Title
Understanding continuous and discrete stochastic models by coarse-graining

Author
Wu, Tianqi

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Understanding continuous and discrete stochastic models by coarse-graining

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Mathematics

by

Tianqi Wu

2022

ABSTRACT OF THE DISSERTATION

Understanding continuous and discrete stochastic models by coarse-graining

by

Tianqi Wu

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2022

Professor Georg Menz, Chair

This thesis will examine three stochastic models using the idea of coarse-graining:

(1) A quantitative hydrodynamic limit of the Kawasaki dynamics via a "two-spatial-scale" approach, refining the original two-scale approach of Grunewald, Otto, Villani, and Westdickenberg.

(2) A quantitative ergodic theorem of the infinite-swapping process via a "two-time-scale" approach, adapted from the approach of Menz and Schlichting to the setting of an inhomogeneous diffusion.

(3) A sharp leading order asymptotic for the diameter of a long-range percolation graph via concentration inequalities, which improves a previous result by Coppersmith, Gamarnik, and Sviridenko.

A common key feature in these problems is the presence of multiple "levels" (or *scales*) in space or time. The solution generally involves understanding the characteristic behavior at each level and then combining the information about different levels together.

The dissertation of Tianqi Wu is approved.

Timothy Derek Austin

Marek Biskup

Christina Kim

Georg Menz, Committee Chair

University of California, Los Angeles

2022

Contents

ACKNOWLEDGMENTS

I would like to thank Professor Georg Menz, my Ph.D. advisor, for sparking my interest in the mathematics of probability, for his patient guidance in my research training, and for providing many opportunities for my professional development. I am also grateful for Professor Marek Biskup for his continuous support through these challenging times and for guiding me into new directions of research. I also want to thank the other committee members, Professor Tim Austin and Professor Inwon Kim, for their time and endorsement as well as their instructive lectures on entropy and gradient flows.

I would also like to thank all of my co-authors. In particular, I want to thank Professor Felix Otto for sharing valuable insights on the hydrodynamic limit and its connection to gradient flows, both at the Max-Planck Institute and over Zoom. I also want to thank Professor Wenpin Tang for discussions on simulated annealing while he was at UCLA and Professor André Schlichting from whom I learned much of the techniques used in our work on the infinite swapping algorithm.

Chapter 2 is a major revision of the preprints [DMOW18a] and [DMOW18b]. It is the result of a collaboration by Deniz Dizdar, Georg Menz, Felix Otto, and myself. Chapter 3 is an expanded version of the submitted preprint [MSTW20]. It is the result of a collaboration by Georg Menz, André Schlichting, Wenpin Tang, and myself. Chapter 4 consists of preliminary findings of an ongoing project by Biskup and myself. Some of the works reported in these chapters were partially supported by Max-Planck Institute for Mathematics in the Sciences, Leipzig, Germany and the National Science Foundation grants DMS-1407558 and DMS-1712632.

I want to take this opportunity to thank all of my family, friends, and teachers, who had supported me in one way or another before and during these years of my Ph.D. studies, without whom I would not be here today. As a believer in the gospel of Jesus Christ, I owe my ultimate gratitude to the Lord whose grace sustained me even in my worst failures.

VITA Spring 2022

Tianqi Wu

EDUCATION

Ph.D. candidate in Mathematics, Department of Mathematics, University of California, Los Angeles. Advisor: Georg Menz. (2015 – present)

S.B. in Mathematics, Massachusetts Institute of Technology. (2011 – 2015)

PREPRINTS

G. Menz, A. Schlichting, W. Tang, and T. Wu, Ergodicity of the infinite swapping algorithm at low temperature. Submitted, preprint available at arXiv:1811.10174

D. Dizdar, G. Menz, F. Otto, and T. Wu, The quantitative hydrodynamic limit of the Kawasaki dynamics. In preparation, preprint available at arXiv:1807.09850

D. Dizdar, G. Menz, F. Otto, and T. Wu, Towards a quantitative hydrodynamic limit of the Kawasaki dynamics. In preparation, preprint available at arXiv:1807.09857

Chapter 1

Introduction

A common theme in probability theory is the emergence of new types of patterns in an appropriately "zoomed out" view, or on a *coarse-grained* scale. The emergent new order may be either deterministic or stochastic:

- For example, the *law of large numbers* says that given a sequence of i.i.d. integrable random variables $X_1, X_2, \cdots$, the sample average $\frac{1}{N}(X_1 + X_2 + \cdots + X_N)$ converges almost surely to the expected value $\mathbb{E}X_1$ as $N \to \infty$.
- On the other hand, if we take away the expected value from each X_i and then change the scale factor $\frac{1}{N}$ to $\frac{1}{\sqrt{N}}$, then we obtain a different type of convergence result known as the *central limit theorem*: the rescaled sample average of the fluctuations, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i - \mathbb{E}X_i)$, converges to a normal distribution $\mathcal{N}(0, \operatorname{Var} X_1)$ as $N \to \infty$.

The intuition behind the law of large numbers is the difficulty for many independent sources of randomness to work together: in a large i.i.d. sample, the individual fluctuations tend to cancel each other, leaving only their common statistical properties in the average. In an opposite manner, the *Poisson limit theorem*, also known as the *law of rare events*, is an example of how rare fluctuations on a small scale can accumulate to an occurrence with high probability on a large scale:

- According to the Poisson limit theorem, a sequence of binomial distributions $\operatorname{Binom}(n, p_n)$ satisfying $np_n \to \lambda$ converges to a Poisson distribution $\operatorname{Poi}(\lambda)$ as $n \to \infty$.

- As a consequence, suppose the probability of getting a phone call in a small time interval Δt is approximately $\lambda \Delta t$ as $\Delta t \to 0$, then the waiting time between phone calls is exponentially distributed with mean $1/\lambda$, i.e. the phone calls follow a Poisson process of rate λ.

In this dissertation, we apply the idea of coarse-graining illustrated in these basic examples to study three more complex stochastic models involving multiple scales in space or time. The solution generally involves understanding the characteristic behavior at each scale and then combining the information about different scales together.

1.1 Quantitative hydrodynamic limit of the Kawasaki dynamics

Our first topic is concerned with a *dynamical* version of the law of large numbers, called the *hydrodynamic limit*. This type of results say that under a suitable time-space rescaling, a deterministic dynamics emerges as the typical macroscopic behavior of a random evolution on a microscopic lattice as the system size goes to infinity. The particular microscopic evolution we study is the Kawasaki dynamics of 1-dimensional lattice systems of continuous, unbounded spins. It is a spin-exchange dynamics preserving the mean spin and, in the hydrodynamic limit, converges to a non-linear diffusion equation. On a qualitative level, this convergence had been established in [Fri87], [GPV88], and [LY93], but it is not apparent how to obtain quantitative rates of convergence from their methods.

The first step toward a quantitative theory was made in [GOVW09] by introducing the *two-scale approach*. In a nutshell, one chooses a *mesoscopic scale* between the microscopic and macroscopic scales and then defines a mesoscopic dynamics that is close to both the microscopic dynamics and macroscopic dynamics. The key insight behind the choice of the mesoscopic dynamics in [GOVW09] is a gradient flow interpretation of the microscopic and

macroscopic dynamics, leading to the definition of the mesoscopic evolution in terms of an appropriately *coarse-grained* gradient flow structure. However, this perspective was not fully carried out in [GOVW09] due to technical reasons. In our work, we capitalize a lot more on the idea of choosing a mesoscopic evolution with the natural gradient flow structure, which leads to improved error estimates in the system size.

Let us now give more details. We consider a spin system consisting of N real-valued spins located on the 1-dimensional periodic lattice $\{1, \dots, N\}$. The associated *Hamiltonian* $H : \mathbb{R}^N \to \mathbb{R}$ for the spin values only has single-site potentials $\psi : \mathbb{R} \to \mathbb{R}$ and no interaction term:

$$H(x_1, x_2, \dots, x_n) := \sum_{i=1}^{N} \psi(x_i).$$

The evolution of the spin values is governed via a coupled system of SDEs, called the *Kawasaki dynamics*:

$$dX_t = -A\nabla H(X_t)dt + \sqrt{2A}dB_t,$$

where B_t denotes a standard N-dimensional Brownian motion and $-A$ denotes the (centered) second-order difference operator for the periodic rescaled lattice $\left\{\frac{1}{N}, \dots, 1\right\}$. The presence of the matrix A means a site can only change its spin by distributing the difference to its neighbors. Consequently, the mean spin is conserved and we may restrict the dynamics to the hyperplane $X_N := \{x \in \mathbb{R}^N : \sum_{i=1}^N x_i = 0\}$.

As mentioned before, it is known that in the hydrodynamic limit $N \to \infty$, the Kawasaki dynamics converges to a nonlinear PDE on the 1-dimensional torus $\mathbb{T} = [0, 1]$

$$\partial_t \zeta = \partial_\theta^2 \varphi'(\zeta),$$

where the function $\varphi : \mathbb{R} \to \mathbb{R}$ is the *Cramér transform* of the single-site potential ψ, i.e.

$$\varphi(m) = \sup_{\sigma \in \mathbb{R}} \left(\sigma m - \log \int_{\mathbb{R}} \exp\left(\sigma z - \psi(z)\right) dz \right).$$

Compared with the microscopic dynamics,

- the 1-dimensional periodic lattice is rescaled and embedded into the 1-dimensional torus;
- the Euclidean space X_N is rescaled and embedded into the function space $L^2(\mathbb{T})$, as the subspace of piecewise constant functions;
- the second-order difference operator $-A$ is replaced by the second derivative ∂_θ^2;
- the microscopic Hamiltonian H is replaced by the *macroscopic free energy* $\mathcal{H}(\zeta) := \int_0^1 \varphi(\zeta) d\theta$, with ∇H and φ' as their gradient mappings in their respective spaces;
- and the noise term $\sqrt{2A}dB_t$ has disappeared.

Our aim is to make this statement of convergence *quantitative*, providing a good estimate on the speed of convergence, which is needed in applications. Towards this end we adapt the *two-scale approach* introduced in [GOVW09]. The basic idea is to choose a suitable *mesoscopic scale* between the microscopic and macroscopic scales on which we can define a mesoscopic dynamics η_t that serves as a *coarse-grained* version of both the microscopic dynamics X_t and macroscopic dynamics ζ_t:

- Relative to the microscopic scale of X_N, the mesoscopic observables should become more and more coarse-grained, which helps remove the random fluctuations of the microscopic dynamics by the law of large numbers.
- Relative to the macroscopic scale of $L^2(\mathbb{T})$, the mesoscopic observables should become more and more fine-grained, so that the coarse-grained dynamics approximate the full dynamics on the macroscopic scale.

After a suitable space Y of mesoscopic observables is chosen and embedded in the function space $L^2(\mathbb{T})$, the mesoscopic dynamics η_t will be given in the form of a (deterministic) high-dimensional ODE

$$\frac{d}{dt}\eta_t = -\bar{A}\nabla_{L^2}\bar{H}(\eta_t),$$

where $\bar{H}$ and $\bar{A}$ will be *coarse-grained* versions of H and A for the mesoscopic space Y.

The key feature of this set-up is that all three dynamics can now be viewed as *gradient*

flows, i.e. the evolution of each dynamics reduces some kind of energy in the fastest possible way via some dissipation mechanism (see e.g. [AGS05, San16] for more details, examples, and further references).

- For the macroscopic dynamics ζ_t, the energy functional is $\mathcal{H}$ and the dissipation mechanism is the H^{-1} metric.
- For the mesoscopic dynamics η_t, the energy functional is $\bar{H}$ and the dissipation mechanism is the Euclidean metric on Y given by the inner product $\langle \cdot, \bar{A}^{-1} \cdot \rangle_{L^2}$.
- For the microscopic dynamics, the gradient flow structure appears on the level of the law ρ_t of the process X_t, which evolves according to the Fokker-Planck equation

$$\partial_t \rho_t = \nabla \cdot (\rho_t A \nabla H + A \nabla \rho_t).$$

  The associated energy functional is then given by the *relative entropy* of ρ_t with respect to the *Gibbs equilibrium measure* $\mu(dx) := \frac{1}{Z} \exp(-H(x))dx$,

$$\mathrm{Ent}(\rho_t | \mu) = \int H(x) \rho_t(dx) + \int \log \rho_t(x) \rho_t(dx).$$

  The dissipation mechanism is the *Wasserstein distance* on the space of probability measures over the Euclidean space X_N with the Euclidean metric given by the inner product $x \cdot A^{-1} y$,

$$W_2(\nu, \nu') := \min_{\gamma \in \Pi(\nu, \nu')} \left(\int_{X_N \times X_N} |x - x'|_{A^{-1}}^2 \, d\gamma \right)^{\frac{1}{2}},$$

  where $\Pi(\nu, \nu')$ is the set of all coupling of measures ν and ν' and $|\cdot|_{A^{-1}}$ denotes the Euclidean norm induced by the inner product $x \cdot A^{-1} y$.

It is known that the convergence of gradient flows may be deduced from the Γ-convergence of the associated energy functionals together with the convergence of the dissipation mechanisms in the proper sense (see e.g. [SS04, Ser11, Mie16]). Consequently, the gradient flow interpretation provides insights for the convergence of these dynamics. In terms of the energy functionals:

- The convergence of the coarse-grained Hamiltonian $\bar{H}$ to the macroscopic free energy $\mathcal{H}$ is a version of the *local Cramér theorem*. This input from equilibrium statistical mechanics also has the important consequence that $\bar{H}$ gains convexity from coarse-graining.
- The relative entropy $\mathrm{Ent}(\rho_t|\mu)$ can be interpreted as a *free energy* associated to the ensemble ρ_t, consisting of an energy part $\int H(x)\rho_t(dx)$ coming from the Hamiltonian H and an entropy part $\int \log \rho_t(x)\rho_t(dx)$ coming from the noise B_t.
  - Under coarse-graining onto mesoscopic profiles $y \in Y$, the entropy part vanishes due to the law of large numbers.
  - On the other hand, as the mesoscopic scale becomes more fine-grained, the energy part converges to the coarse-grained Hamiltonian $\bar{H}$ because the Kawasaki dynamics equilibrates faster on smaller spatial scales due to its spin-exchange mechanism. In the rigorous analysis this fact will be quantified with the help of a *log-Sobolev inequality* for the conditional Gibbs measures $\mu(dx|y)$ that is based on the convexity of $\bar{H}$.

In terms of the dissipation mechanisms:

- The operator $-\bar{A}$ is a coarse-grained version of the second-order difference operator $-A$ and therefore should converge to the second derivative ∂_θ^2.
- The dissipation mechanism on the microscopic level has two layers, an outer Wasserstein distance associated with the transportation of probability measures and an inner Euclidean metric associated with the spin-exchange mechanism mediated by the matrix A. Under coarse-graining, the outer metric on the space of probability measures becomes degenerate as randomness vanishes, leaving only the inner metric on X_N associated with the A^{-1} inner product, which should then converge to the Euclidean metric on Y associated with the $\bar{A}^{-1}$ inner product.

To fully take advantage of the gradient flow interpretation, the operation of taking second-order difference should be compatible with coarse-graining, which means the mesoscopic

observables need to be sufficiently smooth. In [GOVW09], the mesoscopic observables were chosen to be *piecewise constant functions*. The lack of regularity then forces one to use an unnatural definition of the coarse-grained operator $\bar{A}$ in the mesoscopic dynamics, leading to sub-optimal error estimates when that approach is taken to completion. In our work, we instead choose the mesoscopic observables to be *splines*, i.e. piecewise polynomials that are smoothly joined together. The extra regularity then allows a more natural definition of the mesoscopic dynamics in terms of its gradient flow structure, leading to overall better error estimates. On the other hand, the smoothness constraints of the splines also make them non-local functions, which makes deducing the main ingredients in the two-scale approach more subtle. As a workaround, we need to introduce another level of mesoscopic observables that are *piecewise polynomials*, on which we can perform the analysis more easily and then transfer the properties established there back to the original level of splines.

1.2 Ergodicity of the infinite swapping algorithm

Our second topic is concerned with the *ergodic properties* of sampling dynamics in a non-convex energy landscape. In many applications in physics, chemistry, engineering, statistics, machine learning, etc, one needs to compute integrals with respect to *Gibbs measures* at low temperature, i.e. probability distributions on $\mathbb{R}^n$ whose probability density is proportional to $\exp(-H(x)/\tau)$, for some energy landscape $H : \mathbb{R}^n \to \mathbb{R}$, and temperature $0 < \tau \ll 1$. Because the dimension n is usually very large, direct integration is not numerically feasible. Instead, by the *ergodic theorem*, one can (with high probability) approximate the integral by the long-time average of one simulation of some random dynamics X_t of a particle moving in the state space $\mathbb{R}^n$, with the property that the distributions of X_t converge to μ as $t \to \infty$. The basic idea is that the dynamics X_t will eventually visit the state space $\mathbb{R}^n$ everywhere with frequency according to μ. For example, the *overdamped Langevin dynamics* X_t governed

by the stochastic differential equation (SDE)

$$dX_t = -\nabla H(X_t) + \sqrt{2\tau} dB_t$$

where B_t is a standard n-dimensional Brownian motion, can be used to sample the Gibbs measure $\mu(dx) := \frac{1}{Z} \exp(-H(x)/\tau)dx$.

However, at low temperatures τ, if the energy landscape H has multiple deep local minima, then a full exploration of the state space becomes very slow because of the *metastability* phenomenon: the particle would quickly go down a nearby valley (due to the gradient term), stay there waiting for an exponentially long time, before it finally jumps over to the neighborhood of another valley (thanks to the random fluctuations of the Brownian motion). The average waiting time for the jump follows the *Arrhenius's law*, originally discovered in the context of chemical reactions:

$$\mathbb{E}[\text{waiting time}] = A(\tau) \exp\left(\frac{\Delta E}{\tau}\right),$$

where the *energy barrier* ΔE is the minimal height that a particle has to climb up in order to cross over from one valley to another in the energy landscape H, and the *pre-exponential factor* $A(\tau)$ is a sub-exponential factor given by the more refined *Eyring-Kramers formula* (see [Ber13] for background).

One way to overcome this slow-down of sampling Gibbs measures at low temperatures due to the dynamics getting trapped at local minima is the *replica exchange* method, also known as *parallel tempering*. In the simplest version, one considers two independent particles governed by the same underlying dynamics, i.e. the overdamped Langevin dynamics in our example, with one particle evolving at the desired low temperature τ_1 and the other particle evolving at some higher temperature τ_2, where $0 < \tau_1 \ll \tau_2 \ll 1$. At some random times, the positions of the two particles are swapped. The advantage of this method is that the low-temperature particle can gather information about the desired low-temperature Gibbs measure while the high-temperature particle can more freely explore the full state space. In

a study of the sampling performance of parallel tempering, the authors of [DLPD12] found that the rate of convergence of the empirical measures (as measured by the large deviation rate function) is a monotone increasing function of the swap rate, which naturally leads them to consider a suitable limiting process as the swap rate goes to infinity, the *infinite swapping algorithm* (isa). This requires a shift of perspective from a *particle-swapped process* to a *temperature-swapped process*: rather than swapping particle positions infinitely fast (which would result in infinitely many jumps in finite time), the isa does not swap particle positions at all but instead swaps temperatures infinitely fast.

In more detail, suppose the goal is to sample the Gibbs measure ν^τ with density proportional to $\exp\left(-\frac{H(x)}{\tau}\right)$, for some smooth non-convex energy landscape $H : \mathbb{R}^n \to \mathbb{R}$, at low temperatures $0 < \tau \ll 1$. Given two different temperatures $0 < \tau_1 \ll \tau_2 \ll 1$, let π be the product measure $\pi(x_1, x_2) := \nu^{\tau_1}(x_1)\nu^{\tau_2}(x_2)$. The isa is defined as the evolution of two particles $X_1(t)$ and $X_2(t)$ governed by the SDEs:

$$\begin{cases} dX_1 = -\nabla H(X_1)\, dt + \sqrt{2\tau_1\rho(X_1, X_2) + 2\tau_2\rho(X_2, X_1)}\, dB_1, \\ dX_2 = -\nabla H(X_2)\, dt + \sqrt{2\tau_2\rho(X_1, X_2) + 2\tau_1\rho(X_2, X_1)}\, dB_2, \end{cases}$$

where B_1, B_2 are independent standard Brownian motions in $\mathbb{R}^n$, and $\rho(x_1, x_2)$ is the relative weight of the configuration (x_1, x_2) versus (x_2, x_1) assigned by

$$\rho(x_1, x_2) := \frac{\pi(x_1, x_2)}{\pi(x_1, x_2) + \pi(x_2, x_1)}.$$

At each instant, this assignment essentially gives the higher temperature τ_2 (resp. the lower temperature τ_1) to the particle whose potential energy H is higher (resp. lower) at that instant (see also [DDN17, Section 3.2]). The crucial feature of the infinite swapping dynamics is that the empirical measure

$$\eta_t := \frac{1}{t}\int_0^t \left(\rho(X_1, X_2)\delta_{(X_1,X_2)} + \rho(X_2, X_1)\delta_{(X_2,X_1)}\right) ds$$

converges weakly to the product measure π as $t \to \infty$ by the ergodic theorem. In particular, the first marginal of the measure η_t approximates the Gibbs measure ν^{τ_1} for t large enough.

In [DLPD12], a large deviation principle was established for the measure η_t, though it is not clear how the rate function depends on the temperatures τ_1, τ_2. Further numerical and heuristic studies in [DDN17] indicate that the isa has an exponential gain in sampling performance over the classical overdamped Langevin dynamics. More recently, the isa was applied to training restricted Boltzmann machines [HNR20] and was shown to be competitive empirically.

In our work, we take the analysis of the isa to the next level via a *functional inequality* approach. It is well-known that *Poincaré* and *log-Sobolev inequalities* yield estimates on the speed of convergence of the time average to the ensemble mean (see e.g. [CG08, WY08]). By rigorously deducing low-temperature asymptotic formulas for the *Poincaré* and *log-Sobolev constants* of the isa, we thus quantify how the speed of convergence depends on the two temperatures τ_1, τ_2, under some standard non-degeneracy assumptions. Compared with the Eyring-Kramers formulas for the classical over-damped Langevin dynamics (see e.g. [BEGK04, BGK05, MS14]), our results for the isa show an exponential gain: the energy barrier in the leading exponential term now only "sees" the higher temperature τ_2 instead of the lower temperature τ_1, at a *polynomial* cost in terms of the ratio τ_2/τ_1.

The proof of our results adapts the "two-time-scale" approach of [MS14] for overdamped Langevin dynamics. The basic idea is to decompose the Gibbs measure ν^τ into two levels:

- a *local measure* on each of the *domains of attraction* Ω_i of the Hamiltonian H, corresponding to the fast convergence to local minima m_i, and
- a *coarse-grained* discrete measure on the set $\{\Omega_1, \Omega_2, \cdots\}$, corresponding to the exponentially long transitions between the local minima.

Consequently, by the law of iterated expectation, the variance (resp. entropy) splits into local variances (resp. entropies) and a coarse-grained variance (resp. entropy). The Poincaré (resp. log-Sobolev) inequality for the full Gibbs measure then follows from analyzing these two levels separately:

- The local variances (resp. entropies) are controlled by the corresponding Poincaré (resp. log-Sobolev) inequalities for the local measures. The two-time-scale heuristic suggests this should not be of leading order. However, the non-convexity of both the Hamiltonian H and the regions Ω_i, combined together with the low temperature τ, makes it difficult to apply the usual methods for proving such inequalities. These obstacles were overcome in [MS14] by a subtle combination of a *Lyapunov argument* and a perturbation argument.
- The coarse-grained variance (resp. entropy) can be expressed in terms of *mean differences* between the local measures. This is where the energy barrier between local minima comes into play and hence will be the dominant contribution to the Eyring-Kramers formula. The proof is based on finding a good transportation interpolation of measures between the regions Ω_i.

In the setting of isa, another layer of complexity is introduced by the swapping between two temperatures: the stationary distribution is now a mixture of products of Gibbs measures at the two different temperatures, $\mu = \frac{1}{2}\left(\nu^{\tau_1} \otimes \nu^{\tau_2} + \nu^{\tau_2} \otimes \nu^{\tau_1}\right)$. In order that the energy barrier only "sees" the higher temperature τ_2, the exponentially long transitions should only happen in the high-temperature component, and thus an estimate for swapping the temperatures of the two components comes into play as a new ingredient.

1.3 Diameter of a long-range percolation graph

Our third topic is concerned with a random graph model inspired by the *small-world* phenomenon exhibited in many real-world networks, i.e. their typical distances are much smaller than their sizes. The term originates from the old paper [Mil67] which suggested that two average Americans are just six acquaintances away from each other (the so-called "six degrees of separation"), and the phenomenon has received renewed interest since advances in communication and transportation technologies, in particular the Internet, have tremendously

increased global connectivity in recent decades.

A natural way to model this phenomenon is a *long-range percolation graph* on the hypercubic lattice $\mathbb{Z}^d$, in which nearest neighbors are connected *a priori* and, independently at random, a new edge is added between each pair of non-adjacent sites x, y with probability $p_{x,y} := 1 - \exp(-\beta|x-y|^{-s}) \approx \beta|x-y|^{-s}$, for some parameter $\beta > 0, s > 0$. The objects of interest are how the *typical graph-theoretical distance* $D(x,y)$ scales with the Euclidean distance $|x-y|$ and, when restricted to a finite box $[N]^d := \{0, 1, \cdots, N\}^d$, how the *diameter* D_N in the graph-theoretical distance scales with the box size N. The motivation is that adding long edges, even quite sparsely, could substantially shorten the typical distance of the network (see e.g. the short article [WS98]).

This type of models was introduced and studied by [BB01] for a one-dimensional case and extended to a multi-dimensional version by [CGS02]. Five regimes of behavior have been identified based on the exponent s: $s < d$, $s = d$, $d < s < 2d$, $s = 2d$, and $s > 2d$.

- In the regimes $s < d$, [BKPS04] showed the graph diameter D_N approaches the (deterministic) number $\lceil d/(d-s) \rceil$ as $N \to \infty$ using the *stochastic dimension* method.
- In the regime $d < s < 2d$, Biskup ([Bis04], [Bis11]) showed a poly-logarithmic scaling for both the graph distance and the graph diameter: $D(x,y) = (\log|x-y|)^{\Delta+o(1)}$ as $|x-y| \to \infty$ and $D_N = (\log N)^{\Delta+o(1)}$ as $N \to \infty$, where $\Delta^{-1} := \log_2(2d/s)$. For the graph distance, this result was improved in the more recent work [BL19] to $D(x,y) = \Theta(1)(\log|x-y|)^{\Delta}$. The main idea is identifying a "binary hierarchy" of edges forming the path between two given sites.
- In the regimes $s > 2d$, [Ber04] showed the graph distance resumes linear scaling with the Euclidean distance.
- The critical regime $s = 2d$, where the model is scale-invariant, is still largely open. In dimension $d = 1$, [DS13] showed that there exists an exponent $\theta(\beta) \in (0,1)$ such that the graph distance $D(0,N) = \Theta(1)N^{\theta(\beta)}$.

Our work focuses on the critical regime $s = d$. For $s = d$, it has been shown in [CGS02] that

$$D_N = \Theta(1) \log N / \log \log N \quad \text{as } N \to \infty.$$

Heuristically, the critical scaling in the connection probabilities suggest the graph is like a "tree" with branching degree $\Theta(\log N)$. For instance, we can partition (or *coarse-grain*) the box $[N]^d$ into $\Theta(\log N)$ many *dyadic annuli* around a given site:

- Within each dyadic annulus, the critical scaling $s = d$ means the sum of the connection probabilities is on the order $\Theta(1)$, and the *law of rare events* takes effect, resulting in an overall degree of connection distributed like $\text{Poi}(\Theta(1))$.
- Across the dyadic annuli, there are $\Theta(\log N)$ many such $\text{Poi}(\Theta(1))$-distributed connection degrees, and the *law of large numbers* takes effect, resulting in a concentration around the expected value of the total connection degree, which is on the order $\Theta(\log N)$.

Therefore, in m steps one would reach $[\Theta(\log N)]^m$ sites, ignoring for the moment any "recoils" back to the sites already visited and multiple "redundant" connections to the same site. Taking $m = d \log N / \log \log N$, then $\Theta(N^d)$ sites would be reached. This suggests the (random) pre-factor $\Theta(1)$ above should be close to the dimension d. In fact, the counting arguments in [CGS02] already give a lower bound $d - o(1)$, but give a much less optimal upper bound. Guided by the "tree" heuristic mentioned before, we show a matching upper bound $d + o(1)$ for the pre-factor by a more systematic counting, so now we know

$$D_N = (d + o(1)) \log N / \log \log N \quad \text{as } N \to \infty.$$

The natural next step is to go beyond the leading order, which requires dealing with the "recoils" and "redundancies" more carefully. In turn, this requires a deeper look at the spatial distribution of the "tree" to see how "homogeneously" the "tree" is distributed in the box.

Chapter 2

Quantitative hydrodynamic limit of the Kawasaki dynamics

We derive a rate of convergence to the hydrodynamic limit of the Kawasaki dynamics for a one-dimensional lattice spin system as considered by Guo, Papanicolaou and Varadhan. We follow the two-scale approach of Grunewald, Otto, Villani, and Westdickenberg. However, we use a different coarse-graining operator that allows us to leverage the gradient flow structure. As a consequence, we obtain a better convergence rate.

2.1 Introduction

The broader context of this work is the derivation of scaling limits for lattice systems. Typically, such a result shows that under a suitable time-space re-scaling, a random evolution of a lattice system converges to a macroscopic evolution as the system size goes to infinity. Two different kinds of limits may be considered. In the hydrodynamic limit (a dynamical version of the law of large numbers), the limiting macroscopic evolution is deterministic and describes the typical macroscopic behavior of the system. In the fluctuation limit (a dynamical version of the central limit theorem), the limiting macroscopic evolution is random and describes the fluctuations around the hydrodynamic limit.

This work is devoted to the hydrodynamic limit of the Kawasaki dynamics of one-dimensional lattice systems of continuous, unbounded spins. The Kawasaki dynamics is a spin-exchange dynamics preserving the mean spin. In the hydrodynamic limit, it converges to a non-linear diffusion equation. On a qualitative level, this convergence was established in [Fri87] using resolvent techniques and in [GPV88] using convergence of martingales and entropy estimates. Our quantitative approach is closer to the [GPV88] method in the sense that we use thermodynamically natural quantities like the relative entropy and its dissipation, and allow for non-convex single site potentials. As an alternative to the martingale method in [GPV88], Lu and Yau introduced the entropy method in [LY93], which is based on a sophisticated Gronwall-type estimate for a relative entropy functional. This method is more straightforward and gives stronger results, but also makes stronger assumptions on the initial data (closeness to hydrodynamic behavior in the sense of relative entropy rather than in the sense of macroscopic observables). All those results were qualitative, and it is not apparent how to make them quantitative.

In the present work we develop a quantitative theory of the hydrodynamic limit of the Kawasaki dynamics by establishing convergence rates. The first step toward a quantitative theory was made in [GOVW09] by introducing the two-scale approach. For a detailed description of the two-scale approach, we refer to Section 2.3. In a nutshell, the two-scale approach introduces an additional mesoscopic scale in-between the microscopic and macroscopic scales. The hydrodynamic limit is then deduced in two steps, first showing the closeness of the stochastic microscopic dynamics to a carefully chosen, deterministic mesoscopic dynamics, and then showing the closeness of that mesoscopic dynamics to the macroscopic dynamics. In [GOVW09], the hydrodynamic limit is still deduced only on a qualitative level, but the main estimate for the first step is already quantitative, and in principle the second step could also be made quantitative with some numerical analysis, which overall would lead to a quantitative result on the hydrodynamic limit.

However, rather than completing the approach of [GOVW09], we instead proceed by applying the two-scale approach with a different choice of the mesoscopic scale. The reason is that the choice of the mesoscopic scale in [GOVW09] would result in error terms with a worse scaling in the system size compared to ours (for details see Remark 2.3.13 and Remark 2.4.6 below). More precisely, [GOVW09] defines the mesoscopic observables by projection onto piecewise constant functions. Due to the lack of regularity, the mesoscopic dynamics has to be defined in an unnatural way, and consequently one has to use a mixed Galerkin procedure, which is not optimal. In the present work, we define the mesoscopic observables by projection onto splines. Because the splines are smooth, the mesoscopic dynamics can be defined more naturally as the Galerkin approximation of the macroscopic dynamics, leading to better error estimates compared to [GOVW09]. On the other hand, because splines do not have a localized basis, deducing the main ingredients of the two-scale approach becomes more subtle. (See Section 2.1.1 for more discussion on the rationale behind the choice of the mesoscopic dynamics.)

The second motivation behind improving the estimates of [GOVW09] is to develop a quantitative theory of the fluctuation limit, which states that the fluctuations of the Kawasaki dynamics converge to the solution of a stochastic diffusion equation. As with the hydrodynamic limit, the fluctuation limit of the Kawasaki dynamics is well understood on a qualitative level (see for example [Spo86, Zhu90, CY92, DGP17]), but there is no quantitative result. A possible line of attack would be to use the two-scale approach. The estimates of [GOVW09] for the distance between the microscopic and mesoscopic dynamics are too weak when using the scaling of the fluctuation limit. Because our error terms scale better, our estimates are still meaningful under this scaling (cf. Theorem 2.3.14).

Another interesting question in this setting is the convergence of the microscopic entropy

to the hydrodynamic entropy. Again, this question is well understood from a qualitative point of view (cf. [Kos01, Fat13]). With the tools provided here, one could hope to make the approach of Fathi [Fat13] quantitative.

2.1.1 Connection to gradient flows

Deducing the hydrodynamic limit is more accessible if both the microscopic and macroscopic dynamics come from gradient flows, i.e. the evolution of each dynamics reduces some kind of energy in the fastest possible way via some dissipation mechanism (see e.g. [AGS05, San16] for more details, examples, and further references). The main idea is that $\Gamma-$convergence of the energy functionals, together with the convergence of the dissipation mechanisms in the proper sense, yields the convergence of the associated gradient flows (see e.g. [SS04, Ser11, Mie16]). This new perspective was applied, for example, in the recent works [FS16, MSW22].

Hence, finding the appropriate gradient flow structure for the microscopic and macroscopic dynamics is beneficial. This task is non-trivial because different gradient flow structures could give rise to the same evolution equation. For example, it was pointed out in [Ott01] that the porous medium equation may be seen both as a H^{-1}-gradient flow of functions and as a Wasserstein gradient flow of number densities. Studying this question led to the recently highlighted insight that the appropriate gradient flow structure arises from the large deviation principle of the underlying microscopic process (see e.g. [ADPZ10, ADPZ13, DLR13, Fat16]), as was implicitly known before (see e.g. line (1.5) in [DG87]).

Let us illustrate the importance of selecting the appropriate gradient flow structure with two examples. The first example is the hydrodynamic limit for interacting Brownian particles on the circle (see [Var91]). The second example is the hydrodynamic limit of the Kawasaki dynamics on a one-dimensional lattice spin system (see e.g. [GPV88]), which is studied in

the present work. The two examples appear quite similar in that they both yield a porous medium type equation in the hydrodynamic limit. However, they differ significantly in terms of the underlying microscopic model: an interacting particle system in the first example and a spin system in the second one. The differences become even more apparent when studying the associated gradient flow structures.

The first example, the hydrodynamic limit of the interacting Brownian particles, can be interpreted as a convergence of gradient flows in the following way.

- In [Var91], on the microscopic level N Brownian particles interact on a circle S. The positions X_i of the particles are given by a coupled system of SDEs with repulsive interaction. Because the evolution is reversible, one can interpret the associated forward Kolmogorov equation as a gradient flow for the relative entropy functional w.r.t. to the Gibbs equilibrium measure in the Wasserstein space of probability measures (see [JKO98]). Here, the inner metric in the Wasserstein distance is given by the Euclidean distance on S^N.
- Because the Brownian particles are indistinguishable, one considers the empirical distribution of the particles, which is obtained by "forgetting" the labels of the particles. The unlabelling of the particles naturally pushes forward the inner metric (see Section 4 in [Ott01]). For the former, the inner metric as described in the first bullet point describes the displacement of the particles (Lagrangian description). For the latter, the inner metric is the discrete Wasserstein distance, i.e. the minimal displacement of the particles required to transport one empirical distribution into another (Eulerian description). The microscopic dynamics relevant for the hydrodynamic limit is then the associated projected evolution of the empirical distributions of the particles.
- As a consequence, one should view the porous medium equation obtained in the hydrodynamic limit as a Wasserstein gradient flow, namely the gradient flow of the macroscopic free energy on the Wasserstein space $M_1(S)$ of number densities on S.

It is worth noting that there are two "levels" of Wasserstein metrics involved here. The "inner" Wasserstein metric is associated with the "movement mechanism" of the dynamics itself, i.e. transporting empirical distributions (for the microscopic dynamics) or transporting number densities (for the macroscopic dynamics). The "outer" Wasserstein metric is associated with the stochastic fluctuations of the dynamics and becomes degenerate, as it is the nature of the hydrodynamic limit to be deterministic. The main takeaway is that the dissipation mechanism for the macroscopic gradient flow is induced by the underlying "movement mechanism" of the microscopic dynamics (in this case, the Wasserstein distance on $M_1(S)$).

Let us now turn to the second example: the interpretation of the hydrodynamic limit of the Kawasaki dynamics as a convergence of gradient flows.

- On the microscopic level, the spin system consists of N real-valued spins located on the discrete one-dimensional torus $\{1, \dots, N\}$. The associated Hamiltonian for the spin values only has single-site potentials and no interaction term (see (2.2.1) below). The evolution of the spin values is governed via a coupled system of SDEs, called the Kawasaki dynamics (see (2.2.4) below). This means that a site can only change its spin by distributing the difference to its neighbors. This spin-exchange mechanism is mediated through the matrix A in (2.2.4). As in the first example, the associated forward Kolmogorov equation (see equation (2.2.9) below) has a gradient flow structure given by the relative entropy w.r.t. to the Gibbs equilibrium measure in the Wasserstein space of probability measures.
- Because a site can only reduce its energy via spin-exchange, the appropriate choice for the inner metric is the inner product $x \cdot A^{-1}y$. Because A is a second-order difference operator, $x \cdot A^{-1}y$ corresponds to a discrete H^{-1} metric. This illustrates another main difference between both examples. The interaction is not mediated by the Hamiltonian but by the dissipation mechanism of the dynamics.
- As a consequence, the porous medium equation obtained in the hydrodynamic limit

of the Kawasaki dynamics should be considered as a (now continuum) H^{-1}-gradient flow.

In [GOVW09], those insights were applied to study the hydrodynamic limit. However, it was not completely carried out. Because of technical reasons, the authors of [GOVW09] did not choose a mesoscopic evolution with the natural $H^{-1}-$ gradient flow structure, leading to suboptimal estimates. In this work, we capitalize more on the idea of using the appropriate gradient flow structure and choose a mesoscopic evolution with the natural $H^{-1}-$ gradient flow structure. While this makes our proof more involved compared to [GOVW09], it leads to estimates with an improved scaling in the systems size N.

Notations and conventions

- We use the letter C to denote a universal generic constant $0 < C < \infty$ that is independent of the dimension N of the underlying lattice.
- We denote with $a \lesssim b$ that $a \leq Cb$. We denote $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$.
- We denote with $a \cdot b$ and $|\cdot|$ the standard Euclidean inner product and norm on $\mathbb{R}^N$.
- Let X be a Euclidean space and $f : X \to \mathbb{R}$. Then we denote with ∇f and $\operatorname{Hess} f$ the gradient and Hessian inherited from the Euclidean structure of X.
- We use dx as a shorthand for the Hausdorff or Lebesgue measure of appropriate dimension.
- $|\cdot|_{H^1}$ denotes the homogeneous H^1 norm.
- $\Phi(z) := z \log z$.
- $[M] := \{1, \dots, M\}$. When indexing over $[M]$, we use the convention $0 = M$.
- $L^2(\mathbb{T})$ denotes the L^2 functions on the torus $\mathbb{T} = [0, 1]$.
- $L^2_0(\mathbb{T})$ denotes the L^2 functions on the torus $\mathbb{T} = [0, 1]$ *with mean zero*.

2.2 Setting and main result

We start with describing the Kawasaki dynamics on the microscopic lattice $\{1, \dots, N\}$.

Definition 2.2.1 (Microscopic Hamiltonian H)**.** *The Hamiltonian $H : \mathbb{R}^N \to \mathbb{R}$ of the system is given by*

$$H(x) := H_N(x) = \sum_{n=1}^{N} \psi(x_n). \tag{2.2.1}$$

Here $\psi : \mathbb{R} \to \mathbb{R}$ is the single-site potential*, assumed to be of the form*

$$\psi(x) = \frac{1}{2} x^2 + ax + b + \delta\psi(x) \tag{2.2.2}$$

for some constants a, b and some function $\delta\psi$ that is bounded in $C^2(\mathbb{R})$, i.e.

$$\|\delta\psi\|_{L^\infty(\mathbb{R})} \leq C \text{ and } \|\delta\psi''\|_{L^\infty(\mathbb{R})} \leq C. \tag{2.2.3}$$

The function ψ may be non-convex and it helps to consider the case of a double-well potential.

Definition 2.2.2 (Microscopic dynamics)**.** *The Kawasaki dynamics X_t is given by the solution of the SDE*

$$dX_t = -A\nabla H(X_t)dt + \sqrt{2A}dB_t. \tag{2.2.4}$$

Here B_t denotes a standard N-dimensional Brownian motion and $-A$ denotes the (centered) second-order difference operator for the periodic rescaled lattice $\left\{\frac{1}{N}, \dots, 1\right\}$. More precisely, the operator A is given by the $N \times N-$matrix

$$A_{i,j} := N^2(-\delta_{i,j-1} + 2\delta_{i,j} - \delta_{i,j+1}). \tag{2.2.5}$$

Remark 2.2.3 (Structure of the operator A)**.** *The second-order difference operator A is of the form $A = D^t D$, where D denotes the periodic rescaled $N \times N$ forward difference matrix given by*

$$D_{i,j} := N(\delta_{i,j-1} - \delta_{i,j}),$$

and D^t denotes the matrix transpose of D.

It follows from the structure of the operator A that the Kawasaki dynamics (2.2.4) conserves the mean spin of the system. Hence, after a translation of the single-site potential ψ, we may restrict the state space $\mathbb{R}^N$ of the Kawasaki dynamics X_t to the hyperplane of zero mean

$$X_N := \left\{ x \in \mathbb{R}^N : \ \frac{1}{N} \sum_{i=1}^{N} x_i = 0 \right\}.$$

We endow the space X_N with the standard Euclidean inner product inherited from $\mathbb{R}^N$

$$\langle x, y \rangle_{X_N} := x \cdot y := \sum_{i=1}^{N} x_i y_i.$$

Additionally, the operator A is positive definite when restricted to X_N. Hence:

Definition 2.2.4 (Euclidean structures on X_N induced by A)**.** *The operator A induces a dual pair of inner products on the state space X_N, given by*

$$\langle x, y \rangle_A := x \cdot Ay, \ \text{ and } \ \langle x, y \rangle_{A^{-1}} := x \cdot A^{-1} y.$$

We denote by $|\cdot|_A$ and $|\cdot|_{A^{-1}}$ the corresponding norms on X_N.

The A/A^{-1}-Euclidean structures can be seen as a discrete version of the H^1/H^{-1} structures. In particular, we have the following well-known discrete analogue of the Poincaré inequality for functions with zero mean.

Lemma 2.2.5 (Discrete Poincaré inequality)**.** *For all integers $N \geq 1$ and all $x \in X_N$,*

$$|x|^2 \lesssim |x|_A^2. \tag{2.2.6}$$

When the state space X_N is endowed with the A^{-1} inner product, the dynamics (2.2.4) can be written in the more suggestive form

$$dX_t = -\nabla_{A^{-1}} H(X_t) dt + \sqrt{2} dB_t^{A^{-1}}, \tag{2.2.7}$$

where $\nabla_{A^{-1}} := A\nabla$ denotes the gradient operation wrt. to the A^{-1} inner product and $B_t^{A^{-1}} := \sqrt{A} B_t$ denotes a Brownian motion on X_N having *identity covariance matrix* wrt. to

the A^{-1} inner product.

As a standard result in the theory of stochastic processes (see for example [Pav14]), the law of the process X_t at time t is characterized via the forward Kolmogorov equation.

Lemma 2.2.6 (Forward Kolmogorov equation)**.** *Assume that the law of initial condition X_0 is absolutely continuous wrt. the $N-1$ dimensional Hausdorff measure $\mathcal{L}^{N-1}$. Let μ denote the Gibbs measure on X_N associated to the Hamiltonian H, i.e. the measure μ is absolutely continuous wrt. the $N-1$-dimensional Hausdorff measure $\mathcal{L}^{N-1}$ with the Radon-Nikodym derivative given by*

$$\frac{d\mu}{d\mathcal{L}^{N-1}}(x) = \frac{1}{Z}\exp\left(-H(x)\right)\mathbb{1}_{x\in X_N}. \tag{2.2.8}$$

Then for all times $t > 0$, the law $\rho(t)$ of the Kawasaki dynamics X_t (2.2.7) is absolutely continuous wrt. the Gibbs measure μ, i.e. $\rho(t) = f(t)\mu$ for some $f(t) \in L^1(\mu)$, and is a weak solution of the Fokker-Planck equation

$$\partial_t \rho_t = \nabla \cdot (\rho_t \nabla_{A^{-1}} H + \nabla_{A^{-1}} \rho_t) \tag{2.2.9}$$

in the sense that for any smooth test function $\xi : X_N \to \mathbb{R}$ it holds

$$\frac{d}{dt}\int \xi(x)\rho_t(dx) = \int -\nabla H \cdot \nabla_{A^{-1}}\xi\, \rho_t(dx) + \int \nabla \cdot \nabla_{A^{-1}}\xi\, \rho_t(dx).$$

In particular, the Gibbs measure μ is the unique stationary distribution of the Kawasaki dynamics (2.2.7)*.*

As a consequence of the forward Kolmogorov equation, the *relative entropy* of the law of X_t wrt. the Gibbs measure μ,

$$\mathrm{Ent}(\rho(t)|\mu) := \int f(t,x)\log f(t,x)\mu(dx),$$

decreases monotonically over time at the rate

$$\frac{d}{dt}\mathrm{Ent}\left(\rho(t)|\mu\right) = -\int |\nabla \log f(t,x)|_A^2\ \rho_t(dx). \tag{2.2.10}$$

The integral on the right hand side is the *Fisher information* for the Kawasaki dynamics, which differs from the standard Fisher information

$$\mathcal{I}_\mu(f(t)) := \int |\nabla \log f(t,x)|^2 f(t) \mu(dx)$$

only in the Euclidean structure being used. Hence, after we use the discrete Poincaré inequality in Lemma 2.2.5 to account for the different Euclidean structures on X_N, the rate of dissipation of the relative entropy is quantified by a *log-Sobolev inequality* (LSI) for the Gibbs measure μ by a standard Gronwall-type argument. In [GOVW09] it was shown that this rate of dissipation is independent of the system size N:

Proposition 2.2.7 (Uniform LSI for μ)**.** *The Gibbs measure μ given by* (2.2.8) *satisfies a LSI with constant $\hat{\varrho} > 0$ uniform in the system size N. More precisely, for any nonnegative test function $g : X_N \to \mathbb{R}$ that satisfies $\int g(x)\mu(dx) = 1$, it holds that*

$$\mathrm{Ent}\,(g\mu|\mu) \leq \frac{1}{2\hat{\alpha}} \mathcal{I}_\mu(g). \tag{2.2.11}$$

Remark 2.2.8 (Gradient flow structure of the microscopic dynamics)**.** *The Fokker-Planck equation* (2.2.9) *can be written in the form*

$$\partial_t \rho_t = \nabla \cdot \left(\rho_t \nabla_{A^{-1}} \frac{\delta E}{\delta \rho}(\rho_t) \right),$$

where $E(\rho)$ is the microscopic free energy *of an ensemble $\rho = f\mu$,*

$$E(\rho) := \int H d\rho + \int \rho \log \rho \, dx = \mathrm{Ent}(\rho|\mu),$$

and $\dfrac{\delta E}{\delta \rho} = H + \log \rho = \log f$ is its first variation. Consequently, on the level of probability densities on X_N, the Kawasaki dynamics X_t may be viewed as an A^{-1}-Wasserstein gradient flow for the convex energy functional $E(\cdot) = \mathrm{Ent}(\cdot|\mu)$, whose unique minimizer is the Gibbs measure μ. As expected for Wasserstein gradient flows, the energy functional E decreases over time at the rate

$$\frac{d}{dt} E(\rho_t) = -\int \left| \nabla_{A^{-1}} \frac{\delta E}{\delta \rho}(\rho) \right|^2_{A^{-1}} d\rho_t.$$

Moreover, the log-Sobolev inequality (2.2.11) *that quantifies this dissipation of the energy functional* $E(\cdot) = \mathrm{Ent}(\cdot|\mu)$ *may also be seen as a statement quantifying its convexity wrt. the Wasserstein distance, since by Theorem 1 in [OV00] the inequality* (2.2.11) *implies the* Talagrand's transportation inequality

$$\mathrm{Ent}\,(\rho|\mu) \geq \frac{\hat{\alpha}}{2} W_2^2(\rho, \mu), \tag{2.2.12}$$

where W_2 *denotes the* L^2*-Wasserstein distance.*

The goal of the present work is to derive quantitative bounds on the hydrodynamic limit of the Kawasaki dynamics $X_t \in X_N$. Hydrodynamic limit means that as $N \to \infty$ the random dynamics X_t defined on the one-dimensional periodic lattice $\{1, 2, \cdots, N\}$ converges to a deterministic dynamics ζ_t on the one-dimensional torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Towards this end, we embed the spaces X_N into the space $L_0^2(\mathbb{T})$ of square-integrable functions of mean zero, by identifying the vector $x \in X_N$ with its corresponding step function on the interval $[0, 1]$.

Convention. *Given* $x \in X_N$*, we identify it with the step function*

$$x(\theta) = x_j, \qquad \theta \in \left[\frac{j-1}{N}; \frac{j}{N}\right).$$

Then the space X_N *is identified with the space of piecewise constant functions on* $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ *with mean* 0*, i.e.*

$$X_N = \left\{ x : \mathbb{T} \longrightarrow \mathbb{R};\ x \text{ is constant on } \left[\frac{j-1}{N}; \frac{j}{N}\right),\ j = 1, .., N, \text{ and } \int_0^1 x(\theta) d\theta = 0 \right\}. \tag{2.2.13}$$

With this identification, $X_N \subset L_0^2(\mathbb{T})$ *and inherits the* L^2 *inner product, which is related to the standard Euclidean inner product on* X_N *by a rescaling*

$$\langle x, y \rangle_{L^2} = \frac{1}{N} x \cdot y.$$

It turns out the L^2 norm is not well-suited to describe the hydrodynamic limit since it is too sensitive to local fluctuations. Therefore we endow the embedded space $X_N \subset L_0^2(\mathbb{T})$

with the weaker homogeneous H^{-1}-norm, which is natural in light of the alternative form of the Kawasaki dynamics in (2.2.7) and the analogy between A^{-1}-norm and H^{-1}-norm.

Definition 2.2.9 (H^{-1}-norm)**.** *If $f : \mathbb{T} \to \mathbb{R}$ is a locally integrable function with mean zero, then*

$$||f||_{H^{-1}}^2 := \int_{\mathbb{T}} w(\theta)^2 d\theta, \quad w' = f, \quad \int_{\mathbb{T}} w(\theta) d\theta = 0.$$

We now describe the limiting macroscopic dynamics ζ_t.

Definition 2.2.10 (Macroscopic free energy)**.** *The macroscopic free energy $\mathcal{H} : L^2(\mathbb{T}) \to \mathbb{R}$ is given by*

$$\mathcal{H}(\zeta) = \int_{\mathbb{T}} \varphi(\zeta(\theta)) d\theta, \tag{2.2.14}$$

where the function $\varphi : \mathbb{R} \to \mathbb{R}$ is the Cramér transform *of the single-site potential ψ, given by*

$$\varphi(m) = \sup_{\sigma \in \mathbb{R}} \left(\sigma m - \log \int_{\mathbb{R}} \exp\left(\sigma z - \psi(z)\right) dz \right). \tag{2.2.15}$$

Accordingly, $\nabla \mathcal{H}(\zeta) = \varphi'(\zeta)$ in the variational sense:

$$\frac{d}{d\varepsilon} \mathcal{H}(\zeta + \varepsilon \xi) = \langle \varphi'(\zeta), \xi \rangle_{L^2} \quad \text{for any } \xi \in L^2(\mathbb{T}). \tag{2.2.16}$$

In particular, the macroscopic free energy $\mathcal{H}$ is convex. Indeed, the integrand $\varphi(m)$ is defined as the Legendre transform of the smooth function $\psi^* : \mathbb{R} \to \mathbb{R}$ given by

$$\psi^*(\sigma) := \log \int_{\mathbb{R}} \exp\left(\sigma z - \psi(z)\right) dz, \tag{2.2.17}$$

which is the *log partition function* associated to the linearly shifted potential $\psi(z) - \sigma z$. It turns out that the perturbed quadratic form (2.2.2) of ψ implies that ψ^* is strongly convex and bounded in C^2 (see e.g. Lemma 2.7.9 below or [GOVW09, Lemma 41]). These properties are then transferred to the conjugate function φ by the Legendre transform:

Lemma 2.2.11. *The function $\varphi : \mathbb{R} \to \mathbb{R}$ given by* (2.2.15) *is smooth and satisfies*

$$0 < \lambda \leq \varphi''(\theta) \leq \Lambda < \infty \quad \textit{for all } \theta \in \mathbb{R}.$$

Moreover, the strong convexity of ψ^* means the mapping $m = (\psi^*)'(\sigma)$ is bijective, so that up to a change in the linear term in the potential ψ, we may assume $(\psi^*)'(0) = 0$ and therefore $\varphi'(0) = 0$. Finally, up to a change in the constant term in ψ, we may also assume $\psi^*(0) = 0$. After applying the Legendre transform, this means we conveniently have that

Assumption 1. *The function φ satisfies $\varphi(0) = \varphi'(0) = 0$. Consequently, the macroscopic free energy $\mathcal{H}(\zeta)$ is minimized at $\zeta = 0$ with $\mathcal{H}(0) = 0$.*

Definition 2.2.12 (Macroscopic dynamics)**.** *The macroscopic dynamics $\zeta(t, \cdot)$ is the unique weak solution of the nonlinear diffusion equation*

$$\frac{\partial \zeta}{\partial t} = \frac{\partial^2}{\partial \theta^2} \varphi'(\zeta) \tag{2.2.18}$$

with initial condition $\zeta(0, \cdot) := \zeta_0$. The precise formulation is deferred to Definition 2.2.15 at the end of this section.

Remark 2.2.13 (Gradient flow structure of the macroscopic dynamics)**.** *The nonlinear diffusion equation* (2.2.18) *can be written in the form*

$$\partial_t \zeta = -\nabla_{H^{-1}} \mathcal{H}(\zeta),$$

where $\nabla_{H^{-1}}$ is the gradient mapping of the first variation $\frac{\delta \mathcal{H}}{\delta \zeta}$ wrt. to the H^{-1} inner product (rather than wrt. to the L^2 inner product as in the formulation of (2.2.16)*). Consequently, the macroscopic dynamics may be viewed as a H^{-1}-gradient flow for the convex energy functional $\mathcal{H}$, which therefore monotonically decreases over time at the rate*

$$\frac{d}{dt} \mathcal{H}(\zeta_t) = -|\nabla_{H^{-1}} \mathcal{H}(\zeta)(t)|^2_{H^{-1}} = -|\varphi'(\zeta_t)|^2_{H^1}. \tag{2.2.19}$$

Now, let us formulate the main result of this work.

Theorem 2.2.14 (Quantitative hydrodynamic limit for the Kawasaki dynamics)**.** *We assume that the single-site potential ψ satisfies* (2.2.2) *and* (2.2.3)*. Let μ denote the Gibbs measure given by* (2.2.8) *and let $\rho(t) = f(t)\mu$ denote the law of the Kawasaki dynamics X_t*

(cf. Lemma 2.2.6). We assume that the initial law $\rho(0) = f(0)\mu$ of X_0 has bounded microscopic entropy in the sense that for some constant $0 < C_{\text{Ent}} < \infty$,

$$\text{Ent}(\rho(0)|\mu) := \int f(0,x) \log f(0,x) \mu(dx) \leq C_{\text{Ent}} N. \tag{2.2.20}$$

Let ζ_t be the deterministic dynamics described by equation (2.2.18). Then there is a constant $0 < C < \infty$ depending only on the constants appearing in (2.2.3) such that for any $T > 0$,

$$\sup_{0 \leq t \leq T} \mathbb{E}|X_t - \zeta_t|^2_{H^{-1}} \leq C\, \mathbb{E}|X_0 - \zeta_0|^2_{H^{-1}} + \frac{C}{N^{\frac{2}{3}}} \Big[T + C_{\text{Ent}} + |\zeta_0|^2_{L^2} + 1 \Big]. \tag{2.2.21}$$

The statement of Theorem (2.2.14) is a quantitative version of the hydrodynamic limit. In [GOVW09], only the error from comparing the microscopic scale to a mesoscopic scale was explicit. That error scaled in [GOVW09] like $\frac{1}{\sqrt{N}}$.

We finish this section by giving the precise formulation of equation (2.2.18) that describes the limiting macroscopic dynamics.

Definition 2.2.15. *We call $\zeta(t,\theta)$ a weak solution of (2.2.18) on $[0,T] \times \mathbb{T}$ if*

$$\zeta \in L^\infty_t(L^2_\theta), \quad \frac{\partial \zeta}{\partial t} \in L^2_t(H^{-1}_\theta), \quad \varphi'(\zeta) \in L^\infty_t(L^2_\theta)$$

and

$$\left\langle \xi, \frac{\partial \zeta}{\partial t} \right\rangle_{H^{-1}} = - \left\langle \xi, \varphi'(\zeta) \right\rangle_{L^2} \quad \text{for all } \xi \in L^2_0(\mathbb{T}), \ \text{ for a.e. } t \in [0,T]. \tag{2.2.22}$$

Here, $L^\infty_t(L^2_\theta)$ (resp. $L^2_t(H^{-1}_\theta)$ is the set of functions $\zeta : [0,T] \times \mathbb{T} \longrightarrow \mathbb{R}$ such that $\int_{\mathbb{T}} \zeta(t,\theta) d\theta = 0$ and $||\zeta(t,\cdot)||_{L^2}$ (resp. $||\zeta(t,\cdot)||_{H^{-1}}$) is essentially bounded in t (resp. in $L^2([0,T])$).

2.3 The two-scale approach

We will use the two-scale approach from [GOVW09] to deduce Theorem 2.2.14. The main idea is to introduce an intermediate dynamics on a mesoscopic scale between the microscopic

dynamics (2.2.4) and the macroscopic dynamics (2.2.18). The hydrodynamic limit is then deduced in two steps; in the first step, one deduces the convergence of the microscopic dynamics to the mesoscopic dynamics (see Theorem 2.3.14 from below); in the second step, one deduces the convergence of the mesoscopic dynamics to the macroscopic dynamics (see Theorem 2.3.15 from below).

The most important ingredient in the two-scale approach is the correct definition of the mesoscopic dynamics. The mesoscopic dynamics emerges from projecting the microscopic observables onto mesoscopic observables. The projection onto mesoscopic observables is done with the help of a coarse-graining operator P. We recall that an element $x \in X_N$ is identified with a function on the torus $\mathbb{T} = [0,1]$ that is piecewise constant with value x_n on $[\frac{n-1}{N}, \frac{n}{N})$, $n = 1, \dots N$ (cf. (2.2.13)). The coarse-graining operator P used in [GOVW09] can be interpreted in the following way. It is the projection of X_N in $L^2(\mathbb{T})$ onto the space of functions that are piecewise constant on the intervals $\left[\frac{m-1}{M}, \frac{m}{M}\right)$, $m = 1, ..., M$. More precisely, this means that first one decomposes the lattice $\{1, \dots, N\}$ into M-many blocks $B(m)$ of size K i.e. $N = MK$ and

$$B(m) = \{m(K-1)+1, \dots, mK\} \qquad \text{for } 1 \leq m \leq M.$$

Then the operator $P : X_N \to \mathbb{R}^M$ in [GOVW09] is given for $x \in X_N$ by

$$P(x) = \left(\frac{1}{K} \sum_{i \in B(1)} x_i, \dots, \frac{1}{K} \sum_{i \in B(M)} x_i \right).$$

The main difference of this work compared to [GOVW09] is that the operator P is now defined as the L^2 projection onto splines of degree 2 instead (see Definition 2.3.1 from below). Because spline functions of degree 2 are $C^1(\mathbb{T})$, the mesoscopic variables are more regular compared to [GOVW09]. This has two important advantages:

- In the first step of the two-scale approach, namely showing the convergence of the microscopic dynamics to the mesoscopic dynamics (see Theorem 2.3.14 below), we get

a better error estimate compared to [GOVW09, Theorem 8].

- The second step of the two-scale approach, namely deducing the convergence of the mesoscopic dynamics to the macroscopic dynamics, becomes significantly easier (see Theorem 2.3.15 below). Instead of a mixed method we can apply a direct Galerkin approximation method.

However, there is a trade-off compared to the argument of [GOVW09]. For deducing the convergence of the microscopic dynamics to the mesoscopic dynamics one needs certain ingredients, among which is a uniform logarithmic Sobolev inequality (LSI) for the conditional Gibbs measures and the strong convexity of the coarse-grained Hamiltonian. Deducing those ingredients becomes more difficult compared to [GOVW09].

We now build up the notion of the mesoscopic dynamics by *coarse-graining* the relevant features of the microscopic dynamics X_t.

Definition 2.3.1 (The coarse-graining operator P)**.** *For $M \in \mathbb{N}$, let $Y = Y_M$ be the space of spline functions of degree L with mean zero on the torus $\mathbb{T} = [0,1]$ corresponding to the mesh $\left\{\frac{m}{M}\right\}_{m\in[M]}$. That is*

$$Y_M := \left\{ y \in C^{L-1}(\mathbb{T}) |\, \forall m \in [M] : y|_{\left(\frac{m-1}{M}, \frac{m}{M}\right)} \text{ polynomial of degree} \leq L, \text{ and} \int_0^1 y(\theta) d\theta = 0 \right\}.$$

In this work, we choose the degree of the splines in Y_M to be $L = 2$. We endow Y_M with the inner product inherited from $L^2(\mathbb{T})$. We define the coarse-graining operator $P : L^2(\mathbb{T}) \to Y_M$ as the L^2-orthogonal projection onto Y_M.

The following basic facts show that splines serve as good approximations for deducing the hydrodynamic limit in the H^{-1} norm.

Lemma 2.3.2 (Penalization of fluctuations by a strong norm)**.** *For any function $\zeta \in L^2_0(\mathbb{T})$,*

$$|\zeta - P\zeta|_{H^{-1}} \lesssim \frac{1}{M} |\zeta - P\zeta|_{L^2} \lesssim \frac{1}{M^2} |\zeta|_{H^1}, \qquad \text{and} \tag{2.3.1}$$

$$|P\zeta|_{H^1} \lesssim |\zeta|_{H^1}, \qquad |P\zeta|_{H^{-1}} \lesssim |\zeta|_{H^{-1}}. \tag{2.3.2}$$

The proof of Lemma 2.3.2 is given in Section 2.6, where we gather and prove facts about splines. The core idea is a Poincaré inequality on an internal length scale $\frac{1}{M}$.

We also need the adjoint operator $P^t : Y_M \to X_N$, defined by

$$\langle Px, y\rangle_{L^2} = x \cdot P^t y, \quad \forall x \in X_N, y \in Y_M.$$

It follows that $NP^t : Y_M \to X_N$ is the L^2-orthogonal projection of Y_M onto X_N. Because the spline space Y_M is not a subspace of X_N for spline degree $L \geq 1$, $PNP^t \neq \mathrm{id}_{Y_M}$ in general (cf. assumption (2) in [GOVW09]). However, once the block size K is large enough, the microscopic space X_N will have enough resolution to fully describe the splines in Y_M, and the back-and-forth projection PNP^t will become close to the identity operator.

Lemma 2.3.3. *Assume $N = KM$ for $K \in \mathbb{N}$. It holds that*

$$\|PNP^t - \mathrm{id}_{Y_M}\| = O\left(\frac{1}{K^2}\right).$$

In particular, if $K = \frac{N}{M}$ is large enough, then $PNP^t : Y_M \to Y_M$ is invertible.

Lemma 2.3.3 is a corollary of Lemma 2.7.3 in Section 2.7.1. The core idea is a Poincaré inequality on an internal length scale $\frac{1}{K}$.

From now on, we assume $N = KM$ for $K \in \mathbb{N}$ large enough so that $PNP^t : Y_M \to Y_M$ is invertible. In particular, this means the coarse-graining operator $P : X_N \to Y_M$ has full range and the orthogonal projection $NP^t : Y_M \to X_N$ is an embedding. Hence:

Definition 2.3.4 (Disintegration of the canonical ensemble μ)**.** *The operator P induces a decomposition of the Gibbs measure μ into a family of conditional measures $\mu(dx|y) := \mu(dx|Px = y)$ on the fibers $P^{-1}(y) \subset X_N$ and a marginal measure $\bar{\mu}(dy)$ on the image Y_M, in the sense that*

$$\int g(x)\mu(dx) = \int\int g(x)\mu(dx|y)\bar{\mu}(dy)$$

for any test function $g : X_N \to \mathbb{R}$.

More explicitly, the conditional measure $\mu(dx|y)$ is a probability measure of the form

$$\mu(dx|y) = \frac{1}{Z}\, \mathbb{1}_{\{Px=y\}}(x)\ \exp\left(-H(x)\right) \mathcal{L}^{N-M}(dx),$$

where $\mathcal{L}^{N-M}$ denotes the $N-M$-dimensional Hausdorff measure on the affine subspace $P^{-1}(y) \subset X_N$. The marginal measure $\bar{\mu}$ is a probability measure of the form

$$\bar{\mu}(dy) = \frac{1}{Z} \exp\left(-N\bar{H}(y)\right) dy,$$

where $\bar{H}$ is the *coarse-grained Hamiltonian* given by (2.3.3) below and dy is the Hausdorff measure on Y_M.

Definition 2.3.5 (Coarse-grained Hamiltonian $\bar{H}$)**.** *The coarse-grained Hamiltonian $\bar{H} : Y_M \to \mathbb{R}$ is given by*

$$\bar{H}(y) := -\frac{1}{N} \log \int_{\{x \in X_N : Px = y\}} \exp\left(-H(x)\right) \mathcal{L}^{N-M}(dx), \tag{2.3.3}$$

where $\mathcal{L}^{N-M}$ denotes the $N-M$-dimensional Hausdorff measure.

It follows from a short calculation that the gradient of $\bar{H}$ is also a coarse-grained version of the gradient of H:

$$NP^t \nabla \bar{H}(y) = \mathbb{E}_{\mu(dx|y)} \nabla H(x), \tag{2.3.4}$$

which serves as a crucial link between the microscopic and mesoscopic dynamics. The main advantage of the coarse-grained Hamiltonian $\bar{H}$ over the original microscopic Hamiltonian H is a convexification resulting from averaging over large blocks, which is a well-known phenomenon in statistical mechanics and will be central to our analysis.

Theorem 2.3.6 (Uniform strong convexity of $\bar{H}$)**.** *There are constants $0 < \lambda, \Lambda, K^* < \infty$ such that for all $K \geq K^*$, M and all $y \in Y_M$ it holds*

$$2\lambda\, \mathrm{id}_{Y_M} \leq \mathrm{Hess}\, \bar{H}(y) \leq 2\Lambda\, \mathrm{id}_{Y_M}$$

in the sense of quadratic forms.

Remark 2.3.7. *Theorem 2.3.6 will be proven in Section 2.7. It should be compared to the similar statement of Lemma 29 in [GOVW09]. The situation here is more subtle. In [GOVW09], the mesoscopic observables are also piecewise constant functions and therefore local functions. In contrast, the mesoscopic observables in our setting are given by continuous splines which are non-local functions. This introduces additional interactions between blocks. We work around this obstacle by first deducing the strong convexity for mesoscopic observables that are piecewise polynomials of degree L, or* discontinuous Galerkin functions *in the jargon of numerical analysis, and then transferring the result back to the spline space* Y_M.

Besides the coarse-grained Hamiltonian $\bar{H}$, we also need a coarse-grained version of the second-order difference operator $-A$.

Definition 2.3.8 (Coarse-grained operator $\bar{A}$)**.** *The coarse-grained second-order difference operator* $-\bar{A}$ *is defined by*

$$\bar{A} := PANP^t.$$

In particular, the coarse-grained operator $\bar{A}$ inherits the positive definiteness of the operator A. Hence:

Definition 2.3.9 (Euclidean structures on Y_M induced by $\bar{A}$)**.** *The operator* $\bar{A}$ *induces a dual pair of inner products on the spline space* Y_M

$$\langle y, z\rangle_{\bar{A}} := \langle y, \bar{A}z\rangle_{L^2} \quad \textit{and} \quad \langle y, z\rangle_{\bar{A}^{-1}} := \langle y, \bar{A}^{-1}z\rangle_{L^2}.$$

We denote by $|\cdot|_{\bar{A}}$ *and* $|\cdot|_{\bar{A}^{-1}}$ *the corresponding norms on* Y_M.

The definition of $-\bar{A}$ as a coarse-graining of the second-order difference operator $-A$ suggests that it is a discrete version of the second derivative adapted to the spline space Y_M (see Lemma 2.5.3 below for a precise statement). Indeed, it turns out that the $\bar{A}$ and $\bar{A}^{-1}$ norms are equivalent to the H^1 and H^{-1} norm on Y_M, respectively.

Lemma 2.3.10. *There exists an integer K^* such that for all $K \geq K^*, M$ and all $y \in Y_M$,*

$$|y|_{\bar{A}} \simeq |y|_{H^1} \quad \textit{and} \quad |y|_{\bar{A}^{-1}} \simeq |y|_{H^{-1}}. \tag{2.3.5}$$

The proof of Lemma 2.3.10 is given in Section 2.6, where we gather and prove facts about splines. This is where we need the degree of the splines in Y_M to be at least $L \geq 1$.

We are now ready to introduce the mesoscopic dynamics.

Definition 2.3.11 (Mesoscopic dynamics)**.** *The mesoscopic dynamics η_t on Y_M is given by a solution of the ordinary differential equation*

$$\frac{d}{dt}\eta_t = -\bar{A}\nabla\bar{H}(\eta_t). \tag{2.3.6}$$

Remark 2.3.12 (Gradient flow structure of the mesoscopic dynamics)**.** *The mesoscopic dynamics may be viewed as a $\bar{A}^{-1}$-gradient flow for the energy functional $\bar{H}$,*

$$\frac{d}{dt}\eta_t = -\nabla_{\bar{A}^{-1}}\bar{H}(\eta_t), \tag{2.3.7}$$

where $\nabla_{\bar{A}^{-1}} := \bar{A}\nabla$ denotes the gradient operation wrt. to the $\bar{A}^{-1}$ inner product. The strong convexity of the energy functional $\bar{H}$ then implies the convergence of all trajectories irrespective of the starting point. More precisely, if η_t and $\tilde{\eta}_t$ are two solutions of (2.3.6)*, then*

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}|\eta_t - \tilde{\eta}_t|^2_{\bar{A}^{-1}} &= -\langle \eta_t - \tilde{\eta}_t, \nabla\bar{H}(\eta_t) - \nabla\bar{H}(\tilde{\eta}_t)\rangle_{L^2} \\ &\leq -2\lambda\,|\eta_t - \tilde{\eta}_t|^2_{L^2}. \end{aligned} \tag{2.3.8}$$

Before moving on, let us take a closer look at the gradient operation $\nabla_{\bar{A}^{-1}}$ on Y_M. It is related to the gradient operation $\nabla_{A^{-1}}$ on X_N by

$$\nabla_{A^{-1}}\xi(Px) = AP^t\bar{A}^{-1}\nabla_{\bar{A}^{-1}}\xi(Px) \quad \text{for all } x \in X_N \tag{2.3.9}$$

for any test function $\xi : Y_M \to \mathbb{R}$. It is easy to check that the operator norm of $AP^t\bar{A}^{-1}$ blows up if one projects onto piecewise constant functions or piecewise linear functions that

are $C^0(\mathbb{T})$ (i.e. splines of degree $L = 0$ or $L = 1$). However, we do get a good control if we project onto splines of degree $L = 2$ (see Lemma 2.4.14 below). This observation was the original motivation to consider the two-scale approach with the coarse-graining operator P given by Definition 2.3.1.

Remark 2.3.13. *In [GOVW09], the coarse-graining operator P was defined as the L^2-orthogonal projection onto piecewise constant functions and one worked around the problem that operator $AP^t\bar{A}^{-1}$ is unbounded by using a less straight-forward definition of $\bar{A}$ as $\bar{A}^{-1} := PA^{-1}NP^t$. That choice led to a sub-optimal error when comparing the microscopic to the mesoscopic evolution (see also Remark 2.4.6 below). Choosing splines of degree $L > 2$ does not improve the error derived with our method further.*

Now, we state the first ingredient of the two-scale approach.

Theorem 2.3.14 (Convergence of the microscopic to the mesoscopic dynamics)**.** *Under the same assumption as in Theorem* (2.2.18)*, let η denote the solution of the mesoscopic equation* (2.3.6)*. Then*

$$\sup_{0\le t\le T} \mathbb{E}|X_t - \eta_t|_{H^{-1}}^2 \lesssim \mathbb{E}|X_0 - \eta_0|_{H^{-1}}^2 + \frac{T}{K} + \frac{1}{M^2}\left(C_{\mathrm{Ent}} + 1\right), \tag{2.3.10}$$

where C_{Ent} is given by (2.2.20)*.*

We prove Theorem 2.3.14 in Section 2.4. The error term $\frac{T}{K}$ on the right hand side of (2.3.10) comes from comparing the stochastic microscopic dynamics to the deterministic mesoscopic dynamics. Its scaling corresponds to what one would expect from the decay of variance in the weak law of large numbers, if we had chosen to project onto piecewise constant functions, in which case y would be a vector whose entries are means of K weakly correlated random variables and η would be interpreted as the expectation of this vector.

Now, let us state the second ingredient in the two-scale approach.

Theorem 2.3.15 (Convergence of the mesoscopic to the macroscopic dynamics)**.** *Let η denote the solution of the mesoscopic dynamics* (2.3.6) *and let ζ denote the solution of the macroscopic dynamics* (2.2.18)*. Then*

$$\sup_{0\leq t\leq T}|\zeta_t-\eta_t|^2_{H^{-1}}+\int_0^T|\zeta_s-\eta_s|^2_{L^2}ds\lesssim|\zeta_0-\eta_0|^2_{H^{-1}}+\frac{T}{K}+\left(\frac{1}{K^2}+\frac{1}{M^2}\right)|\zeta_0|^2_{L^2}.$$

We prove Theorem 2.3.15 in Section 2.5. For the proof we adapt a standard method from numerical analysis, in which the mesoscopic evolution (2.3.6) is interpreted as a Galerkin approximation of the macroscopic evolution (2.2.18). The non-standard part of the argument is that when comparing (2.3.6) to (2.2.18) one gets two additional sources of errors. One source of error comes from approximating the Euclidean structure $\langle\cdot,\cdot\rangle_{H^{-1}}$ by the Euclidean structure $\langle\cdot,\bar{A}^{-1}\cdot\rangle_{L^2}$. The other source of error comes from approximating the gradient of the macroscopic free energy $\mathcal{H}$ by the gradient of the coarse-grained Hamiltonian $\bar{H}$.

We are now ready to give the proof of Theorem 2.2.14.

Proof of Theorem 2.2.14. We choose the initial condition of the mesoscopic dynamics η given by (2.3.6) to be $\eta_0=P\zeta_0$. Combining Theorem 2.3.14 and Theorem 2.3.15 yields the estimate

$$\begin{aligned}\sup_{0\leq t\leq T}\mathbb{E}|X_t-\zeta_t|^2_{H^{-1}}&\leq\sup_{0\leq t\leq T}2\mathbb{E}|X_t-\eta_t|^2_{H^{-1}}+\sup_{0\leq t\leq T}2|\eta_t-\zeta_t|^2_{H^{-1}}\\&\lesssim\mathbb{E}|X_0-P\zeta_0|^2_{H^{-1}}+|\zeta_0-P\zeta_0|^2_{H^{-1}}\\&\quad+\frac{T}{K}+\frac{1}{M^2}(C_{\mathrm{Ent}}+1)+\left(\frac{1}{K^2}+\frac{1}{M^2}\right)|\zeta_0|^2_{L^2}.\end{aligned}$$

Applying (2.3.1) and (2.3.2), and choosing $K=M^2$ yields the desired estimate (2.2.21). □

2.4 Convergence of microscopic dynamics to mesoscopic dynamics

The proof of Theorem 2.3.14 is quite complex. Before proceeding to the rigorous argument let us give some heuristics. Theorem 2.3.14 states that the stochastic microscopic evolution

given by the Kawasaki dynamics in (2.2.7), i.e.

$$dX_t = -\nabla_{A^{-1}} H(X_t)dt + \sqrt{2}dB_t^{A^{-1}},$$

is close in the $H^{-1}-$norm to the mesoscopic deterministic dynamics given by (2.3.7), i.e.

$$\frac{d}{dt}\eta = -\nabla_{\bar{A}^{-1}} \bar{H}(\eta). \tag{2.4.1}$$

The first observation needed is that because the $H^{-1}-$norm is a weak norm (i.e. it involves integration, see Definition 2.2.9) one can control the difference between X_t and the projected process PX_t in this norm (see Lemma 2.3.2). Hence, it suffices to show that the stochastic evolution

$$dPX_t = -P\nabla_{A^{-1}} H(X_t)dt + \sqrt{2}PdB_t^{A^{-1}} \tag{2.4.2}$$

is close to the deterministic mesoscopic dynamics (2.4.1). Because the operator P takes averages over blocks of size K, the noise term $\sqrt{2}PdB_t^{A^{-1}}$ of the projected Kawasaki dynamics (2.4.2) should vanish as $K \to \infty$ by the law of large numbers. It is left to show that the dynamics

$$\frac{d}{dt}PX_t = -P\nabla_{A^{-1}} H(X_t) \tag{2.4.3}$$

is close to the mesoscopic dynamics (2.4.1). By the coarse-graining relation (2.3.4) one sees that the mesoscopic dynamics (2.4.1)can be written as

$$\frac{d}{dt}\eta_t = -P\ \mathbb{E}_\mu \left[\nabla_{A^{-1}} H(x) \mid Px = \eta_t\right], \tag{2.4.4}$$

where the expectation is taken with respect to the canonical ensemble μ conditioned on the mesoscopic profile given by η_t. Let us recall that μ is also the equilibrium distribution of the Kawasaki dynamics (2.2.4) (see Lemma 2.2.6). The nearest-neighbor interaction of the spins mediated by the matrix A means the Kawasaki dynamics X_t equilibrates faster on smaller spatial scales, so we expect that the dynamics (2.4.3) and (2.4.4) are close if the blocks are much smaller than the overall system size N, in other words $\frac{K}{N} = \frac{1}{M} \to 0$. In the rigorous

argument, this fact will be quantified with the help of a uniform LSI which characterizes the speed of the convergence to equilibrium (see Theorem 2.4.12 below).

Let us turn now to the rigorous proof of Theorem 2.3.14. The first ingredient of the proof is an estimate of the second moment of X_t in L^2 norm, which controls the difference in H^{-1} norm between X_t and the projected dynamics PX_t by Lemma 2.3.2.

Proposition 2.4.1 (Second moment estimate)**.** *Under assumption* (2.2.20)*, the Kawasaki dynamics satisfies that*

$$\mathbb{E}|X_t|_{L^2}^2 \lesssim \frac{1}{N}\operatorname{Ent}(\rho(t)|\mu) + \mathbb{E}_\mu|x|_{L^2}^2 \lesssim C_{\mathrm{Ent}} + 1.$$

Remark 2.4.2. *This was shown as part of Proposition 24 in [GOVW09]. A quicker derivation suggested there is that the first inequality can be restated as*

$$W_2^2(\rho(t), \delta_0) \lesssim \operatorname{Ent}(\rho(t)|\mu) + W_2^2(\mu, \delta_0),$$

where W_2 denotes the L^2-Wasserstein distance and δ_0 is the Dirac measure supported at $0 \in X_N$. After first applying a triangle inequality for Wasserstein distance, this follows from Talagrand's transportation inequality (2.2.12)*. The latter is implied by the log-Sobolev inequality for μ* (2.2.11)*, which also implies a Poincaré inequality for μ and yields the second inequality.*

In light of the equivalence between the H^{-1} norm and $\bar{A}^{-1}$ norm (Lemma 2.3.10), it remains to control the difference between the projected microscopic dynamics PX_t and the mesoscopic dynamics η_t in $\bar{A}^{-1}$ norm. This is provided by the following estimate, which constitutes the main part of the proof of Theorem 2.3.14.

Theorem 2.4.3. *Under the same assumptions as in Theorem 2.3.14, there is an integer K^* and $\lambda > 0$ such that for all $K \geq K^*$ and any finite time $T > 0$ it holds*

$$\frac{1}{2}\sup_{0\leq t\leq T}\mathbb{E}|PX_t - \eta_t|_{\bar{A}^{-1}}^2 + \lambda\int_0^T \mathbb{E}|PX_t - \eta_t|_{L^2}^2\, dt \;\leq \mathbb{E}|PX_0 - \eta_0|_{\bar{A}^{-1}}^2 + \frac{2T}{K} + 2C\,\frac{C_{\mathrm{Ent}}}{M^2}. \tag{2.4.5}$$

Remark 2.4.4. *The estimate* (2.4.5) *also shows that the projected Kawasaki dynamics* (2.4.2) *is close to the mesoscopic dynamics* (2.4.1) *using a time-integrated strong norm. This is reminiscent of the well-known phenomenon of parabolic improvement in numerical analysis.*

Remark 2.4.5. *The universal constant* $0 < C < \infty$ *in Theorem 2.4.3 is given by* $C = \frac{\kappa^2\gamma}{4\sigma^2\lambda\alpha^2}$*, where the constants* κ*,* λ*,* γ*,* σ*, and* α *are given by:*

$\cdot\ \kappa := \|\operatorname{Hess} H\|$*, which is bounded independently of* N *by the assumption* (2.2.1)*,* (2.2.2) *and* (2.2.3)*;*

$\cdot\ 2\lambda$ *the lower bound on* $\operatorname{Hess}\bar{H}$ *as in Theorem 2.3.6 from below;*

$\cdot\ \alpha$ *is the constant of the logarithmic Sobolev inequality (LSI) from Theorem 2.4.12 from below;*

$\cdot\ \sigma$ *is the constant from Lemma 2.4.14 from below;*

$\cdot\ \gamma$ *the constant from Lemma 2.4.11 from below.*

Remark 2.4.6. *Theorem 2.4.3 was first derived in Dizdar's diploma thesis [Diz07]. We present below a more streamlined derivation that makes clear the underlying gradient flow structure. Theorem 2.4.3 should be compared with Theorem 8 in [GOVW09]. They arrive at a similar bound for the deviation from hydrodynamic behavior with an additional term scaling like* M^{-1}*. As mentioned before this additional error term occurs due to their choice of the coarse-graining operator* P *as the projection onto piecewise constant functions and the different definition of* $\bar{A}$*.*

Theorem 2.4.3 will be proven in Section 2.4.1. We finish this section with a quick derivation of Theorem 2.3.14 based on the ingredients above.

Proof of Theorem 2.3.14. Combining Proposition 2.4.1 and Theorem 2.4.3 together with Lemma 2.3.2 and Lemma 2.3.10,

$$\begin{aligned}\mathbb{E}|X_t - \eta_t|^2_{H^{-1}} &\leq 2\,\mathbb{E}|X_t - PX_t|^2_{H^{-1}} + 2\,\mathbb{E}|PX_t - \eta_t|^2_{H^{-1}} \\ &\lesssim \frac{1}{M^2}\mathbb{E}|X_t|^2_{L^2} + \mathbb{E}|PX_t - \eta_t|^2_{\bar{A}^{-1}}\end{aligned}$$

$$\lesssim \frac{1}{M^2}(C_{\mathrm{Ent}}+1)+\mathbb{E}|PX_0-\eta_0|^2_{\bar{A}^{-1}}+\frac{T}{K}+C\frac{C_{\mathrm{Ent}}}{M^2}$$
$$\lesssim \mathbb{E}|PX_0-\eta_0|^2_{H^{-1}}+\frac{T}{K}+\frac{1}{M^2}(C_{\mathrm{Ent}}+1).$$

This verifies the estimate (2.3.10). □

2.4.1 Proof of Theorem 2.4.3

Before we proceed, let us first introduce an conditioning for the Kawasaki dynamics X_t on the level of mesoscopic profiles, analogous to the disintegration of the canonical ensemble μ from Definition 2.3.4.

Definition 2.4.7 (Disintegration of the law ρ_t of the Kawasaki dynamics)**.** *The operator P induces a decomposition of the law ρ_t of the Kawasaki dynamics into a family of conditional measures $\rho_t(dx|y) := \rho_t(dx|Px=y)$ on the fibers $P^{-1}(y) \subset X_N$ and a marginal measure $\bar{\rho}_t(dy)$ on the image Y_M, in the sense that*

$$\int g(x)\rho_t(dx) = \int\int g(x)\rho_t(dx|y)\bar{\rho}_t(dy) \tag{2.4.6}$$

for any test function $g : X_N \to \mathbb{R}$.

We also need a decomposition of the microscopic observables in X_N into parts corresponding to mesoscopic profiles and microscopic fluctuations.

Definition 2.4.8 (Orthogonal decomposition of the state space X_N)**.** *The operator P induces an orthogonal decomposition of the state space X_N into a subspace corresponding to mesoscopic profiles, $X_N^{\perp} := \operatorname{im} NP^t$, and a subspace corresponding to microscopic fluctuations around these profiles, $X_N^{\parallel} := (\operatorname{im} NP^t)^{\perp} = \ker P \cap X_N$, as*

$$X_N \ni x = x_{\parallel} \oplus x_{\perp} \in X_N^{\parallel} \oplus X_N^{\perp}.$$

For a smooth function $f : X_N \to \mathbb{R}$, its gradient $\nabla_{\parallel}$ along $X_N^{\parallel}$ and its gradient $\nabla_{\perp}$ along $X_N^{\perp}$ are given by $\nabla_{\parallel} f = (\nabla f)_{\parallel}$ and $\nabla_{\perp} f = (\nabla f)_{\perp}$.

The starting point of the analysis is an equation of time evolution for the difference between the projected dynamics PX_t and the mesoscopic dynamics η_t, provided by the forward Kolmogorov equation (2.2.9) for the microscopic dynamics X_t

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\,\mathbb{E}\,|PX_t-\eta_t|^2_{\bar{A}^{-1}} &= \int\left(\frac{d}{dt}-\nabla H(x)\cdot\nabla_{A^{-1}}+\nabla\cdot\nabla_{A^{-1}}\right)\left(\frac{1}{2}|Px-\eta_t|^2_{\bar{A}^{-1}}\right)\rho_t(dx)\\ &= \mathbb{E}\,\langle-\frac{d}{dt}\eta_t, PX_t-\eta_t\rangle_{\bar{A}^{-1}}-\int\nabla H(x)\cdot AP^t\bar{A}^{-1}(Px-\eta_t)\,\rho_t(dx)\\ &\quad+\int\nabla\cdot(AP^t\bar{A}^{-1}(Px-\eta_t))\,\rho_t(dx),\end{aligned}\tag{2.4.7}$$

where we used the relation (2.3.9) between $\nabla_{A^{-1}}$ and $\nabla_{\bar{A}^{-1}}$. Let us first look at the last integral of the right hand side of (2.4.7). This is a purely *entropic* term coming from the projected Brownian motion $PdB_t^{A^{-1}}$ (see (2.4.2)), whose covariance matrix can be easily calculated to be $\dfrac{\mathrm{id}_Y}{N}$ wrt. the $\bar{A}^{-1}$ inner product. Indeed, the divergence term inside the integral equals

$$\mathrm{tr}_X(AP^t\bar{A}^{-1}P)=\mathrm{tr}_Y(PAP^t\bar{A}^{-1})=\mathrm{tr}_Y\left(\frac{\mathrm{id}_Y}{N}\right)=\frac{\dim Y}{N}.$$

This is a constant $\frac{M-1}{N}\approx\frac{1}{K}$ that scales like the variance of the average of K i.i.d. random variables. It accounts for the discrepancy that the Kawasaki dynamics (2.2.4) has noise whereas the mesoscopic dynamics (2.3.6) is deterministic.

Having dealt with the noise term, we apply disintegration of measure (2.4.6) to the law ρ_t in (2.4.7) and get

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\,\mathbb{E}|PX_t-\eta_t|^2_{\bar{A}^{-1}} &= \frac{M-1}{N}+\mathbb{E}\,\langle-\frac{d}{dt}\eta_t, PX_t-\eta_t\rangle_{\bar{A}^{-1}}\\ &\quad-\int\mathbb{E}_{\rho_t(dx|y)}\nabla H(x)\cdot AP^t\bar{A}^{-1}(y-\eta_t)\,\bar{\rho}_t(dy).\end{aligned}$$

As the expectations are now taken on the level of the mesoscopic profiles of the process X_t, it helps to mentally fix one realization $y=PX_t$. The first expectation on the right hand side is a purely *kinetic* term: it looks like the rate of convergence between the mesoscopic dynamics η_t and an alternate trajectory $\tilde{\eta}_t$ of it that happens to be in state y at time t (cf.

(2.3.8)), if there were a corresponding time derivative

$$\frac{d}{dt}\tilde{\eta}_t = -\nabla_{\bar{A}^{-1}}\bar{H}(\tilde{\eta}_t) = -\nabla_{\bar{A}^{-1}}\bar{H}(y).$$

Remarkably, the coarse-graining relation (2.3.4) between ∇H and $\nabla \bar{H}$ indicates that this missing mesoscopic energy gradient may be supplied by the average microscopic energy gradient, when the *microscopic fluctuations* of the dynamics X_t around the mesoscopic profile y has *reached stochastic equilibrium*:

$$\nabla \bar{H}(y) = \bar{A}^{-1}PA\, \mathbb{E}_{\mu(dx|y)} \nabla H(x). \tag{2.4.8}$$

Moreover, the operator $\bar{A}^{-1}PA : X_N \to Y_M$ in (2.4.8) is exactly the adjoint of the operator $AP^t\bar{A}^{-1} : Y_M \to X_N$ in (2.3.9). These observations lead to the rearranged equation

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\mathbb{E}|PX_t - \eta_t|^2_{\bar{A}^{-1}} = \frac{M-1}{N} - \mathbb{E}\,\langle \nabla \bar{H}(PX_t) - \nabla \bar{H}(\eta_t), PX_t - \eta_t\rangle_{L^2} \\ - \int (\mathbb{E}_{\rho_t(dx|y)}\nabla H(x) - \mathbb{E}_{\mu(dx|y)}\nabla H(x)) \cdot AP^t\bar{A}^{-1}(y - \eta_t)\, \bar{\rho}_t(dy).\end{aligned} \tag{2.4.9}$$

In this form, the first expectation on the right hand side of (2.4.9) now fully resembles (2.3.8) and can be estimated by the uniform convexity of $\bar{H}$ from Theorem 2.3.6:

$$\mathbb{E}\,\langle \nabla \bar{H}(PX_t) - \nabla \bar{H}(\eta_t), PX_t - \eta_t\rangle_{L^2} \geq 2\lambda\, \mathbb{E}\,|PX_t - \eta_t|^2_{L^2}. \tag{2.4.10}$$

It remains to estimate the second integral on the right hand side of (2.4.9). After taking the operator norm of $AP^t\bar{A}^{-1}$ into account (see Lemma 2.4.14 below), this integral quantifies how far away the conditional measure $\rho_t(dx|y)$ is from the conditional Gibbs measure $\mu(dx|y)$ through the *mean difference* $\mathbb{E}_{\rho_t(dx|y)}\nabla H(x) - \mathbb{E}_{\mu(dx|y)}\nabla H(x)$. The latter quantity turns out to be controlled by a bound on the operator norm of $\operatorname{Hess} H$ and a *log Sobolev inequality* (LSI) of the conditional Gibbs measure $\mu(dx|y)$.

Proposition 2.4.9 (Mean difference estimate)**.** *Let $\kappa := \|\operatorname{Hess} H\|$. Suppose $\mu(dx|y)$ given by* (2.2.8) *satisfies a LSI with constant $\varrho > 0$ in the sense of* (2.4.13) *below. Then we have:*

$$|\mathbb{E}_{\rho_t(dx|y)}\nabla H(x) - \mathbb{E}_{\mu(dx|y)}\nabla H(x)|^2 \;\leq\; \frac{\kappa^2}{\alpha^2}\int |\nabla_{||} \log f(t,x)|^2\, \rho_t(dx|y), \tag{2.4.11}$$

where $\nabla_{||}$ is the gradient along the subspace of fluctuations defined in Definition 2.4.8.

Remark 2.4.10. *This is a well-known result, e.g. contained in Lemma 22 in [GOVW09] in the form of a covariance estimate. The proof given there starts with using the Kantorovich-Rubinstein duality to bound the mean difference by the Wasserstein distance. By Theorem 1 in [OV00], the Wasserstein distance is bounded by the relative entropy provided a logarithmic Sobolev inequality is satisfied, which in turn bounds the relative entropy by the Fisher information.*

Moreover, the following discrete analogue of (2.3.1) allows us to pass from the Fisher information involving $\nabla_{||}$ in (2.4.11) to the full Fisher information for the Kawasaki dynamics in (2.2.10).

Lemma 2.4.11 (Penalization of fluctuations by spin-exchange)**.** *There exists constant $\gamma > 0$ such that for $x \in X_N$*

$$|x_{||}|^2 = |x - x_\perp|^2 \leq \frac{\gamma}{M^2} x \cdot Ax, \tag{2.4.12}$$

where $x_{||}, x_\perp$ denote the fluctuation and mesoscopic parts of x, respectively, as defined in Definition 2.4.8.

The proof of Lemma 2.4.11 is given in Section 2.6, where we gather and prove facts about splines. It remains to establish a uniform log Sobolev inequality:

Theorem 2.4.12 (Uniform LSI for $\mu(dx|y)$)**.** *The conditional measure $\mu(dx|y)$ given by* (2.2.8) *satisfies a LSI with constant $\varrho > 0$ uniform in the system size N and the mesoscopic profile y. More precisely, this means that for any nonnegative test function $g : X_N \to \mathbb{R}$ that satisfies $\int g(x)\mu(dx|y) = 1$, it holds that*

$$\mathrm{Ent}\left(g\mu(dx|y)|\mu(dx|y)\right) \leq \frac{1}{2\alpha} \int \frac{|\nabla_{||} g(x)|^2}{g(x)} \mu(dx|y), \tag{2.4.13}$$

where $\nabla_{||}$ is the gradient along the subspace of fluctuations defined in Definition 2.4.8.

Remark 2.4.13. *Theorem 2.4.12 should be compared to [GOVW09, Theorem 14], where a similar statement was deduced for the case $L = 0$ using the* two-scale criterion *for LSI (see*

Lemma 2.8.6 below). As with the proof of Theorem 2.3.6, the non-locality of the projection P onto the spline space Y_M creates difficulties with a direct application of the two-scale criterion, and we have to take a detour through the space of discontinuous Galerkin functions.

The proof of Theorem 2.4.12 is given in Section 2.7. To get an overall estimate for the integral involving the mean difference, we also need to control the operator norm of $AP^t\bar{A}^{-1} : Y_M \to X_N$, which measures the compatibility of projecting and taking second differences.

Lemma 2.4.14 (Interchanging second-order difference with coarse-graining)**.** *There exists a universal constant $\sigma > 0$ and an integer K^* such that for all $K \geq K^*, M$ and all $y \in Y_M$ it holds*

$$|ANP^t\bar{A}^{-1}y|_{L^2} \leq \frac{1}{\sigma}|y|_{L^2}. \tag{2.4.14}$$

The proof of Lemma 2.4.14 is given in Section 2.6, where we gather and prove facts about splines. We are now ready to prove Theorem 2.4.3.

Proof of Theorem 2.4.3. Applying the convexity estimate (2.4.10) to the evolution equation (2.4.9) and using Lemma 2.4.14 and Cauchy-Schwarz on the last integral yields that

$$\begin{aligned}&\frac{d}{dt}\frac{1}{2}\,\mathbb{E}|PX_t - \eta_t|^2_{\bar{A}^{-1}} + 2\lambda\,\mathbb{E}|PX_t - \eta_t|^2_{L^2}\\ &\leq \frac{1}{K} + \int \frac{1}{\sigma}|y - \eta_t|_{L^2}\left|\mathbb{E}_{\rho_t(dx|y)}\nabla H(x) - \mathbb{E}_{\mu(dx|y)}\nabla H(x)\right|_{L^2}\,\bar{\rho}_t(dy),\end{aligned} \tag{2.4.15}$$

where we accounted for different Euclidean structures on X_N. By Lemma 2.4.9, Lemma 2.4.11 and the observation (2.2.10), we have

$$\int |\mathbb{E}_{\rho_t(dx|y)}\nabla H(x) - \mathbb{E}_{\mu(dx|y)}\nabla H(x)|^2_{L^2}\,\bar{\rho}_t(dy) \leq -\frac{\kappa^2}{\alpha^2}\frac{\gamma}{M^2}\frac{1}{N}\frac{d}{dt}\operatorname{Ent}(\rho_t|\mu)\,.$$

Applying Young's inequality and using this estimate, the integral on the right hand side of (2.4.15) is bounded by

$$\lambda\,\mathbb{E}|PX_t - \eta_t|^2_{L^2} - \frac{1}{4\lambda}\frac{\kappa^2\gamma}{\sigma^2\alpha^2}\frac{1}{M^2}\frac{1}{N}\,\frac{d}{dt}\operatorname{Ent}(\rho_t|\mu)\,. \tag{2.4.16}$$

Putting the upper bound (2.4.16) back into (2.4.15) and integrating over the time interval $[0, T]$ yields the desired estimate (2.4.5). □

2.5 Convergence of mesoscopic dynamics to macroscopic dynamics

In this section we state the proof of Theorem 2.3.15. We need to show that the mesoscopic evolution (2.3.6)

$$\frac{d}{dt}\eta_t = -\nabla_{\bar{A}^{-1}}\bar{H}(\eta_t) = -\bar{A}\nabla\bar{H}(\eta_t)$$

converges to the macroscopic evolution (2.2.18)

$$\frac{\partial}{\partial t}\zeta_t = -\nabla_{H^{-1}}\mathcal{H}(\zeta_t) = \frac{\partial^2}{\partial\theta^2}\varphi'(\zeta_t).$$

Formally, this means that one has to exchange the operator $-\bar{A}$ with the operator ∂_θ^2 and the function $\nabla\bar{H}$ with the function φ'. The first exchange is plausible because $-\bar{A}$ is a coarse-grained version of the second-order difference operator $-A$. The second exchange is essentially the consequence of a (local) Cramér theorem: $\bar{H}$ is a coarse-graining of the Hamiltonian H with single-site potential ψ, while φ is the Cramér transform of the same ψ.

The proof of Theorem 2.3.15 is inspired by the Galerkin approximation scheme, a well-known method in numerical analysis. First, we need to show the macroscopic dynamics ζ_t is close to the projected dynamics $P\zeta_t$. Because the H^{-1} norm is a weak norm, this difference is controlled by the spline estimates in Lemma 2.3.2 together with the following a priori energy estimates.

Lemma 2.5.1. *Let ζ_t denote the macroscopic dynamics given by* (2.2.22)*. Then it holds that*

$$\sup_{0\le t\le T}|\zeta_t|_{L^2}^2 \lesssim |\zeta_0|_{L^2}^2 \quad \textit{and} \quad \int_0^\infty |\zeta_t|_{H^1}^2 dt \lesssim |\zeta_0|_{L^2}^2.$$

Lemma 2.5.1 plays a similar role as Proposition 2.4.1. It follows directly from integrating the dissipation (2.2.19) of the macroscopic free energy $\mathcal{H}$ wrt. the H^{-1} gradient flow structure,

after applying the estimates

$$\lambda|\zeta|_{L^2}^2 \le \mathcal{H}(\zeta) \le \Lambda|\zeta|_{L^2}^2 \quad \text{and} \quad \lambda|\zeta|_{H^1} \le |\varphi'(\zeta)|_{H^1} \le \Lambda|\zeta|_{H^1},$$

which are straightforward consequences of the convexity estimates of φ (see Lemma 2.2.11).

It remains to show that the mesoscopic dynamics η_t is close to the projected dynamics $P\zeta_t$. Because of Lemma 2.3.10, we will work with the more natural $\bar{A}^{-1}$ norm instead of H^{-1} norm. Differentiating in time, we have

$$\frac{d}{dt}\frac{1}{2}|\eta_t - P\zeta_t|_{\bar{A}^{-1}}^2 = \left\langle \frac{d}{dt}\eta_t - \frac{d}{dt}P\zeta_t, \eta_t - P\zeta_t \right\rangle_{\bar{A}^{-1}}.$$

This almost looks like the rate of convergence between the mesoscopic dynamics η_t and an alternate trajectory $\tilde{\eta}_t$ of it that happens to be in state $P\zeta_t$ at time t (cf. (2.3.8)), except that the time derivative of the projected macroscopic dynamics

$$\frac{d}{dt}P\zeta_t = P\frac{\partial\zeta}{\partial t} = P\frac{\partial^2}{\partial\theta^2}\varphi'(\zeta_t)$$

is not exactly the same as the time derivative of the hypothetical mesoscopic dynamics

$$\frac{d}{dt}\tilde{\eta}_t = -\nabla_{\bar{A}^{-1}}\bar{H}(\tilde{\eta}_t) = -\bar{A}\nabla\bar{H}(P\zeta_t).$$

These observations lead to the rearranged equation

$$\begin{aligned}
\frac{d}{dt}\frac{1}{2}|\eta_t - P\zeta_t|_{\bar{A}^{-1}}^2 &= \left\langle \frac{d}{dt}\eta_t + \nabla_{\bar{A}^{-1}}\bar{H}(P\zeta_t), \eta_t - P\zeta_t \right\rangle_{\bar{A}^{-1}} + \langle \bar{A}\varphi'(\zeta_t) - \nabla_{\bar{A}^{-1}}\bar{H}(P\zeta_t), \eta - P\zeta_t\rangle_{\bar{A}^{-1}} \\
&\quad + \left\langle -\frac{d}{dt}P\zeta_t - \bar{A}\varphi'(\zeta_t), \eta_t - P\zeta_t \right\rangle_{\bar{A}^{-1}} \\
&= -\langle \nabla\bar{H}(\eta_t) - \nabla\bar{H}(P\zeta_t), \eta - P\zeta_t\rangle_{L^2} + \langle \varphi'(\zeta_t) - \nabla\bar{H}(P\zeta_t), \eta - P\zeta_t\rangle_{L^2} \\
&\quad + \langle \varphi'(\zeta_t), (-\partial_\theta^2\bar{A}^{-1} - \mathrm{id})(\eta_t - P\zeta_t)\rangle_{L^2}. \qquad (2.5.1)
\end{aligned}$$

The first term in (2.5.1) now fully resembles (2.3.8) and can be estimated by the uniform strong convexity of $\bar{H}$ (see Theorem 2.3.6):

$$\langle \nabla\bar{H}(P\zeta_t) - \nabla\bar{H}(\eta_t), \eta_t - P\zeta_t\rangle_{L^2} \le -\lambda|\eta_t - P\zeta_t|_{L^2}^2. \qquad (2.5.2)$$

The second term in (2.5.1) is small because the gradient of the coarse-grained Hamiltonian $\bar{H}$ is a good approximation of the gradient of the macroscopic free energy $\mathcal{H}$.

Lemma 2.5.2 (Closeness of $\nabla\bar{H}$ and $\nabla\mathcal{H}$)**.** *There is an integer K^* such that if $K \geq K^*$ then it holds for all $\zeta \in L^2_0(\mathbb{T})$*

$$\left|\nabla\bar{H}(P\zeta) - \nabla\mathcal{H}(\zeta)\right|_{L^2} \lesssim \left(\frac{1}{K} + \frac{1}{M}\right)|\zeta|_{H^1} + \frac{1}{K^{\frac{1}{2}}}. \tag{2.5.3}$$

We prove Lemma 2.5.2 in Section 2.9. The last term in (2.5.1) is controlled by the following error estimates for exchanging the coarse-grained second-order difference operator $-\bar{A}$ and the second derivative ∂_θ^2.

Lemma 2.5.3 (Discrepancy between $-\bar{A}$ and ∂_θ^2)**.** *There exists an integer K^* such that for all $K \geq K^*, M$ and all $y, \tilde{y} \in Y_M$,*

$$|-\partial_\theta^2\bar{A}^{-1}y|_{L^2} \lesssim |y|_{L^2}, \tag{2.5.4}$$

$$|\langle -\partial_\theta^2\bar{A}^{-1}y, \tilde{y}\rangle_{L^2} - \langle y, \tilde{y}\rangle_{L^2}| \lesssim \frac{1}{K}|y|_{H^{-1}}|\tilde{y}|_{H^1}. \tag{2.5.5}$$

The proof of Lemma 2.5.3 is given in Section 2.6, where we gather and prove facts about splines. The error estimate (2.5.4) is closely related to the error estimate (2.4.14). We are now ready to prove Theorem 2.3.15.

Proof of Theorem 2.3.15. We first bound $P\zeta_t - \zeta_t$. By Lemma 2.3.2 and Lemma 2.5.1,

$$\sup_{0\leq t\leq T}|P\zeta_t - \zeta_t|^2_{H^{-1}} \lesssim \sup_{0\leq t\leq T}\frac{1}{M^2}|\zeta_t|^2_{L^2} \lesssim \frac{1}{M^2}|\zeta_0|^2_{L^2},$$
$$\int_0^T |P\zeta_t - \zeta_t|^2_{L^2}dt \lesssim \int_0^T \frac{1}{M^2}|\zeta_t|^2_{H_1}dt \lesssim \frac{1}{M^2}|\zeta_0|^2_{L^2}.$$

We now bound $\eta_t - P\zeta_t$. By Lemma 2.5.2, the second term in (2.5.1) is estimated as

$$\langle\varphi'(\zeta_t) - \nabla\bar{H}(P\zeta_t), \eta_t - P\zeta_t\rangle_{L^2} \lesssim \left(\left(\frac{1}{K} + \frac{1}{M}\right)|\zeta_t|_{H^1} + \frac{1}{K^{\frac{1}{2}}}\right)|\eta_t - P\zeta_t|_{L^2}. \tag{2.5.6}$$

By Lemma 2.5.3 and Lemma 2.3.2, the last term in (2.5.1) is estimated as

$$\langle\varphi'(\zeta_t), (-\partial_\theta^2\bar{A}^{-1} - \mathrm{id})(\eta_t - P\zeta_t)\rangle_{L^2}$$

$$
\begin{aligned}
&= \langle \varphi'(\zeta_t) - P\varphi'(\zeta_t), -\partial_\theta^2 \bar{A}^{-1}(\eta_t - P\zeta_t)\rangle_{L^2} + \langle P\varphi'(\zeta_t), (-\partial_\theta^2 \bar{A}^{-1} - \mathrm{id})(\eta_t - P\zeta_t)\rangle_{L^2} \\
&\lesssim |\varphi'(\zeta_t) - P\varphi'(\zeta_t)|_{L^2} |\eta_t - P\zeta_t|_{L^2} + \frac{1}{K} |P\varphi'(\zeta_t)|_{H^1} |\eta_t - P\zeta_t|_{H^{-1}} \\
&\lesssim \frac{1}{M} |\varphi'(\zeta_t)|_{H^1} |\eta_t - P\zeta_t|_{L^2} + \frac{1}{K} |\varphi'(\zeta_t)|_{H^1} |\eta_t - P\zeta_t|_{L^2}. \qquad (2.5.7)
\end{aligned}
$$

Combining the estimates (2.5.2), (2.5.6), and (2.5.7) for equation (2.5.1) and applying Young's inequality yields that

$$
\frac{d}{dt} \frac{1}{2} |\eta_t - P\zeta_t|^2_{\bar{A}^{-1}} + \frac{\lambda}{2} |\eta_t - P\zeta_t|^2_{L^2} \lesssim \frac{1}{K} + \left(\frac{1}{K^2} + \frac{1}{M^2} \right) \left(|\zeta_t|^2_{H^1} + |\varphi'(\zeta_t)|^2_{H^1} \right).
$$

Integrating in time from 0 to T, applying the energy estimates in Lemma 2.5.1, and exchanging $\bar{A}^{-1}$ norm with H^{-1} norm (see Lemma 2.3.10), we get

$$
\sup_{0 \le t \le T} \frac{1}{2} |\eta_t - P\zeta_t|^2_{H^{-1}} + \frac{\lambda}{2} \int_0^T |\eta_t - P\zeta_t|^2_{L^2} dt \lesssim \frac{T}{K} + \left(\frac{1}{K^2} + \frac{1}{M^2} \right) |\zeta_0|^2_{L^2}.
$$

Combining the estimates for $\eta_t - P\zeta_t$ and $P\zeta_t - \zeta_t$ yields Theorem 2.3.15. □

2.6 Properties of spline approximations

In this section we gather and prove the facts about splines $y \in Y_M$ needed in this work, namely Lemma 2.3.2, Lemma 2.3.10, Lemma 2.4.11, Lemma 2.4.14, Lemma 2.5.3.

2.6.1 Penalization of fluctuations around spline profiles

In this subsection we prove those auxiliary results that show fluctuations around spline profiles are penalized when measured in a weak norm, namely Lemma 2.3.2 and Lemma 2.4.11. We need two auxiliary ingredients. The first is an inverse Sobolev inequality on the space Y_M.

Lemma 2.6.1 (Inverse Sobolev inequality)**.** *For all $y \in Y_M$ holds*

$$
|y|_{H^2} \lesssim M |y|_{H^1} \lesssim M^2 |y|_{L^2}. \qquad (2.6.1)
$$

It is clear that the estimate (2.6.1) holds. The spline spaces Y_M are finite dimensional and norms on finite-dimensional vector spaces are equivalent. The factors M and M^2 comes from a scaling argument i.e. $\frac{1}{M}$ is the only internal length scale. We omit the details of the proof, which consists of a straightforward calculation. The second auxiliary ingredient is a nice basis for the spline functions that forms a "partition of unity" on the torus $\mathbb{T}$.

Definition 2.6.2 (B-spline functions)**.** *The B-spline functions of Y_M are given by*

$$B_j(\theta) = \begin{cases} \frac{M^2}{2}(\theta - \frac{j-2}{M})^2 & for \theta \in [\frac{j-2}{M}, \frac{j-1}{M}), \\ \frac{3}{4} - M^2(\theta - \frac{2j-1}{2M})^2 & for \theta \in [\frac{j-1}{M}, \frac{j}{M}), \\ \frac{M^2}{2}(\theta - \frac{j+1}{M})^2 & for \theta \in [\frac{j}{M}, \frac{j+1}{M}), \\ 0 & else. \end{cases} \tag{2.6.2}$$

Proof of Lemma 2.3.2. The proof of (2.3.1) is based on the following spline interpolation: for $\zeta \in L^2(\mathbb{T})$, we define $I\zeta \in Y_M$ as

$$I\zeta(\theta) = \sum_{j=1}^{M} \zeta\left(\frac{2j-1}{2M}\right) B_j(\theta),$$

where $B_j \in Y_M$ is the B-spline basis defined in (2.6.2). We claim that

$$|\zeta - P\zeta|_{L^2} \leq |\zeta - I\zeta|_{L^2} \lesssim \frac{1}{M} |\zeta|_{H^1}, \tag{2.6.3}$$

which establishes the second estimate of (2.3.1), from which the rest follows by duality. To verify (2.6.3), note the first inequality is simply due to the fact that $P\zeta$ is the best L^2 approximation of ζ in Y_M. The second estimate of (2.6.3) is well known in the literature on B-splines (see for example [DL93]). For the convenience of the reader we give a short proof of this fact. Using the fact that the B_j have small support and sum to 1, we obtain for $\theta \in \left(\frac{m-1}{M}, \frac{m}{M}\right)$:

$$\zeta(\theta) - I\zeta(\theta) \overset{(2.6.2)}{=} \sum_{j=0}^{2} \left(\zeta(\theta) - \zeta\left(\frac{2m-3+2j}{2M}\right)\right) B_{m-1+j}(\theta).$$

We use Young's inequality, the Fundamental Theorem of Calculus and the Cauchy-Schwarz inequality to deduce an estimate:

$$\int_{\frac{m-1}{M}}^{\frac{m}{M}} |\zeta(\theta) - I\zeta(\theta)|^2 \, d\theta$$

$$
\begin{aligned}
&\leq 3\sum_{j=0}^{2}\int_{\frac{m-1}{M}}^{\frac{m}{M}}\left(\zeta(\theta)-\zeta\left(\frac{2m-3+2j}{2M}\right)\right)B_{m-1+j}(\theta)\\
&\leq \int_{\frac{m-1}{M}}^{\frac{m}{M}} 3\,\frac{2}{M}\left(\int_{\frac{2m-3}{2M}}^{\frac{2m+1}{2M}}|\zeta'(\widetilde{\theta})|^2 d\widetilde{\theta}\right)\,(B_{m-1}^2(\theta)+B_m^2(\theta)+B_{m+1}^2(\theta))\,d\theta\\
&\leq \frac{6}{M^2}\left|\sum_{j=1}^{M}B_j^2\right|_{L^\infty}\left(\int_{\frac{2m-3}{2M}}^{\frac{2m+1}{2M}}|\zeta'(\widetilde{\theta})|^2 d\widetilde{\theta}\right).
\end{aligned}
$$

Summing over $m=1,...,M$ yields the second estimate of (2.6.3).

Let us now turn to the verification of (2.3.2). Again, by duality it suffices to show the first inequality only, which follows at once from the next two estimates

$$
|\zeta - I\zeta|_{H^1} \lesssim |\zeta|_{H^1} \qquad \text{and} \qquad |P\zeta - I\zeta|_{H^1} \lesssim |\zeta|_{H^1}. \tag{2.6.4}
$$

The first estimate of (2.6.4) can be deduced with similar calculations as were used for verifying the second estimate of (2.6.3). We leave the details as an exercise. The second estimate of (2.6.4) follows from the inverse Sobolev inequality on Y_M and the estimates of (2.6.3):

$$
|P\zeta - I\zeta|_{H^1} \overset{(2.6.1)}{\lesssim} M|P\zeta - I\zeta|_{L^2} \leq M(|P\zeta-\zeta|_{L^2}+|\zeta-I\zeta|_{L^2}) \overset{(2.6.3)}{\lesssim} |\zeta|_{H^1}.
$$

□

Proof of Lemma 2.4.11. The argument is essentially analogous to the one given for the previous lemma. In the main step, we will use the discrete Poincaré inequality (2.2.6) and define a spline interpolation. Observe that

$$
|x - x_\perp|^2 \leq |x - NP^t y|^2 \quad \text{for all } y \in Y_M
$$

since $\operatorname{im} NP^t = \ker P^\perp$. Let $\beta^j := NP^t B_j$ be the L^2-orthogonal projection onto X_N of the B-spline function B_j given by (2.6.2). For each $x \in X_N$, define a projected spline interpolation

$$
I(x) := \sum_{j=1}^{M}\left(\frac{1}{3K}\sum_{i=(j-2)K+1}^{(j+1)K} x_i\right)\beta^j =: \sum_{j=1}^{M}\hat{x}_j\beta^j.
$$

The properties that $0 \leq B_j \leq 1$, B_j is supported on $[\frac{j-2}{M}, \frac{j+1}{M}]$, and $\sum_{j=1}^{M} B_j = 1$ imply the corresponding properties for β^j: $0 \leq \beta_n^j \leq 1$ for each $n \in [N]$, β^j is supported on

$n \in ((j-2)K, (j+1)K]$, and $\sum_{j=1}^{M} \beta_n^j = 1$. Hence,

$$(x_n - I(x)_n)^2 = \left(\sum_{j=m-1}^{m+1} (x_n - \hat{x}_j)\beta_n^j \right)^2 \leq \sum_{j=m-1}^{m+1} (x_n - \hat{x}_j)^2$$

for $n \in ((m-1)K, mK]$. Summing this inequality over n leads to

$$\begin{aligned}
\sum_{n=1}^{N} (x_n - I(x)_n)^2 &\leq \sum_{m=1}^{M} \sum_{n=(m-2)K+1}^{(m+1)K} (x_n - \hat{x}_m)^2 \\
&\overset{(2.2.6)}{\leq} \sum_{m=1}^{M} C(3K)^2 \sum_{n=(m-2)K+1}^{(m+1)K} (x_n - x_{n-1})^2 \\
&= 3^3 C K^2 \sum_{n=1}^{N} (x_n - x_{n-1})^2 \\
&\overset{(2.2.5)}{=} 3^3 C \frac{K^2}{N^2} x \cdot Ax.
\end{aligned}$$

Thus (2.4.12) holds with $\gamma = 3^3 C$, where C is the constant in (2.2.6). □

2.6.2 Spline approximations involving the operator $\bar{A}$

In this section we prove those auxiliary results which make precise the idea that the operator $-\bar{A}$, due to its definition as a coarse-graining of the second-order difference operator $-A$, should be a discrete version of the second derivative adapted to the spline space Y_M. We begin with showing that the H^1 inner product on Y_M is close to the inner product induced by the positive definite operator $\bar{A}$.

Lemma 2.6.3. *There exists an integer K^* such that for all $K \geq K^*, M$ and all $y, \tilde{y} \in Y_M$*

$$|\langle \tilde{y}, \bar{A}y \rangle_{L^2} - \langle \tilde{y}, y \rangle_{H^1}| \lesssim \frac{1}{N} \left(|\tilde{y}|_{H^1} |y|_{H^2} + |\tilde{y}|_{H^2} |y|_{H^1} \right) \tag{2.6.5}$$

$$\lesssim \frac{M}{N} |\tilde{y}|_{H^1} |y|_{H^1}. \tag{2.6.6}$$

This leads to a quick proof of Lemma 2.3.10 by a duality argument.

Proof of Lemma 2.3.10. Let $z \in H^1(\mathbb{T})$ be arbitrary, then we have

$$\langle y, z \rangle_{L^2} = \langle y, Pz \rangle_{L^2} \leq |y|_{\bar{A}^{-1}} |Pz|_{\bar{A}} \overset{(2.3.5)}{\lesssim} |y|_{\bar{A}^{-1}} |Pz|_{H^1} \overset{(2.3.2)}{\lesssim} |y|_{\bar{A}^{-1}} |z|_{H^1},$$

which implies $|y|_{H^{-1}} \lesssim |y|_{\bar{A}^{-1}}$. To show the opposite inequality, let $w \in Y_M$ be arbitrary, then we have

$$\langle y, \bar{A}^{-1} w \rangle_{L^2} \leq |y|_{H^{-1}} |\bar{A}^{-1} w|_{H^1} \overset{(2.3.5)}{\lesssim} |y|_{H^{-1}} |\bar{A}^{-1} w|_{\bar{A}} = |y|_{H^{-1}} |w|_{\bar{A}^{-1}},$$

which implies $|y|_{\bar{A}^{-1}} \lesssim |y|_{H^{-1}}$. □

To prove Lemma 2.6.3, we will use the following simple facts about finite difference operators.

Lemma 2.6.4 (Finite Difference Operators)**.** *Let $Q : L^2(\mathbb{T}) \to X_N$ be the L^2-orthogonal projection onto X_N (cf. (2.2.13)), then $NP^t = Q$ on Y_M. Let D be a rescaled $N \times N$ forward difference matrix, given by $(Dx)_i = N(x_{i+1} - x_i)$, then $A = D^t D$, where D^t denotes the transpose of D. Let ∂_θ^h be the difference quotient with step size h, given by $\partial_\theta^h y(\theta) = \frac{1}{h}\left(y\left(\theta + h\right) - y\left(\theta\right)\right)$, then*

$$DQ = Q\partial_\theta^{\frac{1}{N}}, \quad D^t Q = -Q\partial_\theta^{-\frac{1}{N}}. \tag{2.6.7}$$

Proof of Lemma 2.6.3. We will prove the estimate (2.6.5), and then the estimate (2.6.6) follows from (2.6.5) by an application of the inverse Sobolev inequality (2.6.1). For $\tilde{y}, y \in Y_M$

$$\begin{aligned} \langle \tilde{y}, \bar{A} y \rangle_{L^2} &= \langle NP^t \tilde{y}, A(NP^t y) \rangle_{L^2} \\ &= \langle DQ\tilde{y}, DQy \rangle_{L^2} \overset{(2.6.7)}{=} \langle \partial_\theta^{\frac{1}{N}} \tilde{y}, Q\partial_\theta^{\frac{1}{N}} y \rangle_{L^2}. \end{aligned} \tag{2.6.8}$$

Using (2.6.8) we get that

$$\begin{aligned} \langle \tilde{y}, \bar{A} y \rangle_{L^2} - \langle \tilde{y}, y \rangle_{H^1} &= \langle \partial_\theta^{\frac{1}{N}} \tilde{y}, Q\partial_\theta^{\frac{1}{N}} y \rangle_{L^2} - \langle \partial_\theta \tilde{y}, \partial_\theta y \rangle_{L^2} \\ &= \langle \partial_\theta^{\frac{1}{N}} \tilde{y} - \partial_\theta \tilde{y}, Q\partial_\theta^{\frac{1}{N}} y \rangle_{L^2} + \langle \partial_\theta \tilde{y}, Q(\partial_\theta^{\frac{1}{N}} y - \partial_\theta y) \rangle_{L^2} \\ &\quad + \langle \partial_\theta \tilde{y}, (Q - \mathrm{id})\partial_\theta y \rangle_{L^2}, \end{aligned}$$

which implies

$$\left| \langle \tilde{y}, \bar{A} y \rangle_{L^2} - \langle \tilde{y}, y \rangle_{H^1} \right| \leq |\partial_\theta^{\frac{1}{N}} \tilde{y} - \partial_\theta \tilde{y}|_{L^2} |\partial_\theta^{\frac{1}{N}} y|_{L^2} + |\partial_\theta \tilde{y}|_{L^2} |\partial_\theta^{\frac{1}{N}} y - \partial_\theta y|_{L^2}$$

$$+ |\partial_\theta \tilde{y}|_{L^2} |(Q - \mathrm{id})\partial_\theta y|_{L^2}.$$

This yields estimate (2.6.5) once we establish the following estimates:

$$|\partial_\theta^{\frac{1}{N}} y|_{L^2} \leq |\partial_\theta y|_{L^2}, \tag{2.6.9}$$

$$|\partial_\theta^{\frac{1}{N}} y - \partial_\theta y|_{L^2} \leq \frac{1}{N} |\partial_\theta^2 y|_{L^2}, \tag{2.6.10}$$

$$|(Q - \mathrm{id})\partial_\theta y|_{L^2} \lesssim \frac{1}{N} |\partial_\theta^2 y|_{L^2}. \tag{2.6.11}$$

Argument for (2.6.9): Observe that

$$\partial_\theta^{\frac{1}{N}} y(\theta) = N \int_\theta^{\theta + \frac{1}{N}} \partial_\theta y(s) ds. \tag{2.6.12}$$

By Cauchy-Schwarz,

$$\int_0^1 \left| \partial_\theta^{\frac{1}{N}} y \right|^2 d\theta \leq \int_0^1 \frac{N^2}{N} \int_\theta^{\theta + \frac{1}{N}} |\partial_\theta y(s)|^2 ds d\theta = \int_0^1 |\partial_\theta y(s)|^2 ds.$$

Argument for (2.6.10): By mean value theorem, for some $\tilde{s} \in (\theta, \theta + \frac{1}{N})$,

$$\left| \partial_\theta^{\frac{1}{N}} y - \partial_\theta y \right| = |\partial_\theta y(\tilde{s}) - \partial_\theta y(\theta)| = \left| \int_\theta^{\tilde{s}} \partial_\theta^2 y(s) ds \right| \leq \int_\theta^{\theta + \frac{1}{N}} |\partial_\theta^2 y(s)| ds.$$

By Cauchy-Schwarz,

$$\int_0^1 \left| \partial_\theta^{\frac{1}{N}} y - \partial_\theta y \right|^2 d\theta \leq \int_0^1 \frac{1}{N} \int_\theta^{\theta + \frac{1}{N}} \left| \partial_\theta^2 y(s) \right|^2 ds d\theta = \frac{1}{N^2} \int_0^1 \left| \partial_\theta^2 y(s) \right|^2 ds.$$

Argument for (2.6.11): This follows from applying a Poincaré inequality on each of the intervals $(\frac{i-1}{N}, \frac{i}{N})$ to the function $\partial_\theta y$.

□

Proof of Lemma 2.4.14. It suffices to show

$$\sigma |ANP^t|_{L^2} \leq |PANP^t y|_{L^2}. \tag{2.6.13}$$

Let $z = ANP^t y$. By Lemma 2.6.4,

$$z = D^t DQy = D^t Q \partial_\theta^{\frac{1}{N}} y = -Q \partial_\theta^{-\frac{1}{N}} \partial_\theta^{\frac{1}{N}} y.$$

Using the integral representation formula (2.6.12), we compute

$$\begin{aligned} z_n &= -N \int_{\frac{n-1}{N}}^{\frac{n}{N}} \partial_\theta^{-\frac{1}{N}} \partial_\theta^{\frac{1}{N}} y(\theta) d\theta \\ &= -N \int_{\frac{n-1}{N}}^{\frac{n}{N}} N \int_{\theta-\frac{1}{N}}^{\theta} \partial_\theta \partial_\theta^{\frac{1}{N}} y(s) ds d\theta \\ &= -N \int_{\frac{n-1}{N}}^{\frac{n}{N}} N \int_{\theta-\frac{1}{N}}^{\theta} N \int_{s}^{s+\frac{1}{N}} \partial_\theta^2 y(t) dt ds d\theta \\ &= -\int_0^1 \int_0^1 \int_0^1 \partial_\theta^2 y \left(\frac{n-u-v+w}{N} \right) dw dv du. \end{aligned} \tag{2.6.14}$$

By definition of the spline space, on each interval $(\frac{j-1}{M}, \frac{j}{M})$,

$$\partial_\theta^2 y(\theta) = \alpha_j \tag{2.6.15}$$

for some constants $\alpha_j \in \mathbb{R}$. Then evaluating the integral (2.6.14) gives

$$z_i = -\alpha_j, \quad \text{for } i = (j-1)K + 2, ..., jK - 1, \tag{2.6.16}$$

$$z_{(j-1)K+1} = -\alpha_j + \frac{1}{6}(\alpha_j - \alpha_{j-1}), \;\; z_{jK} = \alpha_j + \frac{1}{6}(\alpha_j - \alpha_{j+1}). \tag{2.6.17}$$

Thus the function $z = ANP^t y$ is almost piecewise constant on blocks of K. This motivates us to define a spline interpolation for z,

$$I(z) := \sum_{j=1}^{M} -\alpha_j \, B_j,$$

where B_j are the B-spline basis of Y_M. We will show that there is a universal constant $\sigma > 0$, such that

$$\frac{\langle z, I(z) \rangle_{L^2}}{|z|_{L^2} |I(z)|_{L^2}} \geq \sigma. \tag{2.6.18}$$

This implies (2.6.13) by

$$|Pz|_{L^2} \, |I(z)|_{L^2} \geq \langle Pz, I(z) \rangle_{L^2} = \langle z, I(z) \rangle_{L^2} \geq \sigma |z|_{L^2} |I(z)|_{L^2}.$$

Argument for (2.6.18): It follows from (2.6.16) and (2.6.17) that

$$|z|_{L^2}^2 = \frac{1}{N} |z|^2 \leq \frac{K}{N} \sum_{j=1}^{M} \alpha_j^2.$$

Using (2.6.2), we can compute $|I(z)|_{L^2}$:

$$|I(z)|_{L^2}^2 = \sum_{j=1}^{M} \alpha_j \alpha_k \langle B_j, B_k \rangle_{L^2} = \frac{1}{M} \langle \alpha, B\alpha \rangle_{\mathbb{R}^M} \leq \frac{1}{M} \sum_{j=1}^{M} \alpha_j^2,$$

where B is the symmetric matrix

$$B_{jk} = \frac{11}{20}\delta_{j=k} + \frac{13}{60}\delta_{|j-k|=1} + \frac{1}{120}\delta_{|j-k|=2}$$

and the last inequality follows from $\|B\| \leq 1$. Finally, we compute $\langle z, I(z) \rangle_{L^2}$:

$$\begin{aligned}
\langle z, I(z) \rangle_{L^2} &= \sum_{j=1}^{M} -\alpha_j \langle z, B_j \rangle_{L^2} \\
&= \sum_{j,k=1}^{M} \alpha_j \alpha_k \int_{\frac{k-1}{M}}^{\frac{k}{M}} B_j - \sum_{j,k=1}^{M} \frac{1}{6} \alpha_j (\alpha_k - \alpha_{k-1}) \int_{\frac{k-1}{M}}^{\frac{k-1}{M} + \frac{1}{N}} B_j \\
&\quad - \sum_{j,k=1}^{M} \frac{1}{6} \alpha_j (\alpha_k - \alpha_{k+1}) \int_{\frac{k}{M} - \frac{1}{N}}^{\frac{k}{M}} B_j \\
&= \frac{1}{M} \langle \alpha, E\alpha \rangle_{\mathbb{R}^M} + O\left(\frac{1}{N}\right) |\alpha|_{\mathbb{R}^M}^2 \geq \frac{c}{M} \sum_{j=1}^{M} \alpha_j^2,
\end{aligned}$$

where E is the symmetric matrix

$$E_{jk} = \frac{2}{3}\delta_{j=k} + \frac{1}{6}\delta_{|j-k|=1}$$

and the last inequality follows from the strict diagonal dominance of E, once K is large enough, for some universal constant $c > 0$. Putting everything together, we arrive at (2.6.18):

$$\frac{\langle z, I(z) \rangle_{L^2}^2}{|z|_{L^2}^2 \, |I(z)|_{L^2}^2} \geq \frac{c^2}{M^2} M \frac{N}{CK} = \frac{c^2}{C}.$$

□

Proof of Lemma 2.5.3. Argument for (2.5.4): By (2.6.15), (2.6.16), and (2.6.17),

$$\begin{aligned}
|(-\partial_\theta^2 y) - ANP^t y|_{L^2}^2 &= \sum_{j=1}^{M} \frac{1}{N} \frac{1}{6^2} (\alpha_{j-1} - \alpha_j)^2 + \frac{1}{N} \frac{1}{6^2} (\alpha_{j+1} - \alpha_j)^2 \\
&\lesssim \sum_{j=1}^{M} \frac{1}{N} |\alpha_j|^2 = \frac{1}{K} |-\partial_\theta^2 y|_{L^2}^2.
\end{aligned}$$

It follows that

$$|-\partial_\theta^2 y|_{L^2} \leq \left(1+O\left(K^{-\frac{1}{2}}\right)\right)|ANP^t y|_{L^2} \overset{(2.6.13)}{\lesssim} |PANP^t y|_{L^2}.$$

Argument for (2.5.5): we apply Lemma 2.6.3 to $\bar{A}^{-1}y, \tilde{y}$:

$$\begin{aligned}|\langle \tilde{y}, y\rangle_{L^2} - \langle \tilde{y}, -\partial_\theta^2 \bar{A}^{-1} y\rangle_{L^2}| &= |\langle \tilde{y}, \bar{A}(\bar{A}^{-1}y)\rangle_{L^2} - \langle \tilde{y}, \bar{A}^{-1}y\rangle_{H^1}| \\ &\overset{(2.6.5)}{\lesssim} \frac{1}{K}|\bar{A}^{-1}y|_{H^1}|\tilde{y}|_{H^1} \overset{(2.3.5)}{\lesssim} \frac{1}{K}|y|_{H^{-1}}|\tilde{y}|_{H^1}.\end{aligned}$$

□

2.7 Uniform strong convexity of the coarse-grained Hamiltonian

This section is devoted to the proof of Theorem 2.3.6. As mentioned in Remark 2.3.7, because the spline functions in Y_M are non-local due to continuity requirements (for $L \geq 1$), we will instead work with an intermediate space $Y_M^{DG} \supset Y_M$ of *discontinuous Galerkin functions*, prove the analogue of Theorem 2.3.6 for Y_M^{DG} and then transfer the result to Y_M.

Definition 2.7.1 (The space of discontinuous Galerkin functions Y_M^{DG})**.** *The space of discontinuous Galerkin functions of degree $L \in \mathbb{N}$ is defined as*

$$Y_M^{DG} := \left\{ y \in L^2(\mathbb{T}) | \, \forall m \in [M] : y|_{\left(\frac{m-1}{M}, \frac{m}{M}\right)} \text{ polynomial of degree} \leq L \right\}. \tag{2.7.1}$$

We endow Y_M^{DG} with the inner product inherited from $L^2(\mathbb{T})$.

Definition 2.7.2 (The coarse-graining operator Q_M)**.** *Let $Q_M : L^2(\mathbb{T}) \to Y_M^{DG}$ be the orthogonal projection onto Y_M^{DG} in $L^2(\mathbb{T})$. We also define $Q_M^t : Y_M^{DG} \to \mathbb{R}^N$ to be the operator such that*

$$\langle Q_M x, y\rangle_{L^2} = x \cdot Q_M^t y \quad \forall x \in \mathbb{R}^N, y \in Y_M^{DG},$$

i.e. $NQ_M^t : Y_M^{DG} \to X_N$ *is the* L^2*-orthogonal projection of* Y_M^{DG} *onto* $\mathbb{R}^N$*, when* $\mathbb{R}^N$ *is embedded into* $L^2(\mathbb{T})$ *as a space of step functions.*

It follows that the coarse-graining operator Q_M is related to the coarse-graining operator P onto splines (see Definition 2.3.1) via $P = PQ_M$ on $L^2(\mathbb{T})$ and $NP^t = NQ_M^t$ on Y_M. Consequently, the approximation property of splines in Lemma 2.3.3 is a direct consequence of the following analogous approximation property of the discontinuous Galerkin functions.

Lemma 2.7.3. *It holds that*

$$\|Q_M N Q_M^t - \mathrm{id}_{Y_M^{DG}}\| \lesssim \frac{1}{K^2}. \tag{2.7.2}$$

The proof of Lemma 2.7.3 is deferred to Section 2.7.2. From now on, we assume K is large enough so that $Q_M N Q_M^t : Y_M^{DG} \to Y_M^{DG}$ is invertible. In particular, this means the coarse-graining operator $Q_M : X_N \to Y_M$ is onto and the orthogonal projection $NQ_M^t : Y_M \to X_N$ is one-to-one.

Definition 2.7.4 (The coarse-grained Hamiltonian $\bar{H}_{Y_M^{DG}}$)**.** *The coarse-grained Hamiltonian* $\bar{H}_{Y_M^{DG}} : Y_M^{DG} \to \mathbb{R}$ *associated to* Q_M *is given by*

$$\bar{H}_{Y_M^{DG}}(y) := -\frac{1}{N} \log \int_{\{x \in \mathbb{R}^N : Q_M x = y\}} \exp\left(-H_N(x)\right) dx. \tag{2.7.3}$$

Theorem 2.7.5 (Uniform strong convexity of $\bar{H}_{Y_M^{DG}}$)**.** *There are constants* $0 < \lambda, \Lambda, K^* < \infty$ *such that for all* $K \geq K^*$*,* M *and all* $y \in Y_M^{DG}$ *it holds*

$$2\lambda\, \mathrm{id}_{Y_M^{DG}} \leq \mathrm{Hess}_{Y_M^{DG}} \bar{H}_{Y_M^{DG}}(y) \leq 2\Lambda\, \mathrm{id}_{Y_M^{DG}}$$

in the sense of quadratic forms.

Our first step towards proving Theorem 2.7.5 is reducing it to the case $M = 1$. Since the blocks are independent for functions in Y_M^{DG}, the space Y_M^{DG} canonically factors as an orthogonal direct sum:

$$Y_M^{DG} \ni \alpha := \oplus_{m=1}^M \alpha^{(m)} \in \bigoplus_{m=1}^M Y_1^{DG}, \tag{2.7.4}$$

where $\alpha^{(m)} \in Y_1^{DG}$ is obtained by restricting α to the subinterval $(\frac{m-1}{M}, \frac{m}{M})$. Under this identification, the Euclidean structure on Y_M^{DG} also factors:

$$\langle \alpha, \beta \rangle_{Y_M^{DG}} = \frac{1}{M} \sum_{m=1}^{M} \langle \alpha^{(m)}, \beta^{(m)} \rangle_{Y_1^{DG}}. \tag{2.7.5}$$

Given $x \in \mathbb{R}^N$, for $1 \leq m \leq M$, denote $x^{(m)} := (x_{(m-1)K+1}, \ldots, x_{mK}) \in \mathbb{R}^K$. Then in the sense of (2.7.4), the coarse-grained operator Q_M decomposes as

$$Q_M x = \oplus_{m=1}^{M} Q_1 x^{(m)}.$$

Consequently, the coarse-grained Hamiltonian $\bar{H}_{Y_M^{DG}}$ also decomposes:

Lemma 2.7.6. *For $\beta \in Y_1^{DG}$ we define*

$$\psi_J(\beta) = -\frac{1}{J} \log \int_{\{x \in \mathbb{R}^J : Q_1 x = \beta\}} \exp(-H_J(x))\, dx. \tag{2.7.6}$$

Then it holds that for $\alpha \in Y_M^{DG}$

$$\bar{H}_{Y_M^{DG}}(\alpha) = \frac{1}{M} \sum_{m=1}^{M} \psi_K(\alpha^{(m)}). \tag{2.7.7}$$

In light of (2.7.7), proving Theorem 2.7.5 reduces to proving the following:

Theorem 2.7.7. *There are constants $0 < \lambda, \Lambda, J^* < \infty$ such that for all $J \geq J^*$, and all $\beta \in Y_1^{DG}$ it holds*

$$2\lambda \operatorname{id}_{Y_1^{DG}} \leq \operatorname{Hess}_{Y_1^{DG}} \psi_J(\beta) \leq 2\Lambda \operatorname{id}_{Y_1^{DG}}, \tag{2.7.8}$$

in the sense of quadratic forms.

Proof of Theorem 2.7.5. By (2.7.7), for all $y \in Y_M^{DG}$

$$D^k \bar{H}_{Y_M^{DG}}(y) = \frac{1}{M} \bigoplus_{m=1}^{M} D^k \psi_K(y^{(m)}),$$

where D^k denotes the k-th derivative as a multilinear form. Setting $k = 2$ and taking account of the different inner product structures by using (2.7.5) , (2.7.8) yields for large enough K the desired estimate

$$2\lambda \operatorname{id}_{Y_M^{DG}} \leq \operatorname{Hess}_{Y_M^{DG}} \bar{H}_{Y_M^{DG}}(y) \leq 2\Lambda \operatorname{id}_{Y_M^{DG}}.$$

□

We prove Theorem 2.7.7 in Section 2.7.1. The argument is quite technical, making up the bulk of the proof of Theorem 2.3.6. We follow a similar approach as in [GOVW09]: a local Cramér theorem and a local central limit theorem (CLT). The main difference here is that the local Cramér theorem has to be extended to canonical ensembles with multiple constraints, which means that we will have to use a multivariate CLT instead of a univariate CLT.

To finish the proof of Theorem 2.3.6, we need to transfer the strong convexity from $\bar{H}_{Y_M^{DG}}$ to $\bar{H}$. Since $\bar{H}_{Y_M^{DG}}$ and $\bar{H}$ are coarse-grained versions of H on the levels of Y_M^{DG} and Y_M, respectively, and $Y_M \subset Y_M^{DG}$, $\bar{H}$ is itself a coarse-grained version of $\bar{H}_{Y_M^{DG}}$, i.e.

$$N\bar{H}(y) = -\log \int_{z \in Y_M^{\perp}} \exp(-N\bar{H}_{Y_M^{DG}}(y+z)) N^{\frac{LM+1}{2}} \mathcal{L}^{LM+1}(dz), \tag{2.7.9}$$

where $Y_M^{\perp} := \{y \in Y_M^{DG} : Py = 0\}$ and $\dim Y_M^{\perp} = LM + 1$. (The factor $N^{\frac{LM+1}{2}}$ is due to the different Euclidean structures on $\mathbb{R}^N$.)

Lemma 2.7.8. *Let $W \oplus Z$ be an orthogonal decomposition of a finite dimensional Euclidean space. Suppose $F : W \oplus Z \to \mathbb{R}$ is a C^2 function such that $\int_{W\oplus Z} \exp(-F) < \infty$. Let $\bar{F}(z) := -\log \int_W \exp(-F(w,z))\, dw$. Let $c > 0$, then it holds that*

$$\mathrm{Hess}_{W\oplus Z} F \geq c\,\mathrm{id}_{W\oplus Z} \;\Rightarrow\; \mathrm{Hess}_Z \bar{F} \geq c\,\mathrm{id}_Z, \tag{2.7.10}$$

$$\mathrm{Hess}_{W\oplus Z} F \leq c\,\mathrm{id}_{W\oplus Z} \;\Rightarrow\; \mathrm{Hess}_Z \bar{F} \leq c\,\mathrm{id}_Z\,. \tag{2.7.11}$$

In [BL76], it was shown in a very neat way that statement (2.7.10) in Lemma 2.7.8 is simple consequence of the well-known Brascamp-Lieb inequality. Statement (2.7.11) follows from a straightforward computation.

Proof of Theorem 2.3.6. We apply Lemma 2.7.8 with $Z = Y_M$, $W = Y_M^{\perp}$, and $F = N\bar{H}_{Y_M^{DG}}$. The hypotheses of (2.7.10) and (2.7.11) are satisfied due to Theorem 2.7.5. This yields the statement of Theorem 2.3.6. □

2.7.1 Proof of Theorem 2.7.7: a multivariate local Cramér theorem:

The rest of this section is devoted to the proof of Theorem 2.7.7. We prove uniform strong convexity of ψ_J for large J by showing that ψ_J converges as $J \to \infty$ in the uniform C^2-topology to a uniformly strongly convex function. Namely, this will be the Legendre transform of the function which, to each $\beta \in Y_1^{DG}$, associates the specific free energy of a *modified* grand canonical ensemble which makes the conditioning $Q_1 x = \beta$ a *typical* event.

Before we enter into the details, we give a sketch of the argument, which closely follows the argument in [GOVW09]. Using Cramér's trick of exponential shift of measure, we construct for each β a product measure $\nu_{J,\beta}$ on $\mathbb{R}^J$ such that

- the law of each spin is an "exponential shift" of the single-site measure (a perturbed standard Gaussian),
- the expectation of $Q_1 x$ under $\nu_{J,\beta}$ is equal to β, i.e. the conditioning $Q_1 x = \beta$ is a "typical" event.

We refer to the product measure $\nu_{J,\beta}$ as the modified grand canonical ensemble for β. The required shifts of spins can be parameterized by a variable $\hat{\beta}$ that is dual to β.

Because the single-site potential ψ is quadratic plus a perturbation that is bounded in $C^2(\mathbb{R})$, it follows that the specific free energy $\bar{\psi}_J^*(\hat{\beta})$ of $\nu_{J,\beta}$ is convex in $\hat{\beta}$ for large J and its Hessian is uniformly bounded from above and from below. Consequently, the Legendre transform $\bar{\psi}_J(\beta)$ of $\bar{\psi}_J^*(\hat{\beta})$ is uniformly strongly convex. Moreover, the difference $\bar{\psi}_J(\beta) - \psi_J(\beta)$ can be interpreted as the difference between the specific free energies of $\nu_{J,\beta}$ and its restriction to the hyperplane determined by $Q_1 x = \beta$ (which is the "typical event"). Hence, we expect that this difference goes to zero as J grows large.

To verify that this difference indeed converges to zero in the uniform C^2-topology, we first relate it through a Cramér-type representation formula with the evaluation at $0 \in Y_1^{DG}$ of the density of the distribution of $J^{\frac{1}{2}}(Q_1 x - \beta)$ under $\nu_{J,\beta}$. This is done in Lemma 2.7.12. We then establish a kind of uniform C^2 local central limit theorem assuring that this eval-

uated density is bounded from above and from below uniformly in β and that moreover, it is bounded in C^2 as a function of β. These estimates are stated in Proposition 2.7.14 below and constitute the core of our proof. Then, combining the statements of Lemma 2.7.12 and Proposition 2.7.14 allow us to deduce Theorem 2.7.7.

Construction of the modified grand canonical ensembles $\nu_{J,\beta}$

We begin by introducing a family of "exponential shifts" of the law of a single spin. For each $m \in \mathbb{R}$, let μ_m be the probability measure on $\mathbb{R}$ given by the Lebesgue density

$$\frac{d\mu_m}{dz} = \exp(-\psi^*(\hat{z}_m) + \hat{z}_m z - \psi(z)), \tag{2.7.12}$$

where ψ^* is the log partition function defined in (2.2.17) and $\hat{z}_m \in \mathbb{R}$ is chosen so that μ_m has mean m, i.e.

$$\int_{\mathbb{R}} z\mu_m(dz) = (\psi^*)'(\hat{z}_m) = m.$$

We will use the fact that the function ψ^* is uniformly strongly convex and its 2nd and 3rd derivatives are uniformly bounded.

Lemma 2.7.9. *There are $0 < c < C < \infty$ such that it holds that:*

$$0 < 4c < \inf_{m\in\mathbb{R}} \mathrm{Var}(\mu_m) \leq (\psi^*)''(\hat{z}) \leq \sup_{m\in\mathbb{R}} \mathrm{Var}(\mu_m) < \frac{C}{2} < \infty, \tag{2.7.13}$$

$$|(\psi^*)'''(\hat{z})| \leq \sup_{m\in\mathbb{R}} \left| \int (z-m)^3 \mu_m(dz) \right| < \frac{C}{2} < \infty. \tag{2.7.14}$$

where $\mathrm{Var}(\mu_m)$ *denotes the variance of* μ_m.

We omit the proof of this result. It is contained in Lemma 41 in [GOVW09].

To construct $\nu_{j,\beta}$, we need to find a suitable dual variable $\hat{\beta} \in Y_1^{DG}$ and exponentially shift the law of the J spins according to the J-dimensional vector $JQ_1^t\hat{\beta}$, which is the L^2

orthogonal projection of $\hat{\beta}$ onto $\mathbb{R}^J$ (cf. Definition 2.7.2). Let $\psi_J^* : \mathbb{R}^J \to \mathbb{R}$ be the sum of the log partition function ψ^* for J sites,

$$\psi_J^*(x) := \sum_{j=1}^{J} \psi^*(x_j),$$

and let $\bar{\psi}_J^* : Y_1^{DG} \to \mathbb{R}$ be the function

$$\bar{\psi}_J^*(\hat{\beta}) := \frac{1}{J}\psi_J^*(JQ_1^t\hat{\beta}) = \frac{1}{J}\sum_{j=1}^{J} \psi^*\left((JQ_1^t\hat{\beta})_j\right), \tag{2.7.15}$$

which will be interpreted as the specific free energy of $\nu_{j,\beta}$ for the right choice of $\hat{\beta}$. After dealing with the approximation error, property (2.7.13) translates into convexity estimates for the function $\bar{\psi}_J^*(\hat{\beta})$.

Lemma 2.7.10. *Let c and C be as in Lemma 2.7.9. Then there is $J_1 \in \mathbb{N}$ such that for all $J \geq J_1$ and $\hat{\beta}, \hat{\eta}, \hat{\gamma} \in Y_1^{DG}$:*

$$2c \;\leq\; \| \operatorname{Hess} \bar{\psi}_J^*(\hat{\beta}) \| \;\leq\; C\,, \tag{2.7.16}$$

$$\| D^3\bar{\psi}_J^*(\hat{\beta}) \| \;\leq\; C. \tag{2.7.17}$$

We postpone the proof of this result to Section 2.7.2.

By standard convex analysis, the bounds (2.7.16) and (2.7.17) imply that for $J \geq J_1$, the Legendre transform of $\bar{\psi}_J^*$,

$$\bar{\psi}_J(\beta) := \sup_{\hat{\beta} \in Y_1^{DG}} \left(\langle \beta, \hat{\beta} \rangle \;-\; \bar{\psi}_J^*(\hat{\beta}) \right), \tag{2.7.18}$$

is uniformly strongly convex and its Hessian and 3rd derivative are uniformly bounded, and the unique maximizer $\hat{\beta}^{max}(\beta)$ of (2.7.18) satisfies

$$\beta \;=\; \nabla\, \bar{\psi}_J^*(\hat{\beta}^{max}), \quad \hat{\beta}^{max} \;=\; \nabla\, \bar{\psi}_J(\beta). \tag{2.7.19}$$

The vector $\hat{\beta}^{max}$ serves to construct $\nu_{J,\beta}$. Set

$$\hat{m}_\beta := JQ_1^t\hat{\beta}^{max}, \quad \hat{m}_{j,\beta} := (\hat{m}_\beta)_j, \tag{2.7.20}$$

$$m_\beta := \nabla\psi_J^*(\hat{m}_\beta), \quad m_{j,\beta} := (m_\beta)_j, \tag{2.7.21}$$

and define a product measure on $\mathbb{R}^J$ (cf. (2.7.12))

$$\frac{d\nu_{J,\beta}}{dx}(x) := \prod_{j=1}^{J} \frac{d\mu_{m_{j,\beta}}}{dx_j}(x_j). \tag{2.7.22}$$

Then the expected value of $Q_1 x$ under $\nu_{J,\beta}$ is equal to β:

$$\int Q_1 x \, d\nu_{J,\beta} \overset{(2.7.22)}{=} Q_1 m_\beta \overset{(2.7.21)}{=} Q_1 \nabla\psi_J^*(JQ_1^t\hat{\beta}^{max}) \overset{(2.7.15)}{=} \nabla\bar{\psi}_J^*(\hat{\beta}^{max}) \overset{(2.7.19)}{=} \beta.$$

This completes the construction of the modified grand canonical ensemble $\nu_{J,\beta}$.

Uniform C^2 convergence of $\bar{\psi}_J - \psi_J$ to zero

For a given β, the specific free energy of the modified grand canonical ensemble $\nu_{J,\beta}$ is just $\bar{\psi}_J^*(\hat{\beta}^{max}) = \langle \beta, \hat{\beta}^{max} \rangle - \bar{\psi}_J(\beta)$. On the other hand, the specific free energy of the canonical ensemble associated with the restriction of $\nu_{J,\beta}$ to the hyperplane $\{x \,|\, Q_1 x = \beta\}$, where it is highly concentrated anyway for large J, is given by $\langle \beta, \hat{\beta}^{max} \rangle - \psi_J(\beta)$ (we leave the calculation to the reader as an exercise). Consequently, $\bar{\psi}_J(\beta) - \psi_J(\beta)$ measures the difference in free energies and hence we expect it to converge to zero in some sense as $J \to \infty$. As we indicated above, the proof that it converges strong enough for our purposes, i.e. in C^2, begins with a Cramér-type representation formula for the density in $0 \in Y_1^{DG}$ of the distribution of $J^{\frac{1}{2}}(Q_1 x - \beta)$ under $\nu_{J,\beta}$, which is a centered $(L+1)$-dimensional vector of "suitably weighted" sums of independent random variables. (Cf. equation (125) in [GOVW09])

Definition 2.7.11. *Fix once for all an orthonormal basis $\{f_l\}_{l=0,1,\cdots,L}$ on Y_1^{DG} (which are polynomials of degree $\leq L$). From now on we will identify Y_1^{DG} with $\mathbb{R}^{L+1}$ using this basis.*

Lemma 2.7.12. *Denote by $g_{J,\beta}$ the law of the $\mathbb{R}^{L+1}$-valued random variable $J^{\frac{1}{2}}(Q_1 x - \beta)$ where x follows the law $\nu_{J,\beta}$ and let $\mathcal{J}Q := (\det Q_1 Q_1^t)^{\frac{1}{2}}$. The density of $g_{J,\beta}$ at $0 \in \mathbb{R}^{L+1}$ with respect to Lebesgue measure can be represented as follows:*

$$g_{J,\beta}(0) := \frac{dg_{J,\beta}}{d\mathcal{L}^{L+1}}(0) = (J^{\frac{L+1}{2}}\,\mathcal{J}Q)^{-1} \exp\left[J\left(\bar{\psi}_J(\beta) - \psi_J(\beta)\right)\right]. \tag{2.7.23}$$

We postpone the proof of Lemma 2.7.12 to Section 2.7.2. Formula (2.7.23) allows us to transfer the strong convexity of $\bar{\psi}_J$ to the function ψ_J, once we have the following estimates on the Jacobian $(J^{\frac{L+1}{2}}\,\mathcal{J}Q)^{-1}$ (appearing on the right hand side of (2.7.23)) and the density $g_{J,\beta}(0)$ (appearing on left hand side of (2.7.23)).

Lemma 2.7.13. *There is a positive integer $J^* \in \mathbb{N}$ such that for $J \geq J^*$:*

$$\frac{1}{C} \leq J^{\frac{L+1}{2}}\,\mathcal{J}Q \leq C. \tag{2.7.24}$$

This follows from estimate (2.7.32).

Proposition 2.7.14. *There exist a constant $C < \infty$ and a positive integer J_2 such that for all $J \geq J_2$ and all $\beta \in \mathbb{R}^{L+1}$:*

$$\frac{1}{C} \leq g_{J,\beta}(0) \leq C, \tag{2.7.25}$$

$$\|\nabla\, g_{J,\beta}(0)\| \leq C, \tag{2.7.26}$$

$$\|\mathrm{Hess}\; g_{J,\beta}(0)\| \leq C. \tag{2.7.27}$$

This result was proven in [GOVW09] for the case $L = 0$ (cf. equation (126) in [GOVW09]). In the general case considered here, establishing the estimates becomes somewhat more subtle. In particular, a geometric property due to the independence of basis polynomials f_l enters as a new ingredient. The proof also shares some similarities to the proof of the local Cramér theorem in [Men11]. As the proof as a whole becomes quite long we postpone it to the end of this chapter (see Section 2.10). We conclude this section with a derivation of Theorem 2.7.7 from these results.

Proof of Theorem 2.7.7. Rewrite formula (2.7.23) as:

$$\bar{\psi}_J(\beta) - \psi_J(\beta) = \frac{1}{J}\left[\log\left(J^{\frac{L+1}{2}}\,\mathcal{J}Q\right) + \log g_{J,\beta}(0)\right]. \tag{2.7.28}$$

For $J \geq \max\{J^*, J_2\}$, the estimates and (2.7.25) and (2.7.24) thus yield

$$|(\bar{\psi}_J - \psi_J)(\beta)| \leq \frac{\log C + \log C}{J} \tag{2.7.29}$$

$$\|\nabla(\bar{\psi}_J - \psi_J)(\beta)\| \overset{(2.7.28)}{=} \frac{1}{J}\,(g_{J,\beta}(0))^{-1}\ \|\nabla g_{J,\beta}(0)\| \overset{(2.7.25),(2.7.26)}{\leq} \frac{C^2}{J} \tag{2.7.30}$$

$$\begin{aligned}\|(\mathrm{Hess}(\bar{\psi}_J - \psi_J)(\beta)\| &\overset{(2.7.28)}{\leq} \frac{1}{J}\,(g_{J,\beta}(0))^{-1}\ \|\mathrm{Hess}\, g_{J,\beta}(0)\| \\ &\quad + \frac{1}{J}\,(g_{J,\beta}(0))^{-2}\ \|\,\nabla g_{J,\beta}(0)\ \otimes\ \nabla g_{J,\beta}(0)\,\| \\ &\overset{(2.7.25),(2.7.26),(2.7.27)}{\leq} \frac{C^2}{J} + \frac{C^4}{J}. \end{aligned} \tag{2.7.31}$$

This proves convergence of $\psi_J - \bar{\psi}_J$ to zero in $C^2(\mathbb{R}^{L+1})$, indeed with difference of order J^{-1} as $J \to \infty$. Since $\bar{\psi}_J$ is uniformly strongly convex and its Hessian is uniformly bounded if $J \geq J_1$, this proves Theorem 2.7.7. □

2.7.2 Proofs of auxiliary results.

We need two facts of approximation. The first fact is that the operator $Q_1 J Q_1^t$ is close to the identity for large J.

Lemma 2.7.15. *It holds (with implicit constants depending on L):*

$$\|Q_1 J Q_1^t - \mathrm{id}_{Y_1^{DG}}\,\| \lesssim \frac{1}{J^2}. \tag{2.7.32}$$

Proof of Lemma 2.7.15. Since $JQ_1^t : Y_1^{DG} \to \mathbb{R}^J$ is an L^2-orthogonal projection onto piecewise constant functions, for any $f \in Y_1^{DG}$

$$\begin{aligned}\langle(\mathrm{id} - Q_1 J Q_1^t)f, f\rangle_{L^2} &= \langle f, f\rangle_{L^2} - \langle JQ_1^t f, JQ_1^t f\rangle_{L^2} \\ &= |f - JQ_1^t f|_{L^2}^2 \lesssim \frac{1}{J^2}|f'|_{L^2}^2 \lesssim \frac{1}{J^2}|f|_{L^2}^2,\end{aligned}$$

where we used a Poincaré inequality on interval of length $\frac{1}{J}$ for each interval $[\frac{j-1}{J}, \frac{j}{J}]$, and in the last step we used the fact that norms on a finite dimensional space are equivalent (this is where L comes into the implicit constant). □

The second fact we need is that the vectors $\gamma^j \in \mathbb{R}^{L+1}$, $j = 1, 2, \cdots, J$, defined by

$$(JQ_1^t y)_j = y \cdot \gamma^j \quad \text{for all } y \in Y_1^{DG} = \mathbb{R}^{L+1},$$

form a piecewise constant approximation of the smooth curve

$$\gamma : [0,1] \to \mathbb{R}^{L+1}\,,\ \gamma(t) \,=\, (f_0(t), f_1(t), \cdots, f_L(t)). \tag{2.7.33}$$

Lemma 2.7.16. $\forall J \in \mathbb{N}$ *(with implicit constants depending on L),*

$$\max_{1\le j\le J} \sup_{t\in[\frac{j-1}{J},\frac{j}{J}]} |\gamma^j - \gamma(t)| \lesssim \frac{1}{J} \quad \textit{and} \quad \max_{1\le j\le J} |\gamma^j| \lesssim 1. \tag{2.7.34}$$

Proof. The proof of (2.7.34) is similar to that of (2.7.32). Note that

$$\gamma_l^j = (JQ_1^t f_l)_j = J\int_{\frac{j-1}{J}}^{\frac{j}{J}} f_l(\theta)d\theta$$

is the average of f_l over the interval $[\frac{j-1}{J}, \frac{j}{J}]$. Thus, for each $1 \le j \le J$, and $t \in [\frac{j-1}{J}, \frac{j}{J}]$,

$$|\gamma(t) - \gamma^j|^2 = \sum_{l=0}^{L} |f_l(t) - \gamma_l^j|^2 \lesssim \sum_{l=0}^{L} \frac{1}{J^2} |f_l'|_{L^\infty}^2 \lesssim \frac{1}{J^2},$$

where we used mean value theorem. This implies the first bound in (2.7.34), from which the second bound follows. □

Let us now begin proving the auxiliary results in the previous sections.

Proof of Lemma 2.7.3. Because $Y_M^{DG} = \oplus_{m=1}^M Y_1^{DG}$ (cf. (2.7.4)), Lemma 2.7.3 follows directly from Lemma 2.7.15. □

Proof of Lemma 2.7.10. For $\hat{\eta} \in \mathbb{R}^{L+1}$ we have:

$$\langle \hat{\eta}, \text{Hess } \bar{\psi}_J^*(\hat{\beta})\, \hat{\eta} \rangle \;\; = \;\; \frac{1}{J} \sum_{j=1}^{J} (\psi^*)'' \left((JQ_1^t \hat{\beta})_j \right) \left[(JQ_1^t \hat{\eta})_j \right]^2.$$

Using (2.7.13) we obtain:

$$\frac{4c}{J} |JQ_1^t \hat{\eta}|^2 \le \langle \hat{\eta}, \text{Hess } \bar{\psi}_J^*(\hat{\beta})\, \hat{\eta} \rangle \le \frac{C}{2J} |JQ_1^t \hat{\eta}|^2.$$

By the approximation of Lemma 2.7.15,

$$\frac{1}{J} |JQ_1^t \hat{\eta}|^2 = \langle JQ_1^t \hat{\eta}, JQ_1^t \hat{\eta} \rangle_{L^2} = \langle Q_1 JQ_1^t \hat{\eta}, \hat{\eta} \rangle_{L^2} \overset{(2.7.32)}{=} \left(1 + O\left(\frac{1}{J^2} \right) \right) |\hat{\eta}|^2.$$

This shows (2.7.16) for large J. Concerning (2.7.17), we find for $\hat{\eta} \in \mathbb{R}^{L+1}$:

$$
\begin{aligned}
|\, D^3 \bar{\psi}_J^*(\hat{\beta})(\hat{\eta},\hat{\eta},\hat{\eta})\,| &= \left| \frac{1}{J} \sum_{j=1}^{J} (\psi^*)'''\left((JQ_1^t\hat{\beta})_j \right) \left[(JQ_1^t\hat{\eta})_j \right]^3 \right| \\
&\leq \max_{j=1,\ldots,J} \left| \hat{\eta}\cdot\gamma^j \right| \frac{1}{J} \sum_{j=1}^{J} \left| (\psi^*)'''\left((JQ_1^t\hat{\beta})_j \right) \right| \left[(JQ_1^t\hat{\eta})_j \right]^2 .
\end{aligned}
$$

We then appeal to (2.7.14) and the uniform bound (2.7.34) and proceed just as above. □

Proof of Lemma 2.7.12. So far it has always been understood that by dx, etc. we mean the Hausdorff measure of appropriate dimension. We will be a bit more careful during the next computation and write out the measures in detail where it seems helpful. Let ζ be a measurable test function defined on $\mathbb{R}^{L+1}$. The proof of identity (2.7.23) essentially boils down to an application of the co-area formula for Q_1.

$$
\begin{aligned}
\int_{\mathbb{R}^{L+1}} \zeta(u)\, \frac{dg_{J,\beta}}{d\mathcal{L}^{L+1}}(u)\, \mathcal{L}^{L+1}(du) &= \int_{\mathbb{R}^J} \zeta\left(J^{\frac{1}{2}}(Q_1 x - \beta) \right) \frac{d\nu_{J,\beta}}{d\mathcal{L}^J}(x)\, \mathcal{L}^J(dx) \\
&= \int_{\mathbb{R}^{L+1}} (\mathcal{J}Q)^{-1} \zeta\left(J^{\frac{1}{2}}(\widetilde{\beta} - \beta) \right) \int_{\mathbb{R}^J_{1,\widetilde{\beta}}} \frac{d\nu_{J,\beta}}{d\mathcal{L}^J}(x) \mathcal{H}^{J-L-1}(dx) \mathcal{L}^{L+1}(d\widetilde{\beta}) \\
&\overset{(2.7.22)}{=} \int_{\mathbb{R}^{L+1}} (\mathcal{J}Q)^{-1} \zeta\left(J^{\frac{1}{2}}(\widetilde{\beta} - \beta) \right) \\
&\qquad \int_{\mathbb{R}^J_{1,\widetilde{\beta}}} \exp\left(-\sum_{j=1}^{J} \psi^*(\hat{m}_{j,\beta}) + \sum_{j=1}^{J} \hat{m}_{j,\beta} x_j - H_J(x) \right) \mathcal{H}(dx) \mathcal{L}(d\widetilde{\beta}) \\
&\overset{(2.7.15),(2.7.20)}{=} \int_{\mathbb{R}^{L+1}} (\mathcal{J}Q)^{-1}\, \zeta\left(J^{\frac{1}{2}}(\widetilde{\beta} - \beta) \right) \\
&\qquad \int_{\mathbb{R}^J_{1,\widetilde{\beta}}} \exp\left(-J\bar{\psi}_J^*(\hat{\beta}^{max}) + J\,\langle \widetilde{\beta}, \hat{\beta}^{max} \rangle - H_J(x) \right) \mathcal{H}(dx) \mathcal{L}(d\widetilde{\beta}) \\
&\overset{(2.7.18),(2.7.6)}{=} \int_{\mathbb{R}^{L+1}} (\mathcal{J}Q)^{-1}\, \zeta\left(J^{\frac{1}{2}}(\widetilde{\beta} - \beta) \right) \\
&\qquad \exp\left(J\left(\bar{\psi}_J(\beta) - \psi_J(\widetilde{\beta}) + \langle \widetilde{\beta} - \beta, \hat{\beta}^{max} \rangle \right) \right) \mathcal{L}^{L+1}(d\widetilde{\beta}) \\
&= \int_{\mathbb{R}^{L+1}} \zeta(u) \left(J^{\frac{L+1}{2}} \mathcal{J}Q \right)^{-1} \\
&\qquad \exp\left(J\left(\bar{\psi}_J(\beta) - \psi_J(J^{-\frac{1}{2}}u + \beta) + \langle J^{-\frac{1}{2}}u, \hat{\beta}^{max} \rangle \right) \right) \mathcal{L}^{L+1}(du).
\end{aligned}
$$

The identity (2.7.23) now follows from approximating the Dirac mass δ_0 in $\mathbb{R}^{L+1}$ by continuous test functions ζ_i. □

Proof of Lemma 2.7.13. We note that $J^{\frac{L+1}{2}}\mathcal{J}Q = (\det Q_1 J Q_1^t)^{\frac{1}{2}}$. Estimate (2.7.32) implies that there is $J^* \in \mathbb{N}$ such that for $J \geq J^*$:

$$\frac{1}{C} \leq J^{\frac{L+1}{2}}\mathcal{J}Q \leq C.$$

□

2.8 A uniform log-Sobolev inequality for conditional measures

The purpose of this section is to deduce Theorem 2.4.12, which states that the conditional measures $\mu(dx|Px = y)$ satisfies a uniform logarithmic Sobolev inequality (LSI). The logarithmic Sobolev inequality was first discovered by Gross [Gro75]. It characterizes the speed of convergence to equilibrium of the natural associated drift diffusion process. For more facts about the LSI we refer to the books [Roy99, BGL14] and survey article [Led01]. In Section 2.8.1, we state some basic principles of the LSI and introduce the two-scale criterion which is the key to our argument for deducing Theorem 2.4.12. In Section 2.8.2, we explain how those principles are applied to deduce the uniform LSI for the conditional measure $\mu(dx|Px = y)$.

2.8.1 Basic principles for the LSI

Four different principles underlie our proofs of logarithmic Sobolev inequalities in Section 2.8.2. Three of these are standard results that have proven to be useful for establishing LSI in many cases and that have been known for a long time. The fourth principle is a more specialized criterion that has been successfully applied for deducing LSI for spin systems. It will guide our main strategy of proof while the other results are needed to verify the assumptions of the criterion. Let us forget for a moment the precise definitions of X_N and H and present the basic principles of the LSI in the setting of Euclidean spaces. Let X be

an Euclidean space or affine subspaces of some Euclidean space. With ∇ and $|\cdot|$ we denote the gradient and norm that is derived from the Euclidean structure of X. We write $\mathcal{P}(X)$ for the space of Borel probability measures on X.

Definition 2.8.1 (LSI). *Let $\Phi(z) := z \log z$. We say that $\nu \in \mathcal{P}(X)$ satisfies a logarithmic Sobolev inequality* (LSI) *with constant $\rho > 0$ if for all smooth functions $h : X \to \mathbb{R}_+$ it holds that*

$$\mathrm{Ent}(h\nu|\nu) := \int \Phi(h)\, \nu(dx) \,-\, \Phi\left(\int h\, \nu(dx)\right) \,\leq\, \frac{1}{\rho} \int \frac{1}{2h}\, |\nabla h|^2\, \nu(dx).$$

In this case, we also use the notation $\mathrm{LSI}(\nu) \geq \rho$.

The following tensorization principle has been known ever since the notion of LSI came up (see [Gro75]). It is the basic reason for why LSI is well-suited for high-dimensional systems.

Lemma 2.8.2 (Tensorization principle). *Given $\nu_n \in \mathcal{P}(X_N)$ for $n = 1, ..., N$. Then $\mathrm{LSI}(\nu_n) \geq \rho_n$ for all $n = 1, ..., N$ implies:*

$$\mathrm{LSI}\left(\bigotimes_{n=1}^{N} \nu_n\right) \geq \min_n \rho_n.$$

We next recall two fundamental criteria for proving logarithmic Sobolev inequalities. The first one is a simple perturbation result due to Holley and Stroock [HS87].

Lemma 2.8.3 (Holley-Stroock). *We assume that $\nu \in \mathcal{P}(X)$ satisfies $\mathrm{LSI}(\nu) \geq \rho$. For a bounded function $\delta\psi : X \to \mathbb{R}$, define a measure $\widetilde{\nu} \in \mathcal{P}(X)$ that is absolutely continuous with respect to ν via*

$$\frac{d\widetilde{\nu}}{d\nu}(x) \,=\, \frac{1}{Z}\, \exp[-\delta\psi(x)].$$

Then $\mathrm{LSI}(\widetilde{\nu}) \geq \rho \exp\left[-2\,\mathrm{osc}(\delta\psi)\right]$. Here $\mathrm{osc}(\delta\psi) = \sup_X \delta\psi \,-\, \inf_X \delta\psi$ stands for the total oscillation of the perturbation.

The second criterion is due to Bakry and Émery [BÉ85]. It says that a uniformly strongly convex Hamiltonian implies LSI.

Lemma 2.8.4 (Bakry-Émery). *Let $\nu \in \mathcal{P}(X)$ be absolutely continuous with respect to the Hausdorff measure $\mathcal{H}$ on X. If the Hamiltonian H of the measure ν, given by*

$$H(x) := -\log \frac{d\nu}{d\mathcal{H}}(x),$$

is twice continuously differentiable and uniformly strongly convex with lower bound λ, i.e.

$$\forall\, x \in X \quad \forall\, v \in T_x X \quad \langle v, \operatorname{Hess} H(x)\, v\rangle_{T_x X} \geq \lambda\, |v|^2_{T_x X},$$

then $\mathrm{LSI}(\nu) \geq \lambda$.

Proofs of the facts mentioned so far can be found for example in [GZ03] or in the nice introduction to both spectral gap and logarithmic Sobolev inequalities [Led01]. As pointed out above, we will in addition need the two-scale criterion that was presented in [OR07] and which is also contained, in a slightly different formulation, in [GOVW09]. We first define a decomposition of measures analogous to Definition 2.3.4 in the setting of a product space.

Definition 2.8.5. *Let $\nu \in \mathcal{P}(X_1 \times X_2)$ be a measure with smooth positive probability density function with respect to Hausdorff measure. We decompose ν into a family of conditional measures $\{\nu(dx_1|x_2)\}_{x_2 \in X_2} \subset \mathcal{P}(X_1)$ and the corresponding marginal measure $\bar{\nu} \in \mathcal{P}(X_2)$. This decomposition is such that for all measurable $h : X_1 \times X_2 \to \mathbb{R}$:*

$$\int_{X_1 \times X_2} h\, d\nu = \int_{X_2} \int_{X_1} h(x_1, x_2)\, \nu(dx_1|x_2)\, \bar{\nu}(dx_2).$$

The two-scale criterion reads as follows.

Lemma 2.8.6 (Two-scale criterion for LSI). *Let $\nu \in \mathcal{P}(X_1 \times X_2)$ be a measure with twice continuously differentiable Hamiltonian H. Assume that there exist constants $\rho_1, \rho_2 > 0$ such that*

$$\mathrm{LSI}(\nu(dx_1|x_2)) \geq \rho_1 \quad \textit{for all } x_2 \in X_2,$$

$$\mathrm{LSI}(\bar{\nu}) \geq \rho_2.$$

Moreover, assume that

$$\frac{1}{\rho_1}\frac{1}{\rho_2}\sup_{X_1\times X_2}|\nabla_{X_1}\nabla_{X_2}H(x)|^2 \;=\; \kappa<\infty. \tag{2.8.1}$$

Here

$$|\nabla_{X_1}\nabla_{X_2}H(x)| \;=\; \sup\{\langle \mathrm{Hess}\,H(x)\,u,v\rangle|\,u\in T_xX_1, v\in T_xX_2, |u|=|v|=1\},$$

which is finite if Hess H *is bounded. Then*

$$\mathrm{LSI}(\nu) \;\geq\; \frac{1}{2}\left(\rho_1+(1+\kappa)\rho_2 \;-\; \sqrt{(\rho_1+(1+\kappa)\rho_2)^2-4\rho_1\rho_2}\right).$$

Lemma 2.8.6 says that LSI for conditional measures and corresponding marginal may - under the coupling assumption (2.8.1) - be combined to yield a LSI for the full measure. A proof of the two-scale criterion can be found in [OR07] where it is stated as Theorem 2.

2.8.2 Uniform LSI for conditional measures

In this section we explain how the basic principles of Section 2.8.1 are used to deduce Theorem 2.4.12. The proof adapts the strategy in [GOVW09], which covered the case for $L=0$, when Y_M is the space of piecewise constant functions. However, for $L\geq 1$, due to the non-locality of the spline functions in Y_M, we need to modify the strategy in [GOVW09] by introducing an intermediate step. Namely, we first deduce a uniform LSI on the level of Y_M^{DG}, the space of piecewise polynomials (which have the important property of being local), and then apply two-scale criterion to get back a uniform LSI on the level of Y_M.

Theorem 2.8.7 (Uniform LSI for $\mu(dx|Q_Mx=y)$)**.** *Let Y_M^{DG} be the space of piecewise polynomials (see* (2.7.1) *or Definition 2.7.1). Let $Q_M: L^2(\mathbb{T})\to Y_M^{DG}$ denote the orthogonal projection onto Y_M^{DG} in $L^2(\mathbb{T})$. Then the conditional measures $\mu(dx|Q_Mx=y)$ satisfy LSI(ϱ_Q) with a constant $\varrho_Q>0$ uniform in N,M and $y\in Y_M^{DG}$.*

Remark 2.8.8. *Both in Theorem 2.4.12 and in Theorem 2.8.7, the case $L=0$ corresponds to the result in [GOVW09]. In Theorem 2.8.7, the space of observables is of course of dimension $(L+1)M$, not M.*

Proof of Theorem 2.8.7. Because the coarse-graining operator Q_M is local and the Hamiltonian H has no interaction between different sites, the conditional measure $\mu(dx|Q_M x = y)$ factors as

$$\mu^N(dx|Q_M x = \alpha) = \bigotimes_{m=1}^{M} \mu^K(dx^{(m)}|Q_1 x^{(m)} = \alpha^{(m)}),$$

where for clarity we denote μ^N instead of μ the Gibbs measure on $\mathbb{R}^N$. (Cf. Section 2.7.1 for the notation $x^{(m)}, \alpha^{(m)}$). By the tensorization principle (cf. Lemma 2.8.2) it suffices to show that the conditional measure

$$\mu_\alpha^K(dx) := \mu^K(dx|Q_1 x = \alpha)$$

satisfy a uniform LSI for K large enough. The strategy now is to apply the two-scale criterion for LSI (cf. Lemma 2.8.6), which requires us to decompose the state space as an orthogonal sum of two spaces, one describing the mesoscopic profile and the other describing the fluctuations around the profile.

From now on we assume that K is of the form $K = RJ$, where $J \in \mathbb{N}$ is large but fixed and $R \in \mathbb{N}$ is arbitrary. The key observation here is that Y_1^{DG} is a linear subspace of Y_R^{DG}. This gives an orthogonal decomposition of the state space $Q_1^{-1}(\alpha)$ as

$$Q_1^{-1}(\alpha) \ni x = x_\parallel \oplus x_\perp \in V \oplus W_\alpha,$$

where

$$V := \ker Q_R \cap \mathbb{R}^K, \quad W_\alpha := \operatorname{im} K Q_R^t \cap Q_1^{-1}(\alpha).$$

Now, following Definition 2.8.5 we decompose μ_α^K with respect to the factorization (2.8.7) into the family of conditional measures $\left\{\mu_\alpha^K(dx_\parallel|x_\perp)\right\}_{x_\perp} \subset \mathcal{P}(V)$ and the marginal $\bar{\mu}_\alpha^K(dx_\perp) \in \mathcal{P}(W_\alpha)$. In order to apply the two-scale criterion, it remains to show that there are $\rho_1, \rho_2 > 0$, independent of R (and hence K) and α, such that

$$\forall x_\perp \in W_\alpha : \operatorname{LSI}(\mu_\alpha^K(dx_\parallel|x_\perp)) \geq \rho_1, \tag{2.8.2}$$

and

$$\mathrm{LSI}(\bar{\mu}_\alpha^K(dx_\perp)) \geq \rho_2. \tag{2.8.3}$$

Additionally, we have to show that

$$\frac{1}{\rho_1}\frac{1}{\rho_2}\sup_{Q_1^{-1}(\alpha)} |\nabla_V \nabla_{W_\alpha} H_K|^2 \leq \kappa < \infty, \tag{2.8.4}$$

for some constant κ that is independent of K and α.

Argument for (2.8.2): Let $\beta = Q_R x_\perp$, then $\mu_\alpha^K(dx_\parallel|x_\perp)$ factors as

$$\mu_\alpha^K(dx_\parallel|x_\perp) = \mu^K(dx|Q_R x = \beta) = \bigotimes_{r=1}^{R} \mu^J(dx^{(r)}|Q_1 x^{(r)} = \beta^{(r)}).$$

The Hamiltonian of the conditional measure $\mu^J(dx^{(r)}|Q_1 x^{(r)} = \beta^{(r)})$ is just a restriction of the Hamiltonian H_J given in (2.2.1). From the explicit form of H_J, we get by a combination of Holley-Stroock (cf. Lemma 2.8.3) and Bakry-Émery (cf. Lemma 2.8.4) that

$$\mathrm{LSI}\left(\mu^J(dx^{(r)}|Q_1 x^{(r)} = \beta^{(r)})\right) \geq \exp\left(-2J\,\mathrm{osc}(\delta\psi)\right) =: \rho_1 > 0.$$

Then tensorization principle yields (2.8.2).

Argument for (2.8.3): The strategy is to show that the Hamiltonian of the marginal measure $\bar{\mu}_\alpha^K(dx_\perp)$ is uniformly strongly convex. The desired statement (2.8.3) follows then from the Bakry-Émery criterion. We will need the following approximation result, which we will verify at the end of this argument.

Lemma 2.8.9. *For all* $u \in \mathrm{im}\, NQ_M^t$,

$$|Q_M u|_{L^2}^2 = \left(1 + O\left(\frac{1}{K^2}\right)\right)|u|_{L^2}^2. \tag{2.8.5}$$

We start with observing the the Hamiltonian $\hat{H}_{W_\alpha}$ of the measure $\bar{\mu}_\alpha^K(dx_\perp)$ is given by

$$\hat{H}_{W_\alpha}(x_\perp) := -\log\frac{d\bar{\mu}_\alpha^K}{dx_\perp}(x_\perp) = -\log\frac{1}{Z}\int_V \exp(-H_K(x_\parallel + x_\perp))dx_\parallel.$$

Here as usual, $dx_\parallel$ is the Hausdorff measure of appropriate dimension on V. Using definition (2.7.3) of $\bar{H}_{Y_R^{DG}}$, the last identity yields

$$\hat{H}_{W_\alpha}(x_\perp) = K\bar{H}_{Y_R^{DG}}(Q_R x_\perp) + \log Z, \tag{2.8.6}$$

where Z is a constant that accounts for different normalization constants. From Theorem 2.7.5, we know that $\bar{H}_{Y_R^{DG}}$ is uniformly strongly convex, provided J is large enough. Now, using (2.8.6) we will transfer the convexity from $\bar{H}_{Y_R^{DG}}$ to $\hat{H}_{W_\alpha}$. Applying Theorem 2.7.5, we get by the chain rule that for $J \geq J^*$ and arbitrary $u \in T_{x_\perp} W_\alpha \subset \operatorname{im} KQ_R^t$,

$$\begin{aligned}
\langle u, \operatorname{Hess} \hat{H}_{W_\alpha}(x_\perp)\, u\rangle_{T_{x_\perp} W_\alpha} &= K\langle Q_R u, \operatorname{Hess} \bar{H}_{Y_R^{DG}}(Q_R x_\perp)\, Q_R u\rangle_{L^2} \\
&\geq 2K\lambda |Q_R u|_{L^2}^2 \\
&\overset{(2.8.5)}{=} 2K\lambda \left(1 + O\left(\frac{1}{J^2}\right)\right) |u|_{L^2}^2 \\
&= 2\lambda \left(1 + O\left(\frac{1}{J^2}\right)\right) |u|_{T_{x_\perp} W_\alpha}^2,
\end{aligned}$$

where we used estimate (2.8.5) with $K = RJ$ in place of $N = MK$. This yields the uniform strong convexity of $\hat{H}_{W_\alpha}$. Thus, the Bakry-Émery criterion (cf. Lemma 2.8.4) implies (2.8.3) with constant $\rho_2 := \lambda$.

Proof of Lemma 2.8.9. By assumption $u = NQ_M^t \beta$ for some $\beta \in Y_M^{DG}$, so

$$NQ_M^t (Q_M N Q_M^t)^{-1} Q_M u = NQ_M^t \beta = u.$$

Using this and (2.7.2), we get

$$\begin{aligned}
&\langle Q_M u, Q_M u\rangle_{L^2} \\
&= \langle (Q_M N Q_M^t)^{-1} Q_M u, Q_M u\rangle_{L^2} + \langle (\mathrm{id}_{Y_M^{DG}} - (Q_M N Q_M^t)^{-1}) Q_M u, Q_M u\rangle_{L^2} \\
&\geq \langle NQ_M^t (Q_M N Q_M^t)^{-1} Q_M u, u\rangle_{L^2} - \|\mathrm{id}_{Y_M^{DG}} - (Q_M N Q_M^t)^{-1}\| |Q_M u|_{L^2}^2 \\
&\geq \langle u, u\rangle_{L^2} - O\left(\frac{1}{K^2}\right) |u|_{L^2}^2 = \left(1 + O\left(\frac{1}{K^2}\right)\right) |u|_{L^2}^2.
\end{aligned}$$

The reverse inequality is obvious. □

Argument for (2.8.4): This follows from the uniform C^2 bound of H_K.

Overall, we may hence apply Lemma 2.8.6 which yields that for $J \geq J^*$:

$$\mathrm{LSI}(\mu_\alpha^K) \geq \frac{1}{2}\left(\rho_1 + (1+\kappa)\lambda - \sqrt{(\rho_1 + (1+\kappa)\lambda)^2 - 4\rho_1\lambda}\right),$$

which is bounded from below uniformly in K.

□

With Theorem 2.8.7 at hand, the proof of Theorem 2.4.12 consists of another application of the two-scale criterion (see Lemma 2.8.6), which is very similar to the one in proof of Theorem 2.8.7: there we introduced an intermediate level by prescribing a global polynomial as the constraint, here we will introduce an intermediate level by prescribing a spline on the same mesh as the constraint.

Proof of Theorem 2.4.12. The key observation here is that Y_M is a linear subspace of Y_M^{DG}, which gives the orthogonal decomposition of the state space $P^{-1}(y) \subset X_N$ as

$$P^{-1}(y) \ni x = x_\parallel \oplus x_\perp \in V \oplus W_y, \tag{2.8.7}$$

where

$$V := \ker Q_M \cap X_N, \quad W_y := \mathrm{im}\, N Q_M^t \cap P^{-1}(y).$$

Now, following Definition 2.8.5 we decompose $\mu^N(dx|Px = y)$ with respect to the factorization (2.8.7) into the family of conditional measures $\left\{\mu^N(dx_\parallel|x_\perp)\right\}_{x_\perp} \subset \mathcal{P}(V)$ and the marginal $\bar{\mu}^N(dx_\perp|Px_\perp = y) \in \mathcal{P}(W_y)$. In order to apply the two-scale criterion, it remains to show that there are $\rho_{DG}, \bar{\rho} > 0$, independent of M (and hence N) and y, such that

$$\forall x_\perp \in W_y : \mathrm{LSI}(\mu^N(dx_\parallel|x_\perp)) \geq \rho_{DG}, \tag{2.8.8}$$

and

$$\mathrm{LSI}(\bar{\mu}^N(dx_\perp|Px_\perp = y)) \geq \bar{\rho}. \tag{2.8.9}$$

Additionally, we have to show that

$$\frac{1}{\rho_{DG}}\frac{1}{\bar{\rho}}\sup_{P^{-1}(y)}|\nabla_V\nabla_{W_y}H_N|^2 \leq \kappa < \infty, \tag{2.8.10}$$

for some constant κ that is independent of N and y.

Argument for (2.8.8): This follows from Theorem 2.8.7 and the representation

$$\mu^N(dx_\parallel|x_\perp) = \mu^N(dx|Q_Mx = Q_Mx_\perp).$$

Argument for (2.8.9): The proof follows the same strategy of the proof of (2.8.3) in the proof of Theorem 2.8.7: compute the Hamiltonian $\hat{H}_{W_y}$ of the measure $\bar{\mu}^N(dx_\perp|Px_\perp = y)$, relate it to $\bar{H}_{Y_M^{DG}}$, transfer the uniform strong convexity of the latter to the former, and apply Bakry-Émery criterion to get the desired statement with $\bar{\rho} := \lambda$. We leave the details as an exercise.

Argument for (2.8.10): This follows from the uniform C^2 bound of H_N.

Overall, we may hence apply Lemma 2.8.6 which yields that for $K \geq K^*$:

$$\mathrm{LSI}(\mu^N(dx|Px = y)) \geq \rho,$$

where the constant

$$\rho := \frac{1}{2}\left(\rho_{DG} + (1+\kappa)\lambda - \sqrt{(\rho_{DG} + (1+\kappa)\lambda)^2 - 4\rho_{DG}\lambda}\right).$$

is uniformly bounded from below in N. □

2.9 Closeness of the gradients of the coarse-grained Hamiltonian and the macroscopic free energy

Our goal is to show that the gradient of the coarse-grained Hamiltonian $\bar{H}_{Y_M} : Y_M \to \mathbb{R}$ in (2.3.3) converges to the gradient of the macroscopic free energy $\mathcal{H} : L^2(\mathbb{T}) \to \mathbb{R}$ in (2.2.14). As in the proofs of Theorems 2.3.6 and 2.4.12, the argument will go through the intermediate space Y_M^{DG} in (2.7.1). The argument will involve approximations by several mesoscopic free energies on the spaces Y_M and Y_M^{DG} introduced in this section. We begin with the observation that the macroscopic Hamiltonian $\mathcal{H}$ has the dual formulation:

$$\mathcal{H}(y) = \sup_{\hat{y} \in L^2(\mathbb{T})} \Big(\langle y, \hat{y} \rangle_{L^2} - \varphi^*(\hat{y}) \Big),$$

where $\varphi^* : L^2(\mathbb{T}) \to \mathbb{R}$ is the function given by

$$\varphi^*(\hat{y}) := \int_0^1 \psi^*(\hat{y}(\theta)) d\theta \tag{2.9.1}$$

and the supremum is attained by $\hat{y} = \varphi'(y)$.

This motivates the following definition of the mesoscopic free energy $\mathcal{H}_{Y_M^{DG}}$ on the space Y_M^{DG}.

Definition 2.9.1. *Let $\mathcal{H}_{Y_M^{DG}} : Y_M^{DG} \to \mathbb{R}$ be the function given by*

$$\mathcal{H}_{Y_M^{DG}}(z) := \sup_{\hat{z} \in Y_M^{DG}} \left(\langle z, \hat{z} \rangle_{L^2} - \varphi_N^*(\hat{z}) \right),$$

where the function $\varphi_N^* : Y_M^{DG} \to \mathbb{R}$ is a discretized version of φ^* in (2.9.1), given by

$$\varphi_N^*(\hat{z}) := \frac{1}{N} \sum_{i=1}^{N} \psi^* \left(N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \hat{z}(s) ds \right) \overset{(2.7.15)}{=} \frac{1}{M} \sum_{m=1}^{M} \bar{\psi}_K^*(\hat{z}^{(m)}), \tag{2.9.2}$$

where $\hat{z}^{(m)} \in Y_1^{DG}, m = 1, \cdots, M$, are obtained by restricting $\hat{z} \in Y_M^{DG}$ to subintervals (cf. (2.7.4)).

The main ingredient of the proof is Lemma 2.9.2 below which states that $\bar{H}_{Y_M^{DG}}$ is close to $\mathcal{H}_{Y_M^{DG}}$ in C^2 for large K, which has been essentially established in Section 2.7. The rest

of the proof consists of arguing that $\bar{H}_{Y_M}$ is close to $\bar{H}_{Y_M^{DG}}$ and $\mathcal{H}$ is close to $\mathcal{H}_{Y_M^{DG}}$. The fact that $\bar{H}$ is close to $\bar{H}_{Y_M^{DG}}$ follows from formula (2.7.9),

$$\bar{H}(y) = -\frac{1}{N}\log\int_{Y_M^{\perp}} \exp(-N\bar{H}_{Y_M^{DG}}(y+z))N^{\frac{LM+1}{2}}\mathcal{L}^{LM+1}(dz),$$

and the fact that $\bar{H}_{Y_M^{DG}}$ is uniformly strongly convex (see Lemma 2.9.3), and so the integral on the right hand side concentrates around the minimum of $\bar{H}_{Y_M^{DG}}$ for large K. The fact that $\mathcal{H}$ is close to $\mathcal{H}_{Y_M^{DG}}$ follows from the observation that as $N \to \infty$ the function φ_N^* given by (2.9.2) converges to the function φ^* and that as $M \to \infty$ the spline space $Y_M \subset L^2(\mathbb{T})$ approximates the full space $L^2(\mathbb{T})$. In the next section we collect the technical details for this argument.

2.9.1 Auxiliary results

The first auxiliary result is that $\bar{H}_{Y_M^{DG}}$ converges to $\mathcal{H}_{Y_M^{DG}}$ in C^2 as $K \to \infty$:

Lemma 2.9.2. *There exists K^* such that for $K \geq K^*$ and for all M and $z \in Y_M^{DG}$,*

$$\left|\bar{H}_{Y_M^{DG}}(z) - \mathcal{H}_{Y_M^{DG}}(z)\right| \lesssim \frac{1}{K}, \tag{2.9.3}$$

$$\left\|\nabla\bar{H}_{Y_M^{DG}}(z) - \nabla\mathcal{H}_{Y_M^{DG}}(z)\right\| \lesssim \frac{1}{K}, \tag{2.9.4}$$

$$\|\operatorname{Hess}\bar{H}_{Y_M^{DG}}(z) - \operatorname{Hess}\mathcal{H}_{Y_M^{DG}}(z)\| \lesssim \frac{1}{K}. \tag{2.9.5}$$

Proof of Lemma 2.9.2. By Lemma 2.7.6 and Definition 2.9.1,

$$\bar{H}_{Y_M^{DG}}(z) = \frac{1}{M}\sum_{m=1}^{M}\psi_K(z^{(m)}), \quad \mathcal{H}_{Y_M^{DG}}(z) = \frac{1}{M}\sum_{m=1}^{M}\bar{\psi}_K(z^{(m)}).$$

Taking into account the different Euclidean structures on Y_M^{DG} and Y_1^{DG} as in the proof of Theorem 2.7.5, we see that the estimate (2.9.3) follows from (2.7.29), the estimate (2.9.4) follows from (2.7.30) and the estimate (2.9.5) follows from (2.7.31). □

The next auxiliary result is that the coarse-grained Hamiltonians $\bar{H}_{Y_M}$ and $\bar{H}_{Y_M^{DG}}$ and the free energies $\mathcal{H}_{Y_M^{DG}}$ and $\mathcal{H}$ are uniformly strongly convex. We summarize those results in the following lemma.

Lemma 2.9.3. *There are constants $0 < \lambda < \Lambda < \infty$ and K_0 such that if $K \geq K_0$ then for all $z \in Y_M^{DG}$,*

$$\lambda \operatorname{id}_{Y_M^{DG}} \leq \operatorname{Hess} \bar{H}_{Y_M^{DG}}(z) \leq \Lambda \operatorname{id}_{Y_M^{DG}}, \tag{2.9.6}$$

$$\lambda \operatorname{id}_{Y_M^{DG}} \leq \operatorname{Hess} \mathcal{H}_{Y_M^{DG}}(z) \leq \Lambda \operatorname{id}_{Y_M^{DG}}, \tag{2.9.7}$$

$$\lambda \operatorname{id}_{Y_M^{DG}} \leq \operatorname{Hess} \varphi_N^*(z) \leq \Lambda \operatorname{id}_{Y_M^{DG}}. \tag{2.9.8}$$

Under the same conditions, for all $z \in Y_M$,

$$\lambda \operatorname{id}_{Y_M} \leq \operatorname{Hess} \bar{H}_{Y_M}(z) \leq \Lambda \operatorname{id}_{Y_M}. \tag{2.9.9}$$

Finally, for all $z \in L^2(\mathbb{T})$,

$$\lambda \operatorname{id}_{L^2} \leq \operatorname{Hess} \varphi^*(z) \leq \Lambda \operatorname{id}_{L^2}, \tag{2.9.10}$$

$$\lambda \operatorname{id}_{L^2} \leq \operatorname{Hess} \mathcal{H}(z) \leq \Lambda \operatorname{id}_{L^2}. \tag{2.9.11}$$

All inequalities are in the sense of quadratic forms.

Proof of Lemma 2.9.3. The estimate (2.9.6) is given by Theorem 2.7.5. The estimate (2.9.8) follows from (2.7.16), from which the estimate (2.9.7) follows by basic properties of Legendre transform. The estimate (2.9.9) is given by Theorem 2.3.6. The estimates (2.9.10) and (2.9.11) follow from the uniform strong convexity of ψ^*, φ^* and uniform bound of $(\psi^*)'', (\varphi^*)''$ (cf. (2.7.13)). □

The next auxiliary statement shows a nice relation between the Hamiltonians $\bar{H}_{Y_M}$ and $\bar{H}_{Y_M^{DG}}$. Recall that $Y_M^\perp := \{z \in Y_M^{DG} : Pz = 0\}$.

Lemma 2.9.4. *For each $y \in Y_M$, there is a unique $\bar{z}^* \in Y_M^\perp$ such that*

$$\bar{H}_{Y_M^{DG}}(y + \bar{z}^*) = \inf_{z \in Y_M^\perp} \bar{H}_{Y_M^{DG}}(y + z). \tag{2.9.12}$$

It follows that $\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) \in Y_M$. Moreover,

$$|\nabla \bar{H}(y) - P \nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*)|_{L^2} \lesssim \frac{1}{K^{\frac{1}{2}}}. \tag{2.9.13}$$

Proof of Lemma 2.9.4. The statement (2.9.12) follows from the strong convexity (2.9.6) of $\bar{H}_{Y_M^{DG}}$, applied to the affine subspace $y + Y_M^\perp$, and the second statement then follows directly from this variational characterization of $\bar{z}^*$. Let us now turn to the verification of (2.9.13). Taking gradient in (2.7.9) yields that

$$\nabla \bar{H}(y) = \int_{Y_M^\perp} P \nabla \bar{H}_{Y_M^{DG}}(y+z) \nu(dz|y),$$

where $\{\nu(\cdot|y)\}$ is the family of probability measures on $Y_M^\perp$ given by

$$\nu(dz|y) = \frac{1}{Z_y} \exp\left(-N \bar{H}_{Y_M^{DG}}(y+z)\right) \mathcal{L}^{LM+1}(dz)$$

for some normalization constant Z_y. Now, the convexity bound (2.9.6) implies that

$$\begin{aligned}
\left| P\nabla \bar{H}_{Y_M^{DG}}(y+z) - P\nabla \bar{H}_{Y_M^{DG}}(y+\bar{z}^*) \right|^2 &\leq \Lambda^2 \left| z - \bar{z}^* \right|^2 \\
&\leq \frac{\Lambda^2}{\lambda} \langle z - \bar{z}^*, \nabla \bar{H}_{Y_M^{DG}}(y+z) - \nabla H_{Y_M^{DG}}(y+\bar{z}^*) \rangle_{L^2} \\
&= \frac{\Lambda^2}{\lambda} \langle z - \bar{z}^*, \nabla_z \bar{H}_{Y_M^{DG}}(y+z) \rangle_{L^2},
\end{aligned}$$

where we used $\nabla H_{Y_M^{DG}}(y+\bar{z}^*) \in Y_M$ in the last step. Thus, by Cauchy-Schwarz and integration by parts,

$$\begin{aligned}
|\nabla \bar{H}(y) - P\nabla \bar{H}_{Y_M^{DG}}(y+\bar{z}^*)|^2 &\leq \int_{Y_M^\perp} \left| P\nabla_y \bar{H}_{Y_M^{DG}}(y+z) - P\nabla_y \bar{H}_{Y_M^{DG}}(y+\bar{z}^*) \right|^2 \nu(dz|y) \\
&\lesssim \int_{Y_M^\perp} \langle z - \bar{z}^*, \nabla_z \bar{H}_{Y_M^{DG}}(y+z) \rangle_{L^2} \ \nu(dz|y) \\
&= -\frac{1}{N} \frac{1}{Z_y} \int_{Y_M^\perp} \langle z - \bar{z}^*, \nabla_z \exp(-N \bar{H}_{Y_M^{DG}}(y+z)) \rangle_{L^2} \mathcal{L}^{LM+1}(dz) \\
&= \frac{1}{N} \int_{Y_M^\perp} \nabla \cdot (z - \bar{z}^*) \, \nu(dz|y) \\
&= \frac{\dim Y_M^\perp}{N} = \frac{LM+1}{N},
\end{aligned}$$

which implies the desired estimate (2.9.13). □

Let us introduce a mesoscopic free energy $\mathcal{H}_{Y_M}$ on the spline space Y_M.

Definition 2.9.5. *Let $\mathcal{H}_{Y_M} : Y_M \to \mathbb{R}$ be the function given by*

$$\mathcal{H}_{Y_M}(y) = \sup_{\hat{y} \in Y_M} \left(\langle y, \hat{y} \rangle_{L^2} - \varphi_N^*(\hat{y}) \right),$$

where $\varphi_N^*(\hat{y})$ is given by (2.9.2).

The next auxiliary statement shows a nice relation between the mesoscopic free energies $\mathcal{H}_{Y_M}$ and $\mathcal{H}_{Y_M^{DG}}$.

Lemma 2.9.6. *For each $y \in Y_M$, there is a unique $z^* \in Y_M^\perp$ such that*

$$\mathcal{H}_{Y_M^{DG}}(y + z^*) = \inf_{z \in Y_M^\perp} \mathcal{H}_{Y_M^{DG}}(y + z). \tag{2.9.14}$$

It follows that $\nabla \mathcal{H}_{Y_M^{DG}}(y + z^*) \in Y_M$. Moreover,

$$\mathcal{H}_{Y_M}(y) = \mathcal{H}_{Y_M^{DG}}(y + z^*) \quad \textit{and} \quad \nabla \mathcal{H}_{Y_M}(y) = \nabla \mathcal{H}_{Y_M^{DG}}(y + z^*). \tag{2.9.15}$$

Proof of Lemma 2.9.6. The unique existence of z^* follows directly from the strong convexity of $\mathcal{H}_{Y_M^{DG}}$, and the second statement then follows directly from this variational characterization of z^*. Moreover, $\mathcal{H}_{Y_M^{DG}}$ and $\mathcal{H}_{Y_M}$ are defined as Legendre transforms of the same free energy φ_N^* in different mesoscopic spaces:

$$\begin{aligned} \mathcal{H}_{Y_M^{DG}}(y + z^*) &= \sup_{\hat{z} \in Y_M^{DG}} \langle y + z^*, \hat{z} \rangle_{L^2} - \varphi_N^*(\hat{z}) \\ &\geq \sup_{\hat{z} \in Y_M} \langle y, \hat{z} \rangle_{L^2} - \varphi_N^*(\hat{z}) = \mathcal{H}_{Y_M}(y). \end{aligned}$$

By basic properties of Legendre transform, the maximization problem in the first line has the unique solution $\hat{z} = \nabla \mathcal{H}_{Y_M^{DG}}(y + z^*)$ and the maximization problem in the second line has the unique solution $\hat{z} = \nabla \mathcal{H}_{Y_M}(y)$. But since $\nabla \mathcal{H}_{Y_M^{DG}}(y + z^*) \in Y_M$, it must also solve the maximization problem in the second line. This verifies (2.9.15). □

As a consequence of the results above, we deduce the following nice relation between the mesoscopic free energies $\bar{H}_{Y_M^{DG}}$ and $\mathcal{H}_{Y_M^{DG}}$.

Corollary 2.9.7. *Let $\bar{z}^*$ be as in* (2.9.12) *and $\bar{z}$ as in* (2.9.14)*, then*

$$|\bar{z}^* - z^*|_{L^2} \lesssim \frac{1}{K}, \tag{2.9.16}$$

and

$$|\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) - \nabla \mathcal{H}_{Y_M^{DG}}(y + z^*)|_{L^2} \lesssim \frac{1}{K}. \tag{2.9.17}$$

Proof. Argument for (2.9.16): Using the convexity bound (2.9.6) of $\bar{H}_{Y_M^{DG}}$, the fact that $\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) \in Y_M$ and $\nabla \mathcal{H}_{Y_M^{DG}}(y + z^*) \in Y_M$ (cf. Lemmas 2.9.4 and 2.9.6), and the closeness between $\bar{H}_{Y_M^{DG}}$ and $\mathcal{H}_{Y_M^{DG}}$ in C^2 (cf. Lemma 2.9.2) , we obtain

$$\begin{aligned}
|\bar{z}^* - z^*|_{L^2}^2 &\leq \frac{1}{\lambda} \langle \nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) - \nabla \bar{H}_{Y_M^{DG}}(y + z^*), \bar{z}^* - z^* \rangle_{L^2} \\
&= \frac{1}{\lambda} \langle \nabla \mathcal{H}_{Y_M^{DG}}(y + z^*) - \nabla \bar{H}_{Y_M^{DG}}(y + z^*), \bar{z}^* - z^* \rangle_{L^2} \\
&\overset{(2.9.4)}{\lesssim} \frac{1}{\lambda} \frac{1}{K} |\bar{z}^* - z^*|_{L^2},
\end{aligned}$$

which yields the desired estimate.

Argument for (2.9.17): This follows from

$$\begin{aligned}
|\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) - \nabla \bar{H}_{Y_M^{DG}}(y + z^*)|_{L^2} &\overset{(2.9.6)}{\leq} \Lambda |\bar{z}^* - z^*|_{L^2} \overset{(2.9.16)}{\lesssim} \frac{1}{K}, \\
|\nabla \bar{H}_{Y_M^{DG}}(y + z^*) - \nabla \mathcal{H}_{Y_M^{DG}}(y + z^*)|_{L^2} &\overset{(2.9.4)}{\lesssim} \frac{1}{K}.
\end{aligned}$$

□

Let us now introduce another mesoscopic free energy $\hat{\mathcal{H}}_{Y_M}$ on the spline space Y_M.

Definition 2.9.8. *Let $\hat{\mathcal{H}}_{Y_M} : Y_M \to \mathbb{R}$ be the function given by*

$$\hat{\mathcal{H}}_{Y_M}(y) = \sup_{\hat{y} \in Y_M} \left(\langle y, \hat{y} \rangle_{L^2} - \varphi^*(\hat{y}) \right),$$

where $\varphi^*(\hat{y})$ is defined in (2.9.1)*.*

The next auxiliary result shows that $\nabla \mathcal{H}_{Y_M}$ and $\nabla \hat{\mathcal{H}}_{Y_M}$ are close.

Lemma 2.9.9. *It holds that for any $y \in Y_M$*

$$|\nabla \mathcal{H}_{Y_M}(y) - \nabla \hat{\mathcal{H}}_{Y_M}(y)|_{L^2} \lesssim \frac{1}{K}|y|_{L^2}. \tag{2.9.18}$$

Proof. By basic properties of the Legendre transform, we have the duality relations

$$\begin{aligned} \nabla \mathcal{H}_{Y_M}(y) = \hat{y}_N \in Y_M \text{ and } P\left(\nabla \varphi_N^*\right)(\hat{y}_N) = y, \\ \nabla \hat{\mathcal{H}}_{Y_M}(y) = \hat{y} \in Y_M \quad \text{and} \quad P\left(\nabla \varphi^*\right)(\hat{y}) = y, \end{aligned} \tag{2.9.19}$$

In particular, $\hat{y}_N - \hat{y} \in Y_M$ and $\nabla \varphi_N^*(\hat{y}_N) - \nabla \varphi^*(\hat{y}) \perp Y_M$. Together with the convexity bound (2.9.10) of φ^*, this gives

$$\begin{aligned} \lambda |\hat{y}_N - \hat{y}|_{L^2}^2 &\leq \langle \nabla \varphi^*(\hat{y}_N) - \nabla \varphi^*(\hat{y}), \hat{y}_N - \hat{y} \rangle_{L^2} \\ &= \langle \nabla \varphi^*(\hat{y}_N) - \nabla \varphi_N^*(\hat{y}_N), \hat{y}_N - \hat{y} \rangle_{L^2} \\ &\leq |\nabla \varphi^*(\hat{y}_N) - \nabla \varphi_N^*(\hat{y}_N)|_{L^2} |\hat{y}_N - \hat{y}|_{L^2}. \end{aligned} \tag{2.9.20}$$

Using the definitions (2.9.1) and (2.9.2) of φ^* and φ_N^*, we find

$$|\nabla \varphi^*(\hat{y}_N) - \nabla \varphi_N^*(\hat{y}_N)|_{L^2}^2 \tag{2.9.21}$$

$$= \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left| (\psi^*)'(\hat{y}_N(\theta)) - (\psi^*)' \left(N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \hat{y}_N(s) ds \right) \right|^2 d\theta \tag{2.9.22}$$

$$\overset{(2.7.13)}{\lesssim} \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left| \hat{y}_N(\theta) - N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \hat{y}_N(s) ds \right|^2 d\theta \tag{2.9.23}$$

$$\lesssim \frac{1}{N^2} \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |\hat{y}_N'(\theta)|^2 d\theta \tag{2.9.24}$$

$$= \frac{1}{N^2} |\hat{y}_N|_{H^1}^2 \overset{(2.6.1)}{\lesssim} \frac{M^2}{N^2} |\hat{y}_N|_{L^2}^2, \tag{2.9.25}$$

where we used a Poincaré inequality on an interval of length $\frac{1}{N}$ and then an inverse Sobolev inequality (2.6.1) on Y_M from Section 2.6. By the convexity bound (2.9.8) of φ_N^* and the fact that $\nabla \varphi_N^*(0) = (\psi^*)'(0) = 0$ (see the discussion preceding Assumption 1),

$$\begin{aligned} \lambda |\hat{y}_N|_{L^2}^2 &\leq \langle \nabla \varphi_N^*(\hat{y}_N) - \nabla \varphi_N^*(0), \hat{y}_N \rangle_{L^2} \\ &= \langle P \nabla \varphi_N^*(\hat{y}_N), \hat{y}_N \rangle_{L^2} \end{aligned}$$

$$\overset{(2.9.19)}{=} \langle y, \hat{y}_N \rangle_{L^2} \leq |y|_{L^2} |\hat{y}_N|_{L^2},$$

and hence $|\hat{y}_N|_{L^2} \lesssim |y|_{L^2} \leq |x|_{L^2}$. Combining this with (2.9.20) yields the desired estimate.

□

The last auxiliary result shows that $\nabla \hat{\mathcal{H}}_{Y_M}$ and $\nabla \mathcal{H}$ are close.

Lemma 2.9.10. *It holds that for any $x \in L^2_0(\mathbb{T})$*

$$|\nabla \hat{\mathcal{H}}_{Y_M}(Px) - \nabla \mathcal{H}(x)|_{L^2} \lesssim \frac{1}{M} |x|_{H^1}. \tag{2.9.26}$$

Proof of Lemma 2.9.10. By basic properties of the Legendre transform, we have the duality relations

$$\begin{aligned} \nabla \hat{\mathcal{H}}_{Y_M}(Px) = \hat{y} \in Y_M \text{ and } \quad & P\left(\nabla \varphi^*\right)(\hat{y}) = Px, \\ \nabla \mathcal{H}(x) = \hat{x} \in L^2 \text{ and } \quad & \left(\nabla \varphi^*\right)(\hat{x}) = x. \end{aligned}$$

In particular, $\hat{y} - P\hat{x} \in Y_M$ and $\nabla \varphi^*(\hat{y}) - \nabla \varphi^*(\hat{x}) \perp Y_M$. Together with the convexity bound (2.9.10), this gives

$$\begin{aligned} \lambda |\hat{y} - \hat{x}|^2_{L^2} &\leq \langle \nabla \varphi^*(\hat{y}) - \nabla \varphi^*(\hat{x}), \hat{y} - \hat{x} \rangle_{L^2} \\ &= \langle \nabla \varphi^*(\hat{y}) - \nabla \varphi^*(\hat{x}), P\hat{x} - \hat{x} \rangle_{L^2} \\ &\leq \Lambda |\hat{y} - \hat{x}|_{L^2} |\hat{x} - P\hat{x}|_{L^2}. \end{aligned} \tag{2.9.27}$$

Now, by (2.6.3) from Section 2.6 and the uniform bound on φ'',

$$\begin{aligned} |\hat{x} - P\hat{x}|_{L^2} \lesssim \frac{1}{M} |\hat{x}|_{H^1} &= \frac{1}{M} |\nabla \mathcal{H}(x)|_{H^1} \\ &\lesssim \frac{1}{M} |\partial_\theta x|_{L^2} = \frac{1}{M} |x|_{H^1}, \end{aligned}$$

since $|\nabla \mathcal{H}(x)|_{H^1} = |\varphi'(x)|_{H^1} = |\partial_\theta \varphi'(x)|_{L^2} = |\varphi''(x) \partial_\theta x|_{L^2}$. Combined with (2.9.27), this gives the desired estimate.

□

2.9.2 Proof of Lemma 2.5.2

Using the auxiliary results that were provided in Section 2.9.1, Theorem 2.5.2 is straightforward to prove.

Proof of Theorem 2.5.2. For any $\zeta \in L^2_0(\mathbb{T})$ and $y = P\zeta$,

$$\begin{aligned}
|\nabla \bar{H}(P\zeta) - \nabla \mathcal{H}(\zeta)|_{L^2} \leq\ & |\nabla \bar{H}(y) - P\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*)|_{L^2} \\
& + |P\nabla \bar{H}_{Y_M^{DG}}(y + \bar{z}^*) - P\nabla \mathcal{H}_{Y_M^{DG}}(y + z^*)|_{L^2} \\
& + |P\nabla \mathcal{H}_{Y_M^{DG}}(y + z^*) - \nabla \mathcal{H}_{Y_M}(y)|_{L^2} \\
& + |\nabla \mathcal{H}_{Y_M}(y) - \nabla \hat{\mathcal{H}}_{Y_M}(y)|_{L^2} \\
& + |\nabla \hat{\mathcal{H}}_{Y_M}(y) - \nabla \mathcal{H}(\zeta)|_{L^2}.
\end{aligned}$$

The first term on the right hand side is estimated by (2.9.13). The second term is estimated by (2.9.17). The third term vanishes by (2.9.15). The fourth term is estimated by (2.9.18). The fifth term is estimated by (2.9.26). Summing up yields the desired estimate (2.5.3). □

2.10 Proof of Proposition 2.7.14: a multivariate local CLT

We now begin with the rather long and technical proof of Proposition 2.7.14. We recommend the interested reader to first read the proof of Proposition 31 in [GOVW09]. As in the usual proof of the (local) central limit theorem, we use independence and the Fourier transform to obtain an explicit formula for $g_{J,\beta}(0)$. This is the starting point of our further analysis.

Lemma 2.10.1. *Let*

$$h(m, z) := e^{-imz} \int_{\mathbb{R}} e^{izx}\, \mu_m(dx) \tag{2.10.1}$$

be the characteristic function of the centered version of the measure μ_m given by (2.7.12).

Then $g_{J,\beta}(0)$ can be represented as

$$g_{J,\beta}(0) \;=\; \left(\frac{1}{2\pi}\right)^{L+1} \int_{\mathbb{R}^{L+1}} \prod_{j=1}^{J} h(m_{j,\beta}\,,\; J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\,d\xi. \tag{2.10.2}$$

Proof of Lemma 2.10.1. Applying Fourier transform,

$$\begin{aligned}
(2\pi)^{L+1} g_{J,\beta}(0) \;&=\; \int_{\mathbb{R}^{L+1}} \int_{\mathbb{R}^{L+1}} \exp\left(i\,\xi\cdot u\right) g_{J,\beta}(u)\,du\,d\xi \\
&= \int_{\mathbb{R}^{L+1}} \int_{\mathbb{R}^{J}} \exp\Big(i\,\xi\cdot\; J^{\frac{1}{2}}(Q_1 x-\beta)\Big)\,\nu_{J,\beta}(dx)\,d\xi \\
&= \int_{\mathbb{R}^{L+1}} \int_{\mathbb{R}^{J}} \exp\left(i\,\xi\cdot\; J^{\frac{1}{2}}(Q_1 x-\int Q_1\tilde{x}\nu_{J,\beta}(d\tilde{x}))\right)\nu_{J,\beta}(dx)\,d\xi \\
&\overset{(2.7.22)}{=} \int_{\mathbb{R}^{L+1}} \prod_{j=1}^{J} \int_{\mathbb{R}} \exp\Big(i\,J^{-\frac{1}{2}}\,(JQ_1^t\xi)_j\,(x_j-m_{j,\beta})\Big)\;\mu_{m_{j,\beta}}(dx_j)\,d\xi \\
&\overset{(2.10.1)}{=} \int_{\mathbb{R}^{L+1}} \prod_{j=1}^{J} h(m_{j,\beta}\,,\; J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\,d\xi,
\end{aligned}$$

as desired. □

To continue from formula (2.10.2) we need two ingredients. The first ingredient is a collection of elementary properties of the function h.

Lemma 2.10.2. *We have the following bounds and decay properties for the function h and its derivatives:*

$$|h(m,z)|\le 1. \tag{2.10.3}$$

Given $\varepsilon>0$, there is $C_\varepsilon<\infty$ (uniform in m), such that for $|z|\ge\varepsilon$:

$$|h(m,z)| \;\le\; \frac{1}{1+|z|\,C_\varepsilon^{-1}}. \tag{2.10.4}$$

For all $z\in\mathbb{R}$, $m\in\mathbb{R}$:

$$\left|\frac{\partial h}{\partial m}(m,z)\right|\le C(1+|z|),\quad \left|\frac{\partial^2 h}{\partial m^2}(m,z)\right|\le C(1+|z|^2). \tag{2.10.5}$$

There is $\delta_0>0$ such that for $z\in[-\delta_0,\delta_0]$, $m\in\mathbb{R}$, we can express h as

$$h(m,z)=\exp(-z^2\,h_2(m,z)). \tag{2.10.6}$$

Here, h_2 is a (complex-valued) function satisfying

$$h_2(m,0) = \frac{\mathrm{Var}(\mu_m)}{2},\ 0 < c \le Re\ h_2(m,z) \le C\ \forall z \in [-\delta_0, \delta_0], \tag{2.10.7}$$

$$\left|\frac{\partial h_2}{\partial z}(m,z)\right| \le C, \quad \left|\frac{\partial h_2}{\partial m}(m,z)\right| \le C, \quad \left|\frac{\partial^2 h_2}{\partial m^2}(m,z)\right| \le C. \tag{2.10.8}$$

The estimates of Lemma (2.10.2) should not be surprising as $h(m,\cdot)$ is just the Fourier transform of μ_m which belongs to the exponential family of a perturbed standard Gaussian measure. For the proofs, we refer the reader to [GOVW09].

The second ingredient for the proof of Proposition 2.7.14 is a lower bound on the inner products $\xi \cdot \gamma^j$ which enter into the second argument of h. This is new compared to [GOVW09].

Lemma 2.10.3. *Fix $L+1$ disjoint closed subintervals of $[0,1]$ of length $1/(L+2)$, denoted I_k, $1 \le k \le L+1$. There exists a constant $c_\gamma > 0$ and an integer J_γ such that for $J \ge J_\gamma$, each subinterval I_k contains at least $J/(L+2)$ spins and, for every $\xi \in \mathbb{R}^{L+1}$, there exists a subinterval $I_{k(\xi)}$ on which*

$$|\xi \cdot \gamma^j| \ge c_\gamma |\xi| \quad \textit{for every spin } \frac{j}{J} \in I_{k(\xi)}. \tag{2.10.9}$$

The proof is postponed to the end of this section. The strategy for the rest of the proof is to split the integral on the right hand side of (2.10.2) into an inner and an outer part. We will show that for sufficiently small δ and for sufficiently large J (depending on δ)

$$\lim_{J\to\infty} \int_{\left\{|\xi| > J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} |\, h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)|\, d\xi = 0, \tag{2.10.10}$$

$$\int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} |\, h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)|\, d\xi \ \le\ C, \tag{2.10.11}$$

$$\left| \int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\, d\xi \right| \ \ge\ \frac{1}{C}, \tag{2.10.12}$$

$$\lim_{J\to\infty} \left\| \mathrm{Hess} \int_{\left\{|\xi| > J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\, d\xi \right\| = 0, \tag{2.10.13}$$

$$\left\| \text{Hess} \int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\; d\xi \right\| \;\le\; C. \tag{2.10.14}$$

The bounds for $g_{J,\beta}(0)$ in (2.7.25) follows from (2.10.10) - (2.10.12). The bounds for the Hessian in (2.7.27) follows from (2.10.13) and (2.10.14). The bounds for the gradient in (2.7.26) is then immediate from interpolation.

Let us assume for the rest of the proof that $J \ge J_\gamma$. First consider the outer integral from (2.10.10). Recall that the interval $I_{k(\xi)}$ contains at least $J/(L+2)$ spins (cf. Lemma 2.10.3). With this and the decay property (2.10.4) of h in mind, we set $\varepsilon := \delta\, c_\gamma$ and compute:

$$\begin{aligned}
&\int_{\left\{|\xi| > J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} |\, h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)|\, d\xi \\
&\overset{(2.10.3)}{\le} \int_{\left\{|\xi| > J^{\frac{1}{2}}\delta\right\}} \prod_{j:\,\frac{j}{J}\in I_{k(\xi)}} |\, h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)|\, d\xi \\
&\overset{(2.10.4),(2.10.9)}{\le} \int_{\left\{|\xi| > J^{\frac{1}{2}}\delta\right\}} \prod_{j:\,\frac{j}{J}\in I_{k(\xi)}} \frac{1}{1+J^{-\frac{1}{2}}\,|\xi\cdot\gamma^j|\,C_\varepsilon^{-1}}\, d\xi \\
&\overset{(2.10.9)}{\le} \left(\frac{1}{1+\,\varepsilon\, C_\varepsilon^{-1}}\right)^{\frac{J}{L+2}-L-2} J^{\frac{L+1}{2}} \int_{\{|\xi|>\delta\}} \left(\frac{1}{1+\,c_\gamma C_\varepsilon^{-1}|\xi|}\right)^{L+2} d\xi.
\end{aligned}$$

This goes to 0 as $J \to \infty$.

For the inner integral from (2.10.11) we use the representation of h via h_2 in Lemma 2.10.2. For this purpose, we assume from now on that

$$\delta \;\le\; \delta_0 \quad \text{and} \quad \delta \max_{j=1,\dots,J} |\gamma^j| \le \delta_0. \tag{2.10.15}$$

We compute :

$$\begin{aligned}
&\int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} |\, h(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)|\; d\xi \\
&\overset{(2.10.6),(2.10.15)}{=} \int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} |\exp\left(-\,(J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)^2\; h_2(m_{j,\beta}\,,\, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\right)|\; d\xi \\
&\overset{(2.10.9),(2.10.3),(2.10.7)}{\le} \int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \prod_{\left\{j\,|\,\frac{j}{J}\in I_{k(\xi)}\right\}} \exp\left(-c\,(J^{-\frac{1}{2}}\,c_\gamma|\xi|)^2\right)\; d\xi
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\left\{|\xi| \le J^{\frac{1}{2}}\delta\right\}} \exp\left(-c\,J^{-1}\,\frac{J}{L+2}\,c_\gamma^2\,|\xi|^2\right)\,d\xi \\
&\le \int_{\mathbb{R}^{L+1}} \exp\left(-c\,\frac{1}{L+2}\,c_\gamma^2\,|\xi|^2\right)\,d\xi,
\end{aligned}
$$

which is finite. Note that here we really need that the number of spins in $I_{k(\xi)}$ is of order J, whereas arbitrary growth to infinity of this number suffices for the previous estimate.

We next turn to the lower bound (2.10.12):

$$
\begin{aligned}
&\left|\int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} \prod_{j=1}^{J} h(m_{j,\beta}, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\,d\xi\right| \\
&\quad = \left|\int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} \exp\left(-\sum_{j=1}^{J}(J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)^2\,h_2(m_{j,\beta}, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)\right)\,d\xi\right| \\
&\quad \overset{(2.10.7)}{\ge} \int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} \underbrace{\exp\left(-\sum_{j=1}^{J}(J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)^2\,h_2(m_{j,\beta},0)\right)}_{S_1}\,d\xi \\
&\quad - \int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} \exp\left(-\sum_{j=1}^{J}(J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)^2\,h_2(m_{j,\beta},0)\right) \\
&\qquad \underbrace{\left|\exp\left(-\sum_{j=1}^{J}(J^{-\frac{1}{2}}\,\xi\cdot\gamma^j)^2\,[h_2(m_{j,\beta}, J^{-\frac{1}{2}}\,\xi\cdot\gamma^j) - h_2(m_{j,\beta},0)]\right) - 1\right|}_{S_2}\,d\xi.
\end{aligned}
$$

We estimate the terms S_1, S_2 as (using $|\exp(z)-1| \le \exp(|z|)-1$):

$$
\begin{aligned}
\exp\left(-C\,\max_j |\gamma^j|^2|\xi|^2\right) &\overset{(2.10.7)}{\le} S_1 \overset{(2.10.7),(2.10.9)}{\le} \exp\left(-c\frac{1}{L+2}\,c_\gamma^2\,|\xi|^2\right), \\
&\qquad S_2 \overset{(2.10.8),(2.10.15)}{\le} \exp\left(|\xi|^2 \max_j |\gamma^j|^2 C\delta_0\right) - 1.
\end{aligned}
$$

Set $C_1 := c\frac{1}{L+2}\,c_\gamma^2$ and $C_2 := \max_j |\gamma^j|^2 C$. We thus find:

$$
\begin{aligned}
&\int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} S_1 d\xi - \int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} S_1 S_2 d\xi \\
&\qquad \ge \int_{\left\{|\xi|\le J^{\frac{1}{2}}\delta\right\}} e^{-C_2|\xi|^2}\,d\xi - \int_{\mathbb{R}^{L+1}} e^{-C_1|\xi|^2}\left(e^{C_2\delta_0|\xi|^2} - 1\right)\,d\xi.
\end{aligned}
$$

Now, we choose δ_0 (and accordingly δ) small enough to ensure that the first integral dominates the second integral for large J. This implies (2.10.12) for all sufficiently large J.

Let us now turn to the terms that involve derivatives with respect to β. We first compute $\nabla g_{J,\beta}(0)$ and Hess $g_{J,\beta}(0)$ starting from (2.10.2). The interchange of differentiation and integration will be justified by the bounds developed below, relying on pointwise bounds for the integrands. From now on, we write $[j] := (m_{j,\beta}, J^{-\frac{1}{2}}\xi \cdot \gamma^j)$ for short. We have:

$$(2\,\pi)^{L+1}\,\nabla g_{J,\beta}(0) = \int_{\mathbb{R}^{L+1}} \sum_{j=1}^{J} \frac{\partial h}{\partial m}[j] \prod_{n\neq j} h[j]\,\nabla m_{j,\beta}\,d\xi$$

and

$$\begin{aligned}
&(2\,\pi)^{L+1}\,\operatorname{Hess} g_{J,\beta}(0) \qquad (2.10.16)\\
&= \int_{\mathbb{R}^{L+1}} \sum_{j=1}^{J} \frac{\partial^2 h}{\partial m^2}[j] \prod_{n\neq j} h[n]\,\nabla m_{j,\beta} \otimes \nabla m_{j,\beta}\,d\xi\\
&+ \int_{\mathbb{R}^{L+1}} \sum_{j=1}^{J} \sum_{p\neq j} \frac{\partial h}{\partial m}[j]\,\frac{\partial h}{\partial m}[p] \prod_{n\neq j,p} h[n]\,\nabla m_{j,\beta} \otimes \nabla m_{p,\beta}\,d\xi\\
&+ \int_{\mathbb{R}^{L+1}} \sum_{j=1}^{J} \frac{\partial h}{\partial m}[j] \prod_{n\neq j} h[n]\,\operatorname{Hess} m_{j,\beta}\,d\xi.
\end{aligned}$$

We will need the following auxiliary statement.

Lemma 2.10.4. *It holds uniformly in $j = 1, ..., J$ and β that*

$$\|\nabla m_{j,\beta}\| \leq C, \qquad (2.10.17)$$

$$\|\operatorname{Hess} m_{j,\beta}\| \leq C. \qquad (2.10.18)$$

The proof is postponed to the end of this section. From now on we write $\bar{\xi}_j := J^{-\frac{1}{2}}\xi \cdot \gamma^j$ for short. Denote $O := \left\{|\xi| > J^{\frac{1}{2}}\delta\right\}$. The outer integral of (2.10.13) becomes

$$\begin{aligned}
&\left\| \operatorname{Hess} \int_O \prod_{j=1}^{J} h(m_{j,\beta}, J^{-\frac{1}{2}}\xi \cdot \gamma^j)\,d\xi \right\|\\
&\overset{(2.10.16)-(2.10.18)}{\lesssim} \int_O \sum_{j=1}^{J} \left|\frac{\partial^2 h}{\partial m^2}[j]\right| \prod_{n\neq j} |h[n]|\,d\xi + \int_O \sum_{j=1}^{J} \left|\frac{\partial h}{\partial m}[j]\right| \prod_{n\neq j} |h[n]|\,d\xi\\
&\quad + \int_O \sum_{j=1}^{J} \sum_{p\neq j} \left|\frac{\partial h}{\partial m}[j]\right| \left|\frac{\partial h}{\partial m}[p]\right| \prod_{n\neq j,p} |h[n]|\,d\xi
\end{aligned}$$

$$\overset{(2.10.3)-(2.10.5)}{\lesssim} \int_O \sum_{j=1}^{J} \prod_{n\neq j:\, \frac{n}{J}\in I_{k(\xi)}} \frac{1+|\bar\xi_j|^2}{1+|\bar\xi_n|\, C_\varepsilon^{-1}}\, d\xi + \int_O \sum_{j=1}^{J} \prod_{n\neq j:\, \frac{n}{J}\in I_{k(\xi)}} \frac{1+|\bar\xi_j|}{1+|\bar\xi_n|\, C_\varepsilon^{-1}}\, d\xi$$
$$+ \int_O \sum_{j=1}^{J} \sum_{p\neq j} \prod_{n\neq j,p:\, \frac{n}{J}\in I_{k(\xi)}} \frac{(1+|\bar\xi_j|)(1+|\bar\xi_p|)}{1+|\bar\xi_n|\, C_\varepsilon^{-1}}\, d\xi$$
$$\overset{(2.10.9)}{\lesssim} JA^{\frac{J}{L+2}-L-3} \int_O B(\xi)^{L+2}\, d\xi + \max_j |\gamma^j|^2 A^{\frac{J}{L+2}-L-5} \int_O |\xi|^2\, B(\xi)^{L+4}\, d\xi$$
$$+ J^2 A^{\frac{J}{L+2}-L-4} \int_O B(\xi)^{L+2}\, d\xi + J \max_j |\gamma^j|^2 A^{\frac{J}{L+2}-L-6} \int_O |\xi|^2\, B(\xi)^{L+4}\, d\xi$$
$$+ J^{\frac12} \max_j |\gamma^j| A^{\frac{J}{L+2}-L-4} \int_O |\xi|\, B(\xi)^{L+3}\, d\xi,$$

where we write for short

$$A := \frac{1}{1+\varepsilon\, C_\varepsilon^{-1}}, \quad B(\xi) := \frac{1}{1+J^{-\frac12}|\xi|\, c_\gamma\, C_\varepsilon^{-1}}.$$

In the last step, we have collected like terms after application of the estimate (2.10.9). We also used Young's inequality once. Observe that we always left exactly enough of the factors that were at our disposal, i.e. $J/(L+2)-1$ and $J/(L+2)-2$ respectively, inside the integral to ensure integrability. Performing a change of variables just as in the last step of the proof for (2.10.10), we find that the right hand side goes to zero as $J\to\infty$ because exponential decay beats polynomial growth. This proves (2.10.13).

For the inner integral of (2.10.14), we again use the representation via h_2 from (2.10.6). In this case, we have the following formulas for the derivatives with respect to m:

$$\frac{\partial h}{\partial m}(m,z) \;=\; -z^2\, \frac{\partial h_2}{\partial m}(m,z)\, h(m,z),$$
$$\frac{\partial^2 h}{\partial m^2}(m,z) \;=\; \left(-z^2\, \frac{\partial^2 h_2}{\partial m^2}(m,z) \;+\; z^4 \left(\frac{\partial h_2}{\partial m}(m,z) \right)^2 \right) h(m,z).$$

Denote $I := \left\{ |\xi| \le J^{\frac12}\delta \right\}$. Using the bounds from (2.10.8), we find

$$\left\| \operatorname{Hess} \int_I \prod_{j=1}^{J} h(m_{j,\beta}\,,\, J^{-\frac12}\, \xi\cdot\gamma^j)\, d\xi \right\|$$
$$\overset{(2.10.16)-(2.10.18)}{\lesssim} \int_I \sum_{j=1}^{J} \left(\bar\xi_j^2 \left| \frac{\partial^2 h_2}{\partial m^2}[j] \right| + \bar\xi_j^4 \left| \frac{\partial h_2}{\partial m}[j] \right|^2 \right) \prod_{n=1}^{J} |\exp(-\bar\xi_n^2 h_2[n])|\, d\xi$$

$$
\begin{aligned}
&+ \int_I \sum_{j=1}^{J} \sum_{p \neq j} \bar{\xi}_j^2 \, \bar{\xi}_p^2 \left| \frac{\partial h_2}{\partial m}[j] \right| \left| \frac{\partial h_2}{\partial m}[p] \right| \prod_{n=1}^{J} |\exp(-\bar{\xi}_n^2 h_2[n])| \, d\xi \\
&+ \int_I \sum_{j=1}^{J} \bar{\xi}_j^2 \left| \frac{\partial h_2}{\partial m}[j] \right| \prod_{n=1}^{J} |\exp(-\bar{\xi}_n^2 h_2[n])| \, d\xi \\
&\overset{(2.10.8),(2.10.7)}{\lesssim} \max_j |\gamma^j|^2 \int_I |\xi|^2 \exp\left(-cJ^{-1} \sum_{n: \frac{n}{J} \in I_{k(\xi)}} |\xi|^2 \, c_\gamma^2 \right) d\xi \\
&+ (1 + J^{-1}) \max_j |\gamma^j|^4 \int_I |\xi|^4 \exp\left(-cJ^{-1} \sum_{n: \frac{n}{J} \in I_{k(\xi)}} |\xi|^2 \, c_\gamma^2 \right) d\xi \\
&\lesssim \int_{\mathbb{R}^{L+1}} (|\xi|^2 + |\xi|^4) \exp\left(-c \, c_\gamma^2 \, \frac{1}{L+2} |\xi|^2 \right) d\xi,
\end{aligned}
$$

which is finite. This completes the proof of Proposition 2.7.14, up to the verification of Lemma 2.10.3 and Lemma 2.10.4.

Proof of Lemma 2.10.3. After approximation of (2.7.34), this reduces to the following elementary geometric property of the curve γ defined in (2.7.33): for $\xi \in \mathbb{R}^{L+1}$, define

$$
\omega_k(\xi) \, := \, \inf_{t \in I_k} |\xi \cdot \gamma(t)|,
$$

then there exists a constant $c_\gamma > 0$ such that for all $\xi \in S^L$,

$$
\max_{k=1,\dots,L+2} \omega_k(\xi) \, \geq \, 2 \, c_\gamma, \tag{2.10.19}
$$

where S^L denotes the unit sphere in $\mathbb{R}^{L+1}$.

Argument for (2.10.19): The function $\omega_k : \mathbb{R}^{L+1} \to \mathbb{R}$ is continuous for all $k = 1, ..., L+1$, so the same is true for $\omega := \max_k \omega_k$. As S^L is compact, for (2.10.19) it only remains to show that ω is strictly positive on S^L. But for $\xi \in S^L$, $\xi \cdot \gamma$ is a polynomial of degree $\leq L$ that is not identically zero by independence of the basis polynomials f_l. Hence, it has at most L zeros in $[0, 1]$, which implies $\omega(\xi) > 0$ by pigeonhole principle. $\square$

Proof of Lemma 2.10.4. We recall that

$$
m_{j,\beta} = \int_{\mathbb{R}} z \, \exp(-\psi^*(\hat{m}_{j,\beta}) + \hat{m}_{j,\beta} z - \psi(z)) \, dz.
$$

Standard calculation yields

$$\nabla m_{j,\beta} \overset{(2.7.21)}{=} \mathrm{Var}(\mu_{m_{j,\beta}})\, \nabla \hat{m}_{j,\beta},$$
$$\mathrm{Hess}\, m_{j,\beta} = \mathrm{Var}(\mu_{m_{j,\beta}})\, \mathrm{Hess}\, \hat{m}_{j,\beta} + \left(\int (z - m_{j,\beta})^3\, \mu_{m_{j,\beta}}(dz) \right) \nabla \hat{m}_{j,\beta} \otimes \nabla \hat{m}_{j,\beta}.$$

By the uniform estimates on $\mathrm{Var}(\mu_m)$ and $\int (z-m)^3\, \mu_m(dz)$ in (2.7.13) and (2.7.14), it remains to bound $\nabla \hat{m}_{j,\beta}$ and $\mathrm{Hess}\, \hat{m}_{j,\beta}$. Note that

$$\hat{m}_{j,\beta} \overset{(2.7.20)}{=} \langle \hat{\beta}^{max}, \gamma^j \rangle \overset{(2.7.19)}{=} \langle \nabla \bar{\psi}_J(\beta), \gamma^j \rangle = \partial_{\gamma^j} \bar{\psi}_J(\beta),$$

where ∂_η denotes η-directional derivative. Thus for any $\eta \in \mathbb{R}^{L+1}$:

$$\langle \nabla \hat{m}_{j,\beta}, \eta \rangle = \partial_\eta \partial_{\gamma^j} \bar{\psi}_J(\beta) \leq \| \mathrm{Hess}\, \bar{\psi}_J \| |\eta| |\gamma^j|,$$
$$\langle \mathrm{Hess}\, \hat{m}_{j,\beta} \eta, \eta \rangle = \partial_\eta^2 \partial_{\gamma^j} \bar{\psi}_J(\beta) \leq \| D^3 \bar{\psi}_J \| |\eta|^2 |\gamma^j|.$$

Since $|\gamma^j|$ is uniformly bounded (cf. (2.7.34)) and the Hessian and 3rd derivative of $\bar{\psi}_J$ are uniformly bounded, this concludes the proof of (2.10.17) and (2.10.18). $\square$

Chapter 3

Ergodicity of the infinite swapping algorithm

Sampling Gibbs measures at low temperatures is an important but computationally challenging task. Numerical evidence suggests that the infinite-swapping algorithm (isa) is a promising method. The isa can be seen as an improvement of the replica methods. We rigorously analyze the ergodic properties of the isa in the low temperature regime, deducing an Eyring-Kramers formula for the spectral gap (or Poincaré constant) and an estimate for the log-Sobolev constant. Our main results indicate that the effective energy barrier can be reduced drastically using the isa compared to the classical overdamped Langevin dynamics. As a corollary, we derive a deviation inequality showing that sampling is also improved by an exponential factor. Finally, we study simulated annealing for the isa and prove that the isa again outperforms the overdamped Langevin dynamics.

3.1 Introduction

Sampling from Gibbs measures at low temperatures is important in science and engineering. It has a variety of applications including molecular dynamics [And80, CS11] and Bayesian inference [RC05, GSC+13]. Usually, sampling at low temperatures is slow due to the fact that at low temperatures energy barriers in the underlying energy landscape are large. This

traps the stochastic sampling process and slows down sampling.

A lot of effort has been made to accelerate sampling at low temperatures and there are many competing methods. One of them is the replica exchange method which is also known as parallel tempering. In the simplest version of a replica exchange method, one considers two independent copies of the underlying dynamics. One copy evolves at the desired low temperature $\tau_1 > 0$ and the other copy with a higher temperature $1 \gg \tau_2 \gg \tau_1$. At random times the positions of both particles are swapped. This approach has the advantage that the particle at a low temperature correctly samples the low temperature Gibbs measure whereas the particle at a high temperature can explore the full state space discovering the relevant states of the system.

Replica exchange methods and parallel tempering have been applied successfully in many different situations and they seem to accelerate sampling in low-temperature situations quite well. To the best of our knowledge, almost all evaluations of the performance of those methods are empirical and numerical. In an attempt to study the sampling performance of parallel tempering via large deviations, it was discovered that the large deviation rate function is a monotone function of the swapping rate (see [DLPD12]). It means that sampling can only improve as the swapping rate increases. This led to the discovery of the infinite swapping algorithm/process (isa), which can be interpreted as the limit of parallel tempering when swapping the particles infinitely fast (see [DLPD12], or Section 3.2.1 for details). Formally, given the underlying energy landscape $H : \mathbb{R}^n \to \mathbb{R}$, the isa is defined as the evolution of two particles X_t^1 and X_t^2 varying between two different temperatures $0 < \tau_1 \ll \tau_2$, given by the stochastic differential equations (SDEs):

$$\begin{cases} dX_t^1 &= -\nabla H(X_t^1)\, dt + \sqrt{2\tau_1 \rho(X_t^1, X_t^2) + 2\tau_2 \rho(X_t^2, X_t^1)}\, dB_t^1 \,, \\ dX_t^2 &= -\nabla H(X_t^2)\, dt + \sqrt{2\tau_2 \rho(X_t^1, X_t^2) + 2\tau_1 \rho(X_t^2, X_t^1)}\, dB_t^2 \,, \end{cases} \tag{3.1.1}$$

with

$$\rho(x_1, x_2) := \frac{\pi(x_1, x_2)}{\pi(x_1, x_2) + \pi(x_2, x_1)} \quad \text{and} \quad \pi(x_1, x_2) := \frac{1}{Z} \exp\left(-\frac{H(x_1)}{\tau_1} - \frac{H(x_2)}{\tau_2} \right),$$

where Z is the normalizing constant making π a probability measure. Numerical and heuristic studies [DDN17] indicate that there is an exponential gain when using the isa for sampling instead of the classical overdamped Langevin dynamics. However, no rigorous result has been established so far.

In this work we take the analysis of [DDN17] to the next level. We carry out the first rigorous study of the ergodic properties of the isa at low temperatures. Under standard non-degeneracy assumptions, we deduce the low-temperature asymptotics for the Poincaré constant and a good estimate for the log-Sobolev constant of the isa (see Theorem 3.2.3 and Theorem 3.2.4 below). In the context of metastability, those type of formulas are also known as Eyring-Kramers formulas. Comparing our results to the Eyring-Kramers formula for the overdamped Langevin dynamics (see e.g. [BEGK04, BGK05, MS14]) we have an exponential gain: the effective energy barrier of the underlying energy landscape H only sees the higher temperature τ_2. We also give indications that the result of Theorem 3.2.3 is optimal.

To the best of our knowledge, this is the first time that an Eyring-Kramers formula was derived for inhomogeneous diffusions. The reason is that usually, if the diffusion coefficient σ is inhomogeneous, the stationary and ergodic distribution μ is unknown. But for the isa (3.1.1), the ergodic distribution μ is explicitly known. It is given by $\mu(x_1, x_2) = \frac{1}{2}(\pi(x_1, x_2) + \pi(x_2, x_1))$. This makes a rigorous analysis of (3.1.1) feasible.

For the proof of Theorem 3.2.3 and Theorem 3.2.4, we follow the transportation approach of [MS14]. There are several other methods which could be used to deduce the Eyring-Kramers formula for the Poincaré constant. For example, one could consider to adapt the potential theoretic approach (see [BEGK04, BGK05]) or the approach using semiclassical analysis (see [HKN04, HN05, HN06]). However, it seems that only the approach of [MS14] is robust enough to deduce good estimates for the log-Sobolev constant. This is important for our applications to sampling and simulated annealing.

In the first application, we apply the main results to study the sampling properties of the isa and compare it to the overdamped Langevin dynamics. It is well known that the

Poincaré and log-Sobolev constants characterize the rate of convergence to equilibrium of the underlying process. It is also known that Poincaré and log-Sobolev inequalities yield deviation inequalities (see [CG08, WY08] and references therein). Hence, our main results yield a precise quantitative control on the convergence of the time average to the ensemble average, quantifying the ergodic theorem. As a consequence, we conclude that sampling at low temperatures using isa is exponentially faster than using the overdamped Langevin dynamics.

In the second application, we study simulated annealing for the isa and compare it to simulated annealing for the overdamped Langevin dynamics. Simulated annealing (SA) is an umbrella term denoting a particular set of stochastic optimization methods. SA can be used to find the global extremum of a function $H : \mathbb{R}^n \to \mathbb{R}$, in particular when H is non-convex and n is large. Those methods have many applications in different fields, for example in physics, chemistry and operations research (see e.g. [vLA87, KAJ94, Nar99]). The name and inspiration comes from annealing in metallurgy. It is a process that aims to increase the size of the crystals by a process involving heating and controlled cooling. The SA mimics this procedure mathematically. The stochastic version of SA was independently described by Kirkpatrick, Gelatt and Vecchi [KGV83] and Černý [Č85]. See Section 3.4.2 for details on simulated annealing.

Replica exchange and parallel tempering have been successfully applied to simulated annealing (see e.g. [KZ09, LPA$^+$09]). Because the isa has better ergodic properties than parallel tempering, there is big hope that the isa can produce even better results. Additionally, our main results show that the isa mixes much faster than the overdamped Langevin dynamics. Therefore, one expects that the isa also outperforms the overdamped Langevin dynamics for simulated annealing. In this work, we show that this is indeed the case, though it is unclear from our theoretical study whether the isa could compete in practice with state-of-the-art methods for simulated annealing, e.g. methods based on Lévy flights [Pav07] or Cuckoo's search [YD09].

3.2 Setting and main results

We start by discussing how the isa emerges as the weak limit from parallel tempering. Then we introduce the precise setting and non-degeneracy assumptions. After this we present the main results of this work, the Eyring-Kramers formula for the Poincaré constant and a good estimate for the log-Sobolev constant for the isa. We also give indications that the Poincaré constant is optimal. We close this section by discussing two applications: sampling Gibbs measures at low temperatures and simulated annealing.

3.2.1 Infinite-swapping as the weak limit of parallel tempering

Before describing parallel tempering, let us consider a simpler situation: a single diffusion on an energy landscape given by a sufficiently smooth, non-convex Hamiltonian function $H : \mathbb{R}^n \to \mathbb{R}$ at a single temperature $\tau > 0$, given by the SDE

$$d\xi_t = -\nabla H(\xi_t)dt + \sqrt{2\tau}dB_t,$$

where B_t is a standard Brownian motion on $\mathbb{R}^n$. The generator of the diffusion is

$$L_\tau := \tau\Delta - \nabla H \cdot \nabla.$$

The associated Dirichlet form is

$$\mathcal{E}_{\nu^\tau}(f) := \int_{\mathbb{R}^n} (-L_\tau f) f d\nu^\tau = \int_{\mathbb{R}^n} \tau |\nabla f|^2 d\nu^\tau$$

and the Fisher information is

$$\mathcal{I}_{\nu^\tau}(f^2) := 2\mathcal{E}_{\nu^\tau}(f).$$

Under some growth assumptions on H (e.g. those of [MS14, Section 1.2]), this process (known as the overdamped Langevin dynamics) has an invariant measure given by

$$\nu^\tau(x) := \frac{1}{Z^\tau} \exp\left(-\frac{H(x)}{\tau}\right), \tag{3.2.1}$$

where Z^τ is the normalization constant. Due to the non-convexity of H, this process shows metastable behavior at low temperatures τ in the sense of a separation of time scales:

- In the short run, the process converges fast to a local minimum of the energy landscape.
- In the long run, the process stays near a local minimum for an exponentially long time before it jumps to another local minimum.

In the previous work of [MS14], this behavior is captured by explicit, low-temperature asymptotic formulas (known as Eyring-Kramers formulas) for the two constants $\rho, \alpha > 0$ appearing in the following two functional inequalities for the invariant measure ν^τ: the Poincaré inequality (PI(ρ))

$$\mathrm{Var}_{\nu^\tau}(f) := \int (f - \int f d\nu^\tau)^2 d\nu^\tau \leq \frac{1}{\rho} \mathcal{E}_{\nu^\tau}(f)$$

and the log-Sobolev inequality (LSI(α))

$$\mathrm{Ent}_{\nu^\tau}(f^2) := \int f^2 \log \frac{f^2}{\int f^2 d\nu^\tau} d\nu^\tau \leq \frac{1}{\alpha} \mathcal{I}_{\nu^\tau}(f)$$

holding all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$.

In the present work, we extend these results to a non-homogeneous diffusion, the "infinite swapping process". It arises from parallel tempering, which we now introduce. Given two temperatures $0 < \tau_1 < \tau_2 \ll 1$, $\tau_2 > K\tau_1$ for some $K > 1$, define two product measures on $\mathbb{R}^n \times \mathbb{R}^n$

$$\pi^+(x_1, x_2) := \nu^{\tau_1}(x_1)\nu^{\tau_2}(x_2), \;\; \pi^-(x_1, x_2) := \nu^{\tau_2}(x_1)\nu^{\tau_1}(x_2).$$

Let us identify the symbol $\sigma = +, -$ with the identity and swap permutation on $\{1, 2\}$, respectively. Then π^σ is the invariant measure of the following SDE:

$$\begin{cases} dX_1 = -\nabla H(X_1)\, dt + \sqrt{2\tau_{\sigma(1)}}\, dB_1 \,, \\ dX_2 = -\nabla H(X_2)\, dt + \sqrt{2\tau_{\sigma(2)}}\, dB_2 \,, \end{cases}$$

where $B := (B_1, B_2)$ is a standard Brownian motion in $\mathbb{R}^n \times \mathbb{R}^n$. Its generator is

$$L_\sigma := L^{x_1}_{\tau_{\sigma(1)}} + L^{x_2}_{\tau_{\sigma(2)}}$$

and the associated Dirichlet form is

$$\begin{aligned}\mathcal{E}_{\pi^\sigma}(f) &:= \int_{\mathbb{R}^n\times\mathbb{R}^n} (-L_\sigma f) f d\pi^\sigma \\ &= \mathbb{E}^{x_1}_{\nu^{\tau_{\sigma(1)}}} \mathcal{E}^{x_2}_{\nu^{\tau_{\sigma(2)}}}(f) + \mathbb{E}^{x_2}_{\nu^{\tau_{\sigma(2)}}} \mathcal{E}^{x_1}_{\nu^{\tau_{\sigma(1)}}}(f) \\ &= \mathbb{E}_{\pi^\sigma}(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2).\end{aligned}$$

The idea of parallel tempering is to swap between the positions of X_1 and X_2. At random times X_1 is moved to the position of X_2 and vice-versa, so the resulting process is a Markov process with jumps. To guarantee that the invariant measure remains the same, the jump intensity is of the Metropolis form $a\, g(x_1, x_2)$, where the constant 'a' is the swapping rate of the parallel tempering and $g = \min(1, \pi^-/\pi^+)$. The resulting process is denoted by $(X_1^a(t), X_2^a(t))$.

Intuitively, larger values of 'a' lead to faster convergence to equilibrium. However, the process $(X_1^a(t), X_2^a(t))$ is not tight so it does not converge weakly as $a \to \infty$. The key idea of [DLPD12] is to swap the 'temperatures' of (X_1, X_2) instead of swapping the positions. Precisely, they consider the following process

$$\begin{cases} d\overline{X}_1^a = -\nabla H(X_1)\, dt + \sqrt{2\tau_1 \mathbb{1}_{Z^a=0} + 2\tau_2 \mathbb{1}_{Z^a=1}}\, dB_1\,, \\ d\overline{X}_2 = -\nabla H(X_2)\, dt + \sqrt{2\tau_2 \mathbb{1}_{Z^a=0} + 2\tau_1 \mathbb{1}_{Z^a=1}}\, dB_2\,, \end{cases}$$

where Z^a is a jump process which switches from state 0 to state 1 with intensity $a\, g(\overline{X}_1^a, \overline{X}_2^a)$, and from state 1 to state 0 with intensity $a\, g(\overline{X}_2^a, \overline{X}_1^a)$. It was shown in [DLPD12] that as $a \to \infty$, the process $(\overline{X}_1^a(t), \overline{X}_2^a(t)$ converges weakly to the infinite swapping process, whose dynamics is governed by the SDE:

$$\begin{cases} dX_1 = -\nabla H(X_1)\, dt + \sqrt{2a_1(X_1, X_2)}\, dB_1\,, \\ dX_2 = -\nabla H(X_2)\, dt + \sqrt{2a_2(X_1, X_2)}\, dB_2\,, \end{cases} \tag{3.2.2}$$

where the diffusion coefficients a_1, a_2 are given by

$$a_1 := \tau_1\rho^+ + \tau_2\rho^- \quad \text{and} \quad a_2 := \tau_2\rho^+ + \tau_1\rho^-,$$

$$\text{where } \rho^+ := \frac{\pi^+}{\pi^+ + \pi^-} \quad \text{and} \quad \rho^- := \frac{\pi^-}{\pi^+ + \pi^-}.$$

The invariant measure of this process is the symmetric measure

$$\mu := \tfrac{1}{2}(\pi^+ + \pi^-). \tag{3.2.3}$$

The generator of this process is

$$\mathcal{L} := \rho^+ L_+ + \rho^- L_- = -\nabla H(x_1) \cdot \nabla_{x_1} - \nabla H(x_2) \cdot \nabla_{x_2} + a_1 \Delta_{x_1} + a_2 \Delta_{x_2}.$$

The associated Dirichlet form is

$$\mathcal{E}_\mu(f) := \int (-\mathcal{L} f) f d\mu = \frac{1}{2}\mathcal{E}_{\pi^+}(f) + \frac{1}{2}\mathcal{E}_{\pi^-}(f) = \int \sum_{k=1}^{2} a_k |\nabla_{x_k} f|^2 d\mu$$

and the Fisher information is

$$\mathcal{I}_\mu(f^2) := 2\mathcal{E}_\mu(f). \tag{3.2.4}$$

3.2.2 Growth and non-degeneracy assumptions

In this work, we use the same assumptions on the potential H as in [MS14, Section 1.2]. These assumptions are standard in the study of metastability (see e.g. [BEGK04, BGK05]).

Definition 3.2.1 (Morse function)**.** *A smooth function $H : \mathbb{R}^n \to \mathbb{R}$ is a* Morse function*, if the Hessian $\nabla^2 H$ of H is non-degenerate on the set of critical points. That is, for some $1 \leq C_H < \infty$ holds*

$$\forall x \in \mathcal{S} := \big\{x \in \mathbb{R}^n : \nabla H = 0\big\} : \frac{|\xi|}{C_H} \leq \big|\nabla^2 H(x)\xi| \leq C_H |\xi|. \tag{3.2.5}$$

We also make the following growth assumptions on the potential H to ensure the existence of PI and LSI.

Assumption 2 (PI)**.** *$H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \geq 0$ holds*

$$\liminf_{|x| \to \infty} |\nabla H| \quad \geq \quad C_H, \tag{3.2.6}$$

$$\liminf_{|x| \to \infty} \big(|\nabla H|^2 - \Delta H\big) \quad \geq \quad -K_H. \tag{3.2.7}$$

Assumption 3 (LSI). *$H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \geq 0$ holds*

$$\begin{aligned} \liminf_{|x|\to\infty} \frac{|\nabla H(x)|^2 - \Delta H(x)}{|x|^2} &\geq C_H, \\ \inf_x \nabla^2 H(x) &\geq -K_H. \end{aligned}$$

Remark 3.2.2. *Assumption 2 has the following consequences for the potential H:*

- *The condition* (3.2.6) *and $H(x) \geq 0$ ensures that $e^{-\frac{H}{\tau}}$ is integrable and can be normalized to a probability measure on $\mathbb{R}^n$ (see [MS14, Lemma 3.14]). Hence, the probability measures ν^τ (and therefore π^+, π^- and μ) are well defined.*
- *The Morse condition* (3.2.5) *together with the growth condition* (3.2.6) *ensures that the set $\mathcal{S}$ of critical points is discrete and finite. In particular, it follows that the set of local minima is a finite set $\mathcal{M} = \{m_1, \dots, m_N\}$.*
- *Together with the rest of Assumption 2, the Lyapunov-type condition* (3.2.7) *leads to a local PI for the Gibbs measures ν^τ (see [MS14, Theorem 2.9]).*

Similarly, Assumption 3 yields the following consequences for the potential H.

- *It leads to a local LSI for the Gibbs measures ν^τ (see [MS14, Theorem 2.10]).*
- *Assumption 3 implies Assumption 2, which is an indication that LSI is stronger than PI.*

To keep the presentation clear, we also make some non-degeneracy assumptions on the potential H. The *saddle height* $\widehat{H}(m_i, m_j)$ between two local minima m_i, m_j is defined by

$$\widehat{H}(m_i, m_j) := \inf \left\{ \max_{s\in[0,1]} H(\gamma(s)) : \gamma \in \mathcal{C}[0,1],\, \gamma(0) = m_i,\, \gamma(1) = m_j \right\}.$$

Assumption 4. *Let $m_1, \cdots, m_N$ be the positions of the local minima of H.*

(i) m_1 is the unique global minimum of H, and $m_1, \dots, m_N$ are ordered in the sense that there exists $\delta > 0$ such that

$$H(m_N) \geq H(m_{N-1}) \geq \cdots \geq H(m_2) \geq \delta \quad \text{and} \quad H(m_1) = 0. \tag{3.2.8}$$

(ii) *For each $i, j \in [N] := \{1, \dots, N\}$, the saddle height between m_i, m_j is attained at a unique critical point s_{ij} of index one. That is, $H(s_{ij}) = \widehat{H}(m_i, m_j)$, and if $\{\lambda_1, \dots, \lambda_n\}$ are the eigenvalues of $\nabla^2 H(s_{ij})$, then $\lambda_1 =: \lambda^- < 0$ and $\lambda_i > 0$ for $i \in \{2, \dots, n\}$. The point s_{ij} is called the* communicating saddle point *between the minima m_i and m_j.*

(iii) *There exists $p \in [N]$ such that the energy barrier $H(s_{p1}) - H(m_p)$ dominates all the others. That is, there exists $\delta > 0$ such that for all $i \in [N] \setminus \{p\}$,*

$$E_* := H(s_{p1}) - H(m_p) \geq H(s_{i1}) - H(m_i) + \delta.$$

The dominating energy barrier E_* is called the critical depth.

3.2.3 The Eyring-Kramers formula

Our main results are the Eyring-Kramers formula for the Poincaré constant and a good estimate for log-Sobolev constant for the isa. Here a crucial new feature occurs in comparison to the usual overdamped Langevin dynamic. The lower temperature cannot be arbitrarily small and there is an effective restriction on the ratio between the two temperatures τ_1 and τ_2. We comment on this observation in Subsection 3.2.4.

Theorem 3.2.3 (Eyring-Kramers formula for the Poincaré constant for the isa)**.** *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. Let μ be the invariant measure of the infinite swapping process defined by* (3.2.3)*. Suppose that the potential H satisfies Assumptions 2 and 4. Then the Gibbs measure μ satisfies the Poincaré inequality*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\rho}\mathcal{E}_\mu(f),$$

with the constant ρ satisfying

$$\begin{aligned}\frac{1}{\rho} \leq{}& \frac{1}{\sqrt{|\det \nabla^2 H(m_p)|}} \, \frac{2\pi\sqrt{|\det \nabla^2 H(s_{p1})|}}{|\lambda^-(s_{p1})|} \\ &\times \exp\left(\frac{H(s_{p1}) - H(m_p)}{\tau_2}\right)\left(1 + O(\sqrt{\tau_2}\,|\log \tau_2|^{\frac{3}{2}})\right) + O(1)\Phi_n\left(\frac{\tau_2}{\tau_1}\right). \qquad (3.2.9)\end{aligned}$$

Here $\lambda^-(s_{p1})$ is the negative eigenvalue of the Hessian $\nabla^2 H(s_{p1})$ at the communicating saddle point s_{p1}, and $\Phi_n : [1,\infty) \to [0,\infty)$ is the function

$$\Phi_n(x) = \begin{cases} 1 & \text{for } n = 1, \\ 1 + \log x & \text{for } n = 2, \\ 1 + x^{(n-2)/2} & \text{for } n \geq 3. \end{cases} \tag{3.2.10}$$

Theorem 3.2.4 (Estimate for the log-Sobolev constant of the isa)**.** *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. Let μ be the invariant measure of the infinite swapping process defined by* (3.2.3)*. Suppose that the potential H satisfies Assumptions 3 and 4. Then the Gibbs measure μ satisfies the log-Sobolev inequality*

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{\alpha}\mathcal{I}_\mu(f), \tag{3.2.11}$$

with

$$\begin{aligned} \frac{2}{\alpha} \leq 2N^2 \left(\frac{H(m_p)}{\tau_1} + \frac{H(m_p)}{\tau_2}\right) \frac{1}{\sqrt{|\det \nabla^2 H(m_p)|}} \frac{2\pi\sqrt{|\det \nabla^2 H(s_{p1})|}}{|\lambda^-(s_{p1})|} \\ \times \exp\left(\frac{H(s_{p1}) - H(m_p)}{\tau_2}\right)\left(1 + O(\sqrt{\tau_2}\,|\log \tau_2|^{\frac{3}{2}})\right) + O(\tau_1^{-1})\Phi_n\left(\frac{\tau_2}{\tau_1}\right). \end{aligned} \tag{3.2.12}$$

Here, N is the number of local minima of H, $\lambda^-(s_{p1})$ is the negative eigenvalue of the Hessian $\nabla^2 H(s_{p1})$ at the communicating saddle point s_{p1}, and Φ_n is the function defined in (3.2.10)*.*

Remark 3.2.5. *If we can ensure that τ_1 is not too low compared to τ_2, e.g. imposing a condition like*

$$\tau_1 \geq e^{-o\left(\frac{1}{\tau_2}\right)},$$

then the error terms involving $\Phi_n\left(\frac{\tau_2}{\tau_1}\right)$ in (3.2.9) *and* (3.2.12) *become negligible, as can be seen from the form of the function Φ_n. (In fact, this restriction can be entirely dropped in dimension $n = 1$, and relaxed to $\tau_1 \geq e^{-e^{o\left(\frac{1}{\tau_2}\right)}}$ in dimension $n = 2$.) Then in this regime of temperatures τ_1, τ_2, the estimates* (3.2.9) *and* (3.2.12) *for the isa essentially reduce to the corresponding Eyring-Kramers formulas for the overdamped Langevin dynamics* ***at the higher temperature*** *τ_2, given in [MS14, Corollary 2.18]. For the Poincaré constant this is true to the exact pre-factor, and for the LSI constant this is true to the leading exponential*

order. Because we choose $\tau_2 \geq K\tau_1$, this means that the effective energy barrier $H(m_p) - H(m_1)$ is reduced by a factor of $K > 1$.

More precisely, the estimate we give for the LSI constant differs from the one in [MS14, Corollary 2.18] (with temperature set to be τ_2) by two additional factors: first, we have $H(m_p)/\tau_1$ instead of $H(m_p)/\tau_2$ in the pre-factor, which amounts to an additional factor of τ_2/τ_1; and second, we also have a combinatorial factor on the order of N^2. Below, we show the change from $H(m_p)/\tau_2$ to $H(m_p)/\tau_1$ is necessary in a generic one-dimensional case. However, presently we do not know whether the combinatorial factor is necessary. It would be interesting to study whether this factor of N^2 can be removed from the LSI constant.

3.2.4 Dependence on the ratio between temperatures

The following proposition shows that the dependence on τ_2/τ_1 in the Poincaré and LSI constants of the isa is necessary and the formula of Φ_n that describes this dependence is close to being optimal.

Proposition 3.2.6. *If $\tau_2, \tau_1/\tau_2$ are sufficiently small, then for every $\eta > 0$, there exists a constant $C_\eta > 0$ such that*

$$\sup_{f \in H^1(\mu)} \frac{\mathrm{Var}_\mu(f)}{\mathcal{E}_\mu(f)} \gtrsim \begin{cases} C_\eta (\tau_2/\tau_1)^{(1-\eta)(n-2)/2} & \text{for } n \geq 3, \\ \log(\tau_2/\tau_1) & \text{for } n = 2. \end{cases}$$

3.2.5 Optimality of the Eyring-Kramers formula in dimension one

For the overdamped Langevin dynamics, the corresponding Eyring-Kramers formula for Poincaré inequality has been shown to be optimal. For the isa, the Poincaré constant of (3.2.9) is optimal in a generic one-dimensional case. This gives a strong indication of optimality in higher dimensions.

Proposition 3.2.7. *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. Assume $n = 1$, and H has three critical points: two minima $m_1 < m_2$ with $H(m_1) = 0 < \delta \leq H(m_2)$ and a local*

maximum s in between. Then

$$\inf_{f \in H^1(\mu)} \frac{\mathcal{E}_\mu(f)}{\mathrm{Var}_\mu(f)} \leq \rho,$$

where ρ is given by the asymptotic formula (3.2.9).

For the overdamped Langevin dynamics, the corresponding Eyring-Kramers formula for LSI inequality has been shown to be optimal in the one-dimensional case. For the isa, we do not expect the LSI constant of (3.2.12) to be optimal. However, up to some combinatorial factor in N, it has the asymptotic behavior for a generic one-dimensional case.

Proposition 3.2.8. *Assume that $\tau_2 \geq K\tau_1$ for some constant $K > 1$. Assume $n = 1$, and H has three critical points: two minima $m_1 < m_2$ with $H(m_1) = 0 < \delta \leq H(m_2)$ and a local maximum s in between. Then*

$$\inf_{f \in H^1(\mu)} \frac{\mathcal{I}_\mu(f^2)}{\mathrm{Ent}_\mu(f^2)} \lesssim_N \alpha,$$

where α is given by the asymptotic formulas (3.2.12).

3.3 Proofs of main results

3.3.1 A two-time-scale approach to Poincaré and log-Sobolev inequalities

Our overall approach follows that of [MS14], which was via a decomposition of the state space $\mathbb{R}^n$ into an "admissible partition" of metastable regions $\{\Omega_i\}_{i=1}^N$ for the Gibbs measure ν^τ defined in (3.2.1), as described below.

Definition 3.3.1 (Admissible partition)**.** *The family $\{\Omega_i\}_{i=1}^N$ with Ω_i open and connected is called an admissible partition for H if*

(i) for each $i \in [N]$, the local minimum $m_i \in \Omega_i$,

(ii) $\{\Omega_i\}_{i=1}^N$ *forms a partition of* $\mathbb{R}^n$ *up to sets of Lebesgue measure zero,*

(iii) The partition sum of Ω_i *is approximately Gaussian. That is, there exists* $\tau_0 > 0$ *such that for all* $\tau < \tau_0$*, for* $i \in [N]$*,*

$$\begin{aligned}\nu^\tau(\Omega_i)Z^\tau &:= \int_{\Omega_i} \exp\left(-\frac{H(x)}{\tau}\right)dx \\ &= \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + O(\sqrt{\tau}|\log\tau|^{3/2})). \end{aligned} \tag{3.3.1}$$

Remark 3.3.2. A canonical way to obtain an admissible partition for H is to associate to each local minimum m_i for $i \in [N]$ its basin of attraction with respect to the gradient flow of H. That is,

$$\Omega_i = \left\{ y \in \mathbb{R}^N : \lim_{t\to\infty} y_t = m_i,\ \frac{dy_t}{dt} = -\nabla H(y_t),\ y_0 = y \right\}.$$

However, as in [MS14], to facilitate the proof, we choose instead the basins of attraction for the gradient flow of a suitable perturbation of H (see Section 3.3.2).

Suppose $\{\Omega_i\}_{i=1}^N$ is an admissible partition in the sense of Definition 3.3.1. Define local measures on $\mathbb{R}^n$

$$\begin{aligned}\nu_i^\tau(x) &:= \frac{1}{Z_i^\tau}\nu^\tau(x)|_{\Omega_i}, \\ Z_i^\tau &:= \nu^\tau(\Omega_i) = \frac{\sqrt{\det \nabla^2 H(m_1)}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + O(\sqrt{\tau}|\log\tau|^{3/2})).\end{aligned} \tag{3.3.2}$$

This induces a decomposition of the measure μ on $\mathbb{R}^n \times \mathbb{R}^n$ as

$$\mu = \frac{1}{2}(\pi^+ + \pi^-) = \sum_{(i,j)} \frac{1}{2} Z_{ij}^+ \pi_{ij}^+ + \sum_{(i,j)} \frac{1}{2} Z_{ij}^- \pi_{ij}^-, \tag{3.3.3}$$

where $Z_{ij}^+ := Z_i^{\tau_1} Z_j^{\tau_2}, Z_{ij}^- := Z_i^{\tau_2} Z_j^{\tau_1}$ and

$$\begin{aligned}\pi_{ij}^+(x_1, x_2) &:= \frac{1}{Z_{ij}^+}\pi^+(x_1, x_2)|_{\Omega_i\times\Omega_j} = \nu_i^{\tau_1}(x_1)\nu_j^{\tau_2}(x_2), \\ \pi_{ij}^-(x_1, x_2) &:= \frac{1}{Z_{ij}^-}\pi^-(x_1, x_2)|_{\Omega_i\times\Omega_j} = \nu_i^{\tau_2}(x_1)\nu_j^{\tau_1}(x_2).\end{aligned}$$

The following results are read from [MS14, Lemma 2.4 and Corollary 2.8].

Lemma 3.3.3 (Decomposition of variance)**.** *For the mixture representation* (3.3.3) *of the Gibbs measure μ, and a smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, it holds*

$$\operatorname{Var}_\mu(f) = \frac{1}{2}\sum_{(i,j)} Z_{ij}^+ \operatorname{Var}_{\pi_{ij}^+}(f) + \frac{1}{2}\sum_{(i,j)} Z_{ij}^- \operatorname{Var}_{\pi_{ij}^-}(f) \tag{3.3.4}$$

$$+ \frac{1}{4}\sum Z_{ij}^+ Z_{kl}^+ (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{kl}^+}(f))^2 + \frac{1}{4}\sum Z_{ij}^- Z_{kl}^- (\mathbb{E}_{\pi_{ij}^-}(f) - \mathbb{E}_{\pi_{kl}^-}(f))^2 \tag{3.3.5}$$

$$+ \frac{1}{4}\sum Z_{ij}^+ Z_{kl}^- (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{kl}^-}(f))^2, \tag{3.3.6}$$

where the second line is summing over unordered pairs $(i,j) \neq (k,l)$ and the last line is summing over ordered pairs $((i,j),(k,l))$.

Lemma 3.3.4 (Decomposition of entropy)**.** *For the mixture representation* (3.3.3) *of the Gibbs measure μ, and a smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, it holds*

$$\operatorname{Ent}_\mu(f^2) \leq \frac{1}{2}\sum_{(i,j)} Z_{ij}^+ \operatorname{Ent}_{\pi_{ij}^+}(f^2) + \frac{1}{2}\sum_{(i,j)} Z_{ij}^- \operatorname{Ent}_{\pi_{ij}^-}(f^2) \tag{3.3.7}$$

$$+ \frac{1}{2}\sum_{(i,j)} \left(\sum_{(k,l)\neq(i,j)} \frac{Z_{kl}^+}{\Lambda(Z_{ij}^+, Z_{kl}^+)} + \sum_{(k,l)} \frac{Z_{kl}^-}{\Lambda(Z_{ij}^+, Z_{kl}^-)} \right) Z_{ij}^+ \operatorname{Var}_{\pi_{ij}^+}(f) \tag{3.3.8}$$

$$+ \frac{1}{2}\sum_{(i,j)} \left(\sum_{(k,l)\neq(i,j)} \frac{Z_{kl}^-}{\Lambda(Z_{ij}^-, Z_{kl}^-)} + \sum_{(k,l)} \frac{Z_{kl}^+}{\Lambda(Z_{ij}^-, Z_{kl}^+)} \right) Z_{ij}^- \operatorname{Var}_{\pi_{ij}^-}(f) \tag{3.3.9}$$

$$+ \frac{1}{2}\sum \frac{Z_{ij}^+ Z_{kl}^+}{\Lambda(Z_{ij}^+, Z_{kl}^+)} (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{kl}^+}(f))^2 \tag{3.3.10}$$

$$+ \frac{1}{2}\sum \frac{Z_{ij}^- Z_{kl}^-}{\Lambda(Z_{ij}^-, Z_{kl}^-)} (\mathbb{E}_{\pi_{ij}^-}(f) - \mathbb{E}_{\pi_{kl}^-}(f))^2 \tag{3.3.11}$$

$$+ \frac{1}{2}\sum \frac{Z_{ij}^+ Z_{kl}^-}{\Lambda(Z_{ij}^+, Z_{kl}^-)} (\mathbb{E}_{\pi_{ij}^+}(f) - \mathbb{E}_{\pi_{kl}^-}(f))^2, \tag{3.3.12}$$

where the second to last line and the third to last line are summing over unordered pairs $(i,j) \neq (k,l)$ and the last line is summing over ordered pairs $((i,j),(k,l))$.

The local variances appearing in (3.3.4), (3.3.8) and (3.3.9) and the local entropies appearing in (3.3.7) are dealt with by Poincaré and log-Sobolev inequalities for local product measures.

Lemma 3.3.5 (Local PI for π_{ij}^{σ})**.** *Under Assumption 2, given τ_2 small enough, there exists an admissible partition $\{\Omega_i\}_{i=1}^N$ such that for all $\tau \leq \tau_2$, , for all smooth functions f : $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Var}_{\pi_{ij}^{\sigma}}(f) \overset{(3.3.18)}{\leq} O(1)\mathbb{E}_{\pi_{ij}^{\sigma}}(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2).$$

Lemma 3.3.6 (Local LSI for π_{ij}^{σ})**.** *Under Assumption 3, for all smooth functions $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Ent}_{\pi_{ij}^{\sigma}}(f^2) \overset{(3.3.19)}{\leq} O(1)\mathbb{E}_{\pi_{ij}^{\sigma}}(|\nabla_{x_1} f|^2 + |\nabla_{x_2} f|^2).$$

We defer the details of the proof of Lemmas 3.3.5 and 3.3.6 to Section 3.3. They are based on the simple product structure of the measures π_{ij}^{σ} and an adaption of the local Poincaré inequality [MS14, Theorem 2.9] and the local LSI inequality [MS14, Theorem 2.10]. It follows that

$$\begin{aligned} Z_{ij}^{\sigma}\,\mathrm{Var}_{\pi_{ij}^{\sigma}}(f) &\leq O(1)\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j], \\ Z_{ij}^{\sigma}\,\mathrm{Ent}_{\pi_{ij}^{\sigma}}(f) &\leq O(\tau_1^{-1})\mathcal{E}_{\pi^{\sigma}}(f)[\Omega_i \times \Omega_j]. \end{aligned} \tag{3.3.13}$$

Here and below, for a Dirichlet form $\mathcal{E}(f)$, we denote $\mathcal{E}(f)[\Omega]$ to be the Dirichlet integral with region of integration restricted to Ω.

To deal with the mean-differences appearing in (3.3.5) and (3.3.10) - (3.3.11), we will apply the mean-difference estimate from [MS14, Theorem 2.12], which allows us to transport in one of the variables x_1, x_2 at a time from one metastable region Ω_j to another metastable region Ω_k. However, in order to ensure we only get exponential dependence on $1/\tau_2$ rather than $1/\tau_1$ in the Eyring-Kramers formula, we can only transport in the high-temperature variable, and not in the low-temperature variable. This allows us to deal with mean-differences of the type between π_{ij}^{+} and π_{ik}^{+}, or the type between π_{ji}^{-} and π_{ki}^{-}.

Lemma 3.3.7 (Mean-difference estimates for $\pi_{ij}^{+}, \pi_{ik}^{+}$ and for $\pi_{ji}^{-}, \pi_{ki}^{-}$)**.**

$$Z_{ik}^{+}(\mathbb{E}_{\pi_{ij}^{+}} f - \mathbb{E}_{\pi_{ik}^{+}} f)^2 \lessapprox C_{kj}^{\tau_2} \cdot \mathcal{E}_{\pi^{+}}(f)[\Omega_i \times \mathbb{R}^n], \tag{3.3.14}$$

$$Z^-_{ki}(\mathbb{E}_{\pi^-_{ji}}f - \mathbb{E}_{\pi^-_{ki}}f)^2 \lessapprox C^{\tau_2}_{kj} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_i], \tag{3.3.15}$$

where

$$C^{\tau_2}_{kj} := \frac{1}{\sqrt{\det \nabla^2 H(m_k)}} \frac{2\pi\sqrt{\det \nabla^2 H(s_{kj})}}{|\lambda^-(s_{kj})|} \exp\left(\frac{H(s_{kj}) - H(m_k)}{\tau_2}\right).$$

Here and below, $\approx$ (resp. $\lessapprox$) means equality (resp. less than or equal) up to a multiplicative factor of $1 + O(\sqrt{\tau_2}|\log \tau_2|^{3/2})$.

Proof. For the first estimate, applying Cauchy-Schwarz and [MS14, Theorem 2.12], we get

$$\begin{aligned} Z^+_{ik}(\mathbb{E}_{\pi^+_{ij}}f - \mathbb{E}_{\pi^+_{ik}}f)^2 &\leq Z^{\tau_1}_i Z^{\tau_2}_k \mathbb{E}^{x_1}_{\nu^{\tau_1}_i}(\mathbb{E}^{x_2}_{\nu^{\tau_2}_j} f - \mathbb{E}^{x_2}_{\nu^{\tau_2}_k} f)^2 \\ &\lessapprox Z^{\tau_1}_i \mathbb{E}^{x_1}_{\nu^{\tau_1}_i} C^{\tau_2}_{kj} \int \tau_2 |\nabla_{x_2} f|^2 d\nu^{\tau_2}(x_2) \\ &\leq C^{\tau_2}_{kj} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n]. \end{aligned}$$

The second estimate is completely analogous. □

To deal with the remaining mean-differences in (3.3.5) - (3.3.6) and (3.3.10) - (3.3.12), we have another move available, which is to swap the temperatures of the two variables, i.e. to swap between π^+_{ij} and π^-_{ij}. This is the main new technical ingredient compared to [MS14], which come at a cost that is polynomial in the ratio of the higher temperature to the lower temperature, τ_2/τ_1.

Lemma 3.3.8 (Mean-difference estimate for π^+_{ij}, π^-_{ij})**.** *In the same setting as Lemma 3.3.13,*

$$\begin{aligned} (\mathbb{E}_{\pi^+_{ij}}f - \mathbb{E}_{\pi^-_{ij}}f)^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(\tau_2)(\mathbb{E}_{\pi^+_{ij}}|\nabla_{x_2} f|^2 + \mathbb{E}_{\pi^-_{ij}}|\nabla_{x_1} f|^2) \\ + \omega(\tau_2) \sum_{\sigma \in \{+,-\}} \mathbb{E}_{\pi^\sigma_{ij}}(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2) \end{aligned}$$

for any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, where $\Phi_n : [1, \infty) \to [0, \infty)$ is the function

$$\Phi_n(x) = \begin{cases} 1 & \text{for } n = 1, \\ 1 + \log x & \text{for } n = 2, \\ 1 + x^{(n-2)/2} & \text{for } n \geq 3, \end{cases}$$

and $\omega(\tau_2) := O(\sqrt{\tau_2}|\log \tau_2|^{3/2})$.

We defer the proof of this lemma to the next two sections. It follows that

$$\min(Z_{ij}^+, Z_{ij}^-)(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)[\Omega_i \times \Omega_j]. \tag{3.3.16}$$

Using these estimates, we will show that the dominating terms in Lemma 3.3.3 are the mean-differences between π_{ip}^+, π_{11}^+ and between π_{pj}^-, π_{11}^- where i, j are arbitrary and p is the local minimum with the dominating energy barrier.

Lemma 3.3.9. *Let p be the local minimum with the dominating energy barrier. Then for any $i, j \in [N]$, and $\sigma \in \{+, -\}$*

$$Z_{ip}^+ Z_{11}^\sigma (\mathbb{E}_{\pi_{ip}^+}(f) - \mathbb{E}_{\pi_{11}^\sigma}(f))^2 \lessapprox C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f),$$
$$Z_{pj}^- Z_{11}^\sigma (\mathbb{E}_{\pi_{pj}^-}(f) - \mathbb{E}_{\pi_{11}^\sigma}(f))^2 \lessapprox C_{p1}^{\tau_2} \cdot \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f).$$

Moreover, if $\{(i,j)^{\sigma_1}, (k,l)^{\sigma_2}\}$ is one of the following forms

$$\{(i,1)^+, (1,1)^+\}, \{(1,j)^-, (1,1)^-\}, \{(i,1)^+, (1,1)^-\}, \{(1,1)^+, (1,l)^-\},$$

then

$$Z_{ij}^{\sigma_1} Z_{kl}^{\sigma_2} (\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f) - \mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2 \leq \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f).$$

Finally, for any other $\{(i,j)^{\sigma_1}, (k,l)^{\sigma_2}\}$, the term $Z_{ij}^{\sigma_1} Z_{kl}^{\sigma_2} (\mathbb{E}_{\pi_{ij}^{\sigma_1}}(f) - \mathbb{E}_{\pi_{kl}^{\sigma_2}}(f))^2$ is negligible in the sense of being exponentially smaller in $1/\tau_2$ compared to one of the terms above on the right hand side.

Proof. Let Γ be the graph whose vertices are labeled $\cdot_{ij}^\sigma$ and have three kinds of edges:

- "vertical" edges between $\cdot_{ij}^+, \cdot_{ik}^+$;
- "horizontal" edges between $\cdot_{ij}^-, \cdot_{kj}^-$; and
- "swapping" edges between $\cdot_{ij}^+, \cdot_{ij}^-$.

We decompose the mean-difference between any two measures π_{ij}^+, π_{kl}^- as a sum of mean-differences of the types in (3.3.14), (3.3.15), and (3.3.16), corresponding to a sequence of

"moves" on the graph Γ. Given any sequence of moves $v_0 \to v_1 \to \cdots \to v_m$ on graph Γ, we have

$$Z_{v_0} Z_{v_m} (\mathbb{E}_{\pi_{v_0}} f - \mathbb{E}_{\pi_{v_m}} f)^2 = Z_{v_0} Z_{v_m} \left(\sum_{t=1}^{m} \sqrt{\omega_t} \frac{1}{\sqrt{\omega_t}} (\mathbb{E}_{\pi_{v_{t-1}}} f - \mathbb{E}_{\pi_{v_t}} f) \right)^2$$
$$= \sum_{t=1}^{m} \frac{1}{\omega_t} Z_{v_0} Z_{v_m} (\mathbb{E}_{\pi_{v_{t-1}}} f - \mathbb{E}_{\pi_{v_t}} f)^2 \tag{3.3.17}$$

for any $\omega_t > 0, \sum_{t=1}^{m} \omega_t = 1$. After taking into account the weights Z_{ij}^{+}, Z_{kl}^{-}, this leads to the choice of the following three types of sequences of moves for the three types of mean-differences occurring in Lemma 3.3.3:

- Type I sequence: $\cdot_{ij}^{+} \to \cdot_{i1}^{+} \to \cdot_{i1}^{-} \to \cdot_{11}^{-} \to \cdot_{k1}^{-} \to \cdot_{k1}^{+} \to \cdot_{kl}^{+}$;
- Type II sequence: $\cdot_{ij}^{-} \to \cdot_{1j}^{-} \to \cdot_{1j}^{+} \to \cdot_{11}^{+} \to \cdot_{1l}^{+} \to \cdot_{1l}^{-} \to \cdot_{kl}^{-}$; and
- Type III sequence: $\cdot_{ij}^{+} \to \cdot_{i1}^{+} \to \cdot_{i1}^{-} \to \cdot_{11}^{-} \to \cdot_{11}^{+} \to \cdot_{1l}^{+} \to \cdot_{1l}^{-} \to \cdot_{kl}^{-}$.

Let us first look at the decomposition (3.3.17) for a Type I sequence. For the 1st move,

$$Z_{ij}^{+} Z_{kl}^{+} (\mathbb{E}_{\pi_{ij}^{+}}(f) - \mathbb{E}_{\pi_{i1}^{+}}(f))^2 \lessapprox Z_{kl}^{+} C_{j1}^{\tau_2} \cdot \mathcal{E}_{\pi^{+}}(f)[\Omega_i \times \mathbb{R}^n],$$

which is negligible unless $j = p, k = l = 1$. For the 2nd move,

$$Z_{ij}^{+} Z_{kl}^{+} (\mathbb{E}_{\pi_{i1}^{+}}(f) - \mathbb{E}_{\pi_{i1}^{-}}(f))^2 \leq Z_{j}^{\tau_2} Z_{kl}^{+} \cdot \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_{\mu}(f),$$

which is negligible unless $j = k = l = 1$. For the 3rd move,

$$Z_{ij}^{+} Z_{kl}^{+} (\mathbb{E}_{\pi_{i1}^{-}}(f) - \mathbb{E}_{\pi_{11}^{-}}(f))^2 \lessapprox \exp\left(-H(m_i) \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) \right) Z_{j}^{\tau_2} Z_{kl}^{+} C_{i1}^{\tau_2} \cdot \mathcal{E}_{\pi^{-}}(f)[\mathbb{R}^n \times \Omega_1],$$

which is always negligible. The analysis for the remaining three moves are completely symmetric: the 4th move is always negligible, the 5th move is negligible unless $i = j = l = 1$, and the 6th move is negligible unless $l = p, i = j = 1$.

Overall, if $(i, j), (k, l)$ is not one of the exceptions mentioned, we can just assign $\omega_1 = \omega_1 = \cdots = \omega_6 = 1/6$, then the overall sum is negligible. This choice of $(\omega_t)_{t=1}^{6}$ also works in the exceptional cases $k = j = l = 1$ and $i = j = l = 1$ (since we can afford to lose a constant factor because of the $O(1)$).

Lastly, in the exceptional case $j = p, k = l = 1$, we consider a shortened 2-move sequence $\cdot^+_{ip} \to \cdot^+_{i1} \to \cdot^+_{11}$. For the 1st move in this sequence,

$$Z^+_{ip} Z^+_{11} (\mathbb{E}_{\pi^+_{ij}}(f) - \mathbb{E}_{\pi^+_{i1}}(f))^2 \lessapprox C^{\tau_2}_{p1} \cdot \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n],$$

and for the 2nd move in this sequence,

$$\begin{aligned} Z^+_{ip} Z^+_{11} (\mathbb{E}_{\pi^+_{i1}}(f) - \mathbb{E}_{\pi^+_{11}}(f))^2 &\approx Z^{\tau_2}_p \cdot Z^+_{i1} Z^+_{11} (\mathbb{E}_{\pi^+_{i1}}(f) - \mathbb{E}_{\pi^+_{11}}(f))^2 \\ &\lessapprox Z^{\tau_2}_p \cdot \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_\mu(f). \end{aligned}$$

Thus, for this sequence, we can assign $\omega_1 = 1 - Z^{\tau_2}_p \approx 1, \omega_2 = Z^{\tau_2}_p$, then the overall sum is as claimed. The exceptional case $l = p, i = j = 1$ is completely symmetric.

The analysis for Type II and Type III sequences are completely analogous.

□

We can adapt this approach to estimate the terms in Lemma 3.3.4.

Lemma 3.3.10. *Let p be the local minimum with the dominating energy barrier. Then for $i, k, l \in [N]$ and $\sigma \in \{+, -\}$ such that*

$$H(m_i) < H(m_p) \text{ or } i = p, \text{ and } \frac{H(m_i)}{\tau_1} + \frac{H(m_p)}{\tau_2} \geq \frac{H(m_k)}{\tau_{\sigma(1)}} + \frac{H(m_l)}{\tau_{\sigma(2)}},$$

it holds that

$$\begin{aligned} \frac{Z^+_{ip} Z^\sigma_{kl}}{\Lambda(Z^+_{ip}, Z^\sigma_{kl})} (\mathbb{E}_{\pi^+_{ip}}(f) - \mathbb{E}_{\pi^\sigma_{kl}}(f))^2 &\lessapprox \frac{1}{\Lambda\left(\frac{Z^+_{ip}}{Z^\sigma_{kl}}, 1 \right)} \left(C^{\tau_2}_{p1} \mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_\mu(f) \right), \\ \frac{Z^-_{pi} Z^\sigma_{kl}}{\Lambda(Z^-_{pi}, Z^\sigma_{kl})} (\mathbb{E}_{\pi^-_{pi}}(f) - \mathbb{E}_{\pi^\sigma_{kl}}(f))^2 &\lessapprox \frac{1}{\Lambda\left(\frac{Z^-_{pi}}{Z^\sigma_{kl}}, 1 \right)} \left(C^{\tau_2}_{p1} \mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_i] + \Phi_n \left(\frac{\tau_2}{\tau_1} \right) O(1) \mathcal{E}_\mu(f) \right). \end{aligned}$$

Finally, for any other $\{(i,j)^{\sigma_1}, (k,l)^{\sigma_2}\}$, the term $\frac{Z^{\sigma_1}_{ij} Z^{\sigma_2}_{kl}}{\Lambda(Z^{\sigma_1}_{ij}, Z^{\sigma_2}_{kl})} (\mathbb{E}_{\pi^{\sigma_1}_{ij}}(f) - \mathbb{E}_{\pi^{\sigma_2}_{kl}}(f))^2$ is negligible in the sense of being exponentially smaller in $1/\tau_2$ compared to one of the terms above on the right hand side.

Proof. The analysis is similar as in the previous lemma, but now we have to take into account the logarithmic mean, using the estimate

$$\frac{ab}{\Lambda(a,b)} = a \cdot \frac{b}{\Lambda(a/b,1)} \lessapprox a\log(1/a)$$

for $b \lessapprox 1, a \ll 1$. The main difference is that we now need to be more careful to show the transport from $\cdot^+_{ip}$ to $\cdot^+_{11}$ is negligible if $H(m_i) \geq H(m_p)$ and $i \neq p$ by choosing the alternative path: $\cdot^+_{ip} \to \cdot^-_{ip} \to \cdot^-_{1p} \to \cdot^+_{1p} \to \cdot^+_{11}$. □

Proof of Theorem 3.2.3. Combining Lemma 3.3.3, (3.3.13) and Lemma 3.3.9, we get

$$\begin{aligned}
\mathrm{Var}_\mu(f) &\lessapprox \frac{1}{2}\sum_{i,j} O(1)\mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2}\sum_{(i,j)} O(1)\mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
&\quad + 2\cdot\frac{1}{4}\sum_i C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + 2\cdot\frac{1}{4}\sum_j C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] \\
&\quad + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f) \\
&\leq \left(O(1) + C^{\tau_2}_{1p} + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\right)\mathcal{E}_\mu(f),
\end{aligned}$$

as desired.

□

Proof of Theorem 3.2.4. Combining Lemma 3.3.4, (3.3.13), (3.3.13) and Lemma 3.3.10, we get

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &\lessapprox \frac{1}{2}\sum_{(i,j)} O(\tau_1^{-1})\mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2}\sum_{(i,j)} O(\tau_1^{-1})\mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
&\quad + \frac{1}{2}\sum_{(i,j)} 2N^2 O(\tau_1^{-1})\mathcal{E}_{\pi^+}(f)[\Omega_i \times \Omega_j] + \frac{1}{2}\sum_{(i,j)} 2N^2 O(\tau_1^{-1})\mathcal{E}_{\pi^-}(f)[\Omega_i \times \Omega_j] \\
&\quad + \frac{1}{2}\sum_{i\leq p}\left(\sum_{(k,l)^\sigma}\frac{1}{\Lambda\left(\frac{Z^+_{ip}}{Z^\sigma_{kl}},1\right)}\right)\left(C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^+}(f)[\Omega_i \times \mathbb{R}^n] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)\right) \\
&\quad + \frac{1}{2}\sum_{i\leq p}\left(\sum_{(k,l)^\sigma}\frac{1}{\Lambda\left(\frac{Z^-_{pi}}{Z^\sigma_{kl}},1\right)}\right)\left(C^{\tau_2}_{p1}\cdot\mathcal{E}_{\pi^-}(f)[\mathbb{R}^n \times \Omega_j] + \Phi_n\left(\frac{\tau_2}{\tau_1}\right) O(1)\mathcal{E}_\mu(f)\right)
\end{aligned}$$

$$\leq 2N^2 \left(O(\tau_1^{-1}) + H(m_p)(\tau_1^{-1} + \tau_2^{-1}) C_{p1}^{\tau_2} + O(\tau_1^{-1}) \Phi_n \left(\frac{\tau_2}{\tau_1} \right) \right) \mathcal{E}_\mu(f),$$

as desired. □

3.3.2 Local Poincaré and log-Sobolev inequalities: proofs of Lemmas 3.3.5 and 3.3.6

The following decomposition of variance and entropy for a product measure reduces proving Lemmas 3.3.5 and 3.3.6 to proving corresponding estimates for the component measures ν_i^τ. It may be verified by basic properties of variance and entropy.

Lemma 3.3.11 (Variance and entropy for product measure)**.** *Let $\pi = \nu_i \otimes \nu_j$ be a product of two probability measures on open subsets of $\mathbb{R}^n$. For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Var}_\pi(f) = \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Var}_{\nu_i}^{x_1}(f)\big) + \mathrm{Var}_{\nu_j}^{x_2}\big(\mathbb{E}_{\nu_i}^{x_1}(f)\big) \leq \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Var}_{\nu_i}^{x_1}(f)\big) + \mathbb{E}_{\nu_i}^{x_1}\big(\mathrm{Var}_{\nu_j}^{x_2}(f)\big). \tag{3.3.18}$$

For any smooth function $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{>0}$,

$$\mathrm{Ent}_\pi(g) = \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Ent}_{\nu_i}^{x_1}(g)\big) + \mathrm{Ent}_{\nu_j}^{x_2}\big(\mathbb{E}_{\nu_i}^{x_1}(g)\big) \leq \mathbb{E}_{\nu_j}^{x_2}\big(\mathrm{Ent}_{\nu_i}^{x_1}(g)\big) + \mathbb{E}_{\nu_i}^{x_1}\big(\mathrm{Ent}_{\nu_j}^{x_2}(g)\big). \tag{3.3.19}$$

Definition 3.3.12 (Local PI and LSI for ν_i^τ)**.** *We say the local Gibbs measure ν_i^τ satisfies a Poincaré inequality with constant ρ if for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$*

$$\mathrm{Var}_{\nu_i^\tau}(f) \leq \frac{1}{\rho} \mathbb{E}_{\nu_i^\tau} |\nabla f|^2,$$

which we denote PI(ρ). We say ν_i^τ satisfies a log-Sobolev inequality with constant α if for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Ent}_{\nu_i^\tau}(f^2) \leq \frac{2}{\alpha} \mathbb{E}_{\nu_i^\tau} |\nabla f|^2,$$

which we denote LSI(α).

Lemma 3.3.13 (Local PI for ν_i^τ)**.** *Under Assumption 2, given τ_2 small enough, there exists an admissible partition $\{\Omega_i\}_{i=1}^N$ such that for all $\tau \leq \tau_2$, the local Gibbs measures ν_i^τ satisfy PI(ρ) with $\rho^{-1} = O(\tau)$.*

Lemma 3.3.14 (Local LSI for ν_i^τ). *Under Assumption 3, given τ_2 small enough, for the same admissible partition $\{\Omega_i\}_{i=1}^N$, for all $\tau \le \tau_2$, the local Gibbs measures ν_i^τ satisfy LSI(α) with $\alpha^{-1} = O(1)$.*

Lemmas 3.3.13 and 3.3.14 are very similar to [MS14, Theorem 2.9] and [MS14, Theorem 2.10], except now that we have two temperatures $\tau_1 < \tau_2$, we want the regions Ω_i in the admissible partition only depend on the higher temperature τ_2 but not the lower temperature τ_1, so that we can get PI and LSI for the local Gibbs measures $\nu_i^{\tau_1}, \nu_i^{\tau_2}$ **defined on the same regions** Ω_i.

This can be shown by making a small modification to the proof of [MS14, Theorem 2.9, 2.10], which is based on constructing a Lyapunov function. Let us recall the definition of a Lyapunov function and the criterion for PI based on it from [MS14].

Definition 3.3.15 (Lyapunov function, Definition 3.7 in [MS14]). *A smooth function $W_\tau : \Omega_i \to (0, \infty)$ is a Lyapunov function for ν_i^τ if the following hold for $L_\tau := \tau\Delta - \nabla H \cdot \nabla$:*

(i) There exists an open set $U_i \subset \Omega_i$ and constants $b > 0, \lambda > 0$ such that

$$\frac{L_\tau W_\tau}{W_\tau} \le -\lambda + b\mathbb{1}_{U_i} \quad \forall x \in \Omega_i. \tag{3.3.20}$$

(ii) W_τ satisfies Neumann boundary condition on Ω_i in the sense that it satisfies the integration by parts formula

$$\int_{\Omega_i} (-L_\tau W_\tau) g d\nu_i^\tau = \int_{\Omega_i} \nabla g \cdot \nabla W_\tau d\nu_i^\tau. \tag{3.3.21}$$

Lemma 3.3.16 (Lyapunov condition for local PI, Theorem 3.8 in [MS14]). *If there exists a Lyapunov function for ν_i^τ in the sense of Definition 3.3.15 and that the truncated Gibbs measure $\nu_i^\tau|_{U_i}$ satisfies PI(ρ_{U_i}), then the local Gibbs measure ν_i^τ satisfies PI(ρ) with*

$$\rho^{-1} \le \frac{b}{\lambda}\rho_{U_i}^{-1} + \frac{1}{\lambda}\tau.$$

We choose U_i to be a ball centered at the local minimum m_i with a small, fixed radius R_0 such that H is strongly convex on U_i. Then the Bakry-Émery criterion provides the following result.

Lemma 3.3.17 (PI for truncated Gibbs measure, Lemma 3.6 in [MS14]). *The measures* $\nu_i^\tau|_{U_i}$ *satisfy PI*(ρ_{U_i}) *with* $\rho_{U_i}^{-1} = O(\tau)$.

In [MS14], the candidate for the Lyapunov function is $W_\tau = \exp\left(\frac{H}{2\tau}\right)$, so that (see [MS14, equation (3.9)])

$$\frac{L_\tau W_\tau}{W_\tau} = \frac{1}{2}\Delta H(x) - \frac{1}{4\tau}|\nabla H(x)|^2.$$

In order to satisfy the condition (3.3.20), the Hamiltonian H was replace by a perturbed one H_τ such that $\|H - H_\tau\|_\infty = O(\tau)$. In order to satisfy the condition (3.3.21), Ω_i is then chosen to be a basin of attraction with respect to the gradient flow of this perturbed Hamiltonian H_τ. Consequently, the local PI was first deduced for the perturbed Gibbs measure $\frac{1}{Z}\exp\frac{H_\tau}{2\tau}$ on Ω_i, which then implies PI for the original measure via Holley-Stroock perturbation principle. One side effect of this approach is that the region Ω_i depends on the temperature τ, which is unsuitable in our setting with two different temperatures.

We modify this approach as follows: instead of perturbing the Hamiltonian in the Gibbs measure, we only perturb the Hamiltonian in the Lyapunov function. Given $\tau_2 = \varepsilon$ small enough, we will choose a perturbation $H_\varepsilon = H + V_\varepsilon$ where $V_\varepsilon = O(\varepsilon)$, and choose Ω_i to be the basin of attraction with respect to the gradient flow of H_ε. Then, for every $\tau \le \varepsilon$, we choose the Lyapunov function to be $W_\tau = \exp\frac{H_\varepsilon}{2\tau}$. Then (3.3.21) is satisfied by [MS14, Theorem B.1] and

$$\begin{aligned}\frac{L_\tau W_\tau}{W_\tau} &= -\frac{\nabla H \cdot \nabla H_\varepsilon}{2\tau} + \tau\left(\frac{\Delta H_\varepsilon}{2\tau} + \frac{|\nabla H_\varepsilon|^2}{4\tau^2}\right)\\ &= \frac{1}{2}\Delta H_\varepsilon - \frac{1}{4\tau}\left(|\nabla H|^2 - |\nabla V_\varepsilon|^2\right) \le \frac{L_\varepsilon W_\varepsilon}{W_\varepsilon},\end{aligned}$$

where the last inequality holds as long as $|\nabla V_\varepsilon| \le |\nabla H|$. Then once (3.3.20) is verified for $\tau = \varepsilon$, PI for ν_i^τ follows for every $\tau \le \varepsilon$ on the same region Ω_i.

It turns out the same perturbation used in [MS14] works here. Let $\mathcal{S}$ be the set of critical points of H and $\mathcal{M} = \{m_1, m_2, \dots, m_N\}$ be the set of local minima of H.

Lemma 3.3.18 (ε-modification)**.** *Given a function H satisfying Assumption 2, there exist constants $\varepsilon_0, \lambda_0, a, C \in (0,\infty)$ and a family of C^3 functions $\{V_\varepsilon\}_{0<\varepsilon<\varepsilon_0}$ such that for $H_\varepsilon := H + V_\varepsilon$ the following hold:*

(i) V_ε is supported on $\bigcup_{s\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(s)$ and $|V_\varepsilon(x)| \le C\varepsilon$ for all x.

(ii) Lyapunov-type condition: $|\nabla V_\varepsilon(x)| \le |\nabla H(x)|$ for all x and

$$\frac{1}{2}\Delta H_\varepsilon - \frac{1}{4\varepsilon}(|\nabla H|^2 - |\nabla V_\varepsilon|^2) \le -\lambda_0 \quad \text{for all } x \notin \bigcup_{m\in\mathcal{M}} B_{a\sqrt{\varepsilon}}(m). \tag{3.3.22}$$

Proof. The proof of Lemma 3.3.18 will closely follow that of [MS14, Lemma 3.12]. By [MS14, Lemma 3.11], there exist constants $\varepsilon_0, \lambda_0, a \in (0,\infty)$ such that for all $\varepsilon < \varepsilon_0$,

$$\Delta H - \frac{1}{2\varepsilon}|\nabla H|^2 \le -\lambda_0 \quad \text{for all } x \notin \bigcup_{m\in\mathcal{S}} B_{a\sqrt{\varepsilon}}(s).$$

Thus, it suffices to construct V_ε satisfying (i), (ii) such that (iii) holds for all $x \in \bigcup_{s\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(s)$. Moreover, it suffices to construct V_ε separately near each saddle point s. By translation and rotational/reflection symmetry, we may assume WLOG $s = 0$ and $\nabla^2 H(0)$ is a diagonal matrix with nonzero eigenvalues in increasing order $\lambda_1 \le \cdots \le \lambda_l < 0 < \lambda_{l+1} \le \cdots \le \lambda_n$, where $1 \le l \le n$. If $l = n$, then all eigenvalues are negative, so no perturbation is necessary, and we can just set $V_\varepsilon = 0$ in $B_{a\sqrt{\varepsilon}}(0)$.

From now on we assume $l < n$. Choose a constant $\delta > 0$ small enough such that

$$-\tilde{\delta} := (n-2l)\delta + \sum_{i=1}^{l} \lambda_i < 0, \quad \text{and } \delta \le \frac{1}{2}\min_{1\le i\le n}(|\lambda_i|).$$

We introduce a norm $|\cdot|_\delta$ on $\mathbb{R}^n$ by

$$|x|_\delta^2 := \sum_{i=1}^{l} \frac{1}{2}\delta x_i^2 + \sum_{i=l+1}^{n} \frac{1}{2}(\lambda_i - \delta)x_i^2 =: \frac{1}{2}\langle x, Q_\delta x\rangle,$$

where Q_δ is the positive definite symmetric matrix inducing the norm $|\cdot|_\delta$. It is a diagonal matrix with first l eigenvalues all equal to δ and the last $n-l$ eigenvalues equal to $\lambda_i - \delta$, for $i = l+1, \cdots, n$. The norm $|\cdot|_\delta$ is equivalent to the Euclidean norm $|\cdot|$ and satisfies

$$\frac{\delta}{4}|x|^2 \le |x|_\delta^2 \le \frac{\lambda_n - \delta}{2}|x|^2 \le \frac{\lambda_n}{2}|x|^2.$$

The last ingredient for the construction of V_ε is a smooth cutoff function $\xi : [0,\infty) \to \mathbb{R}$ satisfying for $a > 0$ to be specified later

$$\begin{aligned} 0 \le \xi(0) \le a^2\varepsilon, \quad \xi(r) = 0 \quad \text{for } r \ge a^2\varepsilon, \\ \xi'(r) = -1 \quad \text{for } r \le \frac{1}{4}a^2\varepsilon, \quad -1 \le \xi'(r) \le 0 \quad \text{for } r \ge \frac{1}{4}a^2\varepsilon, \\ 0 \le \xi''(r) \le \frac{2}{a^2\varepsilon}. \end{aligned}$$

(For example, let $0 \le \eta \le 2$ be a smooth function supported on $[\frac{1}{4}, 1]$ with total integral equal to 1 and then choose ξ to be the function satisfying $\xi(a^2\varepsilon) = 0, \xi'(a^2\varepsilon) = 0$ and $\xi''(x) = \frac{1}{a^2\varepsilon}\eta\left(\frac{x}{a^2\varepsilon}\right).$)

Now, we define the ε-perturbation at $s = 0$

$$V_\varepsilon(x) := \xi(|x|_\delta^2), \quad \text{and } H_\varepsilon(x) = H(x) + V_\varepsilon(x).$$

Then condition (i) and (ii) holds. It remains to check (iii) holds for $B_{a\sqrt{\varepsilon}}(0)$. It is enough to establish the following estimates: for ε small enough and a large enough, there exists constant $\lambda_0 \in (0,\infty)$ such that for all $x \in B_{\frac{a}{2}\sqrt{\varepsilon}}(0)$,

$$\Delta H_\varepsilon(x) \le \frac{\lambda_0}{2}, \tag{3.3.23}$$

and there exist constants $R_0, C_\Delta, c_\nabla \in (0,\infty)$ such that for all $x \in B_{R_0}(0)$,

$$\Delta H_\varepsilon(x) \le C_\Delta \tag{3.3.24}$$

$$|\nabla H(x)|^2 - |\nabla V_\varepsilon(x)|^2 \ge c_\nabla |x|^2. \tag{3.3.25}$$

Let's verify (iii) assuming these estimates. By (3.3.25), $|\nabla V_\varepsilon| \le |\nabla H(x)|$ in $B_{a\sqrt{\varepsilon}}(0)$. For $x \in B_{\frac{a}{2}\sqrt{\varepsilon}}(0)$, (3.3.23) implies (3.3.22). For $x \in B_{a\sqrt{\varepsilon}}(0) \setminus B_{\frac{a}{2}\sqrt{\varepsilon}}(0)$, (3.3.24) and (3.3.25) imply that

$$\Delta H_\varepsilon(x) \le C_\Delta, \quad |\nabla H(x)|^2 - |\nabla V_\varepsilon(x)|^2 \ge c_\nabla \frac{a^2}{4}\varepsilon.$$

Choosing $a^2 \ge \frac{8(C_\Delta + \lambda_0)}{c_\nabla}$ gives (3.3.22).

It remains to derive the estimates (3.3.23)-(3.3.25). Compared to the proof of [MS14, Lemma 3.12] (cf. equation (3.22), (3.24)), the only difference is (3.3.25). We will make use of the explicit form of the perturbation V_ε:

$$\begin{aligned}\nabla V_\varepsilon(x) &= \xi'(|x|_\delta^2)Q_\delta x,\\ \nabla^2 V_\varepsilon(x) &= \xi''(|x|_\delta^2)Q_\delta x \otimes Q_\delta x + \xi'(|x|_\delta^2)Q_\delta.\end{aligned}$$

Derivation of (3.3.23): for $|x|_\delta \le \frac{a}{2}\sqrt{\varepsilon}$, $\xi'(|x|_\delta^2) = -1$, so by Taylor expansion we get

$$\Delta H_\varepsilon(x) = \Delta H(0) - \operatorname{tr} Q_\delta + O(|x|) = -\tilde{\delta} + O(\sqrt{\varepsilon}) \le -\frac{\tilde{\delta}}{2}$$

for ε small enough.

Derivation of (3.3.24): By Taylor expansion,

$$\begin{aligned}\Delta H_\varepsilon(x) &= \Delta H(0) + (\Delta H(x) - \Delta H(0)) + \xi''(|x|_\delta^2)|Q_\delta x|^2 + \xi'(|x|_\delta^2)\operatorname{tr} Q_\delta\\ &\le \Delta H(0) + O(|x|) + \frac{2}{a^2\varepsilon}\mathbb{1}_{|x|_\delta^2 \le a\sqrt{\varepsilon}} \cdot 2\lambda_n |x|_\delta^2 - 0\\ &\le \Delta H(0) + O(|x|) + 4\lambda_n.\end{aligned}$$

Choosing R_0 small enough, then for $x \in B_{R_0}(0)$, we get (3.3.24).

Derivation of (3.3.25): By Taylor expansion, $|\nabla H(x) - \nabla^2 H(0)x| = O(|x|^2)$, so

$$|\nabla H(x)|^2 \ge (|\nabla^2 H(0)x| - |\nabla H(x) - \nabla^2 H(0)x|)^2 \ge |\nabla^2 H(0)x|^2 - O(|x|^3).$$

Moreover, since $-1 \le \xi' \le 0$,

$$|\nabla V_\varepsilon(x)|^2 = \xi'(|x|_\delta^2)^2 |Q_\delta x|^2 \le |Q_\delta x|^2.$$

Thus,

$$\begin{aligned}|\nabla H(x)|^2 - |\nabla V_\varepsilon(x)|^2 &\ge \langle x, (\nabla^2 H(0)^2 - Q_\delta^2)x\rangle - O(|x|^3)\\ &\ge \tilde{c}|x|^2 - O(|x|^3),\end{aligned}$$

where $\tilde{c}$ is the constant

$$\tilde{c} = \min\left(\min_{1\le i\le l}(\lambda_i^2 - \delta^2), \min_{l+1\le j\le n}(\lambda_j^2 - (\lambda_j - \delta)^2)\right) > 0.$$

Choosing R_0 small enough, then for $x \in B_{R_0}(0)$, we get (3.3.25). □

We call a family $\{H_\varepsilon\}_{0<\varepsilon<\varepsilon_0}$ having properties (i)-(v) of Lemma 3.3.18 a family of ε-modifications of H, and from now on we will fix one such family.

Lemma 3.3.19. *For any $\varepsilon < \varepsilon_0$, the set of local minima of H_ε is $\mathcal{M}$.*

Proof. Property (iii) of Lemma 3.3.18 implies $|\nabla H_\varepsilon| \neq 0$ or $\Delta H_\varepsilon < 0$ outside $\bigcup_{m\in\mathcal{M}} B_{a\sqrt{\varepsilon}}(m)$, so H_ε has no local minima there. Property (ii) of Lemma 3.3.18 implies that inside $B_{R_0}(m)$, $H_\varepsilon = H$ and therefore has a unique local minimum at m for each $m \in \mathcal{M}$. □

For each local minimum m_i of H, let $\Omega_i = \Omega_i^{\tau_2}$ be the associated basin of attraction w.r.t. the deterministic gradient flow defined by H_{τ_2}, that is

$$\Omega_i := \left\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i,\ \frac{dy_t}{dt} = -\nabla H_{\tau_2}(y_t),\ y_0 = y \right\}.$$

Then $(\Omega_i)_{i=1}^N$ is a partition of $\mathbb{R}^n$ up to Lebesgue null sets. The preceding shows ν_i^τ defined by (3.3.2)satisfies PI(ρ) with $\rho^{-1} = O(\tau)$ for all $\tau \le \tau_2$.

Equipped with the Poincaré inequality for ν_i^τ, the log-Sobolev inequality for ν_i^τ is now a simple consequence of the following criterion from [MS14].

Lemma 3.3.20 (Lyapunov condition for local LSI, Theorem 3.15 in [MS14])**.** *Assume that*

(i) There exists a smooth function $W_\tau : \Omega_i \to (0,\infty)$ and constants $\lambda, b > 0$ such that for $L_\tau := \tau\Delta - \nabla H \cdot \nabla$

$$\frac{L_\tau W_\tau}{W_\tau} \le -\lambda|x|^2 + b \quad \forall x \in \Omega_i.$$

(ii) $\nabla^2 H \ge -K_H$ for some $K_H > 0$ and ν_i^τ satisfies PI(ρ).

(iii) W_τ satisfies Neumann boundary condition on Ω_i (see (3.3.21)).

Then ν_i^τ satisfies LSI(α) with

$$\alpha^{-1} \leq 2\sqrt{\frac{\tau}{\lambda}\left(\frac{1}{2} + \frac{b + \lambda\nu_i^\tau(|x|^2)}{\rho\tau}\right)} + \frac{K_H}{\lambda}\left(\frac{1}{2} + \frac{b + \lambda\nu_i^\tau(|x|^2)}{\rho\tau}\right) + \frac{2}{\rho},$$

where $\nu_i^\tau(|x|^2)$ denotes the second moment of ν_i^τ.

Choosing W_τ to be the same we chose for the PI, it is straightforward to check the conditions (i)-(iii) holds and check the second moment $\nu_i^\tau(|x|^2)$ is uniformly bounded. Indeed, the verification of these facts is virtually identical to the counterpart in [MS14] (cf. Lemmas 3.17-3.19). Finally, $\rho^{-1} = O(\tau)$ yields $\alpha^{-1} = O(1)$.

This establishes Lemma 3.3.13 and Lemma 3.3.14 up to the verification that $(\Omega_i)_{i=1}^N$ is an admissible partition, which we now show.

Lemma 3.3.21. *For any $\varepsilon < \varepsilon_0$, let Ω_i^ε be the basin of attraction associated to the minimum m_i w.r.t. the deterministic gradient flow defined by the ε-modified potential H_ε, i.e.*

$$\Omega_i^\varepsilon := \left\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i,\ \frac{dy_t}{dt} = -\nabla H_\varepsilon(y_t),\ y_0 = y \right\}.$$

Then there exists $\tau_0 > 0$ such that for all $\tau < \tau_0$ and all $\varepsilon < \varepsilon_0$,

$$\int_{\Omega_i^\varepsilon} \exp\left(-\frac{H(x)}{\tau}\right) dx = \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + \omega(\tau)).$$

This result is essentially shown in [MS14, Lemma 3.12]. The only change is that we modify the proof to get a result independent of the perturbation size ε, which is needed here due to the presence of two temperatures.

Proof. By property (ii) of Lemma 3.3.18 and Assumption, there exists R_0 small enough such that $H_\varepsilon = H$ on $B_{R_0}(m_i)$ and H is strongly convex on $B_{R_0}(m_i)$:

$$\nabla^2 H(x) \geq c\,\mathrm{id} \quad \text{for all } x \in B_{R_0}(m_i)$$

for some constant $c > 0$. Then $B_{R_0}(m_i) \subset \Omega_i^\varepsilon$. Define for $r_0 > 0$ specified later the ellipsoid

$$E_i := \{x \in \mathbb{R}^n : |\nabla^2 H(m_i)^{1/2}(x - m_i)| \leq \sqrt{2r_0\tau|\log\tau|}\}.$$

For τ small enough, $E_i \subset B_{R_0}(m_i) \subset \Omega_i^\varepsilon$. Moreover, there is a constant $\kappa > 0$ such that $B_{\sqrt{2\kappa r_0 \tau |\log \tau|}}(m_i) \subset E_i$.

We split the integral into

$$\int_{\Omega_i^\varepsilon} \exp\left(-\frac{H(x)}{\tau}\right) dx = \underbrace{\int_{E_i} \exp\left(-\frac{H(x)}{\tau}\right) dx}_{:=I_1} + \underbrace{\int_{\Omega_i^\varepsilon \setminus E_i} \exp\left(-\frac{H(x)}{\tau}\right) dx}_{:=I_2}$$

The main contribution is from the term I_1. By a straightforward calculation (cf. the proof of [MS14, Lemma 3.12]), I_1 has the asymptotic expansion

$$I_1 = \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right) \left(1 - \frac{\Gamma(n/2, r_0|\log\tau|)}{\Gamma(n/2)}\right) \exp O(\sqrt{\tau}|\log\tau|^{3/2}),$$

where $\Gamma(s,x) := \int_x^\infty r^{s-1}x^r dr$ is the incomplete Gamma function. Note that

$$\Gamma(n/2, r_0|\log\tau|) = \int_{r_0|\log\tau|}^\infty r^{\frac{n}{2}-1} e^{-r} dr = o(\tau) \int_{r_0|\log\tau|}^\infty e^{-r/2} dr = O(\tau^{r_0/2}).$$

Thus, for $r_0 \geq 1$,

$$I_1 = \frac{(2\pi\tau)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\tau}\right)(1 + \omega(\tau)).$$

Thus, it remains to show I_2 is small enough in comparison, e.g. $I_2 = \exp\left(-\frac{H(m_i)}{\tau}\right) O(\tau^{(n+1)/2})$.

By [MS14, Lemma 3.14], the growth condition 3.2.6 implies there exists $c_H > 0$ such that for R large enough

$$H(x) \geq \min_{|z|=R} H(z) + c_H(|x| - R) \quad \text{for all } |x| \geq R.$$

By choosing R large enough,

$$H(x) \geq H(m_i) + 1 + c_H(|x| - R) \quad \text{for all } |x| \geq R.$$

We split I_2 into

$$I_2 = \underbrace{\int_{(\Omega_i^\varepsilon \setminus E_i) \cap B_R(0)} \exp\left(-\frac{H(x)}{\tau}\right) dx}_{:=I_3} + \underbrace{\int_{\Omega_i^\varepsilon \setminus B_R(0)} \exp\left(-\frac{H(x)}{\tau}\right) dx}_{:=I_4}.$$

Estimate of I_3: We obtain a pointwise bound for the integrand by dividing into two cases.

Case 1: $x \in \Omega_i^\varepsilon \setminus B_{R_0}(m_i)$, the gradient flow associated to H_ε starting at x must hit $\partial B_{R_0}(m_i)$. By strong convexity of $H_\varepsilon = H$ on $B_{R_0}(m_i)$,

$$H_\varepsilon(x) \geq \min_{z \in \partial B_{R_0}(m_i)} H_\varepsilon(z) \geq \min_{z \in \partial B_{R_0}(m_i)} \frac{c}{2}|z - m_i|^2 + H(m_i) = \frac{cR_0^2}{2} + H(m_i).$$

Thus, by property (i) of Lemma 3.3.18, for ε small enough,

$$H(x) \geq \frac{cR_0^2}{2} - C\varepsilon + H(m_i) \geq \frac{cR_0^2}{4} + H(m_i).$$

Case 2: $x \in B_{R_0}(m_i) \setminus E_i$, then $H_\varepsilon(x) = H(x)$ and the gradient flow associated to H starting at x must hit ∂E_i. By strong convexity of H on $B_{R_0}(m_i)$,

$$H(x) \geq \min_{z \in \partial E_i} H(z) \geq \min_{z \in \partial E_i} \frac{c}{2}|z - m_i|^2 + H(m_i) \geq c\kappa r_0 \tau |\log \tau| + H(m_i).$$

Thus, for τ small enough,

$$I_3 \leq \exp\left[-c\kappa r_0 |\log \tau| - \frac{H(m_i)}{\tau}\right] |B_R(0)| = \exp\left(-\frac{H(m_i)}{\tau}\right) O(\tau^{c\kappa r_0}).$$

Lastly, I_4 is estimated as

$$\begin{aligned} I_4 &\leq \exp\left(-\frac{H(m_i)+1}{\tau}\right) \int_{|x|>R} \exp\left(-c_H \frac{|x| - R}{\tau}\right) dx \\ &= \exp\left(-\frac{H(m_i)}{\tau}\right) \exp\left(-\frac{1}{\tau}\right) O(\tau). \end{aligned}$$

For $r_0 \geq \frac{n+1}{2c\kappa}$, $I_2 = I_3 + I_4 = \exp\left(-\frac{H(m_i)}{\tau}\right) O(\tau^{(n+1)/2})$, so we are done. □

3.3.3 Cost of exchanging temperatures: proof of Lemma 3.3.8

In order to prove Lemma 3.3.8, we observe that the local Gibbs measures ν_i^τ are close to a class of truncated Gaussian measures in the sense of mean-difference (cf. [MS14, Lemma 4.6]).

Definition 3.3.22 (Truncated Gaussian measure)**.** *Given $m \in \mathbb{R}^n$, Σ a symmetric positive definite $n \times n$ matrix, $R > 0$, consider the ellipsoid*

$$E_i^\tau := \{x \in \mathbb{R}^n : (x-m) \cdot \Sigma^{-1}(x-m) \leq R^2\tau\}.$$

The truncated Gaussian measure γ^τ at temperature τ with mean m and covariance Σ on scale R is defined to be

$$\gamma^\tau(x) := \frac{\exp\left(-\frac{1}{2\tau}(x-m)\cdot\Sigma^{-1}(x-m)\right)}{Z_R\sqrt{\tau}^n\sqrt{\det\Sigma}}\mathbb{1}_{E^\tau},$$

where $Z_R := \int_{B_R(0)} \exp(-|x|^2/2)dx = \sqrt{2\pi}^n(1 - O(e^{-R^2}R^{n-2}))$.

Lemma 3.3.23 (Approximation by truncated Gaussian)**.** *For $\tau \leq \tau_2$, let γ_i^τ be the truncated Gaussian measure at temperature τ with mean m_i and covariance $\Sigma_i = (\nabla H^2(m_i))^{-1}$ on scale $R(\tau_2) = |\log\tau_2|^{1/2}$. Then*

$$\frac{d\gamma_i^\tau}{d\nu_i^\tau}(x) = 1 + \omega(\tau_2), \tag{3.3.26}$$

uniformly in the support of γ_i^τ, and for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$

$$(\mathbb{E}_{\nu_i^\tau} f - \mathbb{E}_{\gamma_i^\tau} f)^2 \leq \mathrm{Var}_{\nu_i^\tau}\left(\frac{d\gamma_i^\tau}{d\nu_i^\tau}\right)\mathrm{Var}_{\nu_i^\tau}(f) \leq \omega(\tau_2)\tau\mathbb{E}_{\nu_i^\tau}|\nabla f|^2,$$

where $\omega(\tau_2) := O(\sqrt{\tau_2}|\log\tau_2|^{3/2})$.

We omit the proof of Lemma 3.3.23, which is the same as [MS14, Lemma 4.6] with only minor changes.

Corollary 3.3.24. *For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$\left(\mathbb{E}_{\pi_{ij}^\sigma} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\gamma_j^{\tau_{\sigma(2)}}} f\right)^2 \leq \omega(\tau_2)\mathbb{E}_{\pi_{ij}^\sigma}\left(\tau_{\sigma(1)}|\nabla_{x_1} f|^2 + \tau_{\sigma(2)}|\nabla_{x_2} f|^2\right),$$

where $\omega(\tau_2) := O(\sqrt{\tau_2}|\log\tau_2|^{3/2})$.

Proof. This follows from the previous lemma by writing

$$\begin{aligned}\mathbb{E}_{\pi_{ij}^\sigma} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\gamma_j^{\tau_{\sigma(2)}}} f = &\left(\mathbb{E}_{\nu_i^{\tau_{\sigma(1)}}\otimes\nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\nu_j^{\tau_{\sigma(2)}}} f\right)\\ &+ \left(\mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\nu_j^{\tau_{\sigma(2)}}} f - \mathbb{E}_{\gamma_i^{\tau_{\sigma(1)}}\otimes\gamma_j^{\tau_{\sigma(2)}}} f\right).\end{aligned}$$

□

This reduces our task to proving mean-difference estimate for truncated Gaussian.

Lemma 3.3.25 (Mean-difference estimate for truncated Gaussians at two temperatures)**.** *For any smooth function $f : \mathbb{R}^n \to \mathbb{R}$*

$$(\mathbb{E}_{\gamma_i^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1}} f)^2 \le C_n \|\Sigma_i\| \left(1 + \Phi_n\left(\frac{\tau_2}{\tau_1}\right)\right) \tau_2 \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2,$$

where the function Φ_n is given by (3.2.10)*, and C_n is a constant only depending on n.*

Proof. By change of variables, it suffices to show the first inequality for $m_i = 0, \Sigma_i = \mathrm{id}$. From the Cauchy-Schwarz inequality and the fundamental theorem of calculus, we can deduce

$$\begin{aligned}
(\mathbb{E}_{\gamma_i^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1}} f)^2 &\le \mathbb{E}_{\gamma_i^1}\big(f(\sqrt{\tau_2}X) - f(\sqrt{\tau_1}X)\big)^2 \\
&\le \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1}r}^{\sqrt{\tau_2}r} |\nabla f(s\omega)| ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\
&\le 2(I_1 + I_2),
\end{aligned}$$

where, for some $0 < \kappa \le R$ to be specified later,

$$\begin{aligned}
I_1 &:= \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1}r}^{\sqrt{\tau_2}r} |\nabla f(s\omega)| \mathbb{1}_{s \le \kappa\sqrt{\tau_2}} ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr, \\
I_2 &:= \int_{S^{n-1}} d\omega \int_0^R \left(\int_{\sqrt{\tau_1}r}^{\sqrt{\tau_2}r} |\nabla f(s\omega)| \mathbb{1}_{s > \kappa\sqrt{\tau_2}} ds\right)^2 \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr.
\end{aligned}$$

Estimate for I_2: By Cauchy-Schwarz,

$$\begin{aligned}
I_2 &\le \int_{S^{n-1}} d\omega \int_0^R (\sqrt{\tau_2}r - \sqrt{\tau_1}r) \left(\int_{\kappa\sqrt{\tau_2}}^{R\sqrt{\tau_2}} |\nabla f(s\omega)|^2 \mathbb{1}_{s \le r\sqrt{\tau_2}} ds\right) \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\
&\le \sqrt{\tau_2} \int_{S^{n-1}} d\omega \int_{\kappa\sqrt{\tau_2}}^{R\sqrt{\tau_2}} |\nabla f(s\omega)|^2 \left(\int_{\frac{s}{\sqrt{\tau_2}}}^R \frac{e^{-\frac{r^2}{2}}}{Z_R} r^n dr\right) ds.
\end{aligned}$$

Using integration by parts and standard Gaussian tail bound, for $s \ge \kappa\sqrt{\tau_2}$,

$$\int_{\frac{s}{\sqrt{\tau_2}}}^R e^{-\frac{r^2}{2}} r^n dr \le C_n(1 + \kappa^{-(n-1)}) e^{-\frac{s^2}{2\tau_2}} \left(\frac{s^2}{\tau_2}\right)^{\frac{n-1}{2}},$$

where C_n is a constant only depending on n. This gives

$$I_2 \le C_n(1 + \kappa^{-(n-1)}) \tau_2 \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2.$$

Estimate for I_1: By Cauchy-Schwarz

$$\begin{aligned} I_1 &\leq \int_{S^{n-1}} d\omega \int_0^R \left(\int_0^{\kappa\sqrt{\tau_2}} |\nabla f(s\omega)|^2 s^{n-1} ds \right) \left(\int_{\sqrt{\tau_1} r}^{\sqrt{\tau_2} r} s^{-(n-1)} ds \right) \frac{e^{-\frac{r^2}{2}}}{Z_R} r^{n-1} dr \\ &= \frac{1}{Z_R} \|\nabla f\|^2_{L^2(B_{\kappa\sqrt{\tau_2}}(0))} \int_0^R \left(\int_{\sqrt{\tau_1}}^{\sqrt{\tau_2}} u^{-(n-1)} du \right) r e^{-\frac{r^2}{2}} dr \\ &\leq C_n e^{\frac{\kappa^2}{2}} \tau_2 \mathbb{E}_{\gamma_i^{\tau_2}} |\nabla f|^2 \cdot \Phi_n \left(\frac{\tau_2}{\tau_1} \right), \end{aligned}$$

where C_n is a constant only depending on n. The conclusion now follows if we choose $\kappa = R$ when $R < 1$ and $\kappa = 1$ when $R \geq 1$. □

Corollary 3.3.26. *For any smooth function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*

$$(\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f)^2 \leq \left(1 + \Phi_n \left(\frac{\tau_2}{\tau_1} \right) \right) O(\tau_2) \big(\mathbb{E}_{\pi_{ij}^+} |\nabla_{x_2} f|^2 + \mathbb{E}_{\pi_{ij}^-} |\nabla_{x_1} f|^2 \big).$$

Proof. This follows from the previous lemma and (3.3.26) by writing

$$\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f = (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_2}} f - \mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\gamma_1}} f) + (\mathbb{E}_{\gamma_i^{\tau_1} \otimes \gamma_j^{\tau_1}} f - \mathbb{E}_{\gamma_i^{\tau_2} \otimes \gamma_j^{\tau_1}} f).$$

□

Lemma 3.3.8 follows from Corollary 3.3.24 and 3.3.26.

Remark 3.3.27. *One can show a weaker version of Lemma 3.3.8 by a simpler approach: First we split the mean-difference as*

$$\begin{aligned} (\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 &= (\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\nu_i^{\tau_1} \otimes \nu_j^{\tau_1}} f + \mathbb{E}_{\nu_i^{\tau_1} \otimes \nu_j^{\tau_1}} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \\ &\leq 2\mathbb{E}^{x_1}_{\nu_i^{\tau_1}} (\mathbb{E}^{x_2}_{\nu_j^{\tau_2}} f - \mathbb{E}^{x_2}_{\nu_j^{\tau_1}} f)^2 + 2\mathbb{E}^{x_2}_{\nu_j^{\tau_1}} (\mathbb{E}^{x_1}_{\nu_i^{\tau_1}} f - \mathbb{E}^{x_1}_{\nu_i^{\tau_2}} f)^2. \end{aligned}$$

Now, using the covariance representation of mean-difference and Cauchy-Schwarz

$$\begin{aligned} (\mathbb{E}_{\nu_k^{\tau_2}} g - \mathbb{E}_{\nu_k^{\tau_1}} g)^2 &= \left(\mathbb{E}_{\nu_k^{\tau_2}} g - \mathbb{E}_{\nu_k^{\tau_2}} g \frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right)^2 = \mathrm{Cov}_{\nu_k^{\tau_2}} \left(g, \frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right)^2 \\ &\leq \mathrm{Var}_{\nu_k^{\tau_2}}(g) \, \mathrm{Var}_{\nu_k^{\tau_2}} \left(\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right) \leq O(\tau_2) \mathbb{E}_{\nu_k^{\tau_2}} |\nabla g|^2 \mathbb{E}_{\nu_k^{\tau_1}} \left(\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} \right). \end{aligned}$$

Finally, using the partition size given in (3.3.1) *we have a uniform estimate on the relative density*

$$\frac{d\nu_k^{\tau_1}}{d\nu_k^{\tau_2}} = \frac{\nu_k^{\tau_2}(\Omega_k)}{\nu_k^{\tau_1}(\Omega_k)} e^{-H(x)(\tau_1^{-1}-\tau_2^{-1})} \leq \frac{\nu_k^{\tau_2}(\Omega_k)}{\nu_k^{\tau_1}(\Omega_k)} \leq \left(\frac{\tau_2}{\tau_1}\right)^{\frac{n}{2}} (1 + O(\sqrt{\tau_2}|\log \tau_2|^{3/2})).$$

Applying the preceding with $k = j, g(\cdot) = f(x_1, \cdot)$ *and* $k = i, g(\cdot) = f(\cdot, x_2)$*, respectively, we obtain the estimate*

$$(\mathbb{E}_{\pi_{ij}^+} f - \mathbb{E}_{\pi_{ij}^-} f)^2 \leq \left(\frac{\tau_2}{\tau_1}\right)^{\frac{n}{2}} O(\tau_2)(\mathbb{E}_{\pi_{ij}^+}|\nabla_{x_2} f|^2 + \mathbb{E}_{\pi_{ij}^-}|\nabla_{x_1} f|^2).$$

In comparison to Lemma 3.3.8, the dependence in the ratio τ_2/τ_1 *appearing in front of the Dirichlet form is about one order worse in exponent.*

3.3.4 Optimality of dependence on temperature ratio: proof of Proposition 3.2.6

It suffices to consider test functions of the form $f(x, y) = f(x)$. This is equivalent to replacing μ by its first marginal, which is $\bar{\mu} = \frac{1}{2}(\nu^{\tau_1} + \nu^{\tau_2})$. In this case, $\mathrm{Var}_\mu(f)$ and $\mathcal{E}_\mu(f)$ reduces to

$$\mathrm{Var}_{\bar{\mu}}(f) = \frac{1}{2}(\mathrm{Var}_{\nu^{\tau_1}}(f) + \mathrm{Var}_{\nu^{\tau_2}}(f)) + \frac{1}{4}(\mathbb{E}_{\nu^{\tau_1}} f - \mathbb{E}_{\nu^{\tau_2}} f)^2,$$
$$\mathcal{E}_{\bar{\mu}}(f) = \frac{1}{2}(\tau_1 \mathbb{E}_{\nu^{\tau_1}}|\nabla f|^2 + \tau_2 \mathbb{E}_{\nu^{\tau_2}}|\nabla f|^2).$$

We further restrict f to $C_c(\Omega_1)$. By (3.3.1) and (3.2.8), $\nu^{\tau_1}(\Omega_1), \nu^{\tau_2}(\Omega_1) \approx 1$ once τ_1, τ_2 are small enough, so $\frac{d\nu_1^{\tau_1}}{d\nu^{\tau_1}}, \frac{d\nu_1^{\tau_2}}{d\nu^{\tau_2}} \approx 1$ on Ω_1 (see equation (3.3.2)). Then by applying Young's inequality

$$\mathrm{Var}_{\bar{\mu}}(f) \gtrsim (\mathbb{E}_{\nu^{\tau_1}} f)^2 - 4(\mathbb{E}_{\nu^{\tau_2}} f)^2 \gtrsim (\mathbb{E}_{\nu_1^{\tau_1}} f)^2 - 5(\mathbb{E}_{\nu_1^{\tau_2}} f)^2,$$
$$\mathcal{E}_{\bar{\mu}}(f) \lesssim \tau_1 \mathbb{E}_{\nu_1^{\tau_1}}|\nabla f|^2 + \tau_2 \mathbb{E}_{\nu_1^{\tau_2}}|\nabla f|^2.$$

By change of variables, we may assume $m_1 = 0, \Sigma_1 = (\nabla^2 H(m_1))^{-1} = \mathrm{id}$. We consider a test function of the form

$$f(x) = f_\varepsilon(x) = h(|x|/\sqrt{\varepsilon}),$$

where $h \geq 0$ is a compactly supported, absolutely continuous function and $\tau_1 \leq \varepsilon \leq \tau_2$ is a scaling parameter, both to be specified later. As in the proof of Lemma 3.3.8, we will approximate by truncated Gaussian measures (see Definition 3.3.22). Since $\varepsilon \leq \tau_2$, f_ε is supported in the support of $\gamma_1^{\tau_2}$. By Lemma 3.3.23,

$$\mathrm{Var}_{\bar{\mu}}(f) \gtrsim (\mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon)^2 - 6(\mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon)^2, \tag{3.3.27}$$

$$\mathcal{E}_{\bar{\mu}}(f) \lesssim \tau_1 \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 + \tau_2 \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2, \tag{3.3.28}$$

if τ_2 is small enough. We have:

$$\tau_1 \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 = \frac{\tau_1}{\varepsilon} \mathbb{E}_{\nu_1^{\frac{\tau_1}{\varepsilon}}} |\nabla f_1|^2, \tag{3.3.29}$$

$$\tau_2 \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 = \frac{\tau_2}{\varepsilon} \mathbb{E}_{\gamma_1^{\frac{\tau_2}{\varepsilon}}} |\nabla f_1|^2 \leq \frac{1}{\sqrt{2\pi}^n} (\varepsilon/\tau_2)^{(n-2)/2} \|\nabla f_1\|_{L^2}^2, \tag{3.3.30}$$

$$\mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon = \mathbb{E}_{\gamma_1^{\frac{\tau_2}{\varepsilon}}} f_1 \leq \frac{1}{\sqrt{2\pi}^n} (\varepsilon/\tau_2)^{n/2} \|f_1\|_{L^1}, \tag{3.3.31}$$

and for any $r \geq 0$,

$$\mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon = \mathbb{E}_{\gamma_1^{\frac{\tau_1}{\varepsilon}}} f_1 \geq P_{\gamma_1^{\frac{\tau_1}{\varepsilon}}}(|X| \leq r) \cdot \inf_{|x| \leq r} f_1 \geq \left(1 - n \exp\left(-\frac{r^2}{2n} \frac{\varepsilon}{\tau_1}\right)\right) \cdot \inf_{[0,r]} h. \tag{3.3.32}$$

In the following $R_n > 0$ is the number such that $\exp\left(-\frac{R_n^2}{2n}\right) = \frac{1}{2}$.

Case 1: $n \geq 3$. We choose h to be a compactly supported smooth function such that $h = 1$ on $[0, R_n]$, decreases to 0 on $[R_n, 2R_n]$ and is 0 outside $[0, 2R_n]$. Then

$$\tau_2 \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.3.30)}{\lesssim} (\varepsilon/\tau_2)^{(n-2)/2},\ \mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon \overset{(3.3.31)}{\lesssim} (\varepsilon/\tau_2)^{n/2},\ \mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon \overset{(3.3.32)}{\geq} \frac{1}{2},$$

where the implicit constants only depend on the dimension n and the function h. Since $h' = 0$ on $[0, R_n]$

$$\tau_1 \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 \overset{(3.3.29)}{\leq} \frac{\tau_1}{\varepsilon} \|h'\|_{L^\infty}^2 P_{\nu_1^{\frac{\tau_1}{\varepsilon}}}(|X| \geq R_n) \leq \frac{\tau_1}{\varepsilon} \|h'\|_{L^\infty}^2 C_H e^{-c_H \frac{\varepsilon}{\tau_1}} \lesssim_m (\tau_1/\varepsilon)^m,$$

for every positive integer m, where the constants $c_H, C_H > 0$ only depend on the Hamiltonian H. The second inequality is a consequence of Assumption 2 (see [MS14, Lemma 3.13]). Now,

for any $0 < \eta < \frac{1}{2}$, set $\varepsilon = \tau_1^{1-\eta}\tau_2^{\eta}$, and choose m large enough so that $\eta m \geq (1-\eta)(n-2)/2$, we obtain

$$\mathcal{E}_{\bar{\mu}}(f) \overset{(3.3.28)}{\lesssim_\eta} (\tau_1/\tau_2)^{(1-\eta)(n-2)/2},\ \mathrm{Var}_{\bar{\mu}}(f) \overset{(3.3.27)}{\gtrsim_\eta} (\tau_2/\tau_1)^{(1-\eta)(n-2)/2}\mathcal{E}_{\bar{\mu}}(f),$$

if τ_2, τ_1/τ_2 are both small enough.

Case 2: $n = 2$. Let h be the function given by

$$h(r) = \begin{cases} 1 & \text{for } 0 \leq r \leq r_0, \\ 2(1-r^\alpha) & \text{for } r_0 \leq r \leq 1, \\ 0 & \text{for } r \geq 1, \end{cases}$$

for parameters $0 < \alpha < 1, 0 < r_0 < 1$ satisfying $r_0^\alpha = \frac{1}{2}$, to be specified later. Then h is absolutely continuous, $h' = 0$ on $[0, r_0]$, and by direct computation

$$\|f_1\|_{L^1} \leq \pi\alpha, \quad \|\nabla f_1\|_{L^\infty}^2 = \alpha^2 r_0^{-2}, \quad \|\nabla f_1\|_{L^2}^2 = 3\pi\alpha.$$

We choose $\varepsilon = \tau_2$ and $r_0^{2\frac{\tau_2}{\tau_1}} = R_2^2$ (which is possible once τ_1/τ_2 is small enough). Then:

$$\begin{aligned} \mathbb{E}_{\gamma_1^{\tau_2}} f_\varepsilon &\overset{(3.3.31)}{\leq} \frac{1}{2\pi}\frac{\varepsilon}{\tau_2}\|f_1\|_{L^1} \leq \frac{\alpha}{2},\ \mathbb{E}_{\gamma_1^{\tau_1}} f_\varepsilon \overset{(3.3.32)}{\geq} \frac{1}{2}, \\ \tau_1 \mathbb{E}_{\nu_1^{\tau_1}} |\nabla f_\varepsilon|^2 &\overset{(3.3.29)}{\leq} \frac{\tau_1}{\varepsilon}\|\nabla f_1\|_{L^\infty}^2 \leq \frac{\alpha^2}{R_2^2},\ \tau_2 \mathbb{E}_{\gamma_1^{\tau_2}} |\nabla f_\varepsilon|^2 \overset{(3.3.30)}{\leq} \frac{1}{2\pi}\|\nabla f_1\|_{L^2}^2 = \frac{3\alpha}{2}. \end{aligned}$$

Since $r_0^\alpha = \frac{1}{2}$, $\frac{1}{\alpha} = \frac{1}{2\log 2}\log\left(\frac{\tau_2}{\tau_1 R_2^2}\right)$. Thus

$$\mathcal{E}_{\bar{\mu}}(f) \overset{(3.3.28)}{\lesssim} \frac{\alpha^2}{R_2^2} + \frac{3\alpha}{2}, \quad \mathrm{Var}_{\bar{\mu}}(f) \overset{(3.3.27)}{\gtrsim} \frac{1}{\alpha}\mathcal{E}_{\bar{\mu}}(f) \gtrsim \log\left(\frac{\tau_2}{\tau_1}\right)\mathcal{E}_{\bar{\mu}}(f),$$

if $\tau_2, \tau_1/\tau_2$ are both small enough.

3.3.5 Optimality in one dimension: proofs of Proposition 3.2.7 and Proposition 3.2.8

It suffices to consider test functions of the form $f(x, y) = g(x)g(y)$. This is equivalent to replacing μ by $\pi = \nu^{\tau_1} \otimes \nu^{\tau_2}$. In this case, $\mathrm{Var}_\mu(f), \mathrm{Ent}_\mu(f^2), \mathcal{E}_\mu(f), \mathcal{I}_\mu(f)$ reduce to

$$\mathrm{Var}_\pi(f) = \mathbb{E}_{\nu^{\tau_1}} g^2 \mathbb{E}_{\nu^{\tau_2}} g^2 - (\mathbb{E}_{\nu^{\tau_1}} g)^2 (\mathbb{E}_{\nu^{\tau_2}} g)^2,$$

$$\mathrm{Ent}_\pi(f) = \mathbb{E}_{\nu^{\tau_1}} g^2 \,\mathrm{Ent}_{\nu^{\tau_2}}\, g^2 + \mathbb{E}_{\nu^{\tau_2}} g^2 \,\mathrm{Ent}_{\nu^{\tau_1}}\, g^2,$$
$$\frac{1}{2}\mathcal{I}_\pi(f^2) = \mathcal{E}_\pi(f) = \tau_1 \mathbb{E}_{\nu^{\tau_1}} (g')^2 \mathbb{E}_{\nu^{\tau_2}} g^2 + \tau_2 \mathbb{E}_{\nu^{\tau_1}} g^2 \mathbb{E}_{\nu^{\tau_2}} (g')^2.$$

We represent ν^{τ_i} as the mixture

$$\nu^{\tau_i} = Z_1^{\tau_i}\nu_1^{\tau_i} + Z_2^{\tau_i}\nu_2^{\tau_i} \quad \text{where } \nu_1^{\tau_i} := \nu^{\tau_i}|_{\Omega_1}, \nu_2^{\tau_i} := \nu^{\tau_i}|_{\Omega_2},$$

where $\Omega_1 := (-\infty, s), \Omega_2 := (s, \infty)$. Denote

$$Z_1^{\tau_i} = \nu^{\tau_i}(\Omega_1) \approx 1, \quad Z_2^{\tau_i} = \nu^{\tau_i}(\Omega_2) \approx \frac{\sqrt{H''(m_1)}}{\sqrt{H''(m_2)}} e^{-H(m_2)/\tau_i}.$$

Here and below, $\approx$ (resp. $\lessapprox$) means equality (resp. less than or equal) up to a multiplicative factor of $1 + O(\sqrt{\tau_2}|\log \tau_2|^{3/2})$.

Proof of Proposition 3.2.7: Imposing $\mathbb{E}_{\nu^{\tau_1}} g = 0$, we get

$$\frac{\mathcal{E}_\pi(f)}{\mathrm{Var}_\pi(f)} = \tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}} (g')^2}{\mathbb{E}_{\nu^{\tau_1}} g^2} + \tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}} (g')^2}{\mathbb{E}_{\nu^{\tau_2}} g^2}.$$

We make the following ansatz for g:

$$g(x) = \begin{cases} g(m_1) & \text{for } x \le s - \delta \\ g(m_1) + \frac{g(m_2) - g(m_1)}{\sqrt{2\pi\sigma\tau_2}} \cdot \kappa \int_{s-\delta}^x e^{-(y-s)^2/(2\sigma\tau_2)} dy & \text{for } s - \delta < x < s + \delta \\ g(m_2) & \text{for } x > s + \delta, \end{cases}$$

where σ is a positive constant to be specified later, $\delta = \sqrt{2r_0\tau_2|\log \tau_2|}$ for some positive constant r_0 to be chosen later, and κ is chosen so that g is continuous at $s + \delta$. (This is the same kind of ansatz used in [MS14, Section 2.4].) Then $\kappa = 1 + O(\tau_2^{-r_0/\sigma}) \approx 1$ once r_0 is large enough. Fix such a choice of r_0. For τ_2 small enough, δ is small enough so that

$$\mathbb{E}_{\nu^{\tau_i}} g \approx g(m_1) Z_1^{\tau_i} + g(m_2) Z_2^{\tau_i}.$$

This motivates the choice

$$g(m_1) \approx -1, g(m_2) \approx 1/Z_2^{\tau_1},$$

such that $\mathbb{E}_{\nu^{\tau_1}} g = 0$. Then

$$\mathbb{E}_{\nu^{\tau_2}} g^2 \approx Z_1^{\tau_2} g(m_1)^2 + Z_2^{\tau_2} g(m_2)^2 \approx g(m_2)^2 Z_2^{\tau_2},$$

$$\mathbb{E}_{\nu^{\tau_1}} g^2 \approx Z_1^{\tau_1} g(m_1)^2 + Z_2^{\tau_1} g(m_2)^2 \approx g(m_2)^2 Z_2^{\tau_1}.$$

Finally, we compute the Dirichlet forms. By Taylor expansion of H around s

$$\begin{aligned}
\mathbb{E}_{\nu^{\tau_2}}(g')^2 &\approx \frac{g(m_2)^2}{2\pi\sigma\tau_2}\frac{1}{Z^{\tau_2}}\int_{B_\delta(s)} e^{-(x-s)^2/(\sigma\tau_2)-H(x)/\tau_2}dx \\
&\approx \frac{g(m_2)^2}{2\pi\sigma\tau_2}\frac{\sqrt{H''(m_1)}}{\sqrt{2\pi\tau_2}}e^{-H(s)/\tau_2}\int_{B_\delta(s)} e^{-(x-s)^2/(2\tau_2)(2/\sigma+H''(s))}dx \\
&\approx g(m_2)^2\frac{\sqrt{H''(m_1)}}{2\pi\tau_2}e^{-H(s)/\tau_2}\sqrt{|H''(s)|},
\end{aligned}$$

where we set $\sigma = 1/|H''(s)| = -1/H''(s)$. This implies

$$\tau_2\frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathbb{E}_{\nu^{\tau_2}} g^2} \approx \frac{\sqrt{H''(m_2)|H''(s)|}}{2\pi}e^{(H(m_2)-H(s))/\tau_2} \approx \rho.$$

It remains to show the other term is asymptotically negligible:

$$\begin{aligned}
\mathbb{E}_{\nu_{\tau_1}}(g')^2 &\lessapprox \frac{g(m_2)^2}{2\pi\sigma\tau_2}\frac{1}{Z_{\tau_1}}\int_{B_\delta(s)} e^{-(x-s)^2/(\sigma\tau_2)}dx \cdot \sup_{x\in B_\delta(s)} e^{-H(x)/\tau_1} \\
&\lessapprox \frac{g(m_2)^2}{2\pi}\frac{\sqrt{H''(m_1)|H''(s)|}}{\sqrt{2\tau_1\tau_2}}e^{-(1-\eta)H(s)/\tau_1},
\end{aligned}$$

where $\eta = O(\delta^2)$. Since $\tau_2 > K\tau_1$ for a constant $K > 1$, choosing δ sufficiently small, this implies $\tau_1\frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathbb{E}_{\nu^{\tau_1}} g^2}$ is asymptotically negligible compared to ρ.

Proof of Proposition 3.2.8: In the same set-up as above, imposing $\mathbb{E}_{\nu^{\tau_1}} g^2 = 1$, we get

$$\frac{1}{2}\frac{\mathcal{I}_\pi(f^2)}{\mathrm{Ent}_\pi(f)} \le \tau_1\frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2} + \tau_2\frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2\mathbb{E}_{\nu^{\tau_2}} g^2}.$$

We use the same form of ansatz as before with

$$g(m_1)^2 \approx \frac{Z_2^{\tau_1}}{Z_1^{\tau_1}} \approx \frac{\sqrt{H''(m_1)}}{\sqrt{H''(m_2)}}e^{-H(m_2)/\tau_1}, \quad g(m_2)^2 = \frac{1}{g(m_1)^2}$$

such that $\mathbb{E}_{\nu^{\tau_1}} g^2 = 1$. Then

$$\begin{aligned}
\mathbb{E}_{\nu^{\tau_2}} g^2 &\approx Z_1^{\tau_2} g(m_1)^2 + Z_2^{\tau_2} g(m_2)^2 \approx Z_2^{\tau_2} g(m_2)^2, \\
\mathrm{Ent}_{\nu^{\tau_1}} g^2 &\approx Z_1^{\tau_1} g(m_1)^2 \log g(m_1)^2 + Z_2^{\tau_1} g(m_2)^2 \log g(m_2)^2 \approx \log g(m_2)^2 \approx \frac{H(m_2)}{\tau_1},
\end{aligned}$$

and the same computation as before shows

$$\mathbb{E}_{\nu^{\tau_1}}(g')^2 \lessapprox g(m_2)^2 \frac{\sqrt{H''(m_1)|H''(s)|}}{2\pi\sqrt{2\tau_1\tau_2}} e^{-(1-\eta)H(s)/\tau_1},$$
$$\mathbb{E}_{\nu^{\tau_2}}(g')^2 \approx g(m_2)^2 \frac{\sqrt{H''(m_1)|H''(s)|}}{2\pi\tau_2} e^{-H(s)/\tau_2},$$

where $\eta = O(\delta^2)$. This implies

$$\tau_2 \frac{\mathbb{E}_{\nu^{\tau_2}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2 \mathbb{E}_{\nu^{\tau_2}} g^2} \approx \tau_1 \frac{\sqrt{H''(m_2)|H''(s)|}}{2\pi H(m_2)} e^{(H(m_2)-H(s))/\tau_2} \lesssim \alpha,$$

and that $\tau_1 \frac{\mathbb{E}_{\nu^{\tau_1}}(g')^2}{\mathrm{Ent}_{\nu^{\tau_1}} g^2}$ is asymptotically negligible compared to α.

3.4 Applications of main results

3.4.1 Application to sampling

It is well known that estimates on the Poincaré and log-Sobolev constant yield estimates of the rate of convergence to equilibrium of the underlying process. Applying to the isa, we obtain the following direct consequence of Theorem 3.2.3 and Theorem 3.2.4. We refer to [Sch12, Theorem 1.7] for a proof in the setting of the overdamped Langevin dynamics. The argument directly carries over to the isa.

Corollary 3.4.1. *Let f_t be the relative density of the infinite swapping process* (3.2.2) *at time t. Under the same assumptions as in Theorem 3.2.3 it holds that*

$$\mathrm{Var}_\mu(f_t) \leq e^{-2\rho t} \mathrm{Var}_\mu(f_0),$$

where ρ satisfies the estimate (3.2.9). Under the same assumptions as in Theorem (3.2.4) *it holds that*

$$\mathrm{Ent}_\mu(f_t) \leq e^{-2\alpha t} \mathrm{Ent}_\mu(f_0),$$

where α satisfies the estimate (3.2.12).

Another well-known consequence is that the Poincaré or log-Sobolev constant allows one to quantify the ergodic theorem i.e. to estimate speed of convergence of the time average to the ensemble mean. See [CG08, Proposition 1.2.] and [Wu00, Corollary 4] for a proof in the setting of the overdamped Langevin dynamics. The same argument carries over to the isa.

Corollary 3.4.2. *Let ν denote the initial law of the isa X_t. Under the same assumptions as in Theorem 3.2.3 it holds that for all functions $f : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ such that $\sup |f| = 1$, all $0 < R \leq 1$ and all $t > 0$*

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)\,ds - \int f\,d\mu \geq R\right) \leq \left\|\frac{d\nu}{d\mu}\right\|_{L^2} \exp\left(-\frac{tR^2\rho}{8\,\mathrm{Var}_\mu(f)}\right),$$

where ρ satisfies the estimate (3.2.9).

Under the same assumptions as in Theorem 3.2.4 it holds that for all functions $f \in L^1(\mu)$ and all $R, t > 0$

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)ds - \int f d\mu \geq R\right) \leq \left\|\frac{d\nu}{d\mu}\right\|_{L^2} \exp\big(-t\alpha H^*(R)\big),$$

where α satisfies the estimate (3.2.12) *and*

$$H^*(R) := \sup_{\lambda\in\mathbb{R}} \left\{\lambda R - \log \int \exp\left(\lambda\Big(f - \int f\,d\mu\Big)\right) d\mu\right\}.$$

One consequence of Corollary 3.4.2 is that the isa has an exponential gain compared to the overdamped Langevin dynamics for sampling (see also Remark 3.2.5). See [DLPD12] for the details on the use of the isa to sample from the Gibbs measure $\frac{1}{Z}\exp\big(-\frac{H}{\tau_1}\big)$ at temperature τ_1.

3.4.2 Application to simulated annealing

In this section, we apply the log-Sobolev inequality of Theorem 3.2.4 to the simulated annealing of the isa.

The goal of simulated annealing is to find the global minimum of a function $H : \mathbb{R}^N \to \mathbb{R}$ that is potentially non-convex and lives in a high-dimensional space. Let us explain the main

idea of the stochastic version of simulated annealing. One considers a stochastic process on H that is subject to thermal noise. When simulating this process one lowers the temperature slowly over time. Hereby, the stochastic process gets trapped. Now, the goal is to show that the trapped process converges to the global minimum of H with high probability. This is typically true if the cooling is slow enough. Hence, another goal is to find the best stochastic process with the fastest possible cooling schedule that still allows one to find the global minimum.

Simulated annealing for the overdamped Langevin dynamics was studied in [GH86, Mic92]. As we will see below, the cooling schedule has to be logarithmically slow. This implies long computation times in order to find the global minimum. There are many ways to improve this behavior. Luckily, one has the freedom to choose the underlying stochastic process which is used for simulated annealing. One of the most efficient approach is called Cuckoo search and is based on Lévy flights (see [Pav07, YD09]). Those methods are able to find the global minimum in certain situations with a polynomial cooling schedule. An alternative is to use replica exchange or parallel tempering. As we know from [DLPD12], mixing can only improve when particles are swapped faster, which makes the isa a natural candidate for simulated annealing.

In [Mic92] it was shown that for the overdamped Langevin dynamics the fastest successful cooling schedule is characterized by the Eyring-Kramers formula for the log-Sobolev constant. However, at that time no estimates on the associated log-Sobolev constant for low temperatures were known at that time. Hence, more sophisticated arguments were applied by [HKS89] to replace the log-Sobolev constant by the Poincaré constant showing that the fastest successful cooling schedule is characterized by the critical depth $E_* = H(s_{1p}) - H(m_p)$. Only in 2014, the Eyring-Kramers formula for the log-Sobolev constant was derived in [MS14] which leads to a more direct proof of the same result. This formula was then used by [Mon18] to study simulated annealing for the underdamped Langevin dynamics, showing that the Langevin dynamics is at least as good as the overdamped Langevin dynamics for simulated

annealing. The main result of [HKS89] (see also [Mon18]) is stated as follows.

Theorem 3.4.3 ([HKS89, Mic92]). *Let X_t be given by the classical overdamped Langevin dynamics*

$$dX_t = -\nabla H(X_t)\,dt + \sqrt{2\tau(t)}\,dB_t. \tag{3.4.1}$$

Let $E_* := H(s_{1p}) - H(m_p)$ denote the critical depth of the potential H. Then:

If $\tau(t) \geq \frac{E}{\log t}$ for t large enough with $E > E_*$, then for all $\delta > 0$

$$\mathbb{P}\big(H(X_t) \leq H(m_1) + \delta\big) \underset{t\to\infty}{\to} 1.$$

If $\tau(t) \leq \frac{E}{\log t}$ for t large enough with $0 < E < E_*$, then for δ small enough

$$\limsup_{t\to\infty} \mathbb{P}\big(H(X_t) \leq H(m_1) + \delta\big) < 1.$$

In this section we study simulated annealing for the infinite swapping dynamics given by the following SDE

$$\begin{cases} dX_1 = -\nabla H(X_1)\,dt + \sqrt{2\,\tau_1(t)\,\rho(X_1,X_2) + 2\,\tau_2(t)\,\rho(X_2,X_1)}\,dB_1\,, \\ dX_2 = -\nabla H(X_2)\,dt + \sqrt{2\,\tau_2(t)\,\rho(X_1,X_2) + 2\,\tau_1(t)\,\rho(X_2,X_1)}\,dB_2\,. \end{cases} \tag{3.4.2}$$

We require that for some fixed constant $K > 1$

$$\tau_2(t) = K\tau_1(t) \qquad \text{and} \qquad \tau_1(t) \downarrow 0\,.$$

In Theorem 3.2.3 and Theorem 3.2.4, we showed that the infinite swapping dynamics mixes faster than the overdamped Langevin dynamics. Choosing $\tau_2 = K\tau_1$, the effective critical depth of the potential H is $\frac{E_*}{K}$ compared to E_* for the classical overdamped Langevin dynamics given by (3.4.1). This indicates that the infinite swapping dynamics could outperform the overdamped Langevin dynamics for simulated annealing. The main result of this section shows that this is true.

Theorem 3.4.4. *Assume that the potential H satisfies Assumptions 3 and 4. Let $E_* := H(s_{p1}) - H(m_p)$ be the critical depth of the potential H. For $K > 1$ and $E > \frac{E_*}{K}$, let*

$$\tau_1(t) = \frac{E}{\log(2+t)} \quad \text{and} \quad \tau_2(t) = \frac{KE}{\log(2+t)}. \tag{3.4.3}$$

Let X_1, X_2 be given by (3.4.2) with initial distribution m. Let $m_t(x_1, x_2)$ be the probability density of $(X_1(t), X_2(t))$. Assume the following moment condition for the initial distribution m: for every $p \geq 1$, there exists a constant C_p such that

$$\int \big(H(x_1) + H(x_2)\big)^p dm(x_1, x_2) \leq C_p. \tag{3.4.4}$$

Then for all $\delta > 0$, $\varepsilon > 0$

$$\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) \lesssim \left(\frac{1}{1+t}\right)^{\min\left(\frac{\delta}{E}, \frac{1}{2} - \frac{E_*}{2KE}\right) - \varepsilon}. \tag{3.4.5}$$

3.4.3 Proof of Theorem 3.4.4

With the help of Theorem 3.2.4, i.e. the low-temperature asymptotics of the log-Sobolev constant, the proof of Theorem 3.4.4 follows the arguments in [Mic92, Mon18].

For each $t > 0$, let μ_t be the probability measure given in (3.2.3) at temperatures $\tau_1 = \tau_1(t), \tau_2 = \tau_2(t)$ as defined in (3.4.3), i.e. $\mu_t(x_1, x_2) = \frac{1}{2}(\pi_t(x_1, x_2) + \pi_t(x_2, x_1))$, with

$$\pi_t(x_1, x_2) := \frac{1}{Z_t} \exp\left(-\frac{H(x_1)}{\tau_1(t)} - \frac{H(x_2)}{\tau_2(t)}\right),$$

where Z_t is the normalizing constant making π_t a probability measure. Our first observation is that the mass of the instantaneous equilibrium μ_t concentrates around the global minimum $\min H = 0$ as $t \to \infty$.

Lemma 3.4.5. *If $(\tilde{X}_1(t), \tilde{X}_2(t))$ has law μ_t, then for every $0 < \varepsilon < \delta$, there exists constant C such that*

$$\mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) \leq Ce^{-\frac{\delta - \varepsilon}{\tau_1(t)}} \leq C(2+t)^{-\frac{\delta - \varepsilon}{E}}.$$

Proof. Since $\mu_t(x_1, x_2) = \frac{1}{2}(\pi_t(x_1, x_2) + \pi_t(x_2, x_1))$ and $\min H(x_1, x_2)$ is symmetric,

$$\begin{aligned}\mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) &= \mathbb{P}(\min\{H(\tilde{Y}_1), H(\tilde{Y}_2)\} > \delta) \\ &= \mathbb{P}(H(\tilde{Y}_1) > \delta)\mathbb{P}(H(\tilde{Y}_2) > \delta) \\ &\leq \mathbb{P}(H(\tilde{Y}_1) > \delta),\end{aligned}$$

where $(\tilde{Y}_1, \tilde{Y}_2)$ has law π_t, and $\tilde{Y}_1, \tilde{Y}_2$ are independent. It remains to bound

$$\mathbb{P}(H(\tilde{Y}_1) > \delta) = \frac{\int_{H(x)>\delta} e^{-\frac{H(x)}{\tau_1}} dx}{\int e^{-\frac{H(x)}{\tau_1}} dx}.$$

Under Assumption 3, [MS14, Lemma 3.14] applies and shows H has linear growth at infinity. More specifically, there exists a constant C_H such that for all sufficiently large R,

$$H(x) \geq \min_{|z|=R} H(z) + C(|x| - R) \quad \text{for } |x| > R.$$

In the above, we can choose R large enough so that $\min_{|z|=R} H(z) > \delta$. Then

$$\begin{aligned}\int_{H(x)>\delta} e^{-\frac{H(x)}{\tau_1}} dx &= \int_{H(x)>\delta, |x|<R} e^{-\frac{H(x)}{\tau_1}} dx + \int_{|x|>R} e^{-\frac{H(x)}{\tau_1}} dx \\ &\leq e^{-\frac{\delta}{\tau_1}} \left(|B_R(0)| + \int_{|x|>R} e^{-\frac{C(|x|-R)}{\tau_1}} dx \right) \\ &\leq e^{-\frac{\delta}{\tau_1}} (|B_R(0)| + O(\tau_1)).\end{aligned}$$

On the other hand, there exists $r > 0$ such that $H(x) < \varepsilon$ when $|x| < r$. Then

$$\int e^{-\frac{H(x)}{\tau_1}} dx > \int_{|x|<r} e^{-\frac{H(x)}{\tau_1}} dx > e^{-\frac{\varepsilon}{\tau_1}} |B_r(0)|.$$

Combining these gives the desired estimate. □

Let $(\tilde{X}_1(t), \tilde{X}_2(t))$ be a random vector with law μ_t. By Lemma 3.4.5 and Pinsker's inequality, we have

$$\begin{aligned}\mathbb{P}(\min\{H(X_1(t)), H(X_2(t))\} > \delta) &\leq \mathbb{P}(\min\{H(\tilde{X}_1(t)), H(\tilde{X}_2(t))\} > \delta) + d_{TV}(\mu_t, m_t) \\ &\leq C(2+t)^{-\frac{\delta-\varepsilon}{E}} + \sqrt{2\,\mathrm{Ent}(m_t|\mu_t)}, \qquad (3.4.6)\end{aligned}$$

where

$$\mathrm{Ent}(m_t|\mu_t) := \int \frac{m_t}{\mu_t} \log\left(\frac{m_t}{\mu_t}\right) d\mu_t$$

is the relative entropy of m_t with respect to μ_t. Thus, it remains to bound $\mathrm{Ent}(m_t|\mu_t)$. The following lemma gives an estimate of $\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t)$, the proof of which is in the same spirit of [Mic92, Proposition 3].

Lemma 3.4.6.

$$\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t) \le -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right) + \frac{d}{dt}\left(\frac{1}{\tau_1(t)} + \frac{1}{\tau_2(t)}\right)\mathbb{E}[H(X_1(t)) + H(X_2(t))]. \tag{3.4.7}$$

Proof. First note that

$$\begin{aligned}\frac{d}{dt}\mathrm{Ent}(m_t|\mu_t) &= \int \frac{dm_t}{dt}\log\left(\frac{m_t}{\mu_t}\right)dx + \int m_t \frac{d}{dt}\log\left(\frac{m_t}{\mu_t}\right)dx \\ &= \int \frac{dm_t}{dt}\log\left(\frac{m_t}{\mu_t}\right)dx + \int \frac{dm_t}{dt}dx - \int \frac{m_t}{\mu_t}\frac{d\mu_t}{dt}dx \\ &= \int \frac{dm_t}{dt}\log\left(\frac{m_t}{\mu_t}\right)dx - \int \frac{d\log(\mu_t)}{dt}dm_t. \end{aligned} \tag{3.4.8}$$

We consider the first term in (3.4.8). Observe that m_t satisfies the Fokker-Planck equation

$$\frac{dm_t}{dt} = \nabla_{x_1}\cdot(m_t\nabla_{x_1}H) + \nabla_{x_2}\cdot(m_t\nabla_{x_2}H) + \Delta_{x_1}(a_1 m_t) + \Delta_{x_2}(a_2 m_t).$$

Combining this with the identity $\nabla_{x_i}(a_i\mu_t) = -\mu_t\nabla_{x_i}H$, we get

$$\frac{dm_t}{dt} = \nabla_{x_1}\cdot\left(a_1\mu_t\nabla_{x_1}\left(\frac{m_t}{\mu_t}\right)\right) + \nabla_{x_2}\cdot\left(a_2\mu_t\nabla_{x_2}\left(\frac{m_t}{\mu_t}\right)\right).$$

Integrating by parts, we have

$$\begin{aligned}\int \frac{dm_t}{dt}\log\left(\frac{m_t}{\mu_t}\right)dx &= -\int\left(a_1\left|\nabla_{x_1}\left(\frac{m_t}{\mu_t}\right)\right|^2 + a_2\left|\nabla_{x_2}\left(\frac{m_t}{\mu_t}\right)\right|^2\right)\frac{\mu_t}{m_t}d\mu_t \\ &= -2\mathcal{I}_{\mu_t}\left(\frac{m_t}{\mu_t}\right), \end{aligned} \tag{3.4.9}$$

where $\mathcal{I}_{\mu_t}$ is the Fisher information defined in (3.2.4) for $\mu = \mu_t$. Next we consider the second term in (3.4.8). Using that $\min H = 0$ and that $\tau_1(t), \tau_2(t)$ are decreasing, direct calculation yields

$$-\frac{d\log(\mu_t)}{dt} \le \frac{d}{dt}\left(\frac{1}{\tau_1(t)}\right)\big(H(x_1)\rho(x_1,x_2) + H(x_2)\rho(x_2,x_1)\big)$$

$$+ \frac{d}{dt}\left(\frac{1}{\tau_2(t)}\right)\big(H(x_1)\rho(x_2,x_1) + H(x_2)\rho(x_1,x_2)\big)$$
$$\leq \frac{d}{dt}\left(\frac{1}{\tau_1(t)} + \frac{1}{\tau_2(t)}\right)\big(H(x_1) + H(x_2)\big).$$

Integrating this against dm_t and combining it with (3.4.9) yields (3.4.7). □

The second term on the right hand side of (3.4.7) are controlled via the following lemma.

Lemma 3.4.7. *For any $\varepsilon > 0$, there exists a constant C such that*

$$\mathbb{E}\big[H(X_1(t)) + H(X_2(t))\big] \leq C(1+t)^{\varepsilon}.$$

We omit the proof of Lemma 3.4.7, which closely follows that of [Mic92, Lemma 2], using the moment assumptions on the initial distribution m given in (3.4.4) and growth assumptions on the potential H in Assumption 3.

Lemma 3.4.8. *For any $\varepsilon > 0$, there exists C such that*

$$\operatorname{Ent}(m_t|\mu_t) \leq C\left(\frac{1}{t}\right)^{1-\frac{E^*}{KE}-\varepsilon}.$$

Proof. Using the log-Sobolev inequality in Theorem 3.2.4, the estimate (3.4.7) becomes

$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \leq -2\alpha_t \operatorname{Ent}(m_t|\mu_t) + \frac{2}{E}(2+t)^{-1}\mathbb{E}[H(X_1(t)) + H(X_2(t))],$$

where α_t is the LSI constant in (3.2.11) for $\mu = \mu_t$. From (3.2.12) we see that for any $\varepsilon > 0$, there exists $t_0 > 0$ and $C_1 > 0$ such that for $t > t_0$,

$$2\alpha_t \geq C_1(2+t)^{-\frac{E_*}{KE}-\varepsilon}.$$

Together with Lemma 3.4.7, we get that for $t > t_0$,

$$\frac{d}{dt}\operatorname{Ent}(m_t|\mu_t) \leq -C_1(1+t)^{-\frac{E_*}{E}-\varepsilon}\operatorname{Ent}(m_t|\mu_t) + C_2(1+t)^{-1+\varepsilon}.$$

A standard Gronwall-type argument as in the proof of [Mon18, Lemma 19] then finishes off the estimate: for $0 < \varepsilon < \frac{1}{2}\left(1 - \frac{E_*}{KE}\right)$, let

$$Q(t) = \operatorname{Ent}(m_t|\mu_t) - \frac{2C_2}{C_1}(1+t)^{-1+\frac{E_*}{KE}+2\varepsilon}.$$

Then for t_0 large enough and $t > t_0$,

$$\begin{aligned}
\frac{d}{dt}Q(t) &\leq -C_1(1+t)^{-\frac{E_*}{KE}-\varepsilon}Q(t), \\
Q(t) &\leq Q(t_0)e^{-C_1\int_{t_0}^{t}(1+t)^{-\frac{E_*}{KE}+\varepsilon}}, \\
\mathrm{Ent}(m_t|\mu_t) &\leq \frac{2C_2}{C_1}(1+t)^{-1+\frac{E_*}{KE}+2\varepsilon} + \mathrm{Ent}(m_{t_0}|\mu_{t_0})e^{-\frac{C_1}{\nu}((1+t)^{\nu}-(1+t_0)^{\nu})},
\end{aligned}$$

where $\nu = 1 - \frac{E_*}{KE} - \varepsilon > 0$, and the conclusion follows. □

Combining (3.4.6) and Lemma 3.4.8, we get that for any $\delta > 0, \varepsilon > 0$, there exists a constant C such that

$$\mathbb{P}\Big(\min\big\{H(X_1(t)), H(X_2(t))\big\} > \delta\Big) \leq C\left(\left(\frac{1}{1+t}\right)^{\frac{(\delta-\varepsilon)}{E}} + \left(\frac{1}{1+t}\right)^{\frac{1-\frac{E^*}{KE}-\varepsilon}{2}}\right),$$

which implies (3.4.5).

Chapter 4

Diameter of a long-range percolation graph

Many real-world networks exhibit the small-world phenomenon: their typical distances are much smaller than their sizes. One way to model this phenomenon is a long-range percolation graph on the d-dimensional hypercube $\{0, 1, \cdots, N\}^d$, in which edges are added between far-away vertices with probability falling off as a power of the Euclidean distance. A natural question of long range percolation is how the resulting diameter of the box of size N in graph-theoretical distance scales with N. This question has been intensely studied in the past and the answer depends on the exponent s in the connection probabilities (see e.g. [BKPS04], [CGS02], [Bis04], [Bis11], [Ber04], and [DS13]). In this work we focus on the critical regime $s = d$ studied earlier by [CGS02] and improve the results from bounds to a sharp leading-order asymptotic.

4.1 Model and Current Result

We consider a random graph $G(N)$ on the hypercube $[N]^d \equiv \{0, 1, \cdots, N\}^d$. Let $\|x\|$ denote the L^1 norm of $x \in \mathbb{Z}^d$. Independently from each other, every pair of sites $x, y \in [N]^d$ is connected with probability

$$\begin{cases} 1 & \text{if } \|x - y\| = 1, \\ 1 - \exp\left(-\frac{\beta}{\|x-y\|^d}\right) & \text{otherwise,} \end{cases}$$

where $\beta > 0$ is some fixed parameter. Let $D(x, y)$ denote the graph-theoretical distance between x, y in $G(N)$ and let D_N denote the associated diameter of $G(N)$. The authors of [CGS02] proved the following result:

Theorem 4.1.1. *There exist constants C_1, C_2 (which may depend on β, d) such that*

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{C_1 \log N}{\log\log N} \leq D_N \leq \frac{C_2 \log N}{\log\log N}\right) = 1.$$

Here is a heuristic explanation of this result. The critical exponent $s = d$ in the connection probabilities means that a typical site x in $G(N)$ has $\Theta(\log N)$ neighbors with high probability. This suggests that as m increments, the set $B_m(x)$ grows like a tree with branching degree $\Theta(\log N)$, ignoring any overlap with previously reached sites. Therefore, starting from x, in m steps we should be able to reach $(\Theta(\log N))^m$ sites. At $m = d \log N / \log\log N$, this amounts to $\Theta(N^d)$ sites. This suggests the constants C_1, C_2 can be brought arbitrarily close to d. The arguments in [CGS02] already show that $C_1 = d - \varepsilon$ for any $\varepsilon > 0$, but the value of C_2 is much bigger than d. In this work we show that we can also take $C_2 = d + \varepsilon$ for any $\varepsilon > 0$:

Theorem 4.1.2. *For any $\varepsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\left(D_N \leq \frac{(d+\varepsilon)\log N}{\log\log N}\right) = 1.$$

In particular, $D_N \frac{\log\log N}{\log N} \to d$ as $N \to \infty$ in probability.

4.2 Proof of Theorem 4.1.2

The main idea of the proof is quite simple. Given a site $x \in [N]^d$, let $B_m(x)$ be the set of sites in $[N]^d$ reachable from x in m steps. We want to find some natural numbers m_1, m_2 such that for any two sites x, y, the set $B_{m_1}(x)$ is connected to the set $B_{m_2}(y)$ with high probability. The sum $m_1 + m_2$ then provides an upper bound on the graph distance between x and y in $G(N)$ and we show it can be chosen to be at most $(d + \varepsilon) \log N / \log\log N$.

To do this, we will use a counting argument for the sizes of the sets $B_{m_1}(x)$, $B_{m_2}(y)$. The basic principle is that for any two sets $A_1, A_2 \subset [N]^d$, there are $|A_1||A_2|$ many ways for them to be connected, in correspondence to the pairs $(x_1, x_2) \in A_1 \times A_2$. If the sites in A_1 are separated by distance R from the sites in A_2, then the connection probability for each pair $(x_1, x_2) \in A_1 \times A_2$ is about $1 - \exp\left(-\Theta(R^{-d})\right)$. Under the assumption of independence, the total connection probability between A_1 and A_2 is then about $1 - \exp\left(-\Theta\left(\frac{|A_1||A_2|}{R^d}\right)\right)$. Therefore, the two sets will almost certainly be connected if $|A_1||A_2| \gg R^d$ and will almost certainly fail to be connected if $|A_1||A_2| \ll R^d$. Since the distance between any two sites in $[N]^d$ is at most $\Theta(N)$, this leads us to show that

$$|B_{m_1}(x)| \cdot |B_{m_2}(y)| \geq \Theta(N^{d+\varepsilon}).$$

Recall our "tree" heuristic from the previous section that says $|B_m(x)| = (\Theta(\log N))^m$. Taking logarithms on both sides of the inequality above, this suggests that we take $m_i = \alpha_i d \log N / \log\log N$, for some $\alpha_i \in (0,1)$ with $\alpha_1 + \alpha_2 > 1$, and show our heuristic lower bound

$$|B_m(x)| \geq [\Theta(\log N)]^m$$

holds with high probability. Moreover, our "tree" heuristic also suggests that we establish this by showing the iterative bound

$$|B_{m+1}(x)| \geq \Theta(\log N) \cdot |B_m(x)|$$

holds with high probability.

Let us now proceed with the rigorous proof. The first ingredient is a spatial decomposition of the hypercube $[N]^d$.

Definition 4.2.1. *Given a site $x \in [N]^d$, and an integer $1 < k \leq \lfloor \log_2 N \rfloor$, the k-th dyadic annulus centered at x is the set*

$$H_k(x) := \{y \in [N]^d : 2^{k-1} \leq \|y - x\| < 2^k\}.$$

Note that for each site x, the dyadic annuli $H_k(x)$ are disjoint from each other, and for each integer k between 1 and $\log_2 N$, $|H_k(x)| \geq C_d \cdot 2^{kd}$ for some constant $C_d > 0$.

Definition 4.2.2. *A set $U \subset [N]^d$ is* well-mixed *if there exists some constant $C_U > 0$ such that for all $x \in [N]^d$, $|H_k(x) \cap U_N|, |H_k(x) \cap U_N^c| \geq C_U \cdot 2^{kd}$ for $1 < k \leq \log_2 N$.*

For example, both the lattice $U_N := \{z \in [N]^d : \|z\| \text{ is even}\}$ and its complement $U_N^c = [N]^d \setminus U_N$ are well-mixed.

Given two sites $x, y \in [N]^d$, our strategy is to iteratively build two sets $\tilde{B}_{m_1}(x) \subset U_N$ and $\tilde{B}_{m_2}(y) \subset U_N^c$ by progressively revealing the edge connections in $G(N)$. Roughly speaking, $\tilde{B}_{m_1}(x)$ and $\tilde{B}_{m_2}(y)$ will be "truncated" versions of the "trees" $B_{m_1}(x)$ and $B_{m_2}(y)$ defined earlier, in which a fraction of the edge connections are systematically removed from consideration at each step of branching out. This procedure of truncated growth is a workaround for having to deal with the overlapping of the full "tree" with itself as it branches out. It results in a non-negligible undercounting of site connections but still suffices for obtaining the leading order asymptotic of the graph distance.

We focus only on the construction of $\tilde{B}_{m_1}(x)$, as the construction for $\tilde{B}_{m_2}(y)$ is completely analogous. Before giving the details of the construction procedure, let us fix the following parameters:

- some $\alpha_1, \alpha_2 \in (0, 1)$ such that $\alpha_1 + \alpha_2 > 1$;
- some $\tilde{\alpha} \in (0, 1)$ such that $\tilde{\alpha} > \alpha_1, \alpha_2$;
- some total order on the sites of $[N]^d$;
- some positive integer M.

We begin with defining

$$\partial\tilde{B}_0 := \tilde{B}_0 := \{z \in U_N : \|z - x\| \leq 2^M\}.$$

The set $\tilde{B}_0$ is to be thought of as our "tree" in step 0 and the parameter M controls this initial size. The alternative notation $\partial\tilde{B}_0$ for this set suggests it can also be thought of as

the set of "new sites" added to our "tree" in step 0. Now, suppose our "tree" has been grown to step $m \geq 0$, i.e. we have constructed

- a nested sequence of sets $\tilde{B}_0 \subset \tilde{B}_1 \subset \cdots \tilde{B}_m$ representing the "tree" after step $0, 1, \cdots, m$;
- and the corresponding mutually disjoint sets $\partial\tilde{B}_0, \partial\tilde{B}_1, \cdots, \partial\tilde{B}_m$ representing the collections of "new sites" added to our "tree" in step $0, 1, \cdots, m$.

We then construct a random set $\partial\tilde{B}_{m+1}$ as follows:

- Arrange the sites in $\partial\tilde{B}_m$ in ascending order as $x_1, x_2, \cdots, x_{|\partial\tilde{B}_m|}$.
- For each $x_i \in \partial\tilde{B}_m$, we construct an auxiliary random set ∂S_i that consists of a fraction of the "new sites" connected to x_i in the current step. The set $\partial\tilde{B}_{m+1}$ will then constructed as the union of all the ∂S_i.
- The random sets ∂S_i are constructed sequentially by selecting at most one "new site" from each of the dyadic annuli $H_k(x_i)$. This is where truncation happens.

In more detail, the auxiliary random sets ∂S_i are constructed by the following algorithm. After $\partial S_1, \cdots, \partial S_{i-1}$ have been constructed for $1 \leq i \leq |\partial\tilde{B}_0|$, ∂S_i is constructed as follows:

- Define $S_{i-1} := \bigcup_{j=1}^{i-1} \partial S_j$. (This means $S_0 = \emptyset$.)
- Let $\tilde{H}_k(x_i) := (H_k(x_i) \cap U_N) \setminus (\tilde{B}_0 \sqcup S_{i-1})$.
- For $\tilde{\alpha} \log_2 N \leq k < \log_2 N$, if $\tilde{H}_k(x_i)$ is connected to x_i, let $z_{i,k}$ be the minimal element of $\tilde{H}_k(x_i)$ that is connected to x_i.
- Let ∂S_i be the collection of these $z_{i,k}$'s.

This completes the discussion for the construction of $\partial\tilde{B}_{m+1}$. We then define $\tilde{B}_{m+1} := \tilde{B}_m \sqcup \partial\tilde{B}_{m+1}$ as the "truncated tree" after step $m+1$. Moreover, in order to keep track of the information revealed at each stage of the construction, we define

$$\mathbf{B}_m := (\tilde{B}_0, \tilde{B}_1, \cdots, \tilde{B}_m) \quad \text{and} \quad \mathbf{S}_i := (\mathbf{B}_m, S_0, S_1, \cdots, S_i).$$

With the set-up above, the key fact driving the proof of Theorem 4.1.2 is the following iterative bound:

Lemma 4.2.3. *There exist some constants $c_1, c_2 > 0$ (which may depend on $\beta, d, \tilde{\alpha}$) such that in the procedure above,*

$$\mathbb{P}\left(|\partial \tilde{B}_{m+1}| \geq c_1 \log_2 N \cdot |\partial \tilde{B}_m| \;\middle|\; \mathbf{B}_m \right) \geq 1 - \exp\left(-c_2 \log_2 N \cdot |\partial \tilde{B}_m|\right)$$

on the event that $|\tilde{B}_m| \leq N^{\alpha_1 d + o(1)}$.

Remark 4.2.4. *Note that the requirement $|\tilde{B}_m| \leq N^{\alpha_1 d + o(1)}$ is satisfied as long as $m \leq \alpha_1 d \log N / \log\log N$.*

By construction, we have that

$$|\partial \tilde{B}_{m+1}| = \sum_{i=1}^{|\partial \tilde{B}_m|} |\partial S_i|, \quad |\partial S_i| = \sum_{\tilde{\alpha}\log_2 N \leq k < \log_2 N} \mathbb{1}_{\tilde{H}_k(x_i) \sim x_i},$$

where the expression $S \sim x$ means the set S is connected to the site x. The construction with dyadic annuli give the following uniform lower bound on the probabilities of the events $\{\tilde{H}_k(x_i) \sim x_i\}$.

Lemma 4.2.5. *Consider step m in the construction procedure above. There exists some constant $p > 0$ (which may depend on β, d) such that*

$$\mathbb{P}(\tilde{H}_k(x_i) \sim x_i | \mathbf{S}_{i-1}) \geq p$$

for $\tilde{\alpha}\log_2 N \leq k < \log_2 N$.

Proof. By definition of $\tilde{H}_k(x_i)$, we have that

$$\begin{aligned}
|\tilde{H}_k(x_i)| &\geq |H_k(x_i) \cap U_N| - (|\tilde{B}_m| + |S_{i-1}|) \\
&\geq C_U \cdot 2^{kd} - (|\tilde{B}_m| + |\partial \tilde{B}_m| \cdot (1 - \tilde{\alpha}) \log_2 N) \\
&\geq C_U \cdot 2^{kd} - N^{\alpha_1 d + o(1)} \\
&\geq C_U (1 + o(1)) \cdot 2^{kd},
\end{aligned}$$

where we used the conditions $k \geq \tilde{\alpha}\log_2 N$ and $\tilde{\alpha} > \alpha_1$. Thus, for $\tilde{\alpha}\log_2 N \leq k \leq \log_2 N$,

$$\mathbb{P}(\tilde{H}_k(x_i) \sim x_i | \mathbf{S}_{i-1}) \geq 1 - \exp\left(-\beta \cdot 2^{-kd} \cdot |\tilde{H}_k(x_i)|\right) \geq 1 - e^{-\beta C_U (1 + o(1))} =: p.$$

□

Before proving Lemma 4.2.3, we need one more fact about stochastic domination of random variables. Here and below $\overset{d}{\geq}$ denotes stochastic domination.

Lemma 4.2.6. *Suppose X, Y, Z are discrete random variables, Z is independent from X, and $Y \overset{d}{\geq} Z$ conditioned on X. Then $X + Y \overset{d}{\geq} X + Z$.*

Proof. Conditioned on $X = x$, we couple a copy $Z(x)$ of Z such that conditioned on $X = x$, $Y \geq Z(x)$ almost surely. Define $Z' = Z(x)$ when $X = x$. Then Z' is independent from X and a.s. $X + Y \geq X + Z' \overset{d}{=} X + Z$. □

Proof of Lemma 4.2.3. Conditioned on $\mathbf{S}_{i-1}$, the events $\{\tilde{H}_k(x_i) \sim x_i\}$, where the index k ranges over $[\tilde{\alpha} \log_2 N, \log_2 N)$, are mutually independent. Moreover, each of them has probability at least p by Lemma 4.2.5. Let $Y_{i,k}$ be i.i.d. Bernoulli(p) random variables, independent from the random graph $G(N)$. Then conditioned on $\mathbf{S}_{i-1}$,

$$\partial S_i \overset{d}{\geq} Y_i := \sum_{\tilde{\alpha} \log_2 N \leq k \leq \log_2 N} Y_{i,k}.$$

By Lemma 4.2.6, this implies that conditioned on $\mathbf{B}_m$,

$$|S_{i-1}| + |\partial S_i| \overset{d}{\geq} |S_{i-1}| + Y_i.$$

Iterating this backwards for $1 \leq i \leq |\partial \tilde{B}_m|$, we get that conditioned on $\mathbf{B}_m$,

$$\sum_{i=1}^{|\partial \tilde{B}_m|} |\partial S_i| \overset{d}{\geq} \sum_{i=1}^{|\partial \tilde{B}_m|} Y_i = \sum_{i=1}^{|\partial \tilde{B}_m|} \sum_{\tilde{\alpha} \log_2 N \leq k \leq \log_2 N} Y_{i,k}.$$

Applying the Chernoff bound for i.i.d. sum of Bernoulli random variables, we get

$$\mathbb{P}\left(|\partial \tilde{B}_{m+1}| \leq (1-\delta) p \cdot (1-\tilde{\alpha}) \log_2 N \cdot |\partial \tilde{B}_m| \,\middle|\, \mathbf{B_m} \right) \leq \exp\left(-\frac{\delta^2}{2} p \cdot (1-\tilde{\alpha}) \log_2 N \cdot |\partial \tilde{B}_m| \right).$$

Now take $c_1 = (1-\delta) p \cdot (1-\tilde{\alpha}), c_2 = \frac{\delta^2}{2} p \cdot (1-\tilde{\alpha})$. □

Suppose $m \leq \alpha_1 d \log N / \log\log N$. Let $\mathcal{E}_m$ be the event that $|\partial \tilde{B}_m| \geq c_1 \log_2 N \cdot |\partial \tilde{B}_{m-1}|$. Then on the event $\bigcap_{i=1}^m \mathcal{E}_i$, $|\partial \tilde{B}_m| \geq (c_1 \log_2 N)^m |\tilde{B}_0|$. By Lemma 4.2.3,

$$\mathbb{P}\left(\bigcap_{i=1}^m \mathcal{E}_i \right) \geq 1 - \sum_{i=1}^m \mathbb{P}\left(\mathcal{E}_i^c \,\middle|\, \bigcap_{j=1}^{i-1} \mathcal{E}_i \right)$$

$$\geq 1 - \sum_{i=1}^{m} \exp\left(-\frac{c_2}{c_1}(c_1 \log_2 N)^i |\tilde{B}_0|\right)$$
$$\geq 1 - \exp\left(-\frac{c_2}{c_1}(c_1 \log_2 N)^1 |\tilde{B}_0|\right)(1 + o(1)).$$

Since $|\tilde{B}_0| \geq C_U \cdot 2^{Md}$, we get

$$\mathbb{P}(|\tilde{B}_m| \geq (c_1 \log_2 N)^m) \geq 1 - N^{-|\tilde{B}_0| \cdot c_2 / \log 2}(1 + o(1)) \geq 1 - N^{-\Theta(2^{Md})}.$$

In conclusion, for $m_1 = \alpha_1 d \log N / \log\log N$, if we define $\tilde{B}_{m_1}(x)$ to be the set of sites in U_N reachable from $\tilde{B}_0(x) := \{z \in U_N : \|z - x\| \leq 2^M\}$ with paths inside U_N in m_1 steps, then $|\tilde{B}_{m_1}(x)| \geq (c_1 \log_2 N)^{m_1} \geq N^{\alpha_1 d - o(1)}$ with probability at least $1 - N^{-\Theta(2^{Md})}$.

By the same argument, for $m_2 = \alpha_2 d \log N / \log\log N$, if we define $\tilde{B}_{m_2}(y)$ to be the set of sites in U_N^c reachable from $\tilde{B}_0(y) := \{z \in U_N : \|z - y\| \leq 2^M\}$ with paths inside U_N^c in m_2 steps, then $|\tilde{B}_{m_2}(y)| \geq N^{\alpha_2 d - o(1)}$ with probability at least $1 - N^{-\Theta(2^{Md})}$.

Conditioned on these two events, namely that $|\tilde{B}_{m_1}(x)| \geq N^{\alpha_1 d - o(1)}$ and $|\tilde{B}_{m_2}(y)| \geq N^{\alpha_2 d - o(1)}$, the probability $\tilde{B}_{m_1}(x)$ is connected to $\tilde{B}_{m_2}(y)$ is at least

$$1 - \exp\left(-\Theta(N^{(\alpha_1 + \alpha_2 - 1)d - o(1)})\right) = 1 - \exp\left(-\Theta(N^{\varepsilon_1})\right),$$

where $0 < \varepsilon_1 < \varepsilon := (\alpha_1 + \alpha_2 - 1)d$. Moreover, $\tilde{B}_0(x), \tilde{B}_0(y)$ are reachable from x, y, respectively, in 2^M steps using nearest-neighbor edges. Thus, the overall probability that x and y is connected by a path of at most

$$2^M + m_1 + 1 + m_2 + 2^M = 1 + 2^{M+1} + \frac{(d + \varepsilon)\log N}{\log\log N}$$

is at least

$$1 - N^{-\Theta(2^{Md})} - \exp\left(-\Theta(N^{\varepsilon_1})\right) = 1 - N^{-\Theta(2^{Md})}.$$

Thus, union bound gives

$$\mathbb{P}\left(D_N \leq 1 + 2^{M+1} + \frac{(d + \varepsilon)\log N}{\log\log N}\right) \geq 1 - N^{2d} \cdot N^{-\Theta(2^{Md})}.$$

Choosing M large enough (but constant), the right hand side tends to 1 as N tends to infinity.

4.3 Beyond the leading order

Let us now make an additional assumption to slightly simplify the model, namely that we assume the hypercube $[N]^d$ has periodic boundary condition, so that every site now "looks the same". Following the construction used in the previous section, we define B_m to be the ball of radius m in graph-theoretical distance centered at $\mathbf{0}$, and $\partial B_m := B_m \setminus B_{m-1}$. We may realize the random graph as the growth of a random "tree" starting from $\mathbf{0}$. To keep track of the information revealed at each level of the "tree", let us denote $\mathbf{B}_m := (B_1, B_2, \cdots, B_m)$. We write $\mathbb{E}_{\mathbf{B}_m}[\cdot]$ for $\mathbb{E}[\cdot|\mathbf{B}_m]$ and $\mathbb{P}_{\mathbf{B}_m}[\cdot]$ for $\mathbb{P}[\cdot|\mathbf{B}_m]$. As before, we write $x \sim y$ to mean the site x is connected to the site y and $x \sim S$ to mean the site x is connected to some site belonging to S.

Lemma 4.3.1 (Concentration of $|\partial B_m|$)**.** *For $0 \le \delta \le 1$,*

$$\mathbb{P}\left(1-\delta \le \frac{|\partial B_1|}{\mathbb{E}|\partial B_1|} \le 1+\delta\right) \ge 1 - 2\exp\left(-\frac{\delta^2}{3}\mathbb{E}|\partial B_1|\right)$$

and for $m \ge 1$,

$$\mathbb{P}_{\mathbf{B}_m}\left(1-\delta \le \frac{|\partial B_{m+1}|}{\mathbb{E}_{\mathbf{B}_m}|\partial B_{m+1}|} \le 1+\delta\right) \ge 1 - 2\exp\left(-\frac{\delta^2}{3}\mathbb{E}_{\mathbf{B}_m}|\partial B_{m+1}|\right).$$

Proof. Conditioned on $\mathbf{B}_m$,

$$|\partial B_{m+1}| = \sum_{y \notin B_m} \mathbb{1}_{y \sim \partial B_m}$$

is a sum of independent (but not identically distributed) Bernoulli random variables. The inequalities then follow from the multiplicative form of Chernoff bound for independent sums of Bernoulli random variables. □

In the previous section, we showed by a rough under-counting argument that for $m < \alpha d \log N / \log\log N$, where $0 < \alpha < 1$

$$\mathbb{E}_{\mathbf{B}_m}|\partial B_{m+1}| \ge (\Theta(\log N))^{m+1},$$

with high probability. This means the concentration in Lemma 4.3.1 gets exponentially better with the increment of m, so the main source of fluctuations is in the very first step, $|\partial B_1|$.

This allows us to choose δ (perhaps differently for each m) so that the overall probability of deviation from the "typical" situation in Lemma 4.3.1 is exponentially small in $\log N$ for $m = 0$ and super-exponentially small in $\log N$ for $m \geq 1$, so that the overall error probability is less than N^{-c_0} for some $c_0 > 0$. Moreover, if instead of starting from a single site $\mathbf{0}$, we start from some constant radius ball in L^1-distance centered at $\mathbf{0}$, the exponent c_0 could be improved to any prescribed $c > 0$, while keeping $|\partial B_m|$ within a constant factor and the graph distance m within an additive constant.

This concentration result naturally raises the question: what is the value of $\mathbb{E}_{\mathbf{B}_m}|\partial B_{m+1}|$? A more precise answer to this question would lead to a sharper asymptotic for the graph distance. The iterative construction implies the following simple upper bound

$$\mathbb{E}_{\mathbf{B}_m}|\partial B_{m+1}| \leq |\partial B_m| \cdot \mathbb{E}|\partial B_1|.$$

There are two sources of overcounting in this upper bound:

- Recoil: some sites $x \in B_m$ are revisited.
- Redundancy: some sites $y \in B_m^c$ are multiply connected to ∂B_m.

To better understand the redundancy phenomenon, we introduce the following quantity.

Definition 4.3.2 (Weight of a site relative to a region)**.** *Given a site $y \in [N]^d$ and a region $S \subset [N]^d$, we define the weight of y relative to S to be*

$$\rho_S(y) := \sum_{x \in S, x \neq y} \frac{\beta}{|x - y|^d}.$$

It follows that $\mathbb{P}(y \sim S) = 1 - e^{-\rho_S(y)}$ for $y \notin S$. The elementary inequality

$$(x \wedge 1)(1 - x \wedge 1) \leq 1 - e^{-x} \leq x \wedge 1$$

means that for sites of vanishing weight, i.e. $\rho_S(y) = o(1)$, the connection probability essentially equals to the weight:

$$\mathbb{P}(y \sim S) = \rho_S(y)(1 - o(1)).$$

On the other hand, for sites of heavy weight, the connection probability no longer behaves like a linear function of the weight. Indeed, for regions S that are "expanding away from y", the weight $\rho_S(y)$ essentially equals to the expected degree of y in S,

$$\mathbb{E}[|\{x \sim y : x \in S\}|] = \sum_{x \in S} \left(1 - \exp\left(-\frac{\beta}{|x-y|^d}\right)\right) \approx \sum_{x \in S, x \neq y} \frac{\beta}{|x-y|^d} =: \rho_S(y).$$

which only approximately equals to the connection probability $\mathbb{P}(y \sim S)$ when it is small enough. Thus, the weight serves to identify the "hot spots" for redundant connections.

We are therefore interested in the weight distribution of sites outside B_m relative to ∂B_m. As a start, the additive definition of the weight makes an easy counting argument available. By Fubini theorem for sums,

$$\sum_{y \in B_m^c} \rho_{\partial B_m}(y) = \sum_{x \in \partial B_m} \rho_{B_m^c}(x) \leq |\partial B_m| \cdot \mathbb{E}|\partial B_1| \cdot (1 + o(1)).$$

By the pigeonhole principle, this implies

$$|\{y \in B_m^c : \rho_{\partial B_m}(y) \geq \omega_m\}| \leq \frac{\sum_{y \in B_m^c} \rho_{\partial B_m}(y)}{\omega_m} \leq (\mathbb{E}|\partial B_1|)^{m + \Theta(1)},$$

where we choose the threshold ω_m to be $\Theta((\log N)^{-c})$ for some constant $c > 0$. Together with the bound $|B_m| \leq (\mathbb{E}|\partial B_1|)^m$, this means that as long as $m < \frac{d \log N}{\log \mathbb{E}|\partial B_1|} - \Theta(1)$, most of the sites $[N]^d$ are outside B_m and have vanishing weights relative to ∂B_m, namely in the set

$$V_m := \{y \in B_m^c : \rho_{\partial B_m}(y) < \omega_m\}.$$

This suggests a lower bound strategy of only connecting ∂B_m to sites in V_m. Using the decomposition $|\partial B_{m+1} \cap V_m| = \sum_{y \in V_m} \mathbb{1}_{y \sim \partial B_m}$, we have

$$\mathbb{E}_{\mathbb{B}_m}|\partial B_{m+1} \cap V_m| = \sum_{y \in V_m} \mathbb{P}(y \sim \partial B_m) \geq \sum_{y \in V_m} \rho_{\partial B_m}(y)(1 - \omega_m) = \sum_{x \in \partial B_m} \rho_{V_m}(x)(1 - \omega_m),$$

where we used Fubini theorem for sums again. For each $x \in \partial B_m$,

$$\rho_{V_m}(x) = \rho_{[N]^d}(x) - \rho_{V_m^c}(x) = \mathbb{E}|\partial B_1| - \rho_{V_m^c}(x).$$

It remains to estimate $\rho_{V_m^c}(x)$. The "worst-case" scenario is that V_m^c is tightly clustered around x like a small hypercube, in which case

$$\rho_{V_m^c}(x) = \Theta\left(\log |V_m^c|^{1/d}\right) = \frac{\Theta(m)}{d} \log \mathbb{E}|\partial B_1|.$$

Hypothetically, this worst-case scenario could be achieved uniformly over $x \in \partial B_m$ if the region ∂B_m itself is tightly clustered like a small hypercube. Without further analysis of the spatial distribution of ∂B_m, this crude bound yields that as long as $|B_m| = N^{o(1)}$, or $m = o(\log N / \log\log N)$,

$$\mathbb{E}_{\mathbb{B}_m} |\partial B_{m+1} \cap V_m| \geq |\partial B_m| \cdot \mathbb{E}|\partial B_1| \cdot (1 - o(1)).$$

This is not quite enough to carry the lower bound strategy all the way through. However, intuition suggests that the "tree" B_m should be quite homogeneously distributed in the hypercube $[N]^d$ until a positive fraction of it is filled, and under that assumption, the weight $\rho_{V_m^c}(x)$ should be substantially smaller, so one would expect the lower bound strategy can be continued until the "penultimate steps". These considerations lead us to make the following conjecture:

Conjecture 4.3.3. *There exists constant integer c such that*

$$\lim_{N\to\infty} \mathbb{P}\left(\left| D_N - \frac{d \log N}{\log \mathbb{E}|\partial B_1|} \right| \leq c\right) = 1.$$

Bibliography

[ADPZ10] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: A new micro-macro passage. *Comm. Math. Phys.*, 307, April 2010.

[ADPZ13] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. Large deviations and gradient flows. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 371(2005):20120341, 17, 2013.

[AGS05] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2005.

[And80] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[BB01] I. Benjamini and N. Berger. The diameter of long-range percolation clusters on finite cycles. *Random Structures Algorithms*, 19:102–111, 2001.

[BÉ85] D. Bakry and M. Émery. Diffusions hypercontractives. *Sem. Probab. XIX, Lecture Notes in Math., Springer*, 1123:177–206, 1985.

[BEGK04] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes I: Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004.

[Ber04] N. Berger. A lower bound for the chemical distance in sparse long-range percolation models, 2004.

[Ber13] N. Berglund. Kramers' law: validity, derivations and generalisations. *Markov Process. Related Fields*, 19(3):459–490, 2013.

[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.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[Bis04] M. Biskup. On the scaling of the chemical distance in long-range percolation models. *Ann. Probab.*, 32:2938–2977, 2004.

[Bis11] M. Biskup. Graph diameter in long-range percolation. *Random Structures Algorithms*, 39:210–227, 2011.

[BKPS04] I. Benjamini, H. Kesten, Y. Peres, and O. Schramm. Geometry of the uniform spanning forest: Transitions in dimensions 4, 8, 12. *Ann. Math.*, 160(2):465–491, 2004.

[BL76] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.*, 22(4):366–389, 1976.

[BL19] M. Biskup and J. Lin. Sharp asymptotic for the chemical distance in long-range percolation. *Random Structures Algorithms*, 55:560–583, 2019.

[CG08] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.*, 12:12–29, 2008.

[CGS02] D. Coppersmith, D. Gamarnik, and M. Sviridenko. The diameter of a long range percolation graph. *Random Structures Algorithms*, 21:1–13, 2002.

[CS11] J. D. Chodera and M. R. Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.*, 135(19):194110, 2011.

[CY92] C. C. Chang and H.-T. Yau. Fluctuations of one-dimensional Ginzburg-Landau models in nonequilibrium. *Comm. Math. Phys.*, 145(2):209–234, 1992.

[DDN17] J. Doll, P. Dupuis, and P. Nyquist. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. *Appl. Math. Optim.*, pages 1–42, 2017.

[DG87] D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[DGP17] J. Diehl, M. Gubinelli, and N. Perkowski. The Kardar–Parisi–Zhang equation as scaling limit of weakly asymmetric interacting Brownian motions. *Comm. Math. Phys.*, 354(2):549–589, September 2017.

[Diz07] D. Dizdar. Schritte zu einer optimalen Konvergenzrate im hydrodynamischen Limes der Kawasaki Dynamik (Towards an optimal rate of convergence in the hydrodynamic limit for Kawasaki dynamics). *Diplomarbeit, Universität Bonn*, 2007.

[DL93] R. A. DeVore and G. G. Lorentz. *Constructive Approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, 1993.

[DLPD12] P. Dupuis, Y. Liu, N. Plattner, and J. Doll. On the infinite swapping limit for parallel tempering. *Multiscale Model. Simul.*, 10(3):986–1022, 2012.

[DLR13] H. Duong, V. Laschos, and M. Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM Control Optim. Calc. Var.*, 19, October 2013.

[DMOW18a] D. Dizdar, G. Menz, F. Otto, and T. Wu. The quantitative hydrodynamic limit of the Kawasaki dynamics, 2018.

[DMOW18b] D. Dizdar, G. Menz, F. Otto, and T. Wu. Toward a quantitative theory of the hydrodynamic limit, 2018.

[DS13] J. Ding and A. Sly. Distances in critical long range percolation, 2013.

[Fat13] M. Fathi. A two-scale approach to the hydrodynamic limit Part II: local Gibbs behavior. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(2):625–651, 2013.

[Fat16] M. Fathi. A gradient flow approach to large deviations for diffusion processes. *J. Math. Pures. Appl.*, 106(5):957–993, 2016.

[Fri87] J. Fritz. On the hydrodynamic limit of a one-dimensional Ginzburg-Landau lattice model. The a priori bounds. *J. Stat. Phys.*, 47:551–572, May 1987.

[FS16] M. Fathi and M. Simon. The gradient flow approach to hydrodynamic limits for the simple exclusion process. In *From Particle Systems to Partial Differential Equations III*, volume 162 of *Springer Proceedings in Mathematics & Statistics*, pages 167–184. Springer, Cham, 2016.

[GH86] S. Geman and C.-R. Hwang. Diffusions for global optimization. *SIAM J. Control Optim.*, 24(5):1031–1043, 1986.

[GOVW09] N. Grunewald, F. Otto, C. Villani, and M. Westdickenberg. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(2):302–351, 2009.

[GPV88] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys.*, 118:31–59, 1988.

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[GSC+13] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[GZ03] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[HKN04] B. Helffer, M. Klein, and F. Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. *Mat. Contemp.*, 26:41–85, 2004.

[HKS89] R. A. Holley, S. Kusuoka, and D. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[HN05] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[HN06] B. Helffer and F. Nier. Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: the case with boundary. *Mém. Soc. Math. Fr. (N.S.)*, (105):vi+89, 2006.

[HNR20] H. Hult, P. Nyquist, and C. Ringqvist. Infinite swapping algorithm for training restricted Boltzmann machines. In *Monte Carlo and quasi-Monte Carlo methods*, volume 324, pages 285–307. Springer, Cham, 2020.

[HS87] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.*, 46:1159–1194, 1987.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[KAJ94] C. Koulamas, S. R. Antony, and R. Jaen. A survey of simulated annealing applications to operations research problems. *Omega*, 22(1):41–56, 1994.

[KGV83] S. Kirkpatrick, J. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[Kos01] E. Kosygina. The behavior of the specific entropy in the hydrodynamic scaling limit. *Ann. Probab.*, 29(3):1086–110, 2001.

[KZ09] S. Kannan and M. Zacharias. Simulated annealing coupled replica exchange molecular dynamics–an efficient conformational sampling method. *J. Struct. Biol.*, 166(3):288–294, June 2009.

[Led01] M. Ledoux. Logarithmic Sobolev inequalities for unbounded spin systems revisted. *Sem. Probab. XXXV, Lecture Notes in Math., Springer*, 1755:167–194, 2001.

[LPA+09] Y. Li, V. Protopopescu, N. Arnold, X. Zhang, and A. Gorin. Hybrid parallel tempering and simulated annealing method. *Appl. Math. Comput.*, 212(1):216–228, 2009.

[LY93] S. L. Lu and H. T. Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.*, 156(2):399–433, 1993.

[Men11] G. Menz. LSI for Kawasaki dynamics with weak interaction. *Comm. Math. Phys.*, 307(3):817–860, 2011.

[Mic92] L. Miclo. Recuit simulé sur $\mathbb{R}^n$. étude de l'évolution de l'énergie libre. *Ann. Inst. Henri Poincaré Probab. Stat.*, 28(2):235–266, 1992.

[Mie16] A. Mielke. *On Evolutionary Γ-Convergence for Gradient Systems*, pages 187–249. Springer, Cham, 2016.

[Mil67] S. Milgram. The small world problem. *Psychol. Today*, 2:60–67, 1967.

[Mon18] P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *Probab. Theory Related Fields*, pages 1–34, 2018.

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

[MSTW20] G. Menz, A. Schlichting, W. Tang, and T. Wu. Ergodicity of the infinite swapping algorithm at low temperature, 2020.

[MSW22] A. Montefusco, C. Schutte, and S. Winkelmann. A route to the hydrodynamic limit of a reaction-diffusion master equation using gradient structures, 2022.

[Nar99] K. Nara. Simulated annealing applications. In YH. Song, editor, *Modern Optimisation Techniques in Power Systems. International Series on Microprocessor-Based and Intelligent Systems Engineering.*, volume 20. Springer, Dordrecht, 1999.

[OR07] F. Otto and M. G. Reznikoff. A new criterion for the logarithmic Sobolev inequality and two applications. *J. Funct. Anal.*, 243(1):121–157, 2007.

[Ott01] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.

[OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173:361–400, 2000.

[Pav07] I. Pavlyukevich. Lévy flights, non-local search and simulated annealing. *J. Comput. Phys.*, 226(2):1830–1844, 2007.

[Pav14] G. A. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics*. Springer, New York, 2014. Diffusion processes, the Fokker-Planck and Langevin equations.

[RC05] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 2005.

[Roy99] G. Royer. Une initiation aux inégalités de Sobolev logarithmiques. *Cours Spécialisés, Soc. Math. de France*, 1999.

[San16] F. Santambrogio. Euclidean, Metric, and Wasserstein gradient flows: an overview. *Bull. Math. Sci.*, 7, September 2016.

[Sch12] A. Schlichting. *The Eyring-Kramers formula for Poincaré and logarithmic Sobolev inequalities*. PhD thesis, Universität Leipzig, 2012. Available at `http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-97965`.

[Ser11] S. Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.*, 31:1427, 2011.

[Spo86] H. Spohn. Equilibrium fluctuations for interacting Brownian particles. *Comm. Math. Phys.*, 103(1):1–33, 1986.

[SS04] E. Sandier and S. Serfaty. Gamma-convergence of gradient flows and application to Ginzburg–Landau. *Comm. Pure Appl. Math.*, 57:1627–1672, December 2004.

[Var91] S. R. S. Varadhan. Scaling limits for interacting diffusions. *Comm. Math. Phys.*, 135(2):313–353, 1991.

[Č85] V. Černý. Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *J. Optim. Theory Appl.*, 45(1):41–51, 1985.

[vLA87] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated annealing*, pages 7–15. Springer Netherlands, Dordrecht, 1987.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[Wu00] L. Wu. A deviation inequality for non-reversible Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 36(4):435–445, 2000.

[WY08] L. Wu and N. Yao. Large deviation principles for Markov processes via Φ-Sobolev inequalities. *Electron. Commun. Probab.*, 13:10–23, 2008.

[YD09] X. Yang and S. Deb. Cuckoo search via Lévy flights. In *2009 World Congress on Nature Biologically Inspired Computing (NaBIC)*, pages 210–214, December 2009.

[Zhu90] M. Zhu. Equilibrium fluctuations for one-dimensional Ginzburg-Landau lattice model. *Nagoya Math. J.*, 117:63–92, 1990.