}{\sqrt{N}}\|\psi_{2,k+1}\|_{L^\infty}.\end{aligned} \tag{7.2}$$

Proof of Lemma 7.1. Applying Lemma 3.3 with

$$p = \pi_k, \quad q = \pi_{k+1}, \quad h = \psi_{2,k+1}, \quad x^i = X^i_{k,T}, \quad y^i = X^i_{k+1,0}$$

gives

$$\begin{aligned}\mathrm{Err}_{k+1,0}(\psi_{2,k+1}) &= \Big\|\frac{1}{N}\sum_{i=1}^N \psi_{2,k+1}(X^i_{k+1,0})\Big\|_{L^2(\boldsymbol{P})} \\ &\stackrel{(3.13)}{\leqslant} \frac{1}{\sqrt{N}}\|\psi_{2,k+1}\|_{L^\infty} + \|\psi_{2,k+1}\|_{L^\infty}\Big\|1 - \frac{1}{N}\sum_{\ell=1}^N r_k(X^i_{k,T})\Big\|_{L^2(\boldsymbol{P})} \\ &\quad + \Big\|\frac{1}{N}\sum_{i=1}^N \psi_{2,k+1}(X^i_{k,T}) r_k(X^i_{k,T})\Big\|_{L^2(\boldsymbol{P})}\end{aligned}$$

$$= \frac{1}{\sqrt{N}} \|\psi_{2,k+1}\|_{L^\infty} + \|\psi_{2,k+1}\|_{L^\infty} \operatorname{Err}_{k,T}(r_k) + \operatorname{Err}_{k,T}(\psi_{2,k+1} r_k). \tag{7.3}$$

We will now bound the last two terms on the right of (7.3). For the term $\operatorname{Err}_{k,T}(r_k)$, we apply Lemma 4.6 with

$$\varepsilon = \eta_k, \quad h = r_k, \quad q_{\varepsilon,0} = q_{k,0},$$

to obtain

$$\begin{aligned} \operatorname{Err}_{k,T}(r_k) &\overset{(4.11)}{\leqslant} \frac{1}{2\sqrt{N}} \|r_k\|_{L^\infty} + \Big| \int_{\mathbb{T}^d} r_k \psi_{2,k} \pi_k \, dx \Big| e^{-\lambda_{2,k} T} \operatorname{Err}_{k,0}(\psi_{2,k}) \\ &\quad + \|r_k\|_{L^\infty} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\| \frac{q_{k,0}^i}{\pi_k} \Big\|_{L^\infty(\pi_k)}^{\frac{1}{2}} \end{aligned} \tag{7.4}$$

where we use the fact that $\|r_k\|_{\text{osc}} \leqslant \|r_k\|_{L^\infty}$.

Similarly, for the term $\operatorname{Err}_{k,T}(\psi_{2,k+1} r_k)$, we apply Lemma 4.6 with

$$\varepsilon = \eta_k, \quad h = \psi_{2,k+1} r_k, \quad q_{\varepsilon,0} = q_{k,0}$$

to obtain

$$\begin{aligned} \operatorname{Err}_{k,T}(\psi_{2,k+1} r_k) &\overset{(4.11)}{\leqslant} \Big| \int_{\mathbb{T}^d} \psi_{2,k+1} \psi_{2,k} \pi_{k+1} \, dx \Big| e^{-\lambda_{2,k} T} \operatorname{Err}_{k,0}(\psi_{2,k}) \\ &\quad + \frac{1}{\sqrt{N}} \|r_k\|_{L^\infty} \|\psi_{2,k+1}\|_{L^\infty} \\ &\quad + 2\|r_k\|_{L^\infty} \|\psi_{2,k+1}\|_{L^\infty} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\| \frac{q_{k,0}^i}{\pi_k} \Big\|_{L^\infty(\pi_k)}^{\frac{1}{2}}, \end{aligned} \tag{7.5}$$

where we use the fact that

$$\|r_k \psi_{2,k+1}\|_{\text{osc}} \leqslant 2\|r_k \psi_{2,k+1}\|_{L^\infty} \leqslant 2\|r_k\|_{L^\infty} \|\psi_{2,k+1}\|_{L^\infty}.$$

Plugging (7.4) and (7.5) into (7.3) and using (7.1), (7.2) yields (4.14), completing the proof. □

7.2. **Estimate of c_k.** Of the terms on the right hand side of (7.2), the term $\|q_{k,0}^i/\pi_k\|_{L^\infty(\pi_k)}$ has already been bounded in Lemma 4.8. We will now bound the remaining terms. First, using a local maximum principle we will obtain an L^∞ bound on the second L^2-normalized eigenfunction of L_ε that is uniform in ε. This is our first lemma.

Lemma 7.2. *There exists constant $C_\psi = C_\psi(U, d, C_m) > 0$ independent of ε such that*

$$\sup_{0<\varepsilon\leqslant 1} \|\psi_{2,\varepsilon}\|_{L^\infty(\mathbb{T}^d)} \leqslant C_\psi. \tag{7.6}$$

Lemma 7.2 immediately gives a bound on $\|\psi_{2,k}\|_{L^\infty}$ that is uniform k. Since the proof of Lemma 7.2 is somewhat lengthy, we postpone it to Section 8.4.

We will now show how N and T can be chosen so that we obtain the bound for c_k in (4.15). Now we are equipped to prove Lemma 7.3.

Lemma 7.3 (Estimate of c_k). *Fix $\delta > 0$, let $C_q = C_q(U, 1)$ be the constant from Lemma 4.8 with $T_0 = 1$, and C_ψ, C_r be the constants defined in (7.6) and (3.2) respectively. Define $\tilde{C}_N$ by*

$$\tilde{C}_N \stackrel{\text{def}}{=} 4\Big(C_\psi\Big(1 + \frac{3}{2} C_r\Big)\Big)^2. \tag{7.7}$$

If N and T are chosen such that

$$N \geqslant \tilde{C}_N \frac{M^2}{\delta^2}, \tag{7.8}$$

$$T \geqslant \max\Big\{\frac{1}{\Lambda}\Big(\log\Big(\frac{1}{\delta}\Big) + \frac{\|U\|_{\mathrm{osc}}}{2\eta} + \log M + \log(6C_\psi C_r C_q^{\frac{1}{2}})\Big), 1\Big\}, \tag{7.9}$$

then

$$c_k \leqslant \frac{\delta}{M}, \quad \textit{for every } 2 \leqslant k \leqslant M-1.$$

Proof of Lemma 7.3. We first rewrite (7.2) as

$$c_k = I_1 + I_2 \tag{7.10}$$

where

$$I_1 = \frac{1}{\sqrt{N}} \|\psi_{2,k+1}\|_{L^\infty} \Big(1 + \frac{3}{2}\|r_k\|_{L^\infty}\Big),$$

$$I_2 = 3\|\psi_{2,k+1}\|_{L^\infty} \|r_k\|_{L^\infty} e^{-\Lambda T} \max_{i=1,\dots,N} \Big\| \frac{q^i_{k,0}}{\pi_k} \Big\|^{\frac{1}{2}}_{L^\infty(\pi_k)}.$$

Notice that the choice of T and N gives that

$$I_1 \overset{(3.2),(7.6)}{\leqslant} \frac{1}{\sqrt{N}} C_\psi \Big(1 + \frac{3}{2} C_r\Big) \overset{(7.7),(7.8)}{\leqslant} \frac{\delta}{2M}, \tag{7.11}$$

and

$$e^{-\Lambda T} \overset{(7.9)}{\leqslant} \delta \exp\Big(-\frac{\|U\|_{\mathrm{osc}}}{2\eta}\Big) \frac{1}{M} \frac{1}{6C_r C_\psi C_q^{\frac{1}{2}}}. \tag{7.12}$$

Therefore,

$$I_2 \overset{(7.6),(3.2),(4.16)}{\leqslant} 3C_\psi C_r e^{-\Lambda T} C_q^{\frac{1}{2}} \exp\Big(\|U\|_{\mathrm{osc}}\big(\frac{1}{\eta_k} - 1\big)\Big) \overset{(7.12)}{\leqslant} \frac{\delta}{2M}. \tag{7.13}$$

Using (7.11) and (7.13) in (7.10) concludes the proof. □

7.3. **Estimate of β_k.** Recall from (4.14) the error grows by a factor of β_k at each level, and so to prove Theorem 2.8 we need to ensure $\prod \beta_k$ remains bounded. The main result in this section (Lemma 7.9, below) obtains this bound and shows that the first inequality in (4.15) holds. For simplicity of notation, let

$$\Theta(k, k+1) \overset{\text{def}}{=} \Big| \int_{\mathbb{T}^d} r_k \psi_{2,k} \pi_k \, dx \Big| \cdot \|\psi_{2,k+1}\|_{L^\infty} + \Big| \int_{\mathbb{T}^d} \psi_{2,k+1} \psi_{2,k} \pi_{k+1} \, dx \Big|, \tag{7.14}$$

and note

$$\beta_k = e^{-\lambda_{2,k} T} \Theta(k, k+1).$$

We will bound $\prod_{j=k}^{M-1} \beta_j$ differently when the temperature η_k is low, and when it is high. First, when the temperature is low, the exponential factor $e^{-\lambda_{2,k}T}$ is very close to 1, and does not help much. In this case will show that the product $\prod_{j=k}^{M-1} \Theta(j, j+1)$ stays bounded, by approximating $\Theta(k, k+1)$ in terms of the mass in each well and estimating the mass distribution using small temperature asymptotics. When the temperature is high, the small temperature asymptotics are not valid anymore. However, in this case $\lambda_{2,k}$ is not too small, and can be used to ensure $\beta_k \leqslant 1$ in a relatively short amount of time.

We begin by stating the fact that when ε is small, the second eigenfunction $\psi_{2,\varepsilon}$ is very close to a linear combination of $\mathbf{1}_{\Omega_1}$ and $\mathbf{1}_{\Omega_2}$. To state a precise bound, consider the subspaces $E_\varepsilon, F_\varepsilon \subseteq L^2(\pi_\varepsilon)$ defined by

$$F_\varepsilon \stackrel{\text{def}}{=} \operatorname{span}\{1, \psi_{2,\varepsilon}\}, \quad E_\varepsilon \stackrel{\text{def}}{=} \operatorname{span}\{\mathbf{1}_{\Omega_1}, \mathbf{1}_{\Omega_2}\}. \tag{7.15}$$

We measure closeness of $\psi_{2,\varepsilon}$ to a linear combination of $\mathbf{1}_{\Omega_1}$ and $\mathbf{1}_{\Omega_2}$, by measuring the "distance" between the subspaces E_ε and F_ε defined by

$$d(E_\varepsilon, F_\varepsilon) \stackrel{\text{def}}{=} \|P_{E_\varepsilon} - P_{E_\varepsilon} P_{F_\varepsilon}\| = \|P_{E_\varepsilon} - P_{F_\varepsilon} P_{E_\varepsilon}\|.$$

Here P_{E_ε}, P_{E_ε} are the $L^2(\pi_\varepsilon)$ orthonormal projectors onto E_ε and F_ε respectively. The next result gives an estimate on $d(E_\varepsilon, F_\varepsilon)$.

Proposition 7.4 (Chapter 8, Proposition 2.2 of [Kol00])**.** *Let $\hat\gamma$ be the energy barrier defined in (4.2). For any $\gamma < \hat\gamma$, there exists a constant $\tilde C_\gamma > 0$ such that for all $\varepsilon \leqslant 1$, we have*

$$d(E_\varepsilon, F_\varepsilon) \leqslant \tilde C_\gamma \exp\Big(\frac{-\gamma}{\varepsilon}\Big). \tag{7.16}$$

We will use Proposition 7.4 to estimate the two integration terms appearing in $\Theta(k, k+1)$. The bounds we need are stated in the next two lemmas, and their proofs will be postponed to Section 8.1.

Lemma 7.5. *Let $\varepsilon' < \varepsilon$ and define r_ε by*

$$r_\varepsilon \stackrel{\text{def}}{=} \frac{\pi_{\varepsilon'}}{\pi_\varepsilon}. \tag{7.17}$$

Then,

$$\begin{aligned} \Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\psi_{2,\varepsilon'} r_\varepsilon \pi_\varepsilon \, dx\Big| \leqslant \min\Big\{ &\|r_\varepsilon\|_{L^\infty}^{\frac12}, \ \|r_\varepsilon\|_{L^\infty}^{\frac12} d(E_\varepsilon, F_\varepsilon) \\ &(1 + (\pi_\varepsilon(\Omega_2) - \pi_\varepsilon(\Omega_1))(\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)))^{\frac12}\Big\}. \end{aligned} \tag{7.18}$$

Lemma 7.6. *Let $\varepsilon' < \varepsilon$. Then*

$$\begin{aligned} \Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\pi_{\varepsilon'} \, dx\Big| \leqslant \min\Bigg\{ &\|r_\varepsilon\|_{L^\infty(\pi_\varepsilon)}, \\ &\Bigg(\frac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}} + \frac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\Bigg) \cdot |\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)| + d(E_\varepsilon, F_\varepsilon)\|r_\varepsilon\|_{L^\infty(\pi_\varepsilon)}\Bigg\}. \end{aligned} \tag{7.19}$$

To apply the previous two results we need to ensure the masses in the two wells stay away from 0 (Assumption 4.3), and do not oscillate too much. We will now show that the required oscillation condition hold provided U satisfies Assumption 4.1 holds.

Lemma 7.7. *If U satisfies Assumption 4.1 then there exists a constant C_{BV} such that such that for every $\eta \in (0,1)$, and every $i \in \{1,2\}$ we have*

$$\int_\eta^1 |\partial_\varepsilon \pi_\varepsilon(\Omega_i)| \, d\varepsilon \leqslant C_{\mathrm{BV}}. \tag{7.20}$$

Notice that (7.20) combined with our assumption Assumption 4.3 implies (2.3) holds. This was the condition required to obtain error estimates for ASMC applied to the local mixing model (Theorem 2.2). We prove Lemma 7.7 in Section 8.2, below.

Finally, we require a lower bound on $\lambda_{2,k}$ to ensure that in high temperature regime we can make $\beta_k \leqslant 1$ in a relative short time.

Lemma 7.8. *Suppose U satisfies assumptions 4.1 and 4.2, and recall $\hat{U}$ is the saddle height defined in (4.1). For every $H > \hat{U}$, there exists $A \stackrel{\text{def}}{=} A(H, d, U) > 0$ independent of ε such that for every $\varepsilon < 1$,*

$$\lambda_{2,\varepsilon} \geqslant A \exp\Big(-\frac{H}{\varepsilon}\Big). \tag{7.21}$$

Postponing the proof of Lemma 7.8 to Section 8.3, we now bound $\prod \beta_j$ to obtain the first inequality in (4.15).

Lemma 7.9 (Estimate of β_j)**.** *For any $\alpha > 0$, there exists constant $\tilde{C}_\alpha = \tilde{C}_\alpha(\alpha, U)$ such that for $2 \leqslant j \leqslant M$, if at each step $T \geqslant \tilde{C}_\alpha M^{(1+\alpha)\hat{\gamma}_r}$, then for $2 \leqslant k \leqslant M-1$, the first inequality inequality (4.15) holds with*

$$C_\beta \stackrel{\text{def}}{=} \exp\Big(C_{\text{BV}}\Big(2C_m C_\psi + \frac{1}{2}\Big) + C_\psi C_r + C_r^{\frac{1}{2}}\Big). \tag{7.22}$$

Here C_r, C_m, C_ψ and C_{BV} are the constants defined in (3.2), (4.4), (7.6), and (7.20), respectively.

Proof. Given a fixed $\alpha > 0$, we choose

$$H \stackrel{\text{def}}{=} (1+\alpha)^{\frac{1}{2}} \hat{U} > \hat{U}, \quad \text{and} \quad \gamma \stackrel{\text{def}}{=} \frac{\hat{\gamma}}{(1+\alpha)^{\frac{1}{2}}} < \hat{\gamma}, \tag{7.23}$$

which gives that

$$\frac{H}{\gamma} = (1+\alpha)\hat{\gamma}_r. \tag{7.24}$$

Given H and γ as (7.23), there exists $A_H, C_\gamma > 0$ independent of ε such that

$$A_H \exp(-\frac{H}{\varepsilon}) \overset{(7.21)}{\leqslant} \lambda_{2,\varepsilon} \overset{(4.10)}{\leqslant} C_\gamma \exp\Big(-\frac{\gamma}{\varepsilon}\Big), \quad \text{for all } \varepsilon < 1. \tag{7.25}$$

We choose a critical temperature $\eta_{\text{cr}} > 0$ so that

$$(C_\gamma \vee \tilde{C}_\gamma) \exp\Big(-\frac{\gamma}{\eta_{\text{cr}}}\Big) = \frac{1}{M}. \tag{7.26}$$

Recall that $\tilde{C}_\gamma$ is the constant defined in (7.16). We will prove the first inequality in (4.15) holds by splitting the analysis into two cases.

Case I: $\eta \geqslant \eta_{\text{cr}}$. In this case, for every $k \geqslant 2$ we have $\eta_k \geqslant \eta \geqslant \eta_{\text{cr}}$ and so

$$\exp\Big(-\frac{\gamma}{\eta_k}\Big) \geqslant \frac{1}{(C_\gamma \vee \tilde{C}_\gamma)M}. \tag{7.27}$$

This implies

$$\lambda_{2,k} \overset{(7.24),(7.25)}{\geqslant} A_H\Big(\exp\Big(-\frac{\gamma}{\eta_k}\Big)\Big)^{(1+\alpha)\hat{\gamma}_r} \overset{(7.27)}{\geqslant} A_H\Big(\frac{1}{(C_\gamma \vee \tilde{C}_\gamma)M}\Big)^{(1+\alpha)\hat{\gamma}_r}. \tag{7.28}$$

Therefore, for

$$T \geqslant \tilde{C}_\alpha M^{(1+\alpha)\hat{\gamma}_r}, \quad \tilde{C}_\alpha \stackrel{\text{def}}{=} \frac{1}{A_H}(C_\gamma \vee \tilde{C}_\gamma)^{(1+\alpha)\hat{\gamma}_r} \log(C_\psi C_r + C_r^{\frac{1}{2}}) \tag{7.29}$$

we have that for $k \geqslant 2$,

$$\lambda_{2,k} T \overset{(7.28)}{\geqslant} \log(C_\psi C_r + C_r^{\frac{1}{2}}). \tag{7.30}$$

Therefore, by Lemma 7.6 and Lemma 7.5,

$$
\begin{aligned}
&e^{-\lambda_{2,k}T}\Theta(k,k+1)\\
&\quad\overset{(7.14)}{=} e^{-\lambda_{2,k}T}\Big(\Big|\int_{\mathbb{T}^d} r_k\psi_{2,k}\pi_k\,dx\Big|\|\psi_{2,k+1}\|_{L^\infty}+\Big|\int_{\mathbb{T}^d}\psi_{2,k+1}\psi_{2,k}\pi_{k+1}\,dx\Big|\Big)\\
&\overset{(7.19),(7.18),(7.6)}{\leqslant} e^{-\lambda_{2,k}T}\Big(C_\psi\|r_k\|_{L^\infty}+\|r_k\|_{L^\infty}^{\frac12}\Big)\\
&\quad\overset{(7.30)}{\leqslant}\frac{1}{C_\psi C_r+C_r^{\frac12}}\Big(C_\psi\|r_k\|_{L^\infty}+\|r_k\|_{L^\infty}^{\frac12}\Big)\overset{(3.2)}{\leqslant}1. \qquad (7.31)
\end{aligned}
$$

We conclude that, if $\eta\geqslant\eta_{\mathrm{cr}}$, and T satisfies (7.29), then for every k,

$$\prod_{j=k}^{M-1}\beta_j=\prod_{j=k}^{M-1}\Big(e^{-\lambda_{2,k}T}\Theta(k,k+1)\Big)\overset{(7.31)}{\leqslant}1. \qquad (7.32)$$

Case II: $\eta<\eta_{\mathrm{cr}}$. Define k_0 by

$$k_0\overset{\text{def}}{=}\min\{2\leqslant k\leqslant M-1 \mid \eta_k\leqslant\eta_{\mathrm{cr}}\}.$$

We first consider $k>k_0$, in which case we have $\eta_k<\eta_{\mathrm{cr}}$. Observe that by Proposition 7.4, we have

$$d(E_{\eta_k},F_{\eta_k})\overset{(7.16)}{\leqslant}\tilde C_\gamma\exp\Big(-\frac{\gamma}{\eta_k}\Big)\overset{(7.26)}{\leqslant}\frac{1}{M}. \qquad (7.33)$$

To bound $\Theta(k,k+1)$, we write

$$\Theta(k,k+1)=J_1\|\psi_{2,k+1}\|_{L^\infty}+J_2, \qquad (7.34)$$

where

$$J_1=\Big|\int_{\mathbb{T}^d} r_k\psi_{2,k}\pi_k\,dx\Big|,\quad J_2=\Big|\int_{\mathbb{T}^d}\psi_{2,k+1}\psi_{2,k}\pi_{k+1}\,dx\Big|.$$

Step 1: Estimating J_1 and J_2. We first estimate J_1 and J_2 using Lemma 7.5 and 7.6 respectively. For simplicity, by a slight abuse of notation we write

$$\pi_k(\Omega_i)\overset{\text{def}}{=}\pi_{\eta_k}(\Omega_i),\quad i=1,2.$$

By Lemma 7.6,

$$
\begin{aligned}
J_1&\overset{(7.19)}{\leqslant} d(E_{\eta_k},F_{\eta_k})\|r_k\|_{L^\infty}+\Big(\frac{\sqrt{\pi_k(\Omega_2)}}{\sqrt{\pi_k(\Omega_1)}}+\frac{\sqrt{\pi_k(\Omega_1)}}{\sqrt{\pi_k(\Omega_2)}}\Big)\cdot|\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1)|\\
&\overset{(4.4)}{\leqslant} d(E_{\eta_k},F_{\eta_k})\|r_k\|_{L^\infty}+2C_m\cdot\big|\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1)\big|\\
&\overset{(7.33)}{\leqslant}\frac{C_r}{M}+2C_m\cdot\big|\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1)\big|. \qquad (7.35)
\end{aligned}
$$

By Lemma 7.5, using the fact that $(1+y)^{\frac12}\leqslant1+\frac12y$ when $y>0$, we have

$$
\begin{aligned}
J_2&\overset{(7.18)}{\leqslant}\|r_k\|_{L^\infty}^{\frac12}d(E_{\eta_k},F_{\eta_k})+\Big(1+(\pi_k(\Omega_2)-\pi_k(\Omega_1))(\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1))\Big)^{\frac12}\\
&\overset{(7.33),(3.2)}{\leqslant}\frac{C_r^{\frac12}}{M}+\Big(1+\big|\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1)\big|\Big)^{\frac12}\\
&\leqslant\frac{C_r^{\frac12}}{M}+1+\frac12\big|\pi_{k+1}(\Omega_1)-\pi_k(\Omega_1)\big|. \qquad (7.36)
\end{aligned}
$$

Hence,

$$\begin{aligned}\Theta(k,k+1) &\overset{(7.34)}{=} J_1\|\psi_{2,k+1}\|_{L^\infty} + J_2\\ &\overset{(7.35),(7.36)}{\leqslant} 1 + (2C_mC_\psi + \frac{1}{2})\cdot\big|\pi_{k+1}(\Omega_1) - \pi_k(\Omega_1)\big| + \frac{C_\psi C_r + C_r^{\frac{1}{2}}}{M}.\end{aligned} \tag{7.37}$$

Step 2: Estimating $\prod_{j=k}^{M-1}\Theta(j,j+1)$. By direct computation, for $k \geqslant k_0$,

$$\begin{aligned}\prod_{j=k}^{M-1}\beta_j &\leqslant \prod_{j=k}^{M-1}\Theta(j,j+1)\\ &\overset{(7.37)}{\leqslant}\prod_{j=k}^{M-1}\Big(1 + (2C_mC_\psi + \frac{1}{2})\cdot\big|\pi_{j+1}(\Omega_1) - \pi_j(\Omega_1)\big| + \frac{C_\psi C_r + C_r^{\frac{1}{2}}}{M}\Big)\\ &\overset{\text{AM-GM}}{\leqslant}\Big(1 + \frac{C_\psi C_r + C_r^{\frac{1}{2}}}{M}\\ &\qquad\qquad + \frac{1}{M-k}\sum_{j=k}^{M-1}(2C_mC_\psi + \frac{1}{2})\cdot\big|\pi_{j+1}(\Omega_1) - \pi_j(\Omega_1)\big|\Big)^{M-k}\\ &\overset{(7.20)}{\leqslant}\left(1 + \frac{C_{\mathrm{BV}}(2C_mC_\psi + \frac{1}{2})}{M-k} + \frac{C_\psi C_r + C_r^{\frac{1}{2}}}{M}\right)^{M-k}\overset{(7.22)}{\leqslant} C_\beta,\end{aligned} \tag{7.38}$$

where the last inequality uses the fact that $M-k \leqslant M$.

Next, for the case or $k < k_0$, we observe that $\eta_k \geqslant \eta_{\mathrm{cr}}$. Now using the same argument as in the case $\eta \geqslant \eta_{\mathrm{cr}}$ we see,

$$\prod_{j=k}^{k_0}\beta_j \leqslant 1, \tag{7.39}$$

provided T satisfies (7.29). This implies that

$$\prod_{j=k}^{M-1}\beta_j = \Big(\prod_{j=k}^{k_0}\beta_j\Big)\Big(\prod_{j=k_0}^{M-1}\beta_j\Big)\overset{(7.39)}{\leqslant}\prod_{j=k_0}^{M-1}\beta_j\overset{(7.38)}{\leqslant} C_\beta. \tag{7.40}$$

Combining (7.32), (7.38) and (7.40) completes the proof. □

7.4. **Proof of Lemma 4.7.** Now we prove Lemma 4.7. The proof follows immediately from Lemmas 7.1, 7.9 and 7.3.

Proof of Lemma 4.7. Choose $\alpha > 0$. For any given $\delta > 0$, we take

$$\begin{aligned}N &= \tilde{C}_N\frac{M^2}{\delta^2}\\ T &= \max\Big\{\frac{1}{\Lambda}\Big(\log(\frac{1}{\delta}) + \frac{\|U\|_{\mathrm{osc}}}{2\eta} + \log(M) + \log(6C_rC_\psi C_q^{\frac{1}{2}})\Big),\\ &\qquad 1, \tilde{C}_\alpha M^{(1+\alpha)\hat{\gamma}_r}\Big\}.\end{aligned}$$

Notice that if T, N are chosen as above, then we can find constants $C_\alpha = C_\alpha(\alpha, U) > 0$ so that this choice is consistent with the choice in Lemma 4.7. Using Lemmas 7.1, 7.3 and 7.9 we obtain (4.14) and (4.15) as desired. □

7.5. **Mixing at the First Level (Lemma 4.9).** In this section we prove Lemma 4.9 which bounds $\mathrm{Err}_{2,0}(\psi_{2,2})$. We obtain this bound without relying on the initial mass distribution, but instead using fast mixing at first level.

Proof of Lemma 4.9. The proof is analogous to that of Lemma 7.1 except when applying Lemma 4.6, we choose $q_{\varepsilon,0} = q_{1,T_0}$ with $T_0 = 1$. For $T \geqslant 1$, using (6.15) we obtain

$$\mathrm{Err}_{2,0}(\psi_{2,2}) \leqslant \beta_1 \, \mathrm{Err}_{1,T_0}(\psi_{2,1}) + c_1,$$

where

$$\beta_1 = e^{-\lambda_{2,1}(T-1)}\Big(\Big|\int_{\mathbb{T}^d} r_1\psi_{2,1}\pi_1 \, dx\Big| \cdot \|\psi_{2,2}\|_{L^\infty} + \Big|\int_{\mathbb{T}^d} \psi_{2,2}\psi_{2,1}\pi_2 \, dx\Big|\Big) \overset{(3.2),(7.6)}{\leqslant} e^{-\lambda_{2,1}(T-1)} C_\psi(C_r+1). \tag{7.41}$$

and

$$c_1 = 3\|\psi_{2,2}\|_{L^\infty}\|r_1\|_{\mathrm{osc}}\Big(\frac{1}{2\sqrt{N}} + e^{-\Lambda(T-1)}C_q^{\frac{1}{2}}\Big) + \frac{1}{\sqrt{N}}\|\psi_{2,2}\|_{L^\infty} \overset{(3.2),(7.6)}{\leqslant} C_\psi(1+\frac{3}{2}C_r)\frac{1}{\sqrt{N}} + 3C_\psi C_r C_q^{\frac{1}{2}} e^{-\Lambda(T-1)}. \tag{7.42}$$

Here $C_q = C_q(U,1)$ is from Lemma 4.8 and C_ψ, C_r are the constants defined in (7.6) and (3.2) respectively.

Therefore, for a given δ, we take

$$N \geqslant \tilde{C_N}\frac{1}{\delta^2}, \tag{7.43}$$

$$T \geqslant 1 + \max\Big\{\frac{1}{\Lambda}\Big(\log(\frac{1}{\delta}) + \log(12C_\psi C_r C_q^{\frac{1}{2}})\Big), \frac{1}{\lambda_{2,1}}\Big(\log(\frac{1}{\delta}) + \log(4C_\psi^2(C_r+1))\Big)\Big\}. \tag{7.44}$$

Notice that (7.43) and (7.44) imply that there exists constant $C_1 = C_1(U)$ such that

$$N \geqslant \tilde{C_N}\frac{1}{\delta^2}, \quad T \geqslant C_1\Big(\log(\frac{1}{\delta}) + 1\Big).$$

It remains to check (4.17). Using the fact that

$$\mathrm{Err}_{1,T_0}(\psi_{2,1}) \leqslant \|\psi_{2,1}\|_{L^\infty} \overset{(7.6)}{\leqslant} C_\psi,$$

and (7.41), (7.42), (7.43) and (7.44) we obtain

$$\begin{aligned}\mathrm{Err}_{2,0}(\psi_{2,2}) &\leqslant \beta_1 \, \mathrm{Err}_{1,T_0}(\psi_{2,1}) + c_1 \\ &\leqslant e^{-\lambda_{2,1}(T-1)}C_\psi^2(C_r+1) + C_\psi(1+\frac{3}{2}C_r)\frac{1}{\sqrt{N}} + 3C_\psi C_r C_q^{\frac{1}{2}} e^{-\Lambda(T-1)} \\ &< \frac{\delta}{4} + \frac{\delta}{2} + \frac{\delta}{4} = \delta.\end{aligned}$$

□

8. Proof of Lemmas from Section 7.

In this section, we prove Lemmas 7.2, 7.5, 7.6, 7.7 and 7.8 that were used in Section 7 to prove Lemma 7.9.

8.1. **Estimation of integrals in β_k (Lemma 7.5 and 7.6).** We first prove Lemmas 7.5 and 7.6. Using Proposition 7.4, the proofs of both these lemmas is a direct calculation.

Proof of Lemma 7.5. Step 1: Using the Cauchy-Schwarz inequality and recalling r_ε is defined by (7.17) we obtain

$$\begin{aligned}\Big|\int_{\mathbb{T}^d}\psi_{2,\varepsilon}\psi_{2,\varepsilon'}r_\varepsilon\pi_\varepsilon\,dx\Big| &= \Big|\int_{\mathbb{T}^d}\psi_{2,\varepsilon}\psi_{2,\varepsilon'}\pi_{\varepsilon'}\,dx\Big| \\ &\leqslant \|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})}\|\psi_{2,\varepsilon'}\|_{L^2(\pi_{\varepsilon'})} = \|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})}.\end{aligned}\tag{8.1}$$

It remains to compute $\|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})}$. Clearly,

$$\|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})} = \Big(\int_{\mathbb{T}^d}(\psi_{2,\varepsilon})^2 r_\varepsilon\pi_\varepsilon\,dx\Big)^{\frac{1}{2}} \leqslant \|r_\varepsilon\|^{\frac{1}{2}}_{L^\infty(\pi_\varepsilon)}.\tag{8.2}$$

To prove (7.18), we need a better bound when ε is small.

Step 2: We decompose $\psi_{2,\varepsilon}$ into the sum of the projection into E_ε and $E_\varepsilon^\perp$, where E_ε is defined in (7.15). Explicitly,

$$\psi_{2,\varepsilon}(x) = a_{1,\varepsilon}\mathbf{1}_{\Omega_1} + a_{2,\varepsilon}\mathbf{1}_{\Omega_2} + \upsilon_\varepsilon(x),$$

where $\upsilon_\varepsilon \in E_\varepsilon^\perp$. Thus we can bound $\|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})}$ by

$$\|\psi_{2,\varepsilon}\|_{L^2(\pi_{\varepsilon'})} \leqslant \big\|a_{1,\varepsilon}\mathbf{1}_{\Omega_1} + a_{2,\varepsilon}\mathbf{1}_{\Omega_2}\big\|_{L^2(\pi_{\varepsilon'})} + \|\upsilon_\varepsilon\|_{L^2(\pi_{\varepsilon'})}.\tag{8.3}$$

Step 2.1: Solve $a_{1,\varepsilon}$ and $a_{2,\varepsilon}$. Since $\int_{\mathbb{T}^d}\psi_{2,\varepsilon}(x)\pi_\varepsilon(x)\,dx = 0$, we have

$$a_{1,\varepsilon}\pi_\varepsilon(\Omega_1) + a_{2,\varepsilon}\pi_\varepsilon(\Omega_2) = 0.\tag{8.4}$$

Moreover, since υ_ε is orthogonal to $\mathbf{1}_{\Omega_1}$ and $\mathbf{1}_{\Omega_1}$ in $L^2(\pi_\varepsilon)$, we define

$$b_\varepsilon \stackrel{\text{def}}{=} \sqrt{(a_{1,\varepsilon})^2\pi_\varepsilon(\Omega_1) + (a_{2,\varepsilon})^2\pi_\varepsilon(\Omega_2)} = \sqrt{1 - \|\upsilon_\varepsilon\|^2_{L^2(\pi_\varepsilon)}}.\tag{8.5}$$

Then we solve $a_{1,\varepsilon}$ and $a_{2,\varepsilon}$ using (8.5) and (8.4), whose solutions are

$$\begin{cases} a_{1,\varepsilon} = b_\varepsilon\dfrac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}} \\[2ex] a_{2,\varepsilon} = -b_\varepsilon\dfrac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\end{cases} \quad\text{or}\quad \begin{cases} a_{1,\varepsilon} = -b_\varepsilon\dfrac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}} \\[2ex] a_{2,\varepsilon} = b_\varepsilon\dfrac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}.\end{cases}\tag{8.6}$$

Step 2.2: Now we compute

$$\begin{aligned}\big\|a_{1,\varepsilon}\mathbf{1}_{\Omega_1} + a_{2,\varepsilon}\mathbf{1}_{\Omega_2}\big\|^2_{L^2(\pi_{\varepsilon'})} &= b_\varepsilon^2\Big(\frac{\pi_\varepsilon(\Omega_2)}{\pi_\varepsilon(\Omega_1)}\pi_{\varepsilon'}(\Omega_1) + \frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)}\pi_{\varepsilon'}(\Omega_2)\Big) \\ &= b_\varepsilon^2\Big(1 + (\pi_\varepsilon(\Omega_2) - \pi_\varepsilon(\Omega_1))(\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1))\Big) \\ &\overset{(8.5)}{\leqslant} 1 + (\pi_\varepsilon(\Omega_2) - \pi_\varepsilon(\Omega_1))(\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)).\end{aligned}\tag{8.7}$$

Here the second equality is obtained by substituting

$$\pi_{\varepsilon'}(\Omega_i) = \pi_\varepsilon(\Omega_i) + (\pi_{\varepsilon'}(\Omega_i) - \pi_\varepsilon(\Omega_i)),\quad i = 1, 2,$$

and using the identity $\sum_{i=1}^2(\pi_{\varepsilon'}(\Omega_i) - \pi_\varepsilon(\Omega_i)) = 0$.

On the other hand, using the fact that

$$\|\upsilon_\varepsilon\|_{L^2(\pi_\varepsilon)} = \|P^\perp_{E_\varepsilon}(\psi_{2,\varepsilon})\|_{L^2(\pi_\varepsilon)} = \|(I - P_{E_\varepsilon})(\psi_{2,\varepsilon})\|_{L^2(\pi_\varepsilon)}$$

$$= \|(P_{F_\varepsilon} - P_{E_\varepsilon} P_{F_\varepsilon})(\psi_{2,\varepsilon})\|_{L^2(\pi_\varepsilon)} \leqslant \|P_{F_\varepsilon} - P_{E_\varepsilon} P_{F_\varepsilon}\| = d(E_\varepsilon, F_\varepsilon), \tag{8.8}$$

we obtain

$$\|v_\varepsilon\|_{L^2(\pi_{\varepsilon'})} = \Big(\int_{\mathbb{T}^d} (v_\varepsilon)^2 r_\varepsilon \pi_\varepsilon\Big)^{\frac{1}{2}} \leqslant \|r_\varepsilon\|_{L^\infty}^{\frac{1}{2}} \|v_\varepsilon\|_{L^2(\pi_\varepsilon)} \overset{(8.8)}{\leqslant} \|r_\varepsilon\|_{L^\infty}^{\frac{1}{2}} d(E_\varepsilon, F_\varepsilon). \tag{8.9}$$

Therefore,

$$\begin{aligned}\Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\psi_{2,\varepsilon'} r_\varepsilon \pi_\varepsilon\, dx\Big| &\overset{(8.1),(8.3)}{\leqslant} \big\|a_{1,\varepsilon}\mathbf{1}_{\Omega_1} + a_{2,\varepsilon}\mathbf{1}_{\Omega_2}\big\|_{L^2(\pi_{\varepsilon'})} + \|v_\varepsilon\|_{L^2(\pi_{\varepsilon'})} \\ &\overset{(8.7),(8.9)}{\leqslant} 1 + (\pi_\varepsilon(\Omega_2) - \pi_\varepsilon(\Omega_1))(\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)) + \|r_\varepsilon\|_{L^\infty}^{\frac{1}{2}} d(E_\varepsilon, F_\varepsilon). \end{aligned} \tag{8.10}$$

Combining (8.2) with (8.10) completes the proof of (7.18). ☐

The proof of Lemma 7.6 reuses the arguments in the proof of Lemma 7.5.

Proof of Lemma 7.6. An easy bound can be obatined by directly using Cauchy-Schwarz as follows

$$\Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\pi_{\varepsilon'}\, dx\Big| = \Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon} r_\varepsilon \pi_\varepsilon\, dx\Big| \leqslant \|\psi_{2,\varepsilon}\|_{L^2(\pi_\varepsilon)} \|r_\varepsilon\|_{L^2(\pi_\varepsilon)} \leqslant \|r_\varepsilon\|_{L^\infty}. \tag{8.11}$$

To prove (7.19) we need a more careful bound when ε is small. Using the argument in the proof of Lemma 7.5,

$$\begin{aligned}\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\pi_{\varepsilon'}\, dx &= \int_{\mathbb{T}^d} \psi_{2,\varepsilon}(r_\varepsilon - 1)\pi_\varepsilon\, dx \\ &\overset{(8.3),(8.6)}{=} \int_{\mathbb{T}^d} \Big(b_\varepsilon \frac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}}\mathbf{1}_{\Omega_1} - b_\varepsilon \frac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\mathbf{1}_{\Omega_2} + v_\varepsilon\Big)(r_\varepsilon - 1)\pi_\varepsilon\, dx \\ &= b_\varepsilon\Big(\frac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}} + \frac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\Big) \cdot (\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)) + \int_{\mathbb{T}^d} v_\varepsilon r_\varepsilon \pi_\varepsilon\, dx\end{aligned}$$

where the last equality we use the fact that $v_\varepsilon \in E_\varepsilon^\perp$ and $\sum_{i=1}^2 (\pi_{\varepsilon'}(\Omega_i) - \pi_\varepsilon(\Omega_i)) = 0$.

Therefore,

$$\begin{aligned}\Big|\int_{\mathbb{T}^d} \psi_{2,\varepsilon}\pi_{\varepsilon'}\, dx\Big| \overset{(8.5),(8.8)}{\leqslant} &\Big(\frac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}} + \frac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\Big) \cdot |\pi_{\varepsilon'}(\Omega_1) - \pi_\varepsilon(\Omega_1)| \\ &+ d(E_\varepsilon, F_\varepsilon)\|r_\varepsilon\|_{L^\infty(\pi_\varepsilon)}.\end{aligned} \tag{8.12}$$

Combining (8.11) with (8.12) yields (7.19), as desired. ☐

8.2. **BV bounds on $\pi_\varepsilon(\Omega_i)$ (Lemma 7.7).** In this section we prove Lemma 7.7. As we stated above, we prove Lemma 7.7 by calculating the derivative of $\pi_\varepsilon(\Omega_i)$ with respect to ε. It turns out that the derivative either has a sign or stay bounded as $\varepsilon \to 0$, which implies that $\pi_\varepsilon(\Omega_i)$ is a BV function.

Proof of Lemma 7.7. We only prove (7.20) for $i = 1$. In fact, if (7.20) holds for $i = 1$, then from the identity $\pi_\varepsilon(\Omega_2) = 1 - \pi_\varepsilon(\Omega_1)$, we see that (7.20) holds for $\pi_\varepsilon(\Omega_2)$ for the same constant C_{BV}.

Using the definition of $\pi_\varepsilon(\Omega_1)$, and the fact that $U \in C^6$, we see that $\pi_\varepsilon(\Omega_1)$ is differentiable in ε. We compute

$$\partial_\varepsilon \pi_\varepsilon(\Omega_1) = \partial_\varepsilon \left(\frac{\int_{\Omega_1} \exp(-\frac{U}{\varepsilon})\, dx}{\int_{\Omega_1} \exp(-\frac{U}{\varepsilon})\, dx + \int_{\Omega_2} \exp(-\frac{U}{\varepsilon})\, dx}\right)$$

$$
\begin{aligned}
&= \frac{1}{\varepsilon^2}\frac{\Big(\int_{\Omega_1}\exp(-\frac{U}{\varepsilon})U\,dx\Big)\Big(\int_{\Omega_2}\exp(-\frac{U}{\varepsilon})\,dx\Big)}{\Big(\int_{\Omega_1}\exp(-\frac{U}{\varepsilon})\,dx+\int_{\Omega_2}\exp(-\frac{U}{\varepsilon})\,dx\Big)^2}\\
&\quad -\frac{1}{\varepsilon^2}\frac{\Big(\int_{\Omega_2}\exp(-\frac{U}{\varepsilon})U\,dx\Big)\Big(\int_{\Omega_1}\exp(-\frac{U}{\varepsilon})\,dx\Big)}{\Big(\int_{\Omega_1}\exp(-\frac{U}{\varepsilon})\,dx+\int_{\Omega_2}\exp(-\frac{U}{\varepsilon})\,dx\Big)^2}. \qquad (8.13)
\end{aligned}
$$

We now split the analysis into two cases: $U(x_{\min,1}) = U(x_{\min,2})$ and $U(x_{\min,1}) < U(x_{\min,2})$.

Case I: $U(x_{\min,1}) = U(x_{\min,2})$. We start by estimating the integrals involved in (8.13). According to Proposition B4 and Remark under the proof of Proposition B4, (page 289-290) in [Kol00],

$$
\int_{\Omega_i}\exp\Big(-\frac{U}{\varepsilon}\Big)U\,dx = (2\pi\varepsilon)^{\frac{d}{2}}\frac{\varepsilon}{2}\Big(\mathfrak{c}_i+O(\varepsilon)\Big), \qquad (8.14)
$$

where $\mathfrak{c}_i = \mathfrak{c}_i(U)$ depends on derivatives of U up to order 3 evaluated at $x_{\min,i}$, and is independent of ε. The $O(\varepsilon)$ involves constants that may depend on derivatives of U up to order 6 evaluated at $x_{\min,i}$. On the other hand, Proposition B2 in [Kol00] guarantees

$$
\int_{\Omega_i} e^{-\frac{U}{\varepsilon}}\,dx = (2\pi\varepsilon)^{\frac{d}{2}}\frac{\exp(-\frac{U(x_{\min,i})}{\varepsilon})}{\sqrt{\det\nabla^2U(x_{\min,i})}}(1+O(\varepsilon)), \quad \text{for } i\in 1,2. \qquad (8.15)
$$

Therefore, when ε is sufficiently small, combining (8.14) and (8.15),

$$
\begin{aligned}
\partial_\varepsilon\pi_\varepsilon(\Omega_1) &= \frac{1}{\varepsilon^2}\frac{\Big(\frac{\varepsilon}{2}(\mathfrak{c}_1+O(\varepsilon))(\det(\nabla^2U(x_{\min,2})))^{-\frac12}\Big)}{\Big(\sum_{i=1,2}(\det(\nabla^2U(x_{\min,i})))^{-\frac12}\Big)^2}\\
&\quad -\frac{1}{\varepsilon^2}\frac{\Big(\frac{\varepsilon}{2}(\mathfrak{c}_2+O(\varepsilon))(\det(\nabla^2U(x_{\min,1})))^{-\frac12}\Big)}{\Big(\sum_{i=1,2}(\det(\nabla^2U(x_{\min,i})))^{-\frac12}\Big)^2}.
\end{aligned}
$$

We now discuss two different cases.

Case I.1: $\mathfrak{c}_1\det(\nabla^2U(x_{\min,2})))^{-\frac12}-\mathfrak{c}_2\det(\nabla^2U(x_{\min,1})))^{-\frac12}=0$. In this case,

$$
|\partial_\varepsilon\pi_\varepsilon(\Omega_1)| = O(1), \quad \varepsilon\to 0.
$$

Since the function $\varepsilon\mapsto\partial_\varepsilon\pi_\varepsilon(\Omega_1))$ is a continuous function on $(0,1]$, it must be bounded which implies (7.20).

Case I.2: $\mathfrak{c}_1\det(\nabla^2U(x_{\min,2})))^{-\frac12}-\mathfrak{c}_2\det(\nabla^2U(x_{\min,1})))^{-\frac12}\neq 0$. Without loss of generality we assume $\mathfrak{c}_1\det(\nabla^2U(x_{\min,2})))^{-\frac12}-\mathfrak{c}_2\det(\nabla^2U(x_{\min,1})))^{-\frac12}<0$. In this case there must exist some $\varepsilon_{\rm cr}>0$ such that $\partial_\varepsilon\pi_\varepsilon(\Omega_1))<0$ for all $\varepsilon\in(0,\varepsilon_{\rm cr}]$. Thus, for any $\eta\in(0,\varepsilon_{\rm cr})$,

$$
\int_\eta^{\varepsilon_{\rm cr}}|\partial_\varepsilon\pi_\varepsilon(\Omega_1)|\,d\varepsilon = -\int_\eta^{\varepsilon_{\rm cr}}\partial_\varepsilon\pi_\varepsilon(\Omega_1)\,d\varepsilon = \pi_\eta(\Omega_1)-\pi_{\varepsilon_{\rm cr}}(\Omega_1)\leqslant 1. \qquad (8.16)
$$

For $\varepsilon\in[\varepsilon_{\rm cr},1]$, the function $\varepsilon\mapsto\partial_\varepsilon\pi_\varepsilon(\Omega_1))$ is continuous and hence bounded. This immediately implies (7.20), concluding the proof of Case I.

Case II: $U(x_{\min,1}) < U(x_{\min,2})$. Using (8.14) and (8.15) we see

$$\begin{aligned}\Big(\int_{\Omega_1} e^{-\frac{U}{\varepsilon}} U\,dx\Big)\Big(\int_{\Omega_2} e^{-\frac{U}{\varepsilon}}\,dx\Big) - \Big(\int_{\Omega_2} e^{-\frac{U}{\varepsilon}} U\,dx\Big)\Big(\int_{\Omega_1} e^{-\frac{U}{\varepsilon}}\,dx\Big)&\\ = (2\pi\varepsilon)^d e^{-U(x_{\min,2})/\varepsilon}\Bigg(\frac{O(\varepsilon)}{\sqrt{\det(\nabla^2 U(x_{\min,2}))}}&\\ - \frac{(U(x_{\min,2}) + O(\varepsilon))}{\sqrt{\det(\nabla^2 U(x_{\min,1}))}}(1+O(\varepsilon))\Bigg),&\end{aligned}$$

which is negative when ε is small. Using this in (8.13) implies implies $\partial_\varepsilon \pi_\varepsilon(\Omega_1) < 0$. Using (8.16) and the same argument as in Case I.2 finishes the proof. □

8.3. **Lower bound of the second eigenvalue (Lemma 7.8).** In this section we prove Lemma 7.8. We begin by introducing the notion of Poincare constants. Let $\mathcal{X}$ be an Euclidean space, and μ be a probability measure on $\mathcal{X}$. We say μ satisfies $\mathrm{PI}(\varrho)$ if it satisfies the *Poincaré inequality* with constant ϱ. That is, for all test functions $f \in H^1(\mu)$ we have

$$\mathrm{Var}_\mu(f) \leqslant \frac{1}{\varrho}\int_{\mathcal{X}} |\nabla f|^2\,d\mu.$$

Here $\mathrm{Var}_\mu(f)$ is the variance of f with respect to the measure μ and is defined by

$$\mathrm{Var}_\mu(f) \overset{\text{def}}{=} \int_{\mathcal{X}} \Big(f - \int_{\mathcal{X}} f\,d\mu\Big)^2\,d\mu.$$

Corollary 2.15 from Menz and Schlichting [MS14] provides bounds on the Poincaré constant for the Gibbs measure in our setting.

Proposition 8.1 (Corollary 2.15 in [MS14])**.** *If U satisfies Assumption 4.1 and 4.2 then π_ε satisfies $\mathrm{PI}(\varrho_\varepsilon)$ with*

$$\frac{1}{\varrho_\varepsilon} \lesssim \frac{\pi_\varepsilon(\Omega_1)\pi_\varepsilon(\Omega_2)}{(2\pi\varepsilon)^{\frac{d}{2}-1}}\,\frac{\sqrt{|\det(\nabla^2(U(s_{1,2})))|}}{|\lambda^-(s_{1,2})|}\exp\Big(\frac{U(s_{1,2})}{\varepsilon}\Big)\int_{\mathcal{X}} e^{-U/\varepsilon}\,dx,$$

where $\lambda^-(s_{1,2})$ denotes the negative eigenvalue of the Hessian $\nabla^2(U(s_{1,2}))$ at the communicating saddle $s_{1,2}$.

We note that in [MS14] their domain is the whole space $\mathbb{R}^d$. The proof can easily be modified to work in the setting of the compact torus. Proposition 8.1 immediately implies Lemma 7.8, as we now show.

Proof of Lemma 7.8. Since

$$\int_{\mathcal{X}} |\nabla f|^2 \pi_\varepsilon\,dx = \int_{\mathcal{X}} f L_\varepsilon f\,\pi_\varepsilon\,dx$$

we immediately see $\lambda_{2,\varepsilon} \geqslant \varrho_\varepsilon$. Thus, Proposition 8.1 implies

$$\limsup_{\varepsilon\to 0} -(\varepsilon\log(\lambda_{2,\varepsilon})) \leqslant \limsup_{\varepsilon\to 0} -(\varepsilon\log(\varrho_\varepsilon)) \leqslant \hat{U}.$$

Thus, for every $H > \hat{U} \geqslant \limsup_{\varepsilon\to 0} -(\varepsilon\log(\lambda_{2,\varepsilon}))$, there exists ε_H such that $\lambda_{2,\varepsilon} \geqslant \exp(-\frac{H}{\varepsilon})$ for every $\varepsilon < \varepsilon_H$. Choosing

$$A_H \overset{\text{def}}{=} \min\{\inf_{\varepsilon_H \leqslant \varepsilon < 1}\Big(\lambda_{2,\varepsilon}\exp(\frac{H}{\varepsilon})\Big), 1\}.$$

immediately implies (7.21) as desired. □

8.4. **Uniform boundedness of eigenfunctions (Lemma 7.2).** In this section we prove Lemma 7.2. In the proof, the constant $C = C(U, d, C_m)$ may change from line to line. The main tools are local and global maximum principles. In particular, we first find proper compact neighborhoods of the local minima and use local maximum principle to show that $\psi_{2,\varepsilon}$ is uniformly bounded in ε in these neighborhoods. Then we apply global maximum principle to show the uniform boundedness outside these regions.

We start by a description of those compact neighborhoods. Define the $R_i > 0$, $i \in \{1, 2\}$ as

$$R_i \stackrel{\text{def}}{=} \sup\Big\{r \;\Big|\; B(x_{\min,i}, r) \subseteq \Omega_i, \sup_{x \in B(x_{\min,i},r)} U(x) - U(x_{\min,i}) \leqslant \frac{\hat{\gamma}}{8}\Big\} \tag{8.17}$$

and then define

$$\widetilde{B}_i = B(x_{\min,i}, R_i), \quad B_i = B\Big(x_{\min,i}, \frac{3R_i}{4}\Big).$$

We will show $\psi_{2,\varepsilon}$ is uniformly bounded in ε both on $B_1 \cup B_2$ and $\mathbb{T}^d \setminus (B_1 \cup B_2)$. We first bound $\psi_{2,\varepsilon}$ in the regions B_1 and B_2.

Lemma 8.2. *There exists a constant $C_a = C_a(d, U, C_m)$ and $\tilde{\varepsilon} = \tilde{\varepsilon}(d, U)$ such that for every*

$$0 < \varepsilon \leqslant \min\Big\{\frac{1}{12}\min\{R_1, R_2\}, \tilde{\varepsilon}\Big\} \tag{8.18}$$

we have

$$\forall x \in B_i, \quad |\psi_{2,\varepsilon}(x) - a_{i,\varepsilon}| \leqslant C_a \exp\Big(-\frac{3\hat{\gamma}}{4\varepsilon}\Big), \quad i = 1, 2. \tag{8.19}$$

Here $a_{1,\varepsilon}$ and $a_{2,\varepsilon}$ are defined as in (8.6).

Proof of Lemma 8.2. Fix $i = 1$ or 2, for each $x \in B_i$, there exists $y \in B_i$ such that $x \in B(y, \varepsilon)$. By the triangle inequality, it follows that $B(y, 2\varepsilon) \subseteq \widetilde{B}_i$. Thus,

$$B(y, 2\varepsilon) \subseteq \widetilde{B}_i \subset \Omega_i. \tag{8.20}$$

First notice that the function $\psi_{2,\varepsilon} - a_{i,\varepsilon}$ satisfies

$$(L_\varepsilon - \lambda_{2,\varepsilon})(\psi_{2,\varepsilon} - a_{i,\varepsilon}) = \lambda_{2,\varepsilon} a_{i,\varepsilon}.$$

Thus using [GT01, Corollary 9.21], there exists dimensional constant C such that for every $y \in B_i$ for which $B(y, 2\varepsilon) \subseteq \widetilde{B}_i$, we have

$$\begin{aligned}\sup_{x \in B(y,\varepsilon)} |\psi_{2,\varepsilon}(x) - a_{i,\varepsilon}| \leqslant C\bigg(\Big(\frac{1}{|B(y, 2\varepsilon)|}\int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon}(x) - a_{i,\varepsilon}|^2\, dx\Big)^{\frac{1}{2}} \\ + |\lambda_{2,\varepsilon} a_{i,\varepsilon}|\bigg).\end{aligned} \tag{8.21}$$

Now we bound $\int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon}(x) - a_{i,\varepsilon}|^2\, dx$ that appears on the right hand side of (8.21). Using the fact that when (8.18) holds, for $i = 1, 2$,

$$\begin{aligned}\int_{\mathbb{T}^d} e^{-\frac{U}{\varepsilon}}\, dx &= \Big(\int_{\Omega_i} e^{-\frac{U}{\varepsilon}}\, dx\Big) \cdot \Big(1 + \frac{\int_{\mathbb{T}^d \setminus \Omega_i} e^{-\frac{U}{\varepsilon}}\, dx}{\int_{\Omega_i} e^{-\frac{U}{\varepsilon}}\, dx}\Big) \\ &= \Big(\int_{\Omega_i} e^{-\frac{U}{\varepsilon}}\, dx\Big) \cdot \Big(1 + \frac{1 - \pi_\varepsilon(\Omega_i)}{\pi_\varepsilon(\Omega_i)}\Big) \stackrel{(4.4)}{\leqslant} \Big(\int_{\Omega_i} e^{-\frac{U}{\varepsilon}}\, dx\Big) \cdot (1 + C_m^2)\end{aligned}$$

$$\stackrel{(8.15)}{\leqslant} C(2\pi\varepsilon)^{\frac{d}{2}} e^{-\frac{U(x_{\min,i})}{\varepsilon}}, \tag{8.22}$$

we have that

$$\begin{aligned}
\int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon} - a_{i,\varepsilon}|^2 \, dx &= \int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon} - a_{i,\varepsilon} \mathbf{1}_{\Omega_i}|^2 \, dx \\
&\leqslant \Big(\sup_{z \in B(y,2\varepsilon)} e^{\frac{U(z)}{\varepsilon}} \Big) \int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon} - a_{i,\varepsilon} \mathbf{1}_{\Omega_i}|^2 e^{-\frac{U}{\varepsilon}} \, dx \\
&\stackrel{(8.20)}{\leqslant} \Big(\int_{\mathbb{T}^d} e^{-\frac{U}{\varepsilon}} \, dx \Big) \Big(\sup_{z \in \widetilde{B}_i} e^{\frac{U(z)}{\varepsilon}} \Big) \int_{\Omega_i} |\psi_{2,\varepsilon}(x) - a_{i,\varepsilon} \mathbf{1}_{\Omega_i}|^2 \, d\pi_\varepsilon(x) \\
&\stackrel{(8.22)}{\leqslant} C(2\pi\varepsilon)^{\frac{d}{2}} \Big(\sup_{z \in \widetilde{B}_i} e^{\frac{U(z)-U(x_{\min,i})}{\varepsilon}} \Big) \big\| \psi_{2,\varepsilon} - a_{1,\varepsilon} \mathbf{1}_{\Omega_1} - a_{2,\varepsilon} \mathbf{1}_{\Omega_2} \big\|_{L^2(\pi_\varepsilon)}^2 \\
&\stackrel{(8.8)}{\leqslant} C(2\pi\varepsilon)^{\frac{d}{2}} \Big(\sup_{z \in \widetilde{B}_i} e^{\frac{U(z)-U(x_{\min,i})}{\varepsilon}} \Big) d(E_\varepsilon, F_\varepsilon)^2 \\
&\stackrel{(7.16)}{\leqslant} C(2\pi\varepsilon)^{\frac{d}{2}} \Big(\sup_{z \in \widetilde{B}_i} e^{\frac{U(z)-U(x_{\min,i})}{\varepsilon}} \Big) \exp\Big(-\frac{7\hat{\gamma}}{4\varepsilon} \Big) \\
&\stackrel{(8.17)}{\leqslant} C(2\pi\varepsilon)^{\frac{d}{2}} \exp\Big(-\frac{13\hat{\gamma}}{8\varepsilon} \Big), \qquad (8.23)
\end{aligned}$$

where the second last inequality we use (7.16) with $\gamma = \frac{7}{8}\hat{\gamma}$.

Notice that there exists constant $\tilde{\varepsilon} = \tilde{\varepsilon}(d, \hat{\gamma})$ that whenever $\varepsilon < \tilde{\varepsilon}$,

$$\exp\Big(-\frac{\hat{\gamma}}{8\varepsilon} \Big) < (2\pi\varepsilon)^{\frac{d}{2}}. \tag{8.24}$$

Thus, for $\varepsilon < \tilde{\varepsilon}$,

$$\int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon} - a_{i,\varepsilon}|^2 \, dx \stackrel{(8.23),(8.24)}{\leqslant} C(2\pi\varepsilon)^d \exp\Big(-\frac{3\hat{\gamma}}{2\varepsilon} \Big). \tag{8.25}$$

Therefore, plugging (8.25) into (8.21) gives

$$\begin{aligned}
\sup_{x \in B(y,\varepsilon)} |\psi_{2,\varepsilon}(x) - a_{1,\varepsilon}| &\stackrel{(8.25)}{\leqslant} \Big(\frac{C(2\pi\varepsilon)^d}{|B(y,2\varepsilon)|} \exp\Big(-\frac{3\hat{\gamma}}{2\varepsilon} \Big) \Big)^{\frac{1}{2}} + C|\lambda_{2,\varepsilon} a_{1,\varepsilon}| \\
&\stackrel{(4.10),(8.6)}{\leqslant} \Big(\frac{C(2\pi\varepsilon)^d}{(2\varepsilon)^d} \exp\Big(-\frac{3\hat{\gamma}}{2\varepsilon} \Big) \Big)^{\frac{1}{2}} + C C_m \exp\Big(-\frac{7\hat{\gamma}}{8\varepsilon} \Big) \\
&\leqslant C_a \exp\Big(-\frac{3\hat{\gamma}}{4\varepsilon} \Big),
\end{aligned}$$

which implies (8.19). □

We will now bound $\psi_{2,\varepsilon}$ on $\mathbb{T}^d \setminus (B_1 \cup B_2)$ by first bounding the L_ε-harmonic extensions of $\psi_{2,\varepsilon}$ and then bounding the eigenfunction of L_ε in $\mathbb{T}^d \setminus (B_1 \cup B_2)$ with inhomogeneous Dirichlet boundary conditions specified by $\psi_{2,\varepsilon}$. For simplicity of notation, define

$$\tilde{\Omega} \stackrel{\text{def}}{=} \mathbb{T}^d \setminus (B_1 \cup B_2).$$

Lemma 8.2 can be used to immediately bound the L_ε-harmonic extensions of $\psi_{2,\varepsilon}$.

Lemma 8.3. *For $i = 1, 2$, let $f_{0,\varepsilon}^{(i)}$ be the solution to*

$$\begin{aligned} L_\varepsilon f_{0,\varepsilon}^{(i)}(y) &= 0, && y \in \tilde{\Omega} \\ f_{0,\varepsilon}^{(i)}(y) &= \psi_{2,\varepsilon}(y), && y \in \partial B_i \\ f_{0,\varepsilon}^{(i)}(y) &= 0, && y \in (\partial B_1 \cup \partial B_2) \setminus \partial B_i. \end{aligned}$$

For every ε satisfying (8.18) and every $y \in \tilde{\Omega}$ we have

$$|f_{0,\varepsilon}^{(i)}(y)| \leqslant |a_{i,\varepsilon}| + C_a \exp\Big(-\frac{3\hat{\gamma}}{4\varepsilon}\Big). \tag{8.26}$$

Proof. Observe that $f_{0,\varepsilon}^{(i)}$ satisfies $L_\varepsilon f_{0,\varepsilon}^{(i)}(y) = 0$ on $\tilde{\Omega}$. Thus, by weak maximum principle [Eva10, Section 6.4.1, Theorem 1],

$$\sup_{\tilde{\Omega}} |f_{0,\varepsilon}^{(i)}| = \sup_{\partial B_1 \cup \partial B_2} |f_{0,\varepsilon}^{(i)}| = \sup_{\partial B_1 \cup \partial B_2} |\psi_{2,\varepsilon}| \overset{(8.19)}{\leqslant} |a_{i,\varepsilon}| + C_a \exp\Big(-\frac{3\hat{\gamma}}{4\varepsilon}\Big),$$

which implies the inequality (8.26). □

We now bound the eigenfunction of L_ε in $\tilde{\Omega}$ with inhomogeneous Dirichlet boundary conditions specified by $\psi_{2,\varepsilon}$.

Lemma 8.4. *For $i = 1, 2$, let f_λ solve*

$$\begin{aligned} (L_\varepsilon - \lambda_{2,\varepsilon}) f_\lambda(y) &= 0, && y \in \tilde{\Omega} \\ f_\lambda(y) &= \psi_{2,\varepsilon}, && y \in \partial B_i \\ f_\lambda(y) &= 0, && y \in (\partial B_1 \cup \partial B_2) \setminus \partial B_i. \end{aligned}$$

There exist $\varepsilon_0 > 0$, $T_0' > 0$ such that for

$$\varepsilon \leqslant \min\Big\{\varepsilon_0, \frac{7\hat{\gamma}}{8\log(2C_\gamma T_0')}, \frac{\min\{R_1, R_2\}}{12}, \tilde{\varepsilon}\Big\} \overset{\text{def}}{=} \varepsilon_1, \tag{8.27}$$

we have

$$|f_\lambda(y)| \leqslant 2\Big(|a_{i,\varepsilon}| + C_a \exp\Big(-\frac{3\hat{\gamma}}{4\varepsilon}\Big)\Big), \quad \forall y \in \tilde{\Omega}. \tag{8.28}$$

Here C_a is the constant in (8.19) and C_γ is a constant such that (4.10) holds with $\gamma = 7\hat{\gamma}/8$.

Proof of Lemma 8.4. We only prove when $i = 1$. The proof in the case $i = 2$ is identical. Let f_0 solve

$$\begin{aligned} L_\varepsilon f_0(y) &= 0, && y \in \tilde{\Omega} \\ f_0(y) &= \psi_{2,\varepsilon}, && y \in \partial B_1 \\ f_0(y) &= 0, && y \in \partial B_2. \end{aligned}$$

Let $\delta f_\lambda = f_\lambda - f_0$. Then δf_λ satisfies

$$\begin{aligned} L_\varepsilon \delta f_\lambda(y) &= \lambda_{2,\varepsilon} f_\lambda(y), && y \in \tilde{\Omega} \\ \delta f_\lambda(y) &= 0, && y \in \partial B_1 \cup \partial B_2. \end{aligned}$$

Let τ be the first exit time of X^ε from $B_1 \cup B_2$. We know $g(y) \overset{\text{def}}{=} \boldsymbol{E}^y \tau$ solves the Poisson equation

$$L_\varepsilon g_\varepsilon = 1, \qquad y \in \tilde{\Omega}$$

$$g_\varepsilon = 0, \qquad y \in \partial B_1 \cup \partial B_2.$$

Thus if $M' \stackrel{\text{def}}{=} \sup_{z\in\tilde\Omega} |f_\lambda(z)| < \infty$, the comparison principle immediately implies

$$\sup_{\tilde\Omega} |\delta f_\lambda| \leqslant \lambda_{2,\varepsilon} M' \sup_{\tilde\Omega} g_\varepsilon.$$

Since $f_\lambda = \delta f_\lambda + f_0$, we see

$$M' \leqslant \|f_0\|_{L^\infty(\tilde\Omega)} + \lambda_{2,\varepsilon} M' \|g_\varepsilon\|_{L^\infty(\tilde\Omega)}. \tag{8.29}$$

According to [FW12, Corollary of Lemma 1.9, Chapter 6], for ε smaller than some ε_0, there exist constant T_0 and c such that

$$\sup_{y\in\tilde\Omega} \boldsymbol{E}^y \tau \leqslant T_0 + \frac{\varepsilon^2}{c} < T_0 + \frac{\varepsilon_0^2}{c} \stackrel{\text{def}}{=} T_0'. \tag{8.30}$$

Notice that the choice (8.27) ensures that (8.18) and

$$\varepsilon \leqslant \varepsilon_0, \quad \lambda_{2,\varepsilon} T_0' \leqslant \frac{1}{2}, \tag{8.31}$$

which implies that

$$\begin{aligned} M &\overset{(8.29)}{\leqslant} \frac{\|f_0\|_{L^\infty(\tilde\Omega)}}{1 - \lambda_{2,\varepsilon}\|g_\varepsilon\|_{L^\infty(\tilde\Omega)}} \overset{(8.30),(8.31)}{\leqslant} \frac{\|f_0\|_{L^\infty(\tilde\Omega)}}{1-\lambda_{2,\varepsilon}T_0'} \\ &\overset{(8.26),(8.31)}{\leqslant} 2\Big(|a_{i,\varepsilon}| + C_a \exp\Big(-\frac{3\hat\gamma}{4\varepsilon}\Big)\Big). \end{aligned}$$

□

Proof of Lemma 7.2. We discuss two cases, $\varepsilon \leqslant \varepsilon_1$ and $\varepsilon > \varepsilon_1$, where ε_1 is defined in (8.18).

Case I: $\varepsilon \leqslant \varepsilon_1$. For $y \in B_1 \cup B_2$, we apply Lemma 8.2, to obtain

$$\sup_{y\in B_1\cup B_2} |\psi_{2,\varepsilon}(y)| \overset{(8.25)}{\leqslant} \max\{|a_{1,\varepsilon}|, |a_{2,\varepsilon}|\} + C_a \exp\Big(-\frac{3\hat\gamma}{4\varepsilon}\Big) \leqslant C_m + C_a. \tag{8.32}$$

To obtain the last inequality above we used the fact that

$$\max\{|a_{1,\varepsilon}|, |a_{2,\varepsilon}|\} \overset{(8.5),(8.6)}{\leqslant} \max\Big\{\frac{\sqrt{\pi_\varepsilon(\Omega_2)}}{\sqrt{\pi_\varepsilon(\Omega_1)}}, \frac{\sqrt{\pi_\varepsilon(\Omega_1)}}{\sqrt{\pi_\varepsilon(\Omega_2)}}\Big\} \overset{(4.4)}{\leqslant} C_m. \tag{8.33}$$

For $y \in \tilde\Omega$, we write

$$\psi_{2,\varepsilon} = \psi_{2,\varepsilon}^{(1)} + \psi_{2,\varepsilon}^{(2)},$$

where $\psi_{2,\varepsilon}^{(i)}$ solves that

$$\begin{aligned} (L_\varepsilon - \lambda_{2,\varepsilon})\psi_{2,\varepsilon}^{(i)}(y) &= 0, && y \in \tilde\Omega \\ \psi_{2,\varepsilon}^{(i)}(y) &= \psi_{2,\varepsilon}, && y \in \partial B_i \\ \psi_{2,\varepsilon}^{(i)}(y) &= 0, && y \in (\partial B_1 \cup \partial B_2) \setminus \partial B_i. \end{aligned}$$

Applying Lemma 8.4 to $\psi_{2,\varepsilon}^{(i)}$ gives

$$\sup_{y\in\tilde\Omega} |\psi_{2,\varepsilon}(y)| \overset{(8.28)}{\leqslant} 2\max\{|a_{1,\varepsilon}|, |a_{2,\varepsilon}|\} + C_a \exp\Big(-\frac{3\hat\gamma}{4\varepsilon}\Big) \overset{(8.33)}{\leqslant} 2C_m + C_a. \tag{8.34}$$

Combining (8.34) and (8.32), we obtain that for $0 < \varepsilon \leqslant \varepsilon_1$, there exists $C = C(U, d, C_m)$ independent of ε such that $\|\psi_{2,\varepsilon}\|_{L^\infty(\mathbb{T}^d)} \leqslant C$.

Case II: $\varepsilon > \varepsilon_1$. According to [GT01, Corollary 9.21], for $y \in \mathbb{T}^d$,

$$\begin{aligned}\sup_{x\in B(y,\varepsilon)} |\psi_{2,\varepsilon}(x)| &\leqslant \Big(\frac{C}{|B(y,2\varepsilon)|}\int_{B(y,2\varepsilon)} |\psi_{2,\varepsilon}(x)|^2\,dx\Big)^{\frac12}\\ &\leqslant \Big(\frac{C}{|B(y,2\varepsilon)|}\Big(\sup_{z\in\mathbb{T}^d} e^{\frac{U(z)-U_{\min}}{\varepsilon}}\Big)\int_{\mathbb{T}^d} |\psi_{2,\varepsilon}(x)|^2\,d\pi_\varepsilon(x)\Big)^{\frac12}\\ &= C(\varepsilon_1)^{-\frac d2}\exp\Big(\frac{\|U\|_{\mathrm{osc}}}{2\varepsilon_1}\Big) = C(U,d,C_m).\end{aligned}$$

We conclude from the above two cases that (7.6) holds. ☐

9. Energy valley estimates

9.1. **The Mass Ratio (Lemma 4.4).** In this section we prove Lemma 4.4, whose main idea is that when $\varepsilon \to 0$, the value of integral $\int_{\Omega_i} \exp(-U/\varepsilon)\,dx$ is mainly determined by landscape near the local minima.

Proof of Lemma 4.4. We will prove that there exists $C > 0$ independent of $\eta_{\min}$ such that

$$\sup_{\varepsilon\in[\eta_{\min},\eta_{\max}]} \frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)} \leqslant C, \qquad \sup_{\varepsilon\in[\eta_{\min},\eta_{\max}]} \frac{\pi_\varepsilon(\Omega_2)}{\pi_\varepsilon(\Omega_1)} \leqslant C. \tag{9.1}$$

Combing this with the fact that

$$\sup_{\varepsilon\in[\eta_{\min},\eta_{\max}]} \frac{1}{\pi_\varepsilon(\Omega_i)} \leqslant \sup_{\varepsilon\in[\eta_{\min},\eta_{\max}]} \frac{\pi_\varepsilon(\Omega_1)+\pi_\varepsilon(\Omega_2)}{\pi_\varepsilon(\Omega_i)} \overset{(4.4)}{\leqslant} 1 + C \overset{\text{def}}{=} C_m^2,$$

for $i \in \{1,2\}$, we obtain (4.4) as desired.

To prove (9.1), we note that Assumption 4.1 implies, via Laplace method as in (8.15), that there exists $\varepsilon_2 > 0$ such that for all $\varepsilon < \varepsilon_2$ we have

$$\frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)} \overset{(8.15)}{=} \frac{(\det(\nabla^2 U(x_{\min,1})))^{-\frac12}}{(\det(\nabla^2 U(x_{\min,2})))^{-\frac12}\exp\big(\frac{|U(x_{\min,1})-U(x_{\min,2})|}{\varepsilon}\big)} + O(\varepsilon),$$

The assumption (4.5) further shows that for all $\varepsilon \in [\eta_{\min}, \varepsilon_2]$, we have

$$\frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)} \leqslant \frac{(\det(\nabla^2 U(x_{\min,1})))^{-\frac12}}{(\det(\nabla^2 U(x_{\min,2})))^{-\frac12}} + O(\varepsilon),$$

$$\text{and}\qquad \frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)} \geqslant \frac{(\det(\nabla^2 U(x_{\min,1})))^{-\frac12}}{(\det(\nabla^2 U(x_{\min,2})))^{-\frac12}\exp(C_l)} + O(\varepsilon).$$

Thus, making ε_2 smaller if necessary, for every $\varepsilon \in [\eta_{\min}, \varepsilon_2]$, we have

$$\frac12 \frac{(\det(\nabla^2 U(x_{\min,1})))^{-\frac12}}{(\det(\nabla^2 U(x_{\min,2})))^{-\frac12}\exp(C_l)} \leqslant \frac{\pi_\varepsilon(\Omega_1)}{\pi_\varepsilon(\Omega_2)} \leqslant \frac32 \frac{(\det(\nabla^2 U(x_{\min,1})))^{-\frac12}}{(\det(\nabla^2 U(x_{\min,2})))^{-\frac12}}.$$

Since $\varepsilon \mapsto \pi_\varepsilon(\Omega_1)/\pi_\varepsilon(\Omega_2)$ is continuous and positive on the interval $[\varepsilon_2, \eta_{\max}]$, we obtain (9.1) as desired. ☐

Lemmas 4.4 and 7.7 immediately shows the following corollary, showing the finiteness condition (2.3) holds provided Assumptions 4.1, 4.2 hold and the wells have nearly equal depth.

Corollary 9.1. *Assume the function U satisfies Assumption 4.1, Assumption 4.2 and there exists $\eta_{\min} \geqslant 0$ and $C_\ell < \infty$ such that (4.5) holds. Then for any finite $\eta_{\max} > \eta_{\min}$ the constant C_{LBV} in (2.3) can be bounded above in terms of U, C_ℓ and $\eta_{\max}$, but independent of $\eta_{\min}$.*

Proof of Corollary 9.1. Notice that the inequality (7.20) applied to $U/\eta_{\max}$ gives that there exists constant $\tilde{C}_{\mathrm{BV}}$ independent of η such that for $i = 1, 2$,

$$\int_\eta^{\eta_{\max}} |\partial_\varepsilon \pi_\varepsilon(\Omega_i)|\, d\varepsilon \leqslant \tilde{C}_{\mathrm{BV}}. \tag{9.2}$$

Therefore,

$$\int_\eta^{\eta_{\max}} |\partial_\varepsilon \ln \pi_\varepsilon(\Omega_i)|\, d\varepsilon = \int_\eta^{\eta_{\max}} \frac{|\partial_\varepsilon \pi_\varepsilon(\Omega_i)|}{\pi_\varepsilon(\Omega_i)}\, d\varepsilon \overset{(4.4),(9.2)}{\leqslant} C_m^2 \tilde{C}_{\mathrm{BV}}.$$

Taking $\eta \to 0$ on the left hand side finishes the proof. ☐

9.2. **Uniform boundedness of r_k.** We recall that C_r defined by (3.2) is the maximum of the ratio of the *normalized* densities. Since this may be hard to estimate in practice, we now obtain a bound for C_r in a manner that may be easier to use in practice.

Lemma 9.2. *Suppose M is chosen by (2.4), and choose $\eta_2, \dots, \eta_M$ so that $\eta_M = \eta$ and $1/\eta_1, \dots, 1/\eta_M$ are linearly spaced. If $C_r = C_r(U/\eta_1, \nu)$ is defined by (3.2), then C_r satisfies (3.3).*

Proof of Lemma 9.2. Without loss of generality, we take $U = U_0$, where U_0 is defined in (3.4). Then we have $\eta_1 = 1$ and $U \geqslant 0$.

Observe that for every $k = 1, \dots, M-1$, $x \in \mathcal{X}$, since $U \geqslant 0$, we have

$$r_k(x) \overset{(6.23)}{=} \frac{Z_k}{Z_{k+1}} \exp\Big(-\Big(\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k}\Big)U(x)\Big) \leqslant \frac{Z_k}{Z_{k+1}} = \frac{\int_{\mathcal{X}} \exp\big(\frac{-U}{\eta_k}\big)\, dy}{\int_{\mathcal{X}} \exp\big(\frac{-U}{\eta_{k+1}}\big)\, dy}. \tag{9.3}$$

Now we bound the ratio on the right hand side of (9.3). For $c \geqslant 0$, the constant s_c defined in (3.4) now becomes

$$s_c \overset{\text{def}}{=} \frac{\int_{\{U>c\}} e^{-U}\, dx}{\int_{\{U\leqslant c\}} e^{-U}\, dx} < \infty.$$

Then for $\varepsilon < 1$,

$$\frac{\int_{\{U>c\}} e^{-\frac{U}{\varepsilon}}\, dx}{\int_{\{U\leqslant c\}} e^{-\frac{U}{\varepsilon}}\, dx} \leqslant \frac{\exp\big(c\big(1 - \frac{1}{\varepsilon}\big)\big) \int_{\{U>c\}} e^{-U}\, dx}{\exp\big(c\big(1 - \frac{1}{\varepsilon}\big)\big) \int_{\{U\leqslant c\}} e^{-U}\, dx} = \frac{\int_{\{U>c\}} e^{-U}\, dx}{\int_{\{U\leqslant c\}} e^{-U}\, dx} = s_c. \tag{9.4}$$

Therefore, for $\eta_k < \eta_1$,

$$\begin{aligned}
\int_{\mathcal{X}} e^{-\frac{U}{\eta_k}}\, dx &= \int_{\{U\leqslant c\}} e^{-\frac{U}{\eta_k}}\, dx + \int_{\{U>c\}} e^{-\frac{U}{\eta_k}}\, dx \overset{(9.4)}{\leqslant} (1+s_c) \int_{\{U\leqslant c\}} e^{-\frac{U}{\eta_k}}\, dx \\
&= (1+s_c) \int_{\{U\leqslant c\}} e^{-\frac{U}{\eta_{k+1}} + (\frac{U}{\eta_{k+1}} - \frac{U}{\eta_k})}\, dx \\
&\leqslant (1+s_c) \exp\Big(c\Big(\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k}\Big)\Big)\Big(\int_{\{U\leqslant c\}} e^{-\frac{U}{\eta_{k+1}}}\, dx\Big) \\
&\leqslant (1+s_c) \exp(c\nu)\Big(\int_{\mathcal{X}} e^{-\frac{U}{\eta_{k+1}}}\, dx\Big). \qquad (9.5)
\end{aligned}$$

Here the last inequality is true because the choice of M and η_k (in (2.8) and (4.13) respectively) ensures

$$\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k} \leqslant \nu.$$

Since r_k is always positive, using (9.5) in (9.3), we obtain that for every $c > 0$,

$$\sup_{1\leqslant k\leqslant M-1} \|r_k\|_{L^\infty} \leqslant (1 + s_c)\exp(c\nu).$$

Taking infimum on the right hand side gives (3.2) with C_r defined as in (3.3). □

9.3. **Dimensional dependence of constants for separated energies.**

Proof of Proposition 3.1. We only need to consider the case where $d \geqslant k_0$, where we recall k_0 is the constant in (3.8). According to (3.5) and (3.6), it suffices to show that both C_r and C_{LBV} are independent of d. For C_{LBV}, we note that (3.7) implies that we can take the domain $\Omega_j, j = 1, \ldots, J$ in the form

$$\Omega_j = \tilde{\Omega}_j \times \mathbb{R}^{d-\tilde{d}},$$

where $\tilde{\Omega}_j, j = 1, \ldots, J$ are subsets in $\mathbb{R}^{\tilde{d}}$, corresponding to the domain of measure proportional to $e^{-\tilde{U}_0}$. Then, Fubini's theorem shows that for any $j = 1, \ldots, J$ and any $\varepsilon > 0$, we have

$$\pi_\varepsilon(\Omega_j) = \frac{\int_{\tilde{\Omega}_j} e^{-\tilde{U}_0/\varepsilon}\,dx}{\int_{\mathbb{R}^{\tilde{d}}} e^{-\tilde{U}_0/\varepsilon}\,dx},$$

which is independent of d. Now using (2.3) shows C_{LBV} is also independent of d.

Now it remains to find an upper bound of C_r which is independent of d. Since U_0 is positive, we note

$$\|r_k\|_{L^\infty(\mathcal{X})} \overset{(3.1)}{\leqslant} T_1 \cdot T_2, \quad \text{where} \quad T_1 = \frac{\int_{\mathbb{R}^{\tilde{d}}} e^{-\frac{\tilde{U}_0}{\eta_k}}\,dx}{\int_{\mathbb{R}^{\tilde{d}}} e^{-\frac{\tilde{U}_0}{\eta_{k+1}}}\,dx}, \quad T_2 = \frac{\int_{\mathbb{R}^{d-\tilde{d}}} e^{-\frac{V_0}{\eta_k}}\,dx}{\int_{\mathbb{R}^{d-\tilde{d}}} e^{-\frac{V_0}{\eta_{k+1}}}\,dx}.$$

Given our choice of η_k, we bound T_1 by

$$T_1 \overset{(9.5)}{\leqslant} \inf_{c>0}(1 + s_c(\tilde{U}_0))\exp\Big(\frac{c}{d}\Big) \leqslant \inf_{c>0}(1 + s_c(\tilde{U}_0))e^c,$$

where

$$s_c(\tilde{U}_0) \overset{\text{def}}{=} \frac{\int_{\{\tilde{U}_0>c\}} e^{-\tilde{U}_0}\,dx}{\int_{\{\tilde{U}_0\leqslant c\}} e^{-\tilde{U}_0}\,dx} < \infty,$$

which implies that an upper bound of T_1 only depends on $\tilde{U}_0$ and is independent of d.

Similarly for T_2, we compute

$$T_2 \overset{(9.5)}{\leqslant} \inf_{c>0}(1 + s_c(V_0))\exp\Big(\frac{c}{d}\Big), \quad \text{where} \quad s_c(V_0) \overset{\text{def}}{=} \frac{\int_{\{V_0>c\}} e^{-V_0}\,dx}{\int_{\{V_0\leqslant c\}} e^{-V_0}\,dx}.$$

Using (3.8) when $c > \alpha_u$, we compute

$$s_c(V_0) \overset{(3.8)}{\leqslant} e^{\alpha_u-\alpha_b} \frac{\int_{\{V_0>c\}} e^{-\alpha_0|x-x_0|^{k_0}}\,dx}{\int_{\{V_0\leqslant c\}} e^{-\alpha_0|x-x_0|^{k_0}}\,dx}$$

$$\leqslant e^{\alpha_u-\alpha_b}\frac{\int_{\{\alpha_0|x-x_0|^{k_0}+\alpha_u>c\}} e^{-\alpha_0|x-x_0|^{k_0}}\,dx}{\int_{\{\alpha_0|x-x_0|^{k_0}+\alpha_u\leqslant c\}} e^{-\alpha_0|x-x_0|^{k_0}}\,dx}$$

$$= e^{\alpha_u-\alpha_b}\frac{\int_{\{|x|>(\frac{c-\alpha_u}{\alpha_0})^{1/k_0}\}} e^{-\alpha_0|x|^{k_0}}\,dx}{\int_{\{|x|\leqslant(\frac{c-\alpha_u}{\alpha_0})^{1/k_0}\}} e^{-\alpha_0|x|^{k_0}}\,dx}$$

$$(9.6)\qquad = \frac{e^{\alpha_u-\alpha_b}\Gamma(\frac{d}{k_0}, c-\alpha_u)}{\Gamma(\frac{d}{k_0})-\Gamma(\frac{d}{k_0}, c-\alpha_u)} \leqslant \frac{e^{\alpha_u-\alpha_b}}{\frac{\Gamma(\frac{d}{k_0})}{\Gamma(\frac{d}{k_0},c-\alpha_u)}-1}.$$

By the estimate of incomplete gamma function [Gab79, Satz 4.4.3], when $\frac{d}{k_0}\geqslant 1$ and $c-\alpha_u \geqslant \frac{d}{k_0}$, we have

$$\Gamma\Big(\frac{d}{k_0}, c-\alpha_u\Big) \leqslant \frac{d}{k_0}\exp\big(-(c-\alpha_u)\big)(c-\alpha_u)^{\frac{d}{k_0}-1}.$$

On the other hand, Stirling's formula gives

$$\Gamma\Big(\frac{d}{k_0}\Big) \geqslant C_\Gamma\Big(\frac{d}{k_0}\Big)^{\frac{d}{k_0}-\frac{1}{2}}\exp\Big(-\frac{d}{k_0}\Big),$$

for a positive constant C_Γ. We now choose $c-\alpha_u = \tilde\nu d/k_0$, where $\tilde\nu>1$ is such that

$$(9.7)\qquad \tilde\nu - \log(\tilde\nu) \geqslant \frac{3}{2} + \log\Big(\frac{2}{C_\Gamma}\Big).$$

This gives

$$\frac{\Gamma(\frac{d}{k_0})}{\Gamma(\frac{d}{k_0}, c-\alpha_u)} \geqslant \frac{C_\Gamma\big(\frac{d}{k_0}\big)^{\frac{d}{k_0}-\frac{1}{2}}\exp(-\frac{d}{k_0})}{\frac{d}{k_0}\exp(-(c-\alpha_u))(c-\alpha_u)^{\frac{d}{k_0}-1}}$$

$$\overset{z=d/k_0}{=} C_\Gamma\exp\Big((\tilde\nu-1)z - \frac{1}{2}\log(z) - (z-1)\log(\tilde\nu)\Big)$$

$$(9.8)\qquad \geqslant C_\Gamma\exp\Big(\Big(\tilde\nu-\log(\tilde\nu)-\frac{3}{2}\Big)z\Big) \overset{(9.7)}{\geqslant} 2,$$

where the last inequality we use the fact that $z=\frac{d}{k_0}\geqslant 1$. Therefore, for this choice of c,

$$s_c(V_0) \overset{(9.6),(9.8)}{\leqslant} e^{\alpha_u-\alpha_b},$$

which implies that

$$T_2 \leqslant (1+s_c(V_0))\exp\Big(\frac{\tilde\nu}{k_0}+\frac{\tilde\nu\alpha_u}{d}\Big) \leqslant (1+e^{\alpha_u-\alpha_b})\exp\Big(\frac{\tilde\nu}{k_0}+\tilde\nu\alpha_u\Big),$$

which is independent of d. □

References

[Arr89] S. Arrhenius. On the reaction velocity of the inversion of cane sugar by acids. *Zeitschrift für physikalische Chemie*, 4:226ff, 1889. doi:10.1016/B978-0-08-012344-8.50005-2.

[BGK05] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005. doi:10.4171/JEMS/22.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, 2014.

[CFIN23] A. Christie, Y. Feng, G. Iyer, and A. Novikov. Speeding up Langevin dynamics by mixing, 2023, 2303.18168.

[Che23] S. Chewi. *Log-Concave Sampling*. 2023. URL https://chewisinho.github.io/main.pdf.

[Cho04] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004. doi:10.1214/009053604000000698.

[CP20] N. Chopin and O. Papaspiliopoulos. *An introduction to sequential Monte Carlo*. Springer Series in Statistics. Springer, Cham, 2020. doi:10.1007/978-3-030-47845-2.

[CV22] Z. Chen and S. S. Vempala. Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions. *Theory Comput.*, 18:Paper No. 9, 18, 2022.

[DdFG01] A. Doucet, N. de Freitas, and N. Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 3–14. Springer, New York, 2001. doi:10.1007/978-1-4757-3437-9_1.

[DFY20] M. Damak, B. Franke, and N. Yaakoubi. Accelerating planar Ornstein-Uhlenbeck diffusion with suitable drift. *Discrete Contin. Dyn. Syst.*, 40:4093, 2020. doi:10.3934/dcds.2020173.

[ERY24] B. Engquist, K. Ren, and Y. Yang. Sampling with adaptive variance for multimodal distributions, 2024, 2411.15220. URL https://arxiv.org/abs/2411.15220.

[Eva10] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010. doi:10.1090/gsm/019.

[FW12] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. doi:10.1007/978-3-642-25847-3. Translated from the 1979 Russian original by Joseph Szücs.

[Gab79] W. Gabcke. *Neue Herleitung und explizite Restabschätzung der Riemann-Siegel-Formel*. Dissertation, Universität Götingen, 1979.

[Gey91] C. J. Geyer. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: The 23rd Symposium on the Inference Interface Foundation, 1991*, pages 156–163, 1991.

[GLR18] R. Ge, H. Lee, and A. Risteski. Beyond log-concavity: provable guarantees for sampling multi-modal distributions using simulated tempering Langevin Monte Carlo. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 7858–7867, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/c6ede20e6f597abf4b3f6bb30cee16c7-Abstract.html.

[GLR20] R. Ge, H. Lee, and A. Risteski. Simulated tempering langevin Monte Carlo II: An improved proof using soft markov chain decomposition, 2020, 1812.00793. URL https://arxiv.org/abs/1812.00793.

[GT95] C. J. Geyer and E. A. Thompson. Annealing Markov chain Monte Carlo with applications to ancestral inference. 90(431):909 – 920, 1995. doi:10.1080/01621459.1995.10476590. Cited by: 712; All Open Access, Green Open Access.

[GT01] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[GTC25] W. Guo, M. Tao, and Y. Chen. Provable benefit of Annealed Langevin Monte Carlo for non-log-concave sampling. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=P6IVIoGRRg.

[Has70] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi:10.1093/biomet/57.1.97.

[HIS25] R. Han, G. Iyer, and D. Slepčev. Example code for ASMC, 2025. URL https://codeberg.org/HanRuiyu/Annealed-Sequential-Monte-Carlo.

[HM54] J. M. Hammersley and K. W. Morton. Poor man's Monte Carlo. *J. Roy. Statist. Soc. Ser. B*, 16:23–38; discussion 61–75, 1954. URL http://links.jstor.org/sici?sici=0035-9246(1954)16:1<23:PMMC>2.0.CO;2-O&origin=MSN.

[Kol00] V. N. Kolokoltsov. *Semiclassical analysis for diffusions and stochastic processes*, volume 1724 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000. doi:10.1007/BFb0112488.

[Liu08] J. S. Liu. *Monte Carlo strategies in scientific computing.* Springer Series in Statistics. Springer, New York, 2008.

[LLN19] Y. Lu, J. Lu, and J. Nolen. Accelerating Langevin sampling with birth-death, 2019, 1905.09863. URL `https://arxiv.org/abs/1905.09863`.

[LP17] D. A. Levin and Y. Peres. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2017. doi:10.1090/mbk/107. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.

[LSG24] H. Lee and M. Santana-Gijzen. Convergence bounds for sequential Monte Carlo on multimodal distributions using soft decomposition, 2024, 2405.19553. URL `https://arxiv.org/abs/2405.19553`.

[LSW23] Y. Lu, D. Slepčev, and L. Wang. Birth–death dynamics for sampling: global convergence, approximations and their asymptotics. *Nonlinearity*, 36(11):5731, sep 2023. doi:10.1088/1361-6544/acf988.

[LWZ22] M. Lindsey, J. Weare, and A. Zhang. Ensemble Markov chain Monte Carlo with teleporting walkers. *SIAM/ASA J. Uncertain. Quantif.*, 10(3):860–885, 2022. doi:10.1137/21M1425062.

[MMS23] J. Marion, J. Mathews, and S. C. Schmidler. Finite-sample complexity of sequential Monte Carlo estimators. *Ann. Statist.*, 51(3):1357–1375, 2023. doi:10.1214/23-aos2295.

[MP92] E. Marinari and G. Parisi. Simulated tempering: A new Monte Carlo scheme. 19(6):451 – 458, 1992. doi:10.1209/0295-5075/19/6/002.

[MRR+53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi:10.1063/1.1699114.

[MS14] G. Menz and A. Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.*, 42(5):1809–1884, 2014. doi:10.1214/14-AOP908.

[MS24] J. Mathews and S. C. Schmidler. Finite sample complexity of sequential Monte Carlo estimators on multimodal target distributions. *Ann. Appl. Probab.*, 34(1B):1199–1223, 2024. doi:10.1214/23-aap1989.

[Nea96] R. M. Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6:353–366, 1996. doi:10.1007/BF00143556.

[Nea01] R. M. Neal. Annealed importance sampling. *Stat. Comput.*, 11(2):125–139, 2001. doi:10.1023/A:1008923215028.

[Nea11] R. M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 113–162. CRC Press, Boca Raton, FL, 2011.

[Øks03] B. Øksendal. *Stochastic differential equations.* Universitext. Springer-Verlag, Berlin, sixth edition, 2003. doi:10.1007/978-3-642-14394-6. An introduction with applications.

[PHLa20] E. Pompe, C. Holmes, and K. Ł atuszyński. A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist.*, 48(5):2930–2952, 2020. doi:10.1214/19-AOS1916.

[PJT19] D. Paulin, A. Jasra, and A. Thiery. Error bounds for sequential Monte Carlo samplers for multimodal distributions. *Bernoulli*, 25(1):310–340, 2019. doi:10.3150/17-bej988.

[RBS15] L. Rey-Bellet and K. Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081, may 2015. doi:10.1088/0951-7715/28/7/2081.

[RR55] M. N. Rosenbluth and A. W. Rosenbluth. Monte Carlo calculation of the average extension of molecular chains. *The Journal of Chemical Physics*, 23(2):356–359, 02 1955, https://pubs.aip.org/aip/jcp/article-pdf/23/2/356/18806837/356_1_online.pdf. doi:10.1063/1.1741967.

[SAAG24] D. Sanz-Alonso and O. Al-Ghattas. A first course in Monte Carlo methods, 2024, 2405.16359. URL `https://arxiv.org/abs/2405.16359`.

[SBCCD24] S. Syed, A. Bouchard-Côté, K. Chern, and A. Doucet. Optimised annealed sequential Monte Carlo samplers, 2024, 2408.12057. URL `https://arxiv.org/abs/2408.12057`.

[Sch12] N. Schweizer. Non-asymptotic error bounds for sequential mcmc methods in multimodal settings, 2012, 1205.6733. URL `https://arxiv.org/abs/1205.6733`.

[SW86] R. H. Swendsen and J.-S. Wang. Replica Monte Carlo simulation of spin-glasses. 57(21):2607 – 2609, 1986. doi:10.1103/PhysRevLett.57.2607. Cited by: 1644.

[VCK25] A. Vacher, O. Chehab, and A. Korba. Polynomial time sampling from log-smooth distributions in fixed dimension under semi-log-concavity of the forward diffusion with application to strongly dissipative distributions, 2025, 2501.00565. URL `https://arxiv.org/abs/2501.00565`.

[Vos91] M. Vose. A linear algorithm for generating random numbers with a given distribution. *IEEE Transactions on Software Engineering*, 17(9):972–975, 1991. doi:10.1109/32.92917.

[VW19] S. Vempala and A. Wibisono. Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. doi:10.48550/arXiv.1903.0856.

[WSH09a] D. B. Woodard, S. C. Schmidler, and M. Huber. Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *Ann. Appl. Probab.*, 19(2):617–640, 2009. doi:10.1214/08-AAP555.

[WSH09b] D. B. Woodard, S. C. Schmidler, and M. Huber. Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electron. J. Probab.*, 14:no. 29, 780–804, 2009. doi:10.1214/EJP.v14-638.

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.
Email address: `ruiyuh@andrew.cmu.edu`

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.
Email address: `gautam@math.cmu.edu`

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.
Email address: `slepcev@andrew.cmu.edu`